MODELLING OF A SYSTEM FOR THE DETECTION OF WEAK SIGNALS THROUGH TEXT MINING AND NLP. PROPOSAL OF IMPROVEMENT BY A QUANTUM VARIATIONAL CIRCUIT.

Israel Griol Barres

Director: José Millet Roig

March 2022

Departamento de Ingeniería Electrónica

Departament d'Enginyeria Electrònica

Universitat Politècnica de València

Tesis Doctoral - PhD

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Modelling of a System for the Detection of Weak Signals through Text Mining and NLP. Proposal of Improvement by a Quantum Variational Circuit.

Modelización de un sistema para la detección de señales débiles mediante minería de texto y NLP. Propuesta de mejora mediante circuito variacional cuántico.

Israel Griol Barres

Director: José Millet Roig

March 2022

Departamento de Ingeniería Electrónica

Departament d'Enginyeria Electrònica

Universitat Politècnica de València

Tesis Doctoral - PhD

Index of Contents

Dedicatorias y agradecimientos

Son muchísimas personas los amigos, familiares y profesionales a los que estaré eternamente agradecido por el inestimable apoyo prestado en este proyecto y en la carrera académica que culmina con él.

En primer lugar, me gustaría agradecer a mi director de tesis, José Millet, por su acompañamiento y respaldo incondicional en cada una de las ideas y líneas que han sido exploradas en esta tesis doctoral, junto a su dedicación, palabras de ánimo e inspiración durante todo el proceso.

No puedo olvidarme tampoco de mi socio en data science, Sergio Milla, que ha aportado una gran multitud de ideas y horas de codificación que han dado muchísimo a este trabajo. Muchísimas gracias.

También es importante citar al Departamento de Ingeniería Electrónica, y especialmente a este programa de doctorado, al cual estoy muy agradecido por haberme permitido explorar este camino y haberme dado la posibilidad de conciliar este largo proceso con mi vida profesional y personal.

Agradezco a mi excepcional familia, empezando en primer lugar por mi madre Francisca Barres, que siempre ha estado para mí, a mis hermanos Amadeu y David Griol Barres, a mis cuñadas Teresa y Zoraida y a mis sobrinos Amadeu y Teresa. Todos me habéis aportado mucho y me habéis inspirado a seguir trabajando duro para culminar este trabajo. Gràcies també a tots els oncles, ties i cosins Barres i Griol!

Igualmente, agradezco a mis suegros Peter y Lesley Austin, a mis cuñados James y Joel, a mis cuñadas Eli y Carlota, y a la pequeña Vega por formar parte de mi vida, y por haber seguido este proceso con interés y amor. Y también a toda mi familia inglesa. Gracias Peter también por tu inestimable ayuda en la corrección de los artículos científicos.

Quiero agradecer también el apoyo de tantos y tantos amigos y amigas que han formado parte del camino, en especial a Luis Adalid (esto lo celebraremos como toca) y Susana, Rafa Blanca y María, Luis, Nati, Irene y Sandra, Rafa y Amparo, Julián Mora, Juan Antonio Capilla, Toni Díaz “Piri”, Alberto Montolio, Alex Usina, Vio Orero, Pau Carbonell, Sofia Aparisi, Marcos Luna, Yolanda Requena, Majo Gimeno, Olesea Terna, Felipe Zúñiga, José Lluch, Bianca Dragomir, Pieter De Jong, Salomé Reillo, Celsa Monrós, Marta Esteve, Emilio Gallego, Luis Domínguez, Pablo Campos, Diana Caputo, Salomé González, Ricardo Jordán, Ana Torralba, Javier Moliner, Fermín Pina, Cristina Rodríguez, Facun Manzi y Mel Amato, Belén Pérez, Belén Franch, Isra Yuste y Rocío Botella y tantos más con los que he vivido y seguiré viviendo multitud de aventuras.

Agradezco también a TODOS los compañeros y compañeras de IDEAS-UPV, empezando por la vicerrectora Esther Gómez, la directora de área María Márquez y el resto de compañeros y compañeras, pero con “mención especial” a Lorena Pedrós, Sandra Lucas y Óscar Morant, por ser mi “grupo pa too”, y también a Aurora Alonso, Fran Lis, y a Julián Talón y Marisé Ramírez por siempre animarme a luchar.

Mención especial también a Justo Nieto, mi primer gran mentor en el mundo del emprendimiento, y mentor de muchos, y del que, como visionario que siempre ha sido, surgió la idea de este trabajo.

Y por supuesto, este trabajo va dedicado a todas aquellas personas emprendedoras, en especial a nuestro ecosistema StartUPV, pues son un ejemplo de valentía y lucha, y a los que espero que un día aplicaciones como las que están descritas en este trabajo sean de utilidad para acceder a nuevas oportunidades.

Quiero agradecer a los Cárcel Hernández: Eva, José, Adri, Max y en especial a la pequeña Ada, la niña de diez años más bonita y valiente del universo, por enseñarme a no rendirme nunca. Sois un ejemplo para mí sobre cómo vivir con alegría por muchos obstáculos que se crucen en el camino.

Agradezco a todos los que han contribuido en mi formación, mis compañeros y profesores de las Escuelas Pías Malva-Rosa, del Instituto Mariano Benlliure, de la Escuela Técnica Superior de Ingenieros de Telecomunicación y del Departamento de Ingeniería Electrónica de la Universitat Politècnica de València. Y también al Departamento de Organización de Empresas en el que participo como profesor asociado.

Por supuesto, agradezco a los proyectos en los que he podido colaborar tanto los de EIT Climate-KIC, como el resto, destacando a mis compañeros de Climate Launchpad y el proyecto InnoCENS con su coordinador Huaan Fan de la KTH.

También a mis colegas telecos electrónicos y sus parejas (Héctor Pérez, Rubén Salvador, Àngel Alberich, Ángel González, Víctor García, Carlos Ávila, Toni Sánchez, Albero Zambrano, Lucia Aparisi, David Aparisi y Jorge Alberola) por todo lo que hemos compartido y compartiremos en el futuro. No quiero olvidarme tampoco de las personas con las que compartí la experiencia irlandesa en Analog Devices, destacando a mi antiguo boss Mike Keane y a compañeros tan formidables como Paul Morrow y James Elliman.

Por supuesto, no me olvido de mi familia de Río de Vida, Q, EBSE o la Facultad de Teología del CSTAD. Gracias sobre todo a Jhoel y Linda, Mois Veros, Majo Ortega, Mar Resa, Fernando Sánchez, Osmany Cruz, Antonio Delgado, Emilio Carmona y Noemí García. Un abrazo también especial a Nacho Marqués y María Juan por haberme ayudado tanto en el camino y por la gran labor que realizan día a día.

Tampoco olvido agradecer el apoyo que recibo día a día de mi escuela de taekwondo TaekBoxing. Gracias Miguel Pérez y Daria por cruzaros en el camino.

Aprovecho para dedicar este trabajo a Javier Echaleku, por su magistral manera de investigar y revolucionar el mundo del marketing.

Por supuesto, no habría llegado aquí sin una gran multitud de personas que ya no están, como mi primer profesor Paco Mula (seguramente ahora llevo barba por ti, pero tus zapatos no son fáciles de llenar), a Víctor Bataller, a mis iaios Amadeo y Vicenta, a mi tío José Luis, a mi grandad Veyne, al cosí Pele, al tio Pepito, a mis queridísimos hijos Paddy Daniel y garbancito, y, por supuesto, mi papá Amadeo. Te echo de menos cada día, tu consejo y apoyo, y nuestras partidas de truc.

Por último, todo mi amor y agradecimiento van para dos personas muy especiales. La primera es mi hija Sarah Grace, la más bonita, que me llena de energía y alegría cada día. Espero poder ayudarte, mimarte, acompañarte y disfrutar de tu camino. Gracias por ser mi hija.

Y por supuesto, muchas gracias Rebekah Austin por ser mi alma gemela, mi mujer perfecta. Cualquier éxito que alcance habría sido imposible sin tu esfuerzo, apoyo, servicio y amor. No tengo palabras para expresar lo especial que eres y lo mucho que me aportas día a día. Espero algún día poder compensar todo el sacrificio y amor que has dado por mí.

A todos y todas, incluso los que no he podido nombrar en esta dedicatoria, muchísimas gracias por haberme acompañado en esta apasionante etapa.

Abstract

In this doctoral thesis, a system to detect weak signals related to future transcendental changes is proposed and tested. While most known solutions are based on the use of structured data, the proposed system quantitatively detects these signals using heterogeneous and unstructured information from scientific, journalistic, and social sources.

Predicting new trends in an environment has many applications. For instance, companies and startups face constant changes in their markets that are very difficult to predict. For this reason, developing systems to automatically detect significant future changes at an early stage is relevant for any organization to make right decisions on time.

This work has been designed to obtain weak signals of the future in any field depending only on the input dataset of documents. Text mining and natural language processing techniques are applied to process all these documents. As a result, a map of ranked terms, a list of automatically classified keywords and a list of multi-word expressions are obtained.

The overall system has been tested in four different sectors: solar panels, artificial intelligence, remote sensing, and medical imaging. This work has obtained promising results that have been evaluated with two different methodologies. As a result, the system was able to successfully detect new trends at a very early stage that have become more and more important today.

Quantum computing is a new paradigm for a multitude of computing applications. This doctoral thesis also presents a study of the technologies that are currently available for the physical implementation of qubits and quantum gates, establishing their main advantages and disadvantages and the available frameworks for programming and implementing quantum circuits.

In order to improve the effectiveness of the system, a design of a quantum circuit based on support vector machines (SVMs) is described for the resolution of classification problems. This circuit is specially designed for the noisy intermediate-scale quantum (NISQ) computers that are currently available. As an experiment, the circuit has been tested on a real quantum computer based on superconducting qubits by IBM as an improvement for the text mining subsystem in the detection of weak signals.

The results obtained with the quantum experiment show interesting outcomes with an improvement of close to 20% better performance than conventional systems, but also confirm that ongoing technological development is still required to take full advantage of quantum computing.

Keywords: weak signals of the future, quantum computing, text mining, decision making, natural language processing, predictive models, variational quantum circuits.

Resumen

En esta tesis doctoral se propone y evalúa un sistema para detectar señales débiles (weak signals) relacionadas con cambios futuros trascendentales. Si bien la mayoría de las soluciones conocidas se basan en el uso de datos estructurados, el sistema propuesto detecta cuantitativamente estas señales utilizando información heterogénea y no estructurada de fuentes científicas, periodísticas y de redes sociales.

La predicción de nuevas tendencias en un medio tiene muchas aplicaciones. Por ejemplo, empresas y startups se enfrentan a cambios constantes en sus mercados que son muy difíciles de predecir. Por esta razón, el desarrollo de sistemas para detectar automáticamente cambios futuros significativos en una etapa temprana es relevante para que cualquier organización tome decisiones acertadas a tiempo.

Este trabajo ha sido diseñado para obtener señales débiles del futuro en cualquier campo dependiendo únicamente del conjunto de datos de entrada de documentos. Se aplican técnicas de minería de textos y procesamiento del lenguaje natural para procesar todos estos documentos. Como resultado, se obtiene un mapa con un ranking de términos, una lista de palabras clave clasificadas automáticamente y una lista de expresiones formadas por múltiples palabras.

El sistema completo se ha probado en cuatro sectores diferentes: paneles solares, inteligencia artificial, sensores remotos e imágenes médicas. Este trabajo ha obtenido resultados prometedores, evaluados con dos metodologías diferentes. Como resultado, el sistema ha sido capaz de detectar de forma satisfactoria nuevas tendencias en etapas muy tempranas que se han vuelto cada vez más importantes en la actualidad.

La computación cuántica es un nuevo paradigma para una multitud de aplicaciones informáticas. En esta tesis doctoral también se presenta un estudio de las tecnologías disponibles en la actualidad para la implementación física de qubits y puertas cuánticas, estableciendo sus principales ventajas y desventajas, y los marcos disponibles para la programación e implementación de circuitos cuánticos.

Con el fin de mejorar la efectividad del sistema, se describe un diseño de un circuito cuántico basado en máquinas de vectores de soporte (SVM) para la resolución de problemas de clasificación. Este circuito está especialmente diseñado para los ruidosos procesadores cuánticos de escala intermedia (NISQ) que están disponibles actualmente. Como experimento, el circuito ha sido probado en un computador cuántico real basado en qubits superconductores por IBM como una mejora para el subsistema de minería de texto en la detección de señales débiles.

Los resultados obtenidos con el experimento cuántico muestran también conclusiones interesantes y una mejora en el rendimiento de cerca del 20% sobre los sistemas convencionales, pero a su vez confirman que aún se requiere un desarrollo tecnológico continuo para aprovechar al máximo la computación cuántica.

Palabras clave: señales débiles del futuro, computación cuántica, minería de textos, toma de decisiones, procesamiento del lenguaje natural, modelos predictivos, circuitos cuánticos variacionales.

Resum

En aquesta tesi doctoral es proposa i avalua un sistema per detectar senyals febles (weak signals) relacionats amb canvis futurs transcendentals. Si bé la majoria de solucions conegudes es basen en l'ús de dades estructurades, el sistema proposat detecta quantitativament aquests senyals utilitzant informació heterogènia i no estructurada de fonts científiques, periodístiques i de xarxes socials.

La predicció de noves tendències en un medi té moltes aplicacions. Per exemple, empreses i startups s'enfronten a canvis constants als seus mercats que són molt difícils de predir. Per això, el desenvolupament de sistemes per detectar automàticament canvis futurs significatius en una etapa primerenca és rellevant perquè les organitzacions prenguen decisions encertades a temps.

Aquest treball ha estat dissenyat per obtenir senyals febles del futur a qualsevol camp depenent únicament del conjunt de dades d'entrada de documents. S'hi apliquen tècniques de mineria de textos i processament del llenguatge natural per processar tots aquests documents. Com a resultat, s'obté un mapa amb un rànquing de termes, un llistat de paraules clau classificades automàticament i un llistat d'expressions formades per múltiples paraules.

El sistema complet s'ha provat en quatre sectors diferents: panells solars, intel·ligència artificial, sensors remots i imatges mèdiques. Aquest treball ha obtingut resultats prometedors, avaluats amb dues metodologies diferents. Com a resultat, el sistema ha estat capaç de detectar de manera satisfactòria noves tendències en etapes molt primerenques que s'han tornat cada cop més importants actualment.

La computació quàntica és un paradigma nou per a una multitud d'aplicacions informàtiques. En aquesta tesi doctoral també es presenta un estudi de les tecnologies disponibles actualment per a la implementació física de qubits i portes quàntiques, establint-ne els principals avantatges i desavantatges, i els marcs disponibles per a la programació i implementació de circuits quàntics.

Per tal de millorar l'efectivitat del sistema, es descriu un disseny d'un circuit quàntic basat en màquines de vectors de suport (SVM) per resoldre problemes de classificació. Aquest circuit està dissenyat especialment per als sorollosos processadors quàntics d'escala intermèdia (NISQ) que estan disponibles actualment. Com a experiment, el circuit ha estat provat en un ordinador quàntic real basat en qubits superconductors per IBM com una millora per al subsistema de mineria de text.

Els resultats obtinguts amb l'experiment quàntic també mostren conclusions interessants i una millora en el rendiment de prop del 20% sobre els sistemes convencionals, però a la vegada confirmen que encara es requereix un desenvolupament tecnològic continu per aprofitar al màxim la computació quàntica.

Paraules clau: senyals febles del futur, computació quàntica, mineria de textos, presa de decisions, processament del llenguatge natural, models predictius, circuits quàntics variacionals.

1. Introduction

This doctoral thesis describes the conceptualization and implementation of a system for the detection of weak signals based on a quantitative analysis of input datasets of unstructured text documents in different sectors. The sources of these documents are scientifical papers, newspaper articles and social network posts.

Markets are complex environments in which many factors intervene that complicate the creation of predictive systems, even in the short term [1]. For this reason, companies need to make frequent analysis of their market environment since making the right decisions at the right time can mark their future. Currently, these analyses are done through market trend studies, consulting experts, examining the company's own data, or creating alerts on specific topics.

Decision-making processes require the analysis of a large amount of data. These analyses are complex and require the use of information technology. Concepts such as data mining, pattern detection, business intelligence or data science / analytics have become the tools that organizations use to be able to analyze the huge amounts of new data that are produced every day.

Organizations must constantly innovate to generate the largest possible gap with their competitors. At the same time, their competitors also try to innovate to generate new competitive advantages. Thus, detecting new trends before others provides a wider window of opportunity.

The process of analyzing the information of an environment requires the detection of signals. These signals will allow the identification of new trends that open new opportunities.

The term "weak signal" was introduced by Ansoff in 1975 [2] and could be defined as "a change factor which is hardly discernible in the present but will constitute a strong tendency in the future" [3]. Weak signals are events that are too novel to measure their future impact and to react properly to them [4].

Therefore, a weak signal is very difficult to detect in the present but will become more and more important in the future [5]. If one of the terms or expressions detected by a system becomes a strong signal, it will confirm that this weak signal was pointing at something worth exploring.

Hiltunen defined a model with three different dimensions for future signals: "issue", "signal" and "interpretation". The signal is defined as "the real visibility of a signal", the issue is defined as "a variety of events", and the interpretation is defined as "the meaning that a receiver gives to the sign" [6].

The process for detecting weak signals involves gathering and analyzing massive data. In the past, this action was done manually. Today, these activities are carried out by using machine learning and data mining techniques. This process has become more efficient thanks to better computational possibilities for analyzing online sources [7].

The analysis process involves data engineers who must define new algorithms to reduce both execution times and technology requirements, but at the same time, obtaining good performances. Available hardware technologies have enabled computing a huge number of online sources with a limited amount of time and resources [8].

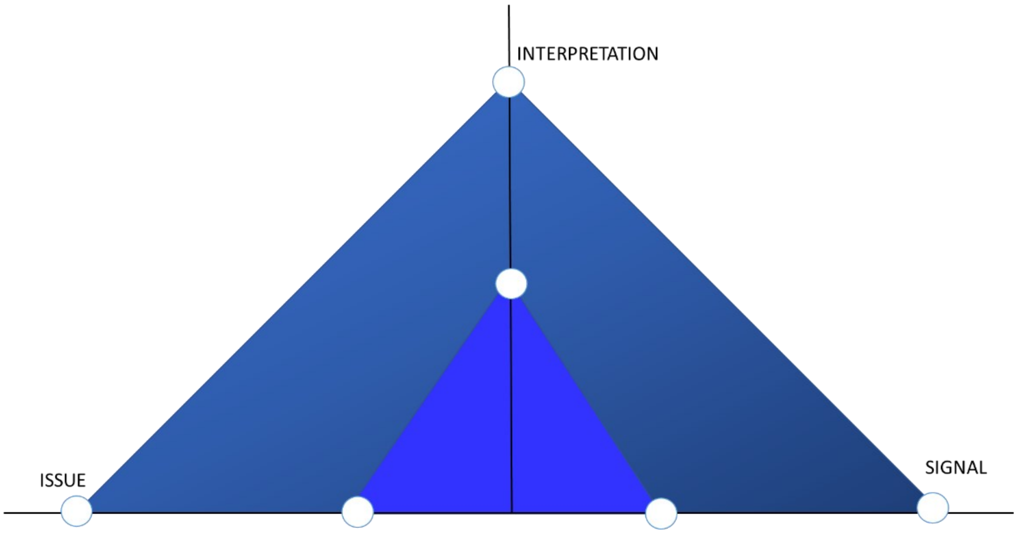


Figure 1. (a) Semiotic model of the sign by Peirce, and (b) semiotic model of the future sign by Hiltunen.

1.1 State of the Art: Context of the fundamental research

Section 2 of the Chapter 2 and Section 1.2 of the Chapter 3 include a full description of the state of the art with the background and related available studies. Section 2 of the Chapter 4 includes a description of the main approaches to quantum technologies. In addition, Section 3 of the Chapter 4 includes a study of the state of the art of the quantum approaches to machine learning.

Although many scholars have created qualitative models and methodologies to define and identify weak signals, there are very few studies on systems based on quantitative models, and on their hardware implementation.

Main studies about detecting weak signals use a single type of information source that is scanned periodically. These systems are based mainly on qualitative or systematic analyses that are only valid for a specific environment or problem.

Some examples of specific models include:

- A model to predict when a new terrorist attack will occur [9].
- A model to predict when a production machine is going to break down using the input data of the vibrations detected from that equipment [10].
- A model to predict where a free space to park the car will be found [11].

- A model to predict the best time to publish a post on social networks, and what are the best words to use [12].

- A model to predict which product will be a top seller next season by studying current and past sales and competitor data [13].

- A model to predict where and when a serial killer will act based on his past behaviors [14].

- A model to predict which film or song to recommend to a user based on past behaviors and current trends [15].

All these models have in common that they cannot be exported to scan other environments, and in many cases, the inputs from these systems include the opinion of experts and other stakeholders.

Other works focus on the use of structured data obtained from the data generated by an organization, or from other external sources, mostly online. For instance, data such as satellite images or temperature can be used for improving precipitation forecast [16].

One of the first works that proposes a quantitative analysis using as an input data source a database of unstructured documents, uses web news to predict the future of the solar panel sector [17]. This work focuses on measuring the overall visibility and number of documents for every signal, discarding one of the dimensions from the model defined by Hiltunen.

The analysis process for unstructured sources, such as text documents, requires the use of text mining techniques. Text mining is a variation of data mining applied to the process of obtaining high-quality information from a text [18]. Processing texts also involves the use of natural language processing with tools such as multiword expressions, which reveal much more complete information than isolated words.

Therefore, after analyzing previous works, it is necessary that the proposed system meets the following characteristics and functionalities:

- The model should carry out a quantitative analysis of the input data.

- The model should detect weak signals in multiple fields only depending on the input dataset.

- The input data of the system must be unstructured.

- The model must take into account all the words in the dataset documents and not only words found in available dictionaries.

- The system input dataset must be composed of documents from different sources.

- The three variables of the Hiltunen model must be considered.

- Text mining and natural language processing techniques must be implemented.

The process for detecting weak signals is not easy, because there are a multitude of documents from different sources that need to be analyzed. Any system implemented for this task must be optimized in both hardware and software to the maximum.

As previously stated, Section 2 of the Chapter 2 and Section 1.2 of the Chapter 3 include a full description of the state of the art with the background and related available studies.

In addition, available conventional technologies are reaching their maximum levels of computational efficiency. The small size of the new circuits is no longer assuring that electrons will move only on their conduction channels, and "tunnel effect" is showing an unpredictive behavior [19]. The main alternative for this problem is the research and development of new technologies based on quantum physics.

Quantum computing is based on the principles of the superposition of matter and quantum entanglement. In addition, quantum computing is based on different computation methods compared to the conventional ones. In theory, it will be possible to store many more states per qubit (the quantum unit of information) and operate with much more efficient algorithms.

One of the main quantum computing applications is the creation of algorithms for artificial intelligence and machine learning [20]. Some studies show that quantum algorithms offer more efficient solutions to problems where computational terms involve a very complex optimization [21] such as, for instance, text mining. In conclusion, a quantum algorithm could improve the efficiency in the detection of weak signals.

Main advantages for quantum computing include improvements in storage [22], design of new materials [23], optimization problems [24], pattern identification [25] or encryption [26].

As previously stated, Section 2 of the Chapter 4 includes a description of the main approaches to quantum technologies. In addition, Section 3 of the Chapter 4 includes a study of the state of the art of the quantum approaches to machine learning.

1.2 Objectives and tasks

The main objective of this doctoral thesis is to **demonstrate the feasibility of the detection of weak signals of the future by the conceptualization of a system and its implementation in the optimal hardware architecture**.

The specific objectives of this doctoral thesis are:

1. To conceptualize a complete system for the detection of weak signals based on quantified factors generated from an unstructured dataset of documents (Article 3 – Chapter 2).

2. To evaluate the applicability of the system in different sectors and to validate the results obtained with experts (Article 2 – Chapter 3).

3. To propose an optimization of the system by adapting it to real quantum computational architectures (Article 1 – Chapter 4).

These objectives have been achieved by the implementation of **the following tasks**:

1. Evaluation and study of state-of-the-art research for detecting weak signals of the future.
2. Creation of a non-linear and NoSQL database of at least 40,000 scientific, journalistic, and social network documents on various topics to form an experimental dataset.
3. Development of a mathematical model for weak signal detection, based on the principles of the Hiltunen model, with a multidimensional basis and applicable to several sectors simultaneously.
4. Implementation of an Extraction, Transformation and Loading (ETL) system to generate system inputs.
5. Development of a text mining model for the classification of keywords in the following classes: noise, weak signals, and strong signals.
6. Development of a natural language processing (NLP) subsystem based on a multi-word expression analysis to obtain more accurate results.
7. Hardware implementation of the system and comparison of results in its execution on CPU and GPU in several fields.
8. Elaboration of a hardware implementation in quantum logic and its execution in a real quantum computer currently available.
9. Publication of results in journals with impact factor, and in prestigious international congresses.

1.3 Main contributions and benefits

The main contributions and benefits from this work are:

1. The model is only dependent on the input dataset of documents. Therefore, it can be applied in every field while other available systems are based on specific models for specific topics. The system has been tested in four different sectors (Chapters 2, 3 and 4).

2. The system is based on a quantified analysis of all the words of every document in the dataset. Therefore, every word is considered as a keyword even if it is not in a standard list of keywords. The classification of keywords is done automatically (Chapter 3).

3. The model analyses three different types of unstructured data sources by using text mining and natural language processing techniques. The architecture of the system has been defined and improved in this doctoral thesis (Chapter 2 and 3).

4. A variational circuit in quantum logic is proposed to improve the efficiency of the system by adapting it to quantum computing (Chapter 4).

Table 1 includes a comparation between the implemented system and the available studies about the detection of weak signals of the future.

Table 1. Comparing implemented system with available studies.

Implemented system	Available previous works
Quantitative analysis of the input data.	Mainly qualitative analysis.
Multiple fields: Model only dependent on the input dataset.	Specific model for a specific topic.
Unstructured input data. Text mining and Natural Language Processing.	Mainly structured data.
All words are considered keywords.	Pre-determined keywords.
Input data from multiple sources.	Mainly one source type.
All 3-dimensions from Hiltunen are considered.	Mainly one or two dimensions are considered.
Hardware optimization by quantum computing.	Mainly theoretical models or conventional software.

1.4 Structure of this document

Chapter 1. Introduction

Includes a brief introduction and motivation of this thesis, state of the art with the context of the fundamental research, definition of objectives and tasks, and main contributions of this work.

List of publications

Includes a list of published articles in journals indexed in JCR, international congresses and international projects.

Chapter 2. Improving strategic decision making by the detection of weak signals in heterogeneous documents by text mining techniques.

This chapter is a published article that includes an experiment to detect weak signals of the future in the field of artificial intelligence. The system includes natural language processing techniques and a new dataset of about 15,000 documents in a period of more than ten years. Results are validated by industry experts. This chapter addresses the Objective 1 of this thesis and develops tasks 1-7 and 9.

Chapter 3. Detecting Weak Signals of the Future: A System Implementation Based on Text Mining and Natural Language Processing.

This chapter is a published article that includes a new experiment to detect weak signals of the future in the field of remote sensing. The system includes natural language processing techniques and a new dataset of about 50,000 documents in a period of more than ten years. Results are validated by industry experts. Improved execution times are obtained by introducing new parallel implementations on GPUs. New validation methodologies are introduced. This chapter addresses the Objective 2 of this thesis and develops tasks 1-7 and 9.

Chapter 4. Variational quantum circuits for machine learning. An application for the detection of weak signals.

This chapter is a published article that includes an analysis of all available quantum computing technologies. In addition, a variational circuit in quantum logic is proposed and applied as part of the system to detect weak signals in the field of medical imaging. This chapter addresses the Objective 3 and develops tasks 2, 7, 8 and 9.

Chapter 5. General discussion of results

This chapter is an analysis of all the process of the design, implementation, and improvements of the system to detect weak signals of the future. In addition, the chapter includes a general discussion of the results obtained in every step of the whole thesis process, and a summary of the most important results by fields of study.

Chapter 6. General conclusions and future work

This chapter includes the general conclusions, limitations, challenges, and the future lines of research.

Appendixes

This section includes a flow chart of the complete system, a list of keywords for DoV and DoD study in the field of remote sensing, and a review of the impact factors of the indexed journals where the research articles have been published.

References

1. S. Brown and K. Eisenhardt, *Competing on the Edge: Strategyas Structured Chaos*, Harvard Business School Press, Boston, **1998,** p. 23.
2. H.I. Ansoff, Managing strategic surprise by response to weak signals, *California Management Review*, **1975**, **18**(2), 21–33. doi:10.2307/41164635.
3. M. Godet, *From Anticipation to Action, a Handbook of Strategic Prospective*, UNESCO Publishing, **1994**, p. 59.
4. Cooper, A.; Voigt, C.; Unterfrauner, E.; Kravcik, M.; Pawlowski, J.; Pirkkalainen, H. Report on Weak Signals Collection. TELMAP, European Commission Seventh Framework Project (IST-257822). **2011.** Deliverable D4.1. pp. 6–7. Available online: https://cordis.europa.eu/docs/projects/cnect/2/257822/080/deliverables/001-D41Weaksignalscollectionfinal.doc (accessed on 25 April 2020).
5. Godet, M. *From Anticipation to Action, a Handbook of Strategic Prospective*; UNESCO Publishing: Paris, France, **1994**; p. 59.
6. E. Hiltunen, The future sign and its three dimensions, *Futures,* **2008**, 40(3), 247–260. doi:10.1016/j.futures.2007.08.021.
7. R. Eckhoff, M. Markus, M. Lassnig and S. Schon, Detecting weak signals with technologies: Overview of current technology enhanced approaches for the detection of weak signals, *Int. J. Trends Econ. Manag. Technol.*, **2014,** 3, 1–7.
8. Wu, X.; Zhu, X.; Wu, G.Q.; Ding, W. Data mining with big data. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 97–107.
9. Koivisto, R.; Kulmala, I.; Gotcheva, N. Weak signals and damage scenarios—Systematics to identify weak signals and their sources related to mass transport attacks. *Technol. Forecast. Soc. Chang.* **2016**, *104*, 180–190.
10. Huang, M.; Liu, Z. Research on Mechanical Fault Prediction Method Based on Multifeature Fusion of Vibration Sensing Data. *Sensors* **2020**, *20*, 6.
11. Awan, F.M.; Saleem, Y.; Minerva, R.; Crespi, N. A Comparative Analysis of Machine/Deep Learning Models for Parking Space Availability Prediction. *Sensors* **2020**, *20*, 322.
12. Kanuri, V.K.; Chen, Y.; Sridhar, S.H. Scheduling Content on Social Media: Theory, Evidence, and Application. *Journal of Marketing*, **2018**, 82, 6, 89-108. DOI: https://doi.org/10.1177/0022242918805411

13. Coyne, K.; Horn, J. Predicting your competitor's reaction. *Harvard Business Review*. **2009**, magazine.
14. Sharma, M. The Development of Serial Killers: A Grounded Theory Study. **2018**. Master Theses in Eastern Illinois University.
15. Schedl, M.; Zamani, H.; Chen, CW. Current challenges and visions in music recommender systems research. *Int J Multimed Info Retr*, **2018**, 7, 95–116. DOI: https://doi.org/10.1007/s13735-018-0154-2
16. Lagasio, M.; Parodi, A.; Pulvirenti, L.; Meroni, A.N.; Boni, G.; Marzano, F.S.; Luini, L.; Venuti, G.; Realini, E.; Gatti, A.; Tagliaferro, G.; Barindelli, S.; Guarnieri, A.M.; Goga, K.; Terzo, O.; Rucci, A.; Passera, E.; Kranzlmueller, D.; Rommen, B. A Synergistic Use of a High-Resolution Numerical Weather Prediction Model and High-Resolution Earth Observation Products to Improve Precipitation Forecast. *Remote Sens.* **2019**, *11*(20), 2387; DOI: https://doi.org/10.3390/rs11202387
17. Yoon, J. Detecting weak signals for long-term business opportunities using text mining of Web news. *Expert Syst. Appl.* **2012**, *39*, 12543–12550.
18. Ishikiriyama, C.S.; Miro, D.; Simoes Gomesa, C.F. Text mining business intelligence: A small sample of what words can say. Procedia Computer Science, **2015**, 55, 261–267. doi:10. 1016/j.procs.2015.07.044.
19. Martín-Palma, R.J. Quantum tunneling in low-dimensional semiconductors mediated by virtual photons. *AIP Adv.* **2020**, *10*, 015145, doi:10.1063/1.5133039
20. Havlíček, V.; Córcoles, A.D.; Temme, K.; Harrow, A.W.; Kandala, A.; Chow, J.M.; Gambetta, J.M. Supervised learning with quantum-enhanced feature spaces. *Nature* **2019**, *567*, 209–212, doi:10.1038/s41586-019-0980-2.
21. Barkoutsos, P.K.I.; Nannicini, G.; Robert, A.; Tavernilli, I.; Woerner, S. Improving variational quantum optimization using CVaR. *Quantum* **2020**, *4*, 256, doi:10.22331/q-2020-04-20-256.
22. Ruiz, I.E.K. Quatum Computing: Practical applications that classical computing cannot solve. Bachelor's Thesis, University Carlos III, Madrid, Spain, **2019**.
23. Liu, H.; Hao Low, G.; Steiger, D.S.; Häner, T.; Reiher, M.; Troyer, M. Prospects of Quantum Computing for Molecular Sciences, *Quantum Phys.* **2021**, arXiv:2102.10081.
24. Bluhm, H.; Foletti, S.; Neder, I.; Rudner, M.; Mahalu, D.; Umansky, V.; Yacoby, A. Dephasing time of GaAs electron-spin qubits coupled to a nuclear bath exceeding 200 µs. *Nat. Phys.* **2010**, *7*, 109–113, doi:10.1038/nphys1856.
25. Schuld, M.; Bocharov, A.; Svore, K.; Wiebe, N. Circuit-centric quantum classifiers. *Phys. Rev. A* **2020**, *101*, 032308, doi:10.1103/PhysRevA.101.032308.
26. Woerner, S.; Egger, D.J. Quantum risk analysis. *Nat. Quantum Inf.* **2019**, *5*, 15, doi:10.1038/s41534-019-0130-6.

List of publications

Articles in journals indexed in JCR

The results of this doctoral thesis have been disseminated in the following articles in journals indexed in JCR:

[Article 1] I Griol, S Milla, A Cebrian, Y Mansoori, J Millet (2021) Variational quantum circuits for machine learning. An application for the detection of weak signals. Applied Sciences 2021, 11, 6427; doi: 10.3390/app11146427 (JCR: Q2)

Available quantum computing technologies are analyzed. A variational circuit in quantum logic is proposed and applied as part of the system to detect weak signals in the field of medical imaging.

Objective 3. Tasks 2, 7, 8 and 9. This article is presented as chapter 4 of this thesis.

[Article 2] I Griol, S Milla, A Cebrian, H Fan, J Millet (2020) Detecting Weak Signals of the Future: A System Implementation Based on Text Mining and Natural Language Processing. Sustainability. 2020, 12(19), 7848; https://doi.org/10.3390/su12197848 (JCR: Q2)

A third experiment is performed in the field of remote sensing to detect weak signals with three different dimensions. The system includes natural language processing techniques and new datasets of about 50,000 documents in a period of more than ten years. Results are validated by industry experts. Improved execution times are obtained by introducing new parallel implementations on GPUs. New validation techniques are introduced.

Objective 2. Tasks 1, 2, 3, 4, 5, 6, 7 and 9. This article is presented as chapter 3 of this thesis.

[Article 3] I Griol, S Milla, J Millet (2020) Improving strategic decision making by the detection of weak signals in heterogeneous documents by text mining techniques. Artificial Intelligence Communications Journal. AI Communications, vol. 32, no. 5-6, pp. 347-360, 2019. DOI: https://doi.org/10.3233/AIC-190625 (JCR: Q4)

A second experiment is developed in the field of artificial intelligence. A new parameter is introduced to measure the spread of keywords. The system includes natural language processing techniques and a new dataset of about 15,000 documents in a period of more than ten years. Results are validated by industry experts.

Objective 1. Tasks 1, 2, 3, 4, 5, 6, 7 and 9. This article is presented as chapter 2 of this thesis.

[Article 4] I Griol, S Milla, J Millet (2019) System implementation to detect weak signals of future using text mining. Revista de Documentación Científica del CSIC. 42(2), abril-junio 2019, e234 ISSN-L:0210-0614. https://doi.org/10.3989/redc.2019.2.1599 (JCR: Q3)

A first experiment is developed in the field of solar panels, with a dataset of about 15,000 documents corresponding to a period of more than ten years. Results are compared with the one-dimensional Yoon system and a high correlation of over 90% is obtained.

Objective 1. Tasks 1, 2, 3, 4, 5 and 9.

International congresses

The international congresses in which results have been shared are listed below:

[Congress 1] ICAART 2019 11th International Conference on Agents and Artificial Intelligence. System Implementation for the Detection of Weak Signals of the Future in Heterogeneous Documents by Text Mining and Natural Language Processing Techniques. 19-21 Feb 2019. Prague (Czech Republic). Published in "Proceedings of the 11th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART, 631-638, 2019, Prague, Czech Republic. ISBN: 978-989-758-350-6". This congress is of special prestige as IEEE is the technical sponsor and reviewer.

[Congress 2] Consortium on Applied Research and Professional Education - CARPE Conference. "A Self-Assessment Tool For Business Models And Automatic Generation Of Business Opportunities" 4-6 Nov 2013. Manchester Metropolitan University, Manchester (United Kingdom).

[Congress 3] 1st Innosim Workshop - Simulation For Innovation "A Self-Assessment Tool For Business Models And Automatic Generation Of Business Opportunities" 24 Mar 2013. Aalto University Of Helsinki (Finland).

[Congress 4] The XXI ISPIM Conference, Paper: "Comparative Analysis of Seven Competitiveness Indexes. A Proposal for Simplified Unification" 7 Jun 2010. Museo Guggenheim Bilbao. Bilbao (Spain).

International projects

The following projects have contributed to the development of this PhD:

[Project 1] Griol Barres, Israel; Millet Roig, José; Morant-Martínez, Oscar. Innocens (573965-EPP-1-2016-1-SE-EPPKA2-CBHE-JP). (15/10/16 - 14/10/19). Proyectos de investigación con coordinador. ERASMUS+ CAPACITY BUILDING.

[Project 2] Millet Roig, José (I.P.); Griol Barres, Israel; Tormo-Carbó, Guillermina; Ramos Peinado, Germán; Llinares Llopis, Raúl. Start Up Accelerator At Upv. (01/01/18 - 01/01/19). Apoyo competitivo a la transferencia y difusión de tecnología. EIT CLIMATE-KIC.

[Project 3] Millet Roig, José (I.P.); Griol Barres, Israel; Cebrián Ferriols, Antonio José; Tormo-Carbó, Guillermina. Pip Programme - Pioneers (APEC0158_2017-1.6.2-215_P066--). (09/01/17 - 01/01/18). Apoyo competitivo a la transferencia y difusión de tecnología. EIT CLIMATE-KIC.

[Project 4] Millet Roig, José (I.P.); Griol Barres, Israel; Tormos Ferrando, Álvaro; Guill Ibáñez, Antonio; Castells, Francisco; Márquez Gómez, María. Start-Up Accelerator Partner Support. (09/01/17 - 01/01/18). Apoyo competitivo a la transferencia y difusión de tecnología. EIT CLIMATE-KIC.

[Project 5] Millet Roig, José (I.P.). Decision Metrics And Finance - Accelerator Support Activities (APDM018_2016-1.4.2-297_P66-18). (01/10/16 - 01/01/17). Apoyo competitivo a la transferencia y difusión de tecnología. EIT CLIMATE-KIC.

Article

Improving strategic decision making by the detection of weak signals in heterogeneous documents by text mining techniques.

Israel Griol-Barres, Sergio Milla and José Millet.

AI Communications 2019, 32, no. 5-6, pp. 347-360; doi: 10.3233/AIC-190625

Published: 4 March 2020

Authors	*Griol-Barres, I.; Milla, S.; Millet, J.*
Title	*Improving strategic decision making by the detection of weak signals in heterogeneous documents by text mining techniques.*
Journal	*AI Communications, **2019,** vol. 32, no. 5-6, pp. 347-360.*
DOI	*10.3233/AIC-190625*
Publisher	IOS PRESS, NIEUWE HEMWEG 6B, 1013 BG AMSTERDAM, NETHERLANDS
ISSN	0921-7126
eISSN	1875-8452
Impact Factor	1.0 (2019) – 1.072 (5 years)

JCR Category	Rank in Category	Quartile in Category
COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	**118 of 136**	**Q4**

CiteScore Highest percentile	54% 103/227 Artificial Intelligence
CiteScore 2020	3.7
SJR 2020	0.337
SNIP 2020	1.281
HIndex	38
Main contributions	This chapter is a published article that includes an experiment to detect weak signals of the future in the field of artificial intelligence. The system includes natural language processing techniques and a new dataset of about 15,000 documents in a period of more than ten years. Results are validated by industry experts. This chapter addresses the Objective 1 of this thesis and develops tasks 1-7 and 9.

Article
Improving strategic decision making by the detection of weak signals in heterogeneous documents by text mining techniques.
Israel Griol-Barres, Sergio Milla and José Millet.
AI Communications 2019, 32, no. 5-6, pp. 347-360; doi: 10.3233/AIC-190625
Published: 4 March 2020

2. Improving strategic decision making by the detection of weak signals in heterogeneous documents by text mining techniques.

Abstract: At present, one of the greatest threats to companies is not being able to cope with the constant changes that occur in the market because they do not predict them well in advance. Therefore, the development of new processes that facilitate the detection of significant phenomena and future changes is a key component for correct decision making that sets a correct course in the company. For this reason, a business intelligence architecture system is hereby proposed to allow the detection of discrete changes or weak signals in the present, indicative of more significant phenomena and transcendental changes in the future. In contrast to work currently available focusing on structured information sources, or at most with a single type of data source, the detection of these signals is here quantitatively based on heterogeneous and unstructured documents of various kinds (scientific journals, newspaper articles and social networks), to which text mining and natural language processing techniques (a multi-word expression analysis) are applied. The system has been tested to study the future of the artificial intelligence sector, obtaining promising results to help business experts in the recognition of new driving factors of their markets and the development of new opportunities.

Keywords: Weak signal of the future, strategic decision making, text mining, business intelligence architecture, unstructured information

1. Introduction

The greatest threat to companies is the pace ofchanges in the market, combined with their inability to manage and anticipate these changes [3]. Markets are complex environments in which it can be very difficultto make the right decisions, and failure to do so can mark the future of the company.

The development of new processes that facilitate decision making in organizations considering data from different internal and external sources, thus becomes more and more important. Since the volume of data tobe managed is increasing, these processes must involvetechniques of capture, transformation, storage, and automatic analysis that reduce the time and means required to analyze the data, while providing a high reliability. Among them, data mining in business intelligence (BI) processes stands out.

At present, information technology plays an important part in business, due to its fundamental role in thecreation of business intelligence. This term covers a broad category of applications and technologies to collect, analyze and provide access to the huge amounts of data stored in a company's database. The term intelligence in "Business Intelligence" is closely related to knowledge [22], in the sense of stored information, or models used by a person or machine to interpret, predict and respond appropriately to the outside world [10].

On the other hand, data mining is based on extracting relevant knowledge from documents from different sources. In the world of business competitiveness, the keyword "future" represents the need to identify potential new business opportunities [40], and many experts are working on evolutionary analysis, pattern detection, data mining methods, theories of disruptive innovation or detection of future signals.

One type of future signal is the weak signal. This term was first coined in 1975 [1] as the detection of evidence of an emerging change within a continuous process of exploring a medium. That is, external or internal events and developments which are still too incomplete to allow an accurate estimation of their impact and/or develop a response to them. However, these changes conceal the potential for more significant phenomena and transcendental changes in the environment, hence the importance of being able to identify and monitor them. If these phenomena evolve to become relevant (strong signals), they have the potential to reinforce a plan of action or to obstruct it. Another definition of the term weak signal [11] is that of "a change factor which is hardly discernible in the present but will constitute a strong tendency in the future".

Although in recent years "weak signals" have been receiving more attention in studies on the future, thereis still no widespread use of the term, and there are authors who have used synonyms such as "seeds of change" [30], "facts emerging" [5,6], "strategy signals" [31] and "early warning signals" [27]. The dilemma of having several definitions led Hiltunen [16] to introduce the concept of future sign, which, at the same time, is based on the semiotic model of the sign [32]; although both are defined through a conceptual framework composed of three dimensions, the Hiltunen model is specific for future signals.

Peirce's triad model consists of an "object", that is,the "portion" of reality that is accessed through the sign itself; a "representamen", or symbolic representation of something; and an "interpretant", related to the interpretation that a culture makes of the sign through itsown knowledge.

From this model, Hiltunen defined a triad on the future sign in these three dimensions: "issue", "signal" and "interpretation", with the aim of deepening the description of weak to strong signals [25], and developed an analysis of the processes of significance in which the future signal is perceived, interpreted and produced. Figure 1 shows a comparison between both models.

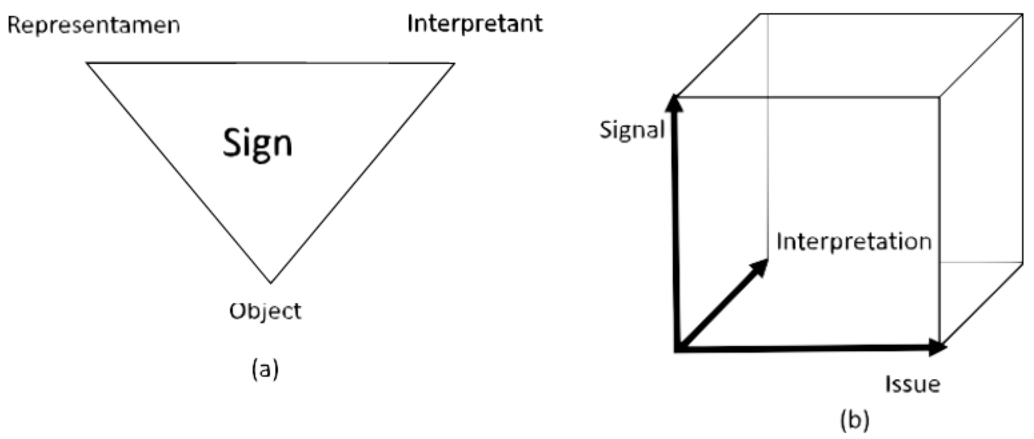


Figure 1. (a) Semiotic model of the sign by Peirce, and (b) semiotic model of the future sign by Hiltunen.

Generally, a weak signal will have low absolute value in all three components, or at least in two of them. An example of the use of this model can be found in a story published in the Helsingin Sanomat, the most important newspaper in Finland, in which the Swedish clothing chain Hennes & Mauritz (H&M) was selling old clothes at the price of new clothes, under the label of "vintage". In actual fact, only 1% of its stores were carrying out this action. From the point of view of the future signal, the amount of signal (i.e. its visibility) was enormous. But the reality, the issue, is that 1% of the stores is not representative. This could be considered a weak signal since only one of the three components is strong, and therefore, it is not clear if this H&M action will represent a future trend or not.

Scanning for weak signals in the past involved gathering and analyzing massive online data manually. Now, the process is facilitated by machine learning and data mining techniques, thanks also to more efficient computational processing to detect weak signals from online sources [8].

This article includes the design and implementation of a system to detect weak signals of the future. The potential scope of users for this system is, firstly, companies (incumbents and startups) but also governments and consumers. The system is used to detect and enlarge windows of opportunities. However, it does not intend to substitute the job of the experts but to help them giving quantified value in their predictions. The system has been tested by the analysis of multiple unstructured documents concerning the sector of artificial intelligence, detecting words and expressions related to new future trends. The system could be applied on other sectors depending on the dataset of documents used in its input.

Section 2 of the paper describes related approaches and the state of the art; while Section 3 covers the implementation of the system, and the results obtained through the experiment are explained in Section 4. Finally, Section 5 presents the main conclusions of the study and suggests future research lines.

2. State of the art

Most of the publications about weak signals are a qualitative analysis related to a specific topic. One instance is the identification of weak signals and their sources related to terrorism and mass transport attacks [24]; another example the influence of maximization in social networks through finding a small set of nodes [29]. A strategic radar [34] which is an integrated framework that uses scenario planning, business analytics and dashboard technology, was proposed to monitor and scan for important external signals including weak signals.

One of the first studies that includes a quantitative analysis is an approach in only one source: web news [41]. The quantification involved measuring the signalpresence in media using degree of visibility and degree of diffusion (distinguishing number of documentsand number of global appearances), and taking into account only their rates of increase with a time weight.

Another work [36] proposes a new methodology that enables the automated identification of weak signals for strategic forecasting. Weak signals are extracted from an organization's environment as represented by internet information. Based on a given hypothesis about the future, related websites are identified and textual information is extracted. The same authors use a latent semantic indexing approach to groupdocuments with similar meaning for the further detection of weak signals [35].

Text mining has also started to extract multiword expressions in multiple applications such as bilingual multiword extraction from free text [7]. Therefore, adding multiword expressions in the detection of futuresignals could provide more accurate results.

An organization operating in a complex and unpredictable environment must be flexible in order to filter this type of information. The implications of a "weak signal" are very difficult to define at an early stage. However, in these complex environments, every organization is forced to make decisions in progressively earlier stages of knowledge. In addition, the time available to react is shortened, while the organization becomes more complex, making the detection of these weak signals of change a priority in order to make theright decisions.

To do this, the organization has to frequently scan and analyze the environment and this capacity of analysis will determine its ability to apply strategic behaviors. In order to be able to monitor an environment composed of a large number of documents, the process of future-oriented meaning must be taken into account [20]: which is defined as the appearance and development of external themes and signals that are interconnected, their interpretation (the transfer of exosignals to endosignals), the recreation of secondary exosignals for communication, and the different waysto act in consequence to these signals and the issues concerned. Therefore, it is a complex process with many interconnections.

As previously stated, the interpretation depends toa large extent on the source in which the data originate. For example, a first exosignal (a signal external tothe organization) could be a research result published in the form of an article in a scientific journal. In the process of significance, this exosignal may be interpreted by a journalist, becoming an endosignal for him, whose subsequent journalistic article becomes a secondary exosignal. In the same way, this signal can be codified by other players who interpret and share it on social networks such as twitter. Therefore, interpreting is an activity in which there are actors who formu late their own endosignals based on the exosignals of the issue itself, followed by secondary exosignals thatare at the same time interpreted and transformed into new exosignals. This process is illustrated in Figure 2. In conclusion, these three different types of information source should be taken into account in the detection ofweak signals.

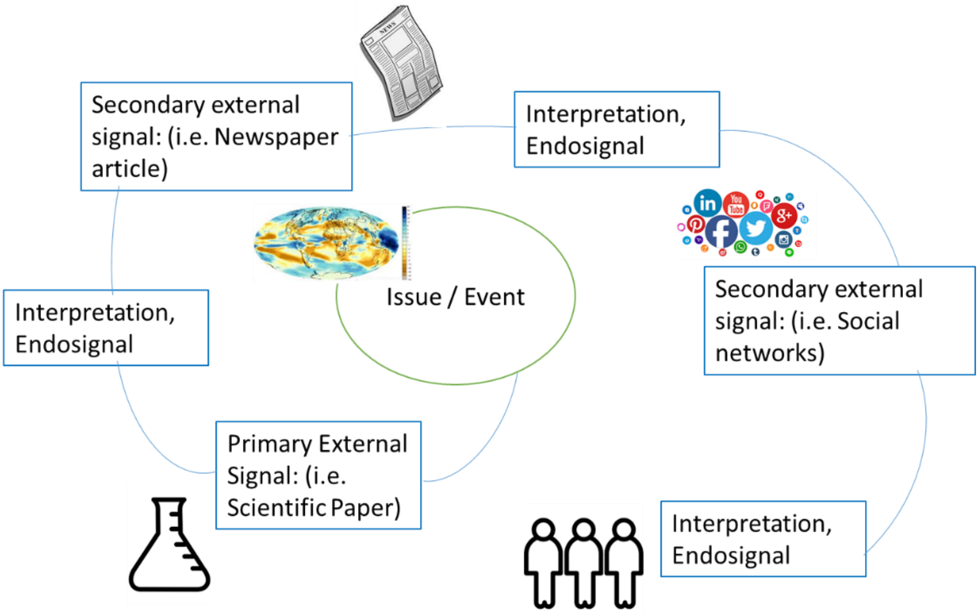


Figure 2. The process of signification in which the diffusion of signals in time is appreciated.

In terms of methodology previous research has focused on techniques for data collection, and text mining or semantic approaches for data processing. In most studies, data assessment and filtering in terms ofeffectiveness or relevance remains reliant on the manual efforts of the experts [14].

Detecting weak signals is a subtle process of observation and analysis because the information regarding these signals is encoded in many data sources. For this reason, any expert system built for the detection of these signals will always contain processes that consume considerable memory and time resources, and should therefore be optimized to the maximum. Most of the studies up to now are qualitative, use a single source, are related to a specific topic or remain relianton the efforts of experts.

This work describes a quantitative analysis of multiple sources and techniques in a system that can be applied to any topic, depending only on the documents used as its input. This approach offers a new opportunity to automatize processes and help experts.

3. System implementation for weak signal detection

3.1. System definition and propositions

A system has been created taking into account the three variables of the Hiltunen semiotic model, including the interpretation variable which is related to the type of information source. The first step was to create annual internal repositories of information obtained from numerous online sources. These repositories were made up of scientific articles, journalistic articles and tweets. Once the repositories were obtained,a system was required to manage a large volume of information from which to extract the necessary knowledge for decision making.

In the first place, an algorithm is required to collectand store a large number of documents from different sources related to a specific topic of interest, in numerous documentary databases, one for each year of the period of study to be able to perform the analysis. Normally, when organizations want to start using business intelligence techniques to take future decisions [9], they already possess internal databases with information about their activities, but in this case, since it isexternal data, it is necessary to create this compilationas a preliminary step.

Secondly, the information in this database must be transformed to store only the useful data, in a manageable format, in a repository or data warehouse. To do this, a second algorithm was created to extract the information, to transform it and to load it into the said data store.

Thirdly, once the information is stored in the data warehouse, another algorithm selects and transforms the information to be treated by mathematical models,through text mining, and event detection from time series data. Fourthly, a quantitative technique for the detection of candidates for weak signals in texts was applied. Finally, natural language processing was appliedto discard false positives.

In short, this study consists of the design of a system that quantitatively measures future signals using text mining techniques, that is, creating a tool that facilitates the analysis of an expert on a sector, assumingthe following premises:

Premise 1. Key words with many occurrences in a collection of documents are important.

Premise 2. Recent occurrences of keywords are more important or relevant than past appearances.

Premise 3. The first two propositions are true forany type of external data source used.

Premise 4. More reliable results are obtained when different data sources are used, taking into account a significance oriented towards the future.

Premise 5. False positives can be discarded by including natural language processing.

3.2. Description of the methodology applied

The proposed experiment is the design and implementation of a system focused on the search for weak signals related to artificial intelligence. The online sources that have been selected for this experiment are in English.

A key stage in this system is the Knowledge Discovery in Databases (KDD) methodology, which refers tothe process of discovering knowledge and potentially useful information within the data contained in a certain information repository [15]. When an organizationuses a KDD system to discover potentially useful information, they usually start the process with their owndatabases of data from their activities, even though thisinfo needs to be transformed in the KDD system. However, in this case a previous stage is necessary to generate a database with the collected information from several online sources: scientific journals, newspapers and social networks.

Figure 3 illustrates the steps of the system designed and implemented for our experiment. The process consists of 6 stages, which are:

1. Collection and integration of information.
2. ETL (Extract, transform and load).
3. Selection, processing and transformation.
4. Text mining focused on detecting weak signals.
5. Semantic Analysis.
6. Interpretation and evaluation.

3.3. Phase 1: Collection and integration ofinformation

For a company, knowledge can come from many different sources, such as information systems, reports, the internet, corporate databases, customers, suppliers or government agencies, or even the knowledgeof employees. During this phase, a preliminary study of the sources from which the necessary informationis obtained was carried out. These sources would be considered as technical and scientific types, journalistic sources and information stemming from social networks.

Once the online sources were analyzed, three were chosen due to several fundamental factors that were taken into account: the data managed by these sources had to be relevant within its type of source, but at the same time, data needed to be in a manageable format and thus easily extracted for storage. The sources selected with these criteria were: Direct Science as scientific source, the New York Times as journalistic source, and finally, Twitter as a social network.

For scientific and technical articles, the following information was extracted: title, author, abstract, keywords, content, conclusions and year of publication.

To collect the information, a Python algorithm was developed to extract data from HTML documents and tweets and to store it in NoSQL-type databases. The extraction of scientific and journalistic documents was possible thanks to Beautiful Soup, which is a Python library designed to analyze HTML documents and to perform the extraction of useful information from websites. However, for the extraction of tweets, Twitter API facilitated this action.

The NoSQL documentary database chosen in this study is MongoDB, an open-source document-oriented database which stores data structures in documentssimilar to JSON, allowing for maximum efficiency inthe integration of data in experiments of this kind [39].Figure 4 shows a graph with the number of documents obtained per year from each of the three sources, related to the topic of artificial intelligence: over 43.000 science papers extracted from ScienceDirect, 1.820 journalistic articles from the New York Times and 58.000 tweets. There are over ten times as many scientific journals as journalistic articles which means that newspapers filter the contents from scientific journals in the propagation of information [17].

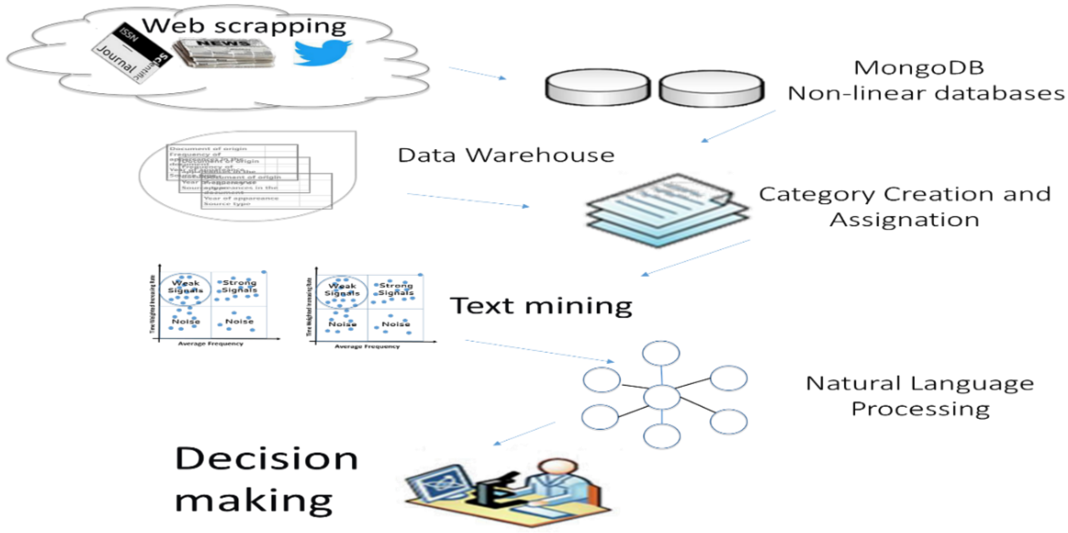


Figure 3. Proposed system for the detection of weak signals

3.4. Phase 2: Extract, transform and load

Once the internal repositories are created, it is necessary to design and implement a data warehouse which supports the storage of a large volume of information. It is important to take into account that this data warehouse must be issue-oriented and variable in time, and that the stored information cannot be lost [18]. At the same time an algorithm needs to be programmed to extract, transform and load the information in the warehouse.

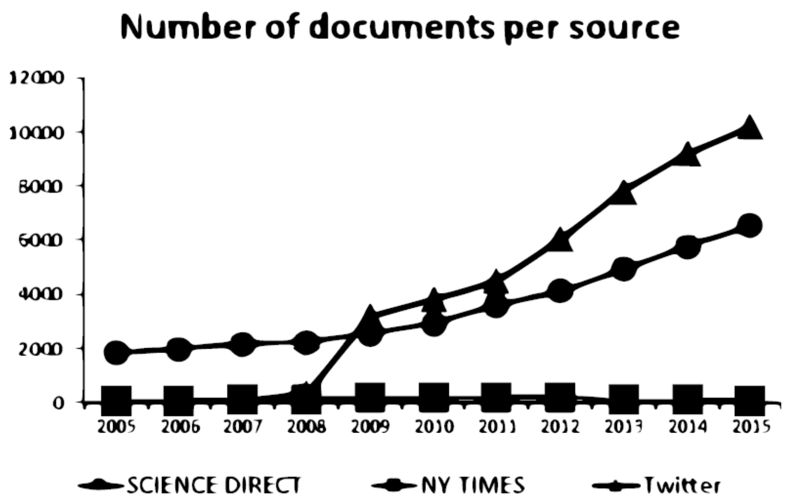


Figure 4. Number of documents per year and source.

The words extracted from the documents collected in the internal repositories of the information system are stored in the data warehouse created, as well as the following properties related to these words: document of origin, frequency of appearance in the document, year of appearance and source to which the document belongs.

Having created the extraction and storage algorithm, it was used to list the documents stored by years and to detect the frequency of appearance of words in the document. This algorithm performs previous checks, to avoid inserting numbers, strange symbols and stop words, i.e., words that contribute neither to the semantics nor to the meaning of the text, and, therefore, can be dismissed as candidates for weak signals. These stop words, generally the most common words in a language, contribute little to overall meaning and are filtered out before further processing of the text. Natural Language Toolkit (NLTK), a leading platform for building Python programs to work with human language data, was used for the elimination of stop words[2]. Some of the stop words included on the algorithm list were: a, able, about, above, according, accordingly, across, actually, after, again, against, all, allows, almost, along, already, also, although, always, am, among, an, and, another, any, anybody, anyhow, alone, be, became, because, become, been, before, behind, being, below, beside, better, etc.

To reduce the time required to insert the words in the database, a stemming phase is performed at this point. Stemming is basically the process of eliminating affixes (suffixes, prefixes, infixes, circumfixes) from a word in order to obtain a word stem. Snowball, whichis a small string processing language designed for creating stemming algorithms, was used to implementthe improved Porter algorithm [38]. This allows us to group words from the same family, reducing the number of insertions in the database. The chosen word to represent this group of words is the one that appears most frequently in the creation phase of the database.

3.5. Phase 3: Selection, processing andtransformation

In order to detect all possible weak signals, every word in the document apart from stop words and stemsis considered a keyword for the experiment.

The next step was to assign categories to each of these keywords. Categories were extracted from thoseprovided in the scientific articles. In this way, every word is included in every category of the articles in which the word appears. A keyword can thus belong to more than one category. Categories included Applied Soft Computing, Brain Research, Biomedical, Nuclear Energy, Expert Systems, Computer and Structures among many others.

Other studies use a standard list of categories and subcategories of different topics, such as for example, the one provided by UNESCO (http://whc.unesco.org/en/factors/). However, creating a dynamic list of categories depending on the sector and the actual contentsof the newest publications, will help in the detection ofthe categories of future signals [23].

3.6. Phase 4: Text mining focused on weak signaldetection

Weak signals in general bear “information on potential change of a system towards an unknown direction” [28]. They are found under a layer of noise which makes them difficult to detect, but once they have passed this layer, they are said to become strong signals. Wild cards are surprising events in which the weak signals become strong, and these events can be predicted with the help of weak signal analysis [4]. Asthe nature of wild cards is that they happen rapidly, thetime to react is short. Figure 5 shows the relation of weak signals to wild cards.

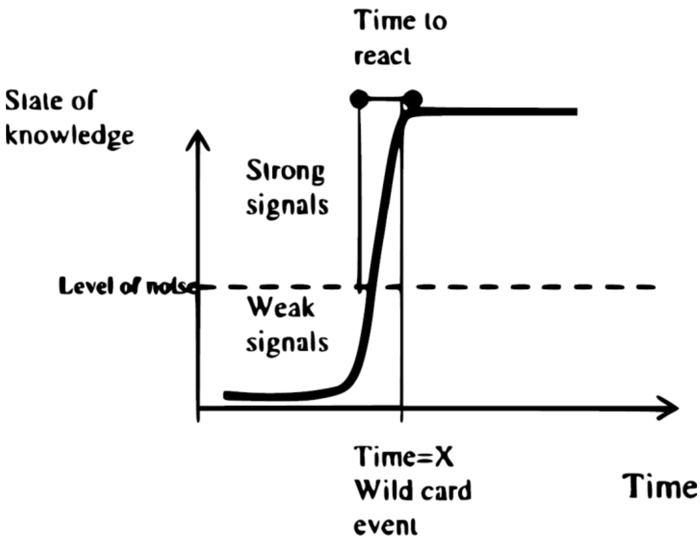


Figure 5. Relation of weak signals to wild cards

Text mining is a variation of data mining applied tothe process of obtaining high-quality information froma text [19]. The process is characterized by structuring the input data, constructing analysis models and analyzing the results obtained. The big difference withrespect to data mining is that patterns in text mining are obtained by processing natural language rather than by processing structured databases. In addition, in thiscase the document databases have been created for thespecific purpose of the study.

In the past few years there have been several studies which use data-mining techniques to extract patterns from time series data [13]. In most cases, the rule for determining when a sensor reading should generate anevent is poorly understood, giving rise to a problem when identifying the time points at which the behavior change occurs. In statistics literature this has been called the change-point detection problem. The standard approach has been to (a) determine the numberof change-points that are to be discovered, and (b) decide the function that will be used for curve fitting in the interval between successive change-points. There are several algorithms such as Batch, Incremental, relative density ratio estimation [26] or direct density ratio estimation [21]. The Batch algorithm works better when data collection precedes the analysis, but all of them lose accuracy when the signal-to-noise ratio de creases. In this algorithm, the data is divided in different segments limited by change-points, and for each ofthe segments the system should select a function that best describes the data.

In order to detect if a word indicates a weak signal, it is necessary to study the increase or decrease of appearances in all the documents and in the document under scrutiny. The three dimensions of the Hiltunen semiotic model are taken into account. To treat “signal” and “issue” dimensions, a curve with increment ratios of appearances is generated. In a series of observations of a particular variable at different moments, its tendency, cyclical factor, seasonality or irregular movement can all be measured. In this case, the trend or increment ratio of the time series is studied becauseit reflects the long-term evolution of the time series. Todo this more accurately the analytical method of minimum squares is used, with the objective of obtaining values of R2 as close as possible to 1, using polynomial, logarithmic or exponential functions.

The "signal" dimension of a future sign is related to the visibility of the future signal. To study the visibility of a sign, the absolute frequency of occurrence of each word in a set of documents (from the three different sources and for every year) is measured. This is defined as the Degree of Visibility (DoV) of the keyword i in period j, as can be seen in Equation (1):

$$DoV_{ij} = \frac{TF_{ij}}{NN_j} \times \{1 - tw \times (n - j)\} \quad (1)$$

where TF_{ij} is the total number of occurrences of theword $\boldsymbol{i}$ in period $\boldsymbol{j}$ (considering all the documents), NN_j is the total number of documents in period $\boldsymbol{j}$, n is the number of periods and tw is a time weight. According to the proposal that new occurrences are the most relevant, the tw it has been defined as 0.05 by a groupof experts [41].

The "issue" dimension indicates the degree of diffusion of the subjects related to weak signals. This dimension is directly related to the frequency of occurrence of each word in each document, since this frequency is generally adopted to measure how general aterm is in a collection of textual information [33].

To measure the Degree of Diffusion (DoD) of theword i in period j, Equation (2) has been used:

$$DoD_{ij} = \frac{DF_{ij}}{NN_j} \times \{1 - tw \times (n - j)\} \quad (2)$$

where DF_{ij} is the number of documents in which theword i appears in period $\boldsymbol{j}$.

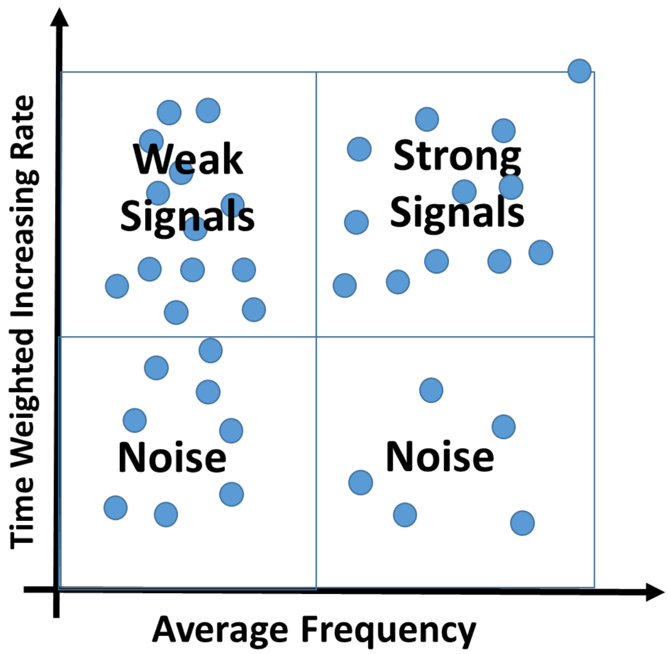


Figure 6. Structure of a keyword map.

In both signal and issue determination, future signals that have the possibility of being weak signals have, from a quantitative point of view, an absolute low occurrence frequency but a high trend fluctuation (a geometric mean of high DoD and DoV but a low occurrence frequency).

These equations provide the average of the increase ratios (DoD and DoV) of each word found in the complete set of documents analyzed, with respect to the frequencies obtained by years. With these data, two graphs are generated: the first is the “Keyword Issue Map”, which represents a map of absolute occurrences of keywords in all the documents (related to DoD); and the second is the “Keyword Emergence Map”, to represent the number of documents in which each keyword appears (related to DoV).

The two keyword map graphs are composed of four quadrants as seen in Fig. 6. Above a threshold weighted increment ratio there are two areas: that of “Strong Signals” above a mean frequency threshold, and that of “Weak Signals” below that threshold on the Y axis, results are considered as noise.

Interpretation is an activity in which endosigns are formulated in the mind of an analyst, based on the exosigns of the issue. Based on this “interpretation” dimension of a sign, a new issue might emerge.

As the first generated exosignals after a new discovery are usually the research results published in the form of an article in a scientific journal, this source has been used to measure the interpretation dimension of a sign.

The Shape of Science is an information visualization project whose aim is to reveal the structure of science. Its interface has been designed to access the bibliometric indicator database of the SCImago Journal portal.

All H-index values from all the documents where a word *i* appears have been added to calculate the Degree of Transmission (DoT) of that word. This measurement is expressed in Equation (3).

$$DoT_i = \sum Hindex_{journal} \tag{3}$$

Then, all DoT values were normalized, and expressed graphically in both DoD and DoV graphs with a different size for each dot, where each one of the dots represents a different word.

3.7. Phase 5: Natural language processing

As a result of the generation of these two graphs (with the addition of the interpretation dimension), a list of words is generated as weak signal candidates.

The next step is to determine the meaning of text selections such as word sequences, a process known as semantic analysis. To this end the context of the words identified as weak signal candidates needs to be analyzed, to avoid the detection of many false positives. A multi-word expressions analysis is a deeper approach that will reveal more accurate information about the detected term of interest. Previous works such as [37] and [42] show the importance of multi-word expressions as a text mining technique.

For this study, a natural language processing tool, namely a multi-word analysis, was applied to the list of candidates for weak signals. The words immediately preceding and immediately following the word identified as a weak signal in all documents were considered. The combinations in which these additional words are in the list of standard stop words were discarded.

3.8. Phase 6: Interpretation and evaluation

Finally, several outputs have been obtained from theexecution of the model:

(i) The Keyword Issue Map with words detected as candidates for weak signals, with their topic categories, ranked by their Degree of Diffusion (DoD) and Transmission (DoT).
(ii) The Keyword Emergence Map with words detected as candidates for weak signals, with theirtopic categories, ranked by their Degree of Visibility (DoV) and Transmission (DoT).
(iii) The table of words detected as candidates for weak signals that appear in both maps; which, as they fulfill the three conditions of a weak signal according to the Hiltunen model, are more likely to point to real weak signals.
(iv) The multi-word study of the words detected inthe previous step, providing more accurate information, and discarding false positives.

This information is given to experts, entrepreneurs,companies and other interested parties in the topic under study, to facilitate their decision making.

4. Results of the experiment applied to the futureof artificial intelligence

4.1. Keyword issue map

Taking as input the documents related to the topic of artificial intelligence as previously mentioned in Fig. 4, a quantitative analysis was launched with the system described.

As previously stated, words that are candidates as weak signals are topics that have an abnormal pattern but are rarely detected. From the perspective of quantitative analysis, these words are likely to have low ab solute document frequency but a high range of fluctuation in the increase of document frequency. Conversely, topics with low absolute document frequencyand high increase rate of document frequency could be strong signals because they are considered to be important and well spread.

The method computed the DoD of each word basedon the document frequency of the word per news article. DoT is also measured by the impact of the journals where the word appears. The keyword issue map is generated by using the average time-weighted increasing rate of document frequency in geometric mean, and the absolute average document frequency of eachword.

Table 1 shows some of the detected words in the Key Issue Map together with the frequency, increase ratio, DoT and some of the categories from the methodology described before. This table shows the keyword analyzed, its related stem, absolute frequency of appearances, and global increase ratio in the whole period; the DoT, and the categories in which this keyword hasbeen detected.

Table 1. Key issue map – degree of diffusion

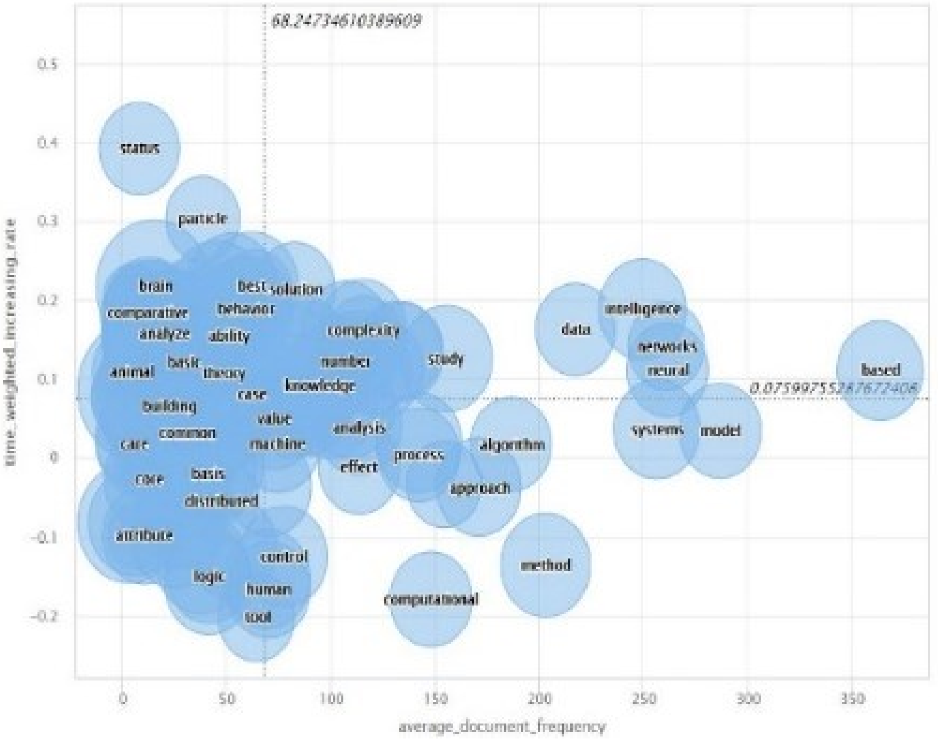
Keyword	Stem	Frequency	Increase ratio	DoT	Category
ability	abil	68.18	0.0394	71.20	Applied soft computing, knowledge based systems
adaptive	adapt	104.64	0.2469	64.70	Applied soft computing, engineering applications
analyze	analyz	26.55	0.0928	57.90	Artificial intelligence, expert systems
animal	anim	6.5455	0.1886	60.40	Artificial intelligence
applications	applic	201.82	0.1323	59.90	Engineering software, medicine, agriculture
approximation	approxim	35.64	−0.0665	62.38	Applied soft computing, approximate reasoning
attribute	attribut	20.64	0.0642	62.93	Knowledge based systems
basic	basic	50.27	0.2087	67.42	Applied soft computing, neurocomputing
basis	basi	59.18	0.1194	69.45	Applied soft computing, neurocomputing
behavior	behavior	85.90	0.0005	68.03	Applied soft computing, robotics
best	best	91.45	0.1946	68.39	Mathematics, energy conversion and management
biological	biolog	59.64	0.0252	67.30	Astronautica, applied soft computing
bottom	bottom	10.27	0.1372	75.00	Artificial intelligence
brain	brain	36.27	0.2525	81.84	Brain research, cognitive systems research, neuroscience
building	build	54.63	0.1861	63.63	Applied mathematics, energy and buildings
capabilities	capabl	26.45	0.2244	64.01	Expert systems with applications
capture	captur	19.36	0.0965	63.21	Engineering applications, expert systems
care	care	10.81	0.0965	60.02	Medicine
classification	classif	140.09	0.1647	55.01	Applied soft computing, medicine
clinical	clinic	30.55	0.1809	68.23	Medicine, biomedicine, expert systems
code	code	13.18	0.1090	57.24	Computing science
cognitive	cognit	82.63	0.0919	69.25	Animal behaviour, applied soft computing, medicine
colony	coloni	43.45	0.1135	57.67	Mathematics, applied soft computing, neurocomputing

In addition, Figure 7a shows a graphic view of the Key Issue Map with the candidate words for weak signals in the first quadrant. The size of the dots correspondswith the DoT variable.

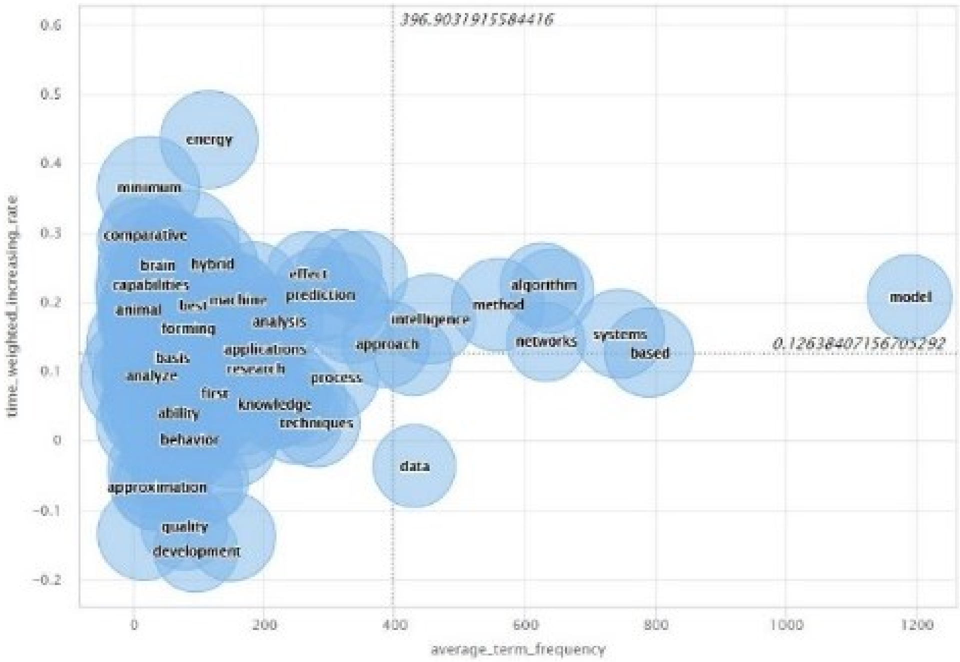
4.2. Keyword emergence map

The method computed DoV of each word based on the document frequency of the word per news article. DOT is also measured by the impact of the journals where the word appears. The keyword emergence map is generated using the average time-weighted increasing rate (geometric mean) and the absolute average term frequency of each word.

Table 2 shows some of the detected words in the Key Emergence Map together with the frequency, increase ratio, DoT and some of the categories according to the methodology described in the previous section.



(a)



(b)

Figure 7. (a) Keyword issue map and (b) Keyword emergence map.

In addition, Figure 7b shows a graphic view of the KeyEmergence Map, with the candidate words for weak signals in the first quadrant. The size of the dots corresponds with the DoT variable.

Table 2. Key emergence map – degree of visibility

Keyword	Stem	Frequency	Increase ratio	DoT	Category
ability	abil	68.18	0.0393	71.20	Applied soft computing, engineering applications
adaptive	adapt	104.64	0.2469	64.68	Applied soft computing, neurocomputing
analysis	analysi	222.27	0.1709	63.85	Applied mathematics and computation, building
analyze	analyz	26.55	0.092	57.87	Expert systems with applications
animal	anim	6.55	0.1886	60.40	Biology
applications	applic	201.81	0.1322	59.98	Engineering software, medicine
approximati	approxim	35.63	−0.0664	62.38	Applied soft computing, neurocomp
onaspects	aspect	33.45	0.1184	64.28	Expert systems with applications
attribute	attribut	20.63	0.0641	62.93	Expert systems with applications
basic	basic basi	50.27	0.2087	67.42	Applied soft computing, expert
		59.18	0.1193	69.45	systemsApplied soft computing,
basisbest	best	91.45	0.1946	68.39	expert systems
biological	biolog	59.63	0.0253	67.30	
bottom	bottom	10.27	0.1373	75.00	Mathematics and computation, energy conversionAstronautica, biology
					Expert systems with applications
brain	brain	36.27	0.2525	81.84	Brain research, cognitive systems,
building	build	54.63	0.1861	63.63	biologyApplied mathematics and
capabilities	capabl	26.45	0.2244	64.01	computation Expert systems with
capture	captur	19.36	0.0965	63.21	applications
care	care	10.81	0.0965	60.03	Engineering applications, environmental
case	case	120.63	0.1337	56.67	sciencesMedicine
classification	classif	140.09	0.1647	55.01	Informatics, medicine, neurocomputing
					Medicine, signal processing
clinical	clinic	30.54	0.1809	68.23	Biomedicine, expert systems
code	code	13.18	0.1090	57.24	Medicine

4.3. Detected candidate for weak signal terms

If in both the Keyword Issue Map and Keyword Emergence Map graphs a word is detected, it means that this word fulfills the criteria to be related to a weak signal according to Hiltunen model. Thereare 92 words appearing simultaneously in both maps, which are: ability, adaptive, analyze, animal, applications, approximation, attribute, basic, basis, best, biological, bottom, brain, building, capabilities, capture, care, classification, clinical, code, cognitive, colony, common, comparative, competitive, conditional, constraint, cooperative, core, current, dataset, definition, design, diagnosis, differences, education, energy, engineering, experimental, forming, future, generation, growth, health, history, hybrid, identity, impact, inspired, internet, limits, list, making, mathematical, measurement, methodology, minimum, multiple, non-linear, number, order, particle, patient, point, potential, presence, principle, procedural, quantitative, rate, regression, review, role, scale, science, self, social, solution, solving, specification, status, styles, success, superior, support, test, transfer, transform, universal, well and yield. Full table is available in [12].

4.4. Multi-word analysis

A multi-word analysis was performed with these words in order to discard false positives and determinemore accurate word expressions related to weak signals. Table 3 shows the results for expressions relatedto the words brain, clinical and cooperative.

Table 3. Multi-word analysis

Brain		Clinical		Cooperative	
Word	Perc.	Word	Perc.	Word	Perc.
Human	26.23	Decision	32.44	Control	18.54
Size	11.81	Practice	8.71	Learning	14.42
Artificial	6.10	Diagnosis	7.44	Robotics	10.98
Body	4.55	Intelligence	6.60	Mobile	9.38
Tumors	4.45	Reasoning	5.90	Intelligent	8.92
Activity	4.36	Trials	5.20	Coevolution	8.01
Function	4.26	Case	5.06	Systems	6.64
Development	4.16	System	4.92	Tasks	5.49
Imaging	4.07	Data	4.78	Strategy	5.26
Functions	3.97	Care	4.63	Hybrid	3.89
Informatics	3.68	Disease	4.49	Network	3.66
Mind	3.48	Important	4.35	Matching	2.75
Processing	3.39			Clustering	2.06
Abstract	3.29				
Particular	3.19				
Computer	3.10				
Intelligence	3.00				
Research	2.90				

Results in artificial intelligence show expressions related to weak signals such as "human brain", "artificial brain", "clinical clustering", "cooperative coevolution", and "cooperative mobile", that can be investigated according to the categories detected previously in the study, for example "Biology" or "Neuroscience". These results are forwarded to the experts in order to help their decision making.

5. Evaluation of the results

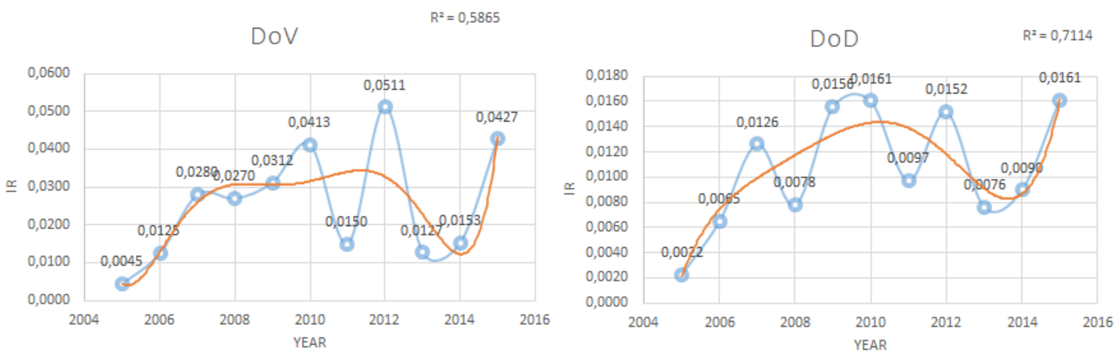
This work provides the architecture of a system todetect weak signals in an efficient, automatic, multi-source, and quantitative manner, and is applicable toevery subject. It thus differs from other works with aqualitative approach about a specific model for a specific subject [24] or [29], or which are quantitative butnot multi-source [41], or which remain reliant on themanual efforts of the experts [14]. Quantitative analysis of multiple sources and techniques open a new opportunity to, at the very least, help experts in their job.Most of the previous works use standard databases of keywords and categories, but in this study everyword is considered to be a keyword, and the set of categories is obtained from the categories of the documentsunder study. In this way, new terms that are not in anydatabase can be detected, providing a more accurate detection.

A previous work [41] validated the results using thesame set of documents which were the input of the system, so obviously, they confirmed their hypothesis.In other applications, a dataset is divided into training, test and validation sets. However, as weak signals havea low absolute frequency, a weak signal detected in thetest set would rarely be also present in the validation set. This evaluation process was therefore discarded.

Using different sources also provides a better mechanism to test the results, in comparison with independent results of documents from only a single source- type. In this study, a ranking of the results taking issue and signal components, and another with the three components, show similar conclusions. Further studies should also determine the weight of different sources depending on the subject or sector to study. These weights can also change through time.

It has already been established that weak signals have an abnormal pattern. From the view of quantitative analysis, these words are likely to have low absolute document frequency but high range of fluctuation in the increase of document frequency as has beenstated previously.

Figure 8 shows the DoV and DoD pattern of the word brain, detected as a term related to a weak signal in the subject of artificial intelligence. The graph shows the expected abnormal pattern with a low frequency. This is a condition for a weak signal, but othermethods need to be added to discard false positives andto generate more useful information. Evaluation of theresults included the same analysis for every keyword.



(a) (b)

Figure 8. (a) DoV and (b) DoD pattern of the word brain.

The first two propositions of the study are that words with many occurrences are more important, and recentoccurrences are more relevant than less recent ones. These propositions are common in previous works such as [41] and also this study. Other works [14,36] or [20] also agree with the value of recent occurrencesfor detecting future signals.

Based on the Hiltunen theoretical model [25], the issue and signal components are applicable to every document (the third proposition), independently ofthe source. The interpretation component depends on the source in which the data originate. As previously stated, an exosignal in the shape of a scientific paper can be interpreted by a journalist, becoming an endosignal for him, who will later codify a journalistic article in the form of a secondary exosignal.

The fourth proposition of this study was that understanding this process can provide more reliable results. Unlike previous works [41] in which only web news are used, this study provides a new variable called Degree of Transmission based on the impact of the scientific papers as primary sources. The Hiltunen model explains that the interpretation component of the sign is also relevant in a weak signal.

The last proposition is that many false positives can be avoided using natural language processing. In this study, a multi-word expressions analysis shows more accurate and interesting information about the detected term of interest, than an analysis based only on single words.

6. Conclusions and future work

In this paper, a quantitative approach to design a system for detecting weak signals of the future is described, with an analysis for artificial intelligence. In contrast to current work focusing on structured information sources, or using only a single type of data source, the detection of these signals is quantitatively based on heterogeneous and unstructured documents of various kinds (scientific journals, newspaper articles and social networks).

In conclusion, this work suggests the possibility of quantifying the detection process of weak signals by text mining using multiple sources. The proposed method can detect weak signals more efficiently than human experts when dealing with the massive textual information available due to the exponential increase in new documents related to any subject.

Although adding more natural language processing will improve the obtained results, the system can help business experts in the recognition of new driving factors of their markets and in the development of new opportunities.

In future work, new topics will be added to evaluate the system. Moreover, in order to obtain more accurate results, other natural language processing tools should be added such as parts-of-speech tagging, bag of words recognition, regular expressions or sentiment analysis. The hardware used for the experiment was an Intel i7 processor, with 2 GB of RAM and the operating system of Ubuntu Server 18.04. For future research, hardware improvement with GPUs will be implemented to obtain a more efficient system.

Acknowledgements

This work is partially supported by EIT ClimateKIC of the European Union (project Accelerator – TC2018B-2.2.5-ACCUPV-P066-1A) and Erasmus InnoCENS (573965-EPP-1-2016-1-SE-EPPKA2-CBHE-JP).

This research is also part of the PhD programme of the Departamento de Ingeniería Electrónica – Department of Electronic Engineering of the Universitat Politècnica de València (Spain).

References

1. H.I. Ansoff, Managing strategic surprise by response to weak signals, *California Management Review* **18**(2) (1975), 21–33. doi:10.2307/41164635.
2. S. Bird, E. Loper and E. Klein, *Natural Language Processing with Python*, O'Reilly Media Inc., 2009.
3. S. Brown and K. Eisenhardt, *Competing on the Edge: Strategy as Structured Chaos*, Harvard Business School Press, Boston, 1998, p. 23.
4. A. Cooper, C. Voigt, E. Unterfrauner, M. Kravcik, J. Pawlowski and H. Pirkkalainen, Report on weak signals collection, in: *TELMAP*, European Commission Seventh Framework Project (IST-257822), Deliverable D4.1, 2011, pp. 6–7.
5. J. Dator, Futures studies as applied knowledge, in: *New Thinking for a New Millennium*, R. Slaughter, ed., Routledge, London, 1996, www.futures.hawaii.edu/dator/futures/ appliedknow.html.
6. J. Dator, Universities without quality and quality without universities, *On the Horizon* **13**(4) (2005), 199–215, www.futures.hawaii.edu/dator/education/DatorOzQual.htm. doi:10.1108/ 10748120510627321.
7. J. Duan, M. Zhang, W. Jingzhong and Y. Xu, A hybrid frame-work to extract bilingual multiword expression from free text, *Expert Systems With Applications* **38**(1) (2011), 314–320. doi:10.1016/j.eswa.2010.06.067.
8. R. Eckhoff, M. Markus, M. Lassnig and S. Schon, Detecting weak signals with technologies: Overview of current technology enhanced approaches for the detection of weak signals, *Int. J. Trends Econ. Manag. Technol.* **3** (2014), 1–7.
9. L. Fink, N. Yogev and A. Even, Business intelligence and organizational learning: An empirical investigation of value creation processes, *Information and Management* **54**(1) (2017), 38–56. doi:10.1016/j.im.2016.03.009.
10. M.A. Fischler and O. Firschein, *Intelligence: The Eye, the Brain and the Computer*, Addison-Wesley, 1987, p. 221.
11. M. Godet, *From Anticipation to Action, a Handbook of Strategic Prospective*, UNESCO Publishing, 1994, p. 59.
12. I. Griol-Barres, S. Milla and J. Roig, DoV DoD in artificial intelligence weak signals detection, *Fig share* (2019), Dataset. doi:10.6084/m9.figshare.7583228.v2.
13. V. Guralnik and J. Srivastava, Event detection from time series data, in: *KDD '99 Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, California, USA, 1999, pp. 33–42. doi:10.1145/312129.312190.
14. K. Haegeman, E. Marinelli, F. Scapolo, A. Ricci and A. Sokolov, Quantitative and qualitative approaches in future-oriented technology analysis (FTA): From combination to integration, *Technol. Forecast. Soc. Chang.* **80** (2013), 386–397. doi:10.1016/j.techfore.2012.10.002.

15. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, USA, 2001, p. 236.
16. E. Hiltunen, The future sign and its three dimensions, *Futures* **40**(3) (2008), 247–260. doi:10.1016/j.futures.2007.08.021.
17. L. Ilmola and O. Kuusi, Filters of weak signals hinder foresight: Monitoring weak signals efficiently in corporate decision-making, *Futures* **38** (2006), 908–924. doi:10.1016/j. futures.2005.12.019.
18. W.H. Inmon, *Building the Data Warehouse*, 4th edn, John Wiley, 2005, p. 156.
19. C.S. Ishikiriyama, D. Miro and C.F. Simoes Gomesa, Text mining business intelligence: A small sample of what words can say, *Procedia Computer Science* **55** (2015), 261–267. doi:10. 1016/j.procs.2015.07.044.
20. K. Jung, *A Study of Foresight Method Based on Text Mining and Complexity Network Analysis*, KISTEP, Seoul, 2010.
21. Y. Kawahara and M. Sugiyama, Change-point detection in time-series data by direct density-ratio estimation, in: *Proceedings of the 2009 SIAM International Conference on Data Mining*, 2009, pp. 389–400. doi:10.1137/1.9781611972795.34.
22. R.A. Khan, KDD for business intelligence, *Journal of Knowledge Management Practice* **13**(2) (2012), 134.
23. J. Kim, M. Han, Y. Lee and Y. Park, Futuristic data-driven scenario building: Incorporating text mining and fuzzy association rule mining into fuzzy cognitive map, *Expert Systems With Applications* **57** (2006), 311–323. doi:10.1016/j.eswa.2016.03.043.
24. R. Koivisto, I. Kulmala and N. Gotcheva, Weak signals and damage scenarios – Systematics to identify weak signals and their sources related to mass transport attacks, *Finland Technological Forecasting and Social Change* **104** (2016), 180–190. doi:10.1016/j.techfore.2015.12.010.
25. O. Kuusi and E. Hiltunen, *The Signification Process of the Future Sign*, Finland Futures Research Centre ebook 4/2007, 2007, p. 7.
26. S. Liu, M. Yamada, N. Collier and M. Sugiyama, Change-point detection in time-series data by relative density-ratio estimation, *Neural Networks* **43** (2013), 72–83. doi:10.1016/j.neunet. 2013.01.012.
27. M. Mannermaa, *Tulevaisuuden Hallinta Skenaariot Strate- giatyoskentelyssa (Managing the Future, Scenarios in Strategy Work)*, WSOY, Porvoo, 1999, p. 227.
28. S. Mendonca, M.P. Cunha, J. Kaivo-Oja and F. Ruff, Wild cards, weak signals and organizational improvisation, FEUNL Working Paper No. 432, 2003.
29. R. MohamadiBaghmolaei, N. Mozafari and A. Hamzeh, Continuous states latency aware influence maximization in social networks, *AI Communications* **30**(2) (2017), 99–116. doi:10. 3233/AIC-170720.
30. G.T. Molitor, Molitor forecasting model: Key dimensions for plotting the patterns of change, *Journal of Future Studies* **8**(1) (2003), 61–72.

31. I.O. Nikander, Early warnings, a phenomenon in project management, Dissertation for the degree of Doctor of Science in Technology, Helsinki University of Technology, 2002, p. 196.
32. C.S. Peirce, Some consequences of four incapacities, *Journal of Speculative Philosophy* **2** (1868), 140–157, https://www. jstor.org/stable/i25665647.
33. G. Salton and C. Buckley, Term-weighting approaches in automatic text retrieval, *Information Processing and Management* **24**(5) (1988), 513–523. doi:10.1016/0306-4573(88)90021-0.
34. P.J.H. Schoemaker, G.S. Day and S.A. Snyder, Integrating organizational networks, weak signals, strategic radars and scenario planning, *Technol. Forecast. Soc. Chang.* **80** (2013), 815– 824. doi:10.1016/j.techfore.2012.10.020.
35. D. Thorleuchter, T. Scheja and D. Van den Poel, Semantic weak signal tracing, *Expert Systems with Applications* **41** (2014), 5009–5016. doi:10.1016/j.eswa.2014.02.046.
36. D. Thorleuchter and D. Van den Poel, Idea mining for web- based weak signal detection, *Futures* **66** (2015), 25–34. doi:10.1016/j.futures.2014.12.007.
37. Y. Tsenga, C. Lin and Y. Linc, Text mining techniques for patent analysis, *Information Processing and Management* **43**(5) (2007), 1216–1247. doi:10.1016/j.ipm.2006.11.011.
38. P. Willett, The Porter stemming algorithm: Then and now, *Program: electronic library and information systems* **40**(3) (2006), 219–223. doi:10.1108/00330330610681295.
39. R. Yangui, A. Nabli and F. Gargouri, Automatic transformation of data warehouse schema to NoSQL data base: Comparative study, *Procedia Computer Science* **96** (2016), 255–264. doi:10. 1016/j.procs.2016.08.138.
40. S.H. Yoo, H.W. Park and K.H. Kim, A study on exploring weak signals of technology innovation using informetrics, *Journal of Technology Innovation* **17**(2) (2009), 109–130.
41. J. Yoon, Detecting weak signals for long-term business opportunities using text mining of web news, *Expert Systems with Applications* **39** (2012), 12543–12550. doi:10.1016/j. eswa.2012.04.059.
42. W. Zhang, T. Yoshida and X. Tang, Text classification based on multi-word with support vector machine, *Knowledge-Based Systems* **21**(8) (2008), 879–886. doi:10.1016/j.knosys.2008.03. 044.

Article

Detecting Weak Signals of the Future: A System Implementation Based on Text Mining and Natural Language Processing

Israel Griol-Barres, Sergio Milla, Antonio Cebrián, Huaan Fan and José Millet.

Sustainability 2020, 12, 7848; doi:10.3390/su12197848

Received: 25 August 2020; Accepted: 21 September 2020; Published: 23 September 2020

Authors	*Griol-Barres, I.; Milla, S.; Cebrián, A.; Fan, H.; Millet, J.*
Title	*Detecting Weak Signals of The Future: A System Implementation to detect new business opportunities based on Text Mining and Natural Language Processing*
Journal	*Sustainability*, **2020**, *12*(19), 7848
DOI	*10.3390/su12197848*
Publisher	MDPI, ST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLAND
ISSN	2071-1050
eISSN	2071-1050
Impact Factor	3.251 (2020), 2.576 (2019) – 3.473 (5 years)

JCR Category	Rank in Category	Quartile in Category
ENVIRONMENTAL SCIENCES SCIE	**124 of 274**	**Q2**
ENVIRONMENTAL STUDIES SSCI	**60 of 125**	**Q2**
GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY SSCI	**6 of 9**	**Q3**
GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY SCIE	**30 of 44**	**Q3**

CiteScore Highest	84% 110/704 Geography, Planning and Development
CiteScore 2020	3.9 (4.0 2021)
SJR 2020	0.612 Q1
SNIP 2020	1.242
HIndex	85
Main contributions	This chapter is a published article that includes a new experiment to detect weak signals of the future in the field of remote sensing. The system includes natural language processing techniques and a new dataset of about 50,000 documents in a period of more than ten years. Results are validated by industry experts. Improved execution times are obtained by introducing new parallel implementations on GPUs. New validation methodologies are introduced. This chapter addresses the Objective 2 of this thesis and develops tasks 1-7 and 9.

Article
Detecting Weak Signals of the Future: A System Implementation Based on Text Mining and Natural Language Processing
Israel Griol-Barres, Sergio Milla, Antonio Cebrián, Huaan Fan and José Millet.
Sustainability 2020, 12, 7848; doi:10.3390/su12197848
Received: 25 August 2020; Accepted: 21 September 2020; Published: 23 September 2020

3. Detecting Weak Signals of the Future: A System Implementation Based on Text Mining and Natural Language Processing

Abstract: Organizations, companies and startups need to cope with constant changes on the market which are difficult to predict. Therefore, the development of new systems to detect significant future changes is vital to make correct decisions in an organization and to discover new opportunities. A system based on business intelligence techniques is proposed to detect weak signals, that are related to future transcendental changes. While most known solutions are based on the use of structured data, the proposed system quantitatively detects these signals using heterogeneous and unstructured information from scientific, journalistic and social sources, applying text mining to analyze the documents and natural language processing to extract accurate results. The main contributions are that the system has been designed for any field, using different input datasets of documents, and with an automatic classification of categories for the detected keywords. In this research paper, results from the future of remote sensors are presented. Remote sensing services are providing new applications in observation and analysis of information remotely. This market is projected to witness a significant growth due to the increasing demand for services in commercial and defense industries. The system has obtained promising results, evaluated with two different methodologies, to help experts in the decision-making process and to discover new trends and opportunities.

Keywords: new sustainable business models; business intelligence; natural language processing; weak signals of the future; predictive models; text mining

1. Introduction

One of the biggest threats for academia, governments, entrepreneurs and companies is the continuous changes in expanding markets. In fact, companies very often show difficulties predicting these variations and acting on time [1].

Markets in general have proven to be unpredictable environments where making right decisions at the right time is very hard, but undoubtedly, doing so can be translated into better results for an organization.

In addition, social entrepreneurs contribute to creating social wealth by addressing social problems and enriching communities and societies. For this reason, an early discovery of a new problem or opportunity in the market could become vital for the success of the entrepreneurial venture [2].

The amount of data managed by any organization is increasing considerably every day. Therefore, these processes of discovering new business opportunities involve both data engineers, and data scientists who must implement automatic analytics that reduce execution times and requirements, while providing a high reliability. For this reason, both software and hardware requirements need to be implemented in the most efficient way.

1.1. Weak Signals

The process of scanning an environment requires the identification of future signals in the shape of events, trends, and their relationships, by analyzing information [3].

A weak signal is a type of future signal [4] which can be defined as the evidence of emerging changes in a continuous process of exploration of a specific environment [5]. Weak signals are events that are too novel to measure their future impact and to react properly to them [6].

These signals have seven common attributes: (1) an idea that will change the environment of an organization, (2) a novelty from the personal perspective of an expert, (3) difficult to detect among other events, (4) a threat or opportunity for the stakeholders, (5) often opposite to the opinion of a big group of experts, (6) it usually takes time before it becomes mainstream, and (7) is an opportunity to learn, grow and evolve [7].

Therefore, a weak signal is hardly discernible in the present but will become an important trend [8]. If one of these signals evolves to the point that it becomes relevant or a strong signal, it will reinforce or obstruct the strategy of an organization.

Although "weak signals" is a term that is becoming popular in research studies, there are other synonym terms in literature, such as "seeds of change" [9], "facts emerging" [10,11], "strategy signals" [12] and "early warning signals" [13]. The semiotic model of the future sign by Hiltunen [14], which follows the same structure of the semiotic model of the sign by Peirce [15], has become the standard to identify the dimensions of future signals, and to understand the process of their amplification.

Hiltunen model defines three different dimensions for future signals: "issue", "signal" and "interpretation". The signal is defined as "the real visibility of a signal", the issue is defined as "a variety of events", and the interpretation is defined as "the meaning that a receiver gives to the sign".

Information technology, thanks to its role in business intelligence, has become relevant in the detection of signals of the future [16]. Business intelligence is the knowledge applied to stored information, to interpret, forecast and act appropriately to the outside world [17]. In conclusion, there are three interrelated concepts: business intelligence as a discipline, data mining as a procedure, and the detection of weak signals of the future as an objective.

This analysis is widely used in many research works, such as, for example, the remote sensing sector [18–20]. Traditionally, the detection of future signals has been an activity carried out by experienced experts. Currently, the amount of available information has increased considerably, and experts are not able to effectively scan environments in a very deep way. In addition, advances in hardware technology have made it feasible to compute a vast amount of data in a short time [21].

1.2. Background and Related Work

The available studies about weak signals use specific types of sources which are scanned regularly. Most of them are qualitative or unsystematic analyses that are based on a specific model for specific environments. A couple of examples are the identification of signals of the future, related to terrorism or mass transport attacks [22] or master planning London's Olympic legacy [23]; sensor-based human activity recognition [24], mechanical fault prediction by sensing vibration [25], or a deep learning analysis to predict parking space availability [26]. Other examples are the influence of maximization in social networks [27], a model to examine what schools would be like if they were a new invention [28], or a deep analysis about charge prediction for criminal cases [29]. In general, these models cannot be used to scan other environments.

Some studies use the opinion of experts and stakeholders as an input for the detection of weak signals. Among those inputs are the feelings of known critics or the behavior of customers [30].

Other works emphasize the use of structured data that is either owned by an organization or accessible via web, such as, for example, a methodology to detect weak signals for strategic forecasting [31]. Weak signals are extracted from the internal repositories of institutions, or available online sources, where online texts are obtained. Another example identifies signals for a better environment understanding, by using a range of storylines with negative and positive valences [32]. New studies propose the use of text mining techniques from online sources, as web news to predict the future of solar panels [33] or documents from the Journal of the Korean Ceramic Society to scan the field of nanotechnology [34]. These are also some of the first research proposals that carry out quantitative analysis, but both use a single type of data. This quantification is focused on measuring the importance of the visibility of events.

In addition, there is an available study that is using keyword network analysis, betweenness centrality for convergence measurement, and minimum spanning tree (MST), which provides a new view to detect future trends analyzing the changes in a network of words [35], and a study based on text-mining techniques to design multidisciplinary group discussions and generate energy policies and technologies (EP + Ts) in South Korea for the future society, from a multidisciplinary perspective [36]. Regarding the triad of the future sign in the Hiltunen model, these studies focus on only one data source, ignoring the interpretation dimension.

Another available study creates a model to evaluate if free text expressed in natural language could serve for the prediction of action selection, in an economic context, modeled as a game [37]. In this study, a classifier is implemented in a model that considers personality attributes, but can only be used in a specific topic, and only uses pre-determined keywords.

The theoretical foundations and methodologies to detect weak signals are developing research areas, and there is a considerable margin of improvement. In other applications beyond identifying future signals, text mining includes natural language processing tools [8]. As the use of the extraction of multiword expressions is currently providing a good performance in other fields, better results could be expected in the field of future signals.

The process for detecting weak signals is complex, because there are a multitude of sources where they could be masked. Any system implemented for this task must be optimized in both hardware and software to the maximum.

This study describes a system to help experts in the detection of weak signals, by a quantitative analysis of multiple sources applied to the field of remote sensing, considering every word as a keyword, depending only on different types of text documents (unstructured data) used as a dataset. The main contributions are that the system has been designed for any field, using different input datasets of documents, and with an automatic classification of categories for the detected keywords.

Table 1 shows a comparison between the main features of the available systems from previous studies and the implemented system proposed in this paper.

This article is structured in the following sections: Section 2 explains, in detail, the design and implementation of the proposed weak signal detection system. In Section 3, the experimental setup to test the system in the field of remote sensing is defined. In Section 4, the results obtained in the experiment for the remote sensing sector are presented. In Section 5, the main findings and limitations of the system are analyzed. Finally, Section 6 synthesizes the conclusions and the lines of future work.

Table 1. Comparing implemented system with available studies.

Available Systems	Implemented System
Mainly qualitative analysis	Quantified analysis
Specific model for a specific topic	Model only dependent on the input dataset
Pre-determined keywords	All words and multi-words expressions are keywords
One single data source and/or expert opinion	Three different types of data sources
Mainly structured data sources	Unstructured data sources (documents and NLP [1])

[1] Natural Language Processing.

2. Description of the Proposed System

In this section, a description of every step of the implemented system for the detection of weak signals is carried out. This system has been designed to detect the three dimensions described by Hiltunen.

In order to monitor a specific sector, organizations use data repositories that have been created internally and are already available [38,39]. However, for this study, the first step has been the creation of a dataset of information from several online sources. One of the main suppositions of the implemented system is that, in order to adapt it to a different sector, it is only necessary to generate a new dataset with information from that field. The data sources used to create the dataset have been scientific articles, newspaper articles and social media posts.

The collected documents are analyzed, and information is extracted from each one of the words from them. Next, categories are assigned to each one of those words, and text mining techniques are used to detect candidates for weak signals. In the last stage, natural language processing techniques are used to discard false positives and to obtain better results.

The seven stages of the implemented system can be seen in Figure 1 and are explained in the following subsections: (i) Definition of the data input, (ii) Creation of an input dataset,
(iii) Extract-Transform-Load (ETL), (iv) Category assignment, (v) Text mining and clustering, (vi) Natural language processing (NLP), and (vii) Interpretation, evaluation and decision making.

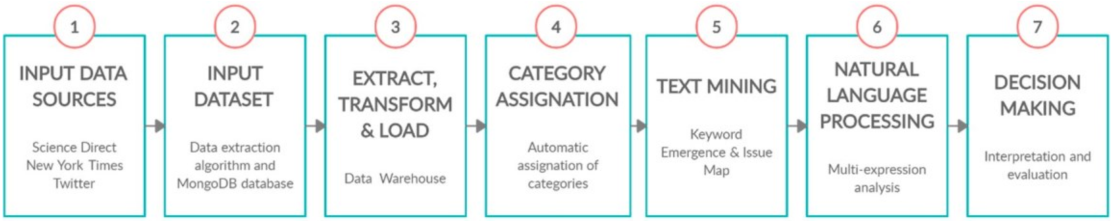


Figure 1. Steps of the system implemented to detect future signals.

2.1. Stage 1: Definition of the Input Data Sources

The signal and issue dimensions depend on the visibility of an event, but the dimension of interpretation depends on the original source of that event. Therefore, to consider the dimension of interpretation, input data should come from documents from different types of sources.

To determine the sources that will be considered in the study, a better understanding of the process of diffusion of a novelty and its potential applications for society is required. Every issue is coded through primary signals called exosignals, that are later interpreted by one or several actors, placing it within a frame of reference to transform it into an interpreter's endosignal, which will be an input signal that will be interpreted again by other actors [14].

A novelty is normally the discovery of a team of engineers and scientists who conduct research. Then, they published their results in a scientific journal. This source is then read by a journalist, who interprets the information and publishes an article in a newspaper or in a web blog. Finally, other users read this post and share the information on social networks, along with their comments, impressions, and sentiments. Figure 2 shows this process.

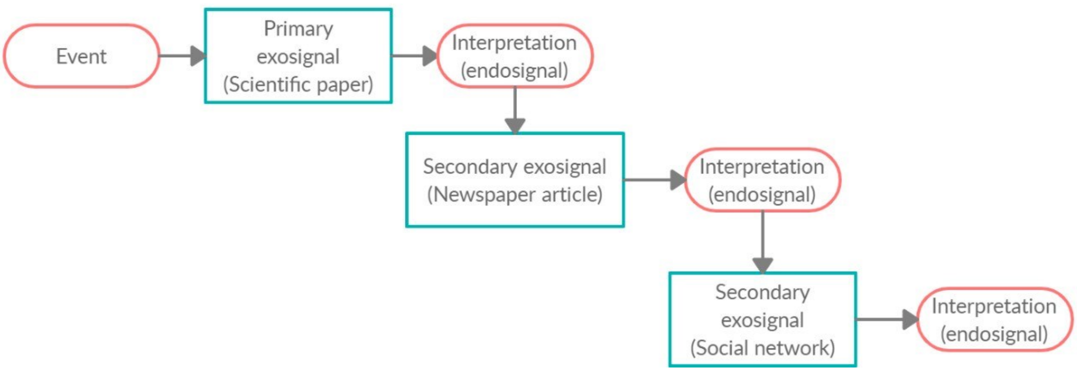


Figure 2. The process of diffusion of signals and their interpretation.

In order to carry out a detailed monitoring of a sector, three different sources of data have been selected: scientific articles extracted from Science Direct, newspaper articles extracted from the New York Times and posts from the social network Twitter.

The starting point is to create a dataset of scientific documents related to the sector under study. To do this, a search of the selected field must be performed in an academic repository. The source selected is Science Direct, a prestigious and widely used website, which provides access to a large database of scientific and medical research. It hosts over 16 million peer-reviewed articles of content from 4279 academic journals, and covers all study areas: Physical Sciences and Engineering, Life Sciences, Health Sciences and Social Sciences and Humanities.

The next step is to detect newspaper articles related to the sector. For this task, The New York Times has been selected to retrieve all articles that mention the sector. The New York Times is an American newspaper with worldwide influence and readership. The newspaper includes an important Science section, in which journalists report about science to the public.

Finally, Twitter has been used as the source for social network posts, because of its flexibility and the numerous recent studies that have incorporated it as a source of knowledge [40–42]. In this study, all Twitter posts sharing one of the detected scientific or newspaper articles have been considered, as well as all the responses and comments to these tweets.

This procedure is thus independent on the field under study, as different datasets of documents can be created for different sectors. Therefore, the system is designed to be applicable to every sector, just generating new input datasets.

2.2. Stage 2: Creation of an Input Dataset

Once the data input is selected, relevant information from Science Direct and New York Times was collected using an algorithm programmed in Python (v3.2.2, Python Software Foundation, Beaverton, OR, USA). This algorithm is created to download data from their respective websites using Beautiful Soup (v3.2.2, open source), a library to analyze HTML (HyperText Markup Language) data.

The algorithm performs a search of a selected field or sector in Science Direct, and extracts the following information for every peer-reviewed paper identified: title, author, summary, keywords, content, conclusions, year of publication and name of the journal. In addition to thousands of journals, Science Direct also contains more than 30,000 books, but the algorithm focuses only on peer-reviewed articles from indexed journals.

The algorithm also performs a search of the same selected field or sector in the New York Times and extracts the following information for every newspaper article: headline, author, section, lead paragraph, content, and year of publication.

After that, the Twitter official API (Twitter Application Programming Interface, v1.1, Twitter, San Francisco, CA, USA) is used to extract every tweet that mentions any of the detected scientific and newspaper articles, together with all comments and retweets. The content of the tweet and the year of publication are stored in the database.

There are multiple options for the implementation of the database for the input dataset. NoSQL (non-structure query languages) technologies have emerged to improve the limitations of the available relational databases [43]. These databases are more oriented to texts and are not modeled in means of tabular relations used in relational databases. In addition, as the dataset is composed of documents from different kinds, document-based databases are generally used [44]. The main operations that are performed in the database are the insertion and selection of items. As these operations are being carried out thousands of times, the execution time of these actions needs to be as short as possible [45]. For this reason, MongoDB (v4.0.8, open source) [46] is a very efficient database technology oriented to store documents and perform text mining operations. Appendix A includes some computational execution times of the system.

2.3. Stage 3: Extract, Transform and Load (ETL)

The next step is the design of a data warehouse to store a large volume of information [47]. Each of the words that make up the documents collected in the previous stage is extracted and treated as a keyword, storing the following information items in the warehouse: source document, the number of occurrences, the year of publication and the source. The only words that are not taken into account are stopwords, which are words that do not have a specific meaning, such as articles, pronouns, prepositions, etc. They are filtered before the processing of natural language data, using the Natural Language Toolkit (NLTK, v3.4.5, open source) leading platform in Python.

Along with the elimination of stopwords, a stemming phase is performed to eliminate the suffixes of each word to obtain their root. Snowball, a small string processing language designed for creating stemming algorithms, is used to implement Porter's process [48]. In this way, the number of insertions in the data warehouse is reduced. The word that represents the word group is the one that appears most frequently in all the documents in the dataset.

2.4. Stage 4: Category Assignation

In Stage 3, a group of categories is designated for every keyword. Two category assignation processes have been carried out. The first one is the assignation of representative layers. The second one is the automatic assignation of categories based on the topics of the documents where the keyword appears.

In the first process, keywords were classified into different layers: "environmental and sustainability factors", "business needs" and "technological components". For example, keywords such as "war" or "oil" were classified as environmental factors, "portable" or "cheap" as business needs, and "batteries" or "hybrid" as product/technological components [33].

The second process is to automatically designate several categories for every word. These categories are automatically assigned, considering keywords, topics, and Special Issues of the documents of the scientific journals, where the word detected as a weak signal is present.

Although some studies use standard category lists [49], one of the main advantages of assigning categories, automatically and dynamically, is that they only depend on the input dataset, therefore being the most relevant ones for the field of study.

2.5. Stage 5: Text Mining

Weak signals generally carry "information on potential change of a system towards an unknown direction" [50].

However, they are very difficult to detect because they could be considered as noise (the trends of these terms are so imperceptible to experts because they do not easily seem to follow any pattern). If these signals exceed a threshold, they become strong signals, known to a vast majority.

In conclusion, weak signals currently exist as small and seemingly insignificant issues that can tell experts and organizations about the changes in the future. A wild card is defined as a surprising event that will have significant consequences in the future [14]. When this event is produced, the weak signal becomes a strong one. Figure 3 shows the process of weak signals becoming strong ones. The goal of this study is to detect weak signals as early as possible to extend the time to react.

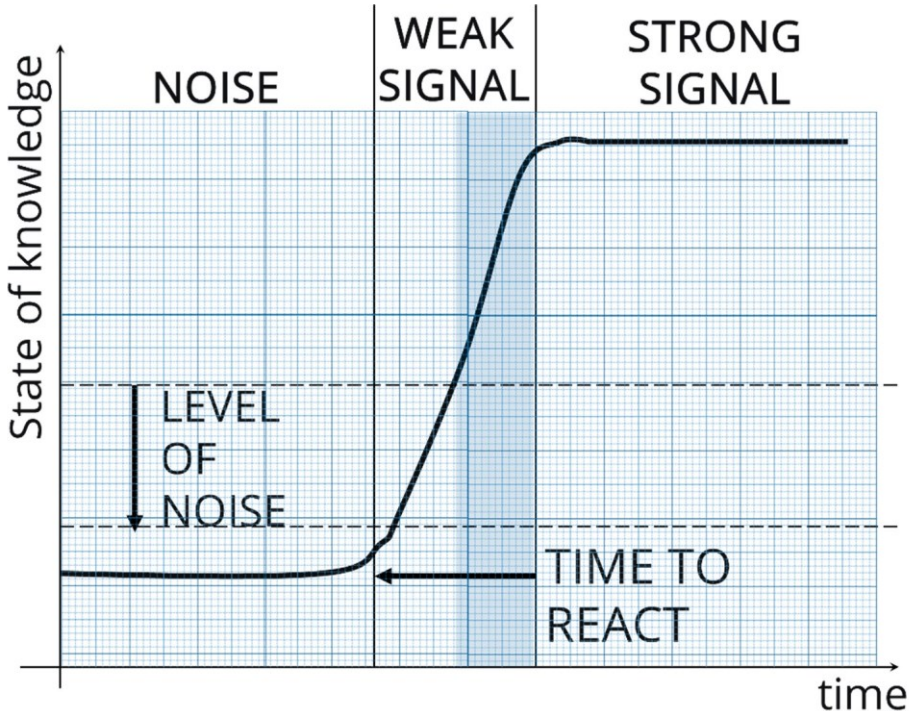


Figure 3. Relation of Weak Signals and the time available to react.

The main difference between text and data mining is that the first one is applied to extract information from textual documents instead of structured sources [51]. The detection of these wild cards is similar to the use of data mining techniques to extract patterns in time series [52]. The main problem is the proper detection of a change point, that is, the identification of the time points where the behavior change is produced. Although all methods lose precision by decreasing the signal/noise ratio, the Batch algorithm offers better results when the dataset precedes the analysis, as in the case under study.

As previously described, the three components of the semiotic model need to be identified: signal, issue and interpretation.

The signal dimension is related to the absolute number of appearances of every word [14]. To measure this dimension, the degree of visibility (*DoV*) is established. First, a value is set by the ratio of the number of appearances and the total number of documents. Then, a factor is introduced to give more importance to the most recent appearances, giving a different weight for every period of a year. In this case study, a dataset of more than fifty thousand documents has been divided into eleven periods of one year (from 2007 to 2017). To carefully set the multiplying factor, a group of business experts has been consulted, defining tw as a time weight of 0.05. The description of the DoV of a word i in a period j is shown in Equation (1).

$$DoV_{ij} = \frac{TF_{ij}}{NN_j} \times \{1 - tw \times (n - j)\} \quad (1)$$

TF_{ij} is the number of appearances of the word i in period j, NN_j is the total number of documents in the period j, while n is the number of periods and tw is a time weight.

The issue dimension is related to the total number of documents where the keyword appears [14]. To measure this dimension, the degree of diffusion (DoD) is established. As in the previous equation, the first step is to set the ratio of the number of documents which contained the keyword and the total number of documents. Then, a factor is introduced to give more importance to the most recent appearances, giving a different weight for every period of a year. This multiplying factor is set, following the same considerations than in the previous case. The description of the DoD of a word i in a period j is shown in Equation (2).

$$DoD_{ij} = \frac{DF_{ij}}{NN_j} \times \{1 - tw \times (n - j)\} \quad (2)$$

DF_{ij} is the number of texts where the word can be found.

Future signals that are candidates for weak signals have an absolute low number of occurrences, but a high fluctuation (a high geometric mean of DoD/DoV but low number of occurrences).

With the calculation of the increase ratios for every word in every document from the input dataset, two graph maps can be generated: the “Keyword Issue Map (KIM)”, a map of DoD with the absolute number of appearances of every word; and the “Keyword Emergence Map (KEM)”, a map of DoV with the number of texts, where every word can be found.

The structure of these two maps is shown in Figure 4. Above a threshold in the time weighted increasing rate axis, two clusters can be identified: the “Strong Signals” area is above an average frequency threshold, and the “Weak Signals” area is below. Below a line on the Y axis, words are identified as noise. This last cluster consists of terms that should be discarded because their appearances are not increased through time, or simply, if their appearances are studied, they do not follow any pattern at all.

The third dimension is interpretation, and this component is related to how the type of document influences the transmission of the signals. The importance of scientific journals has been selected to measure this dimension, because the first exosignals generated are usually the result of a research published in this type of documents.

All available bibliometric indexes for scientific journals have their advantages and limitations, but in general, studies show a high correlation between them, especially in the top of their rankings [53]. There are several journal indexes, such as the Journal Impact Factor (JIF), SCImago Journal Rank or SJR indicator, Eigenfactor Metrics, Scopus h-index or Google h-index. Documents published in journals with a high factor are more influential, and therefore, could accelerate the transformation of weak signals into strong.

Although Impact Factor is widely used, it does not consider that the source of the citations and journals that publish a large amount per year have less potential for a high Impact Factor. Clarivate Analytics has added the Emerging Sources Citation Index to extend the scope of publications in the Web of Science, but this index classifies scientific papers from 2015, and this study is using a dataset of documents from 2005. SCImago Journal Rank indicator is based on the SCOPUS database, which is bigger than the Journal Impact Factor, and places more emphasis on the value of publishing in top-rated journals than on the number of citations of a publication. In addition, SCImago has taken its database and calculated the H-index of every journal, providing a relevant value within the field of research.

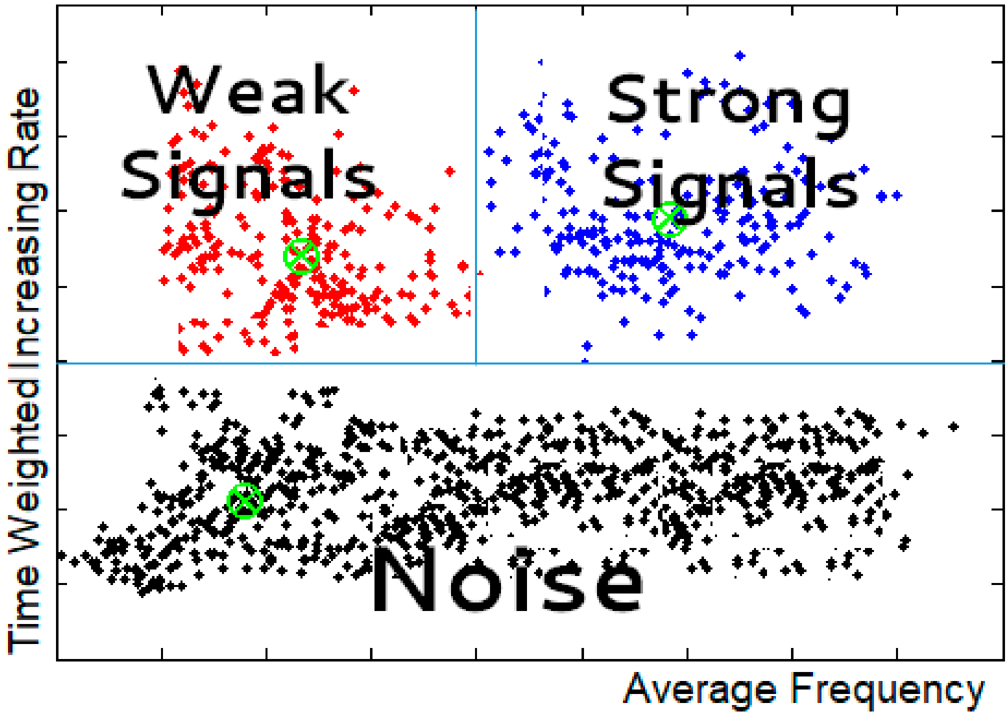


Figure 4. The clusters of a Keyword Map.

The degree of transmission (*DoT*) is measured, considering all h-index values from the journals from all texts where the word *i* can be found, as shown in Equation (3). This index has been selected for several reasons: (i) it is based on a bigger database than Journal Impact Factor, (ii) it is freely available, which means it is easier to access to its content, helping the transmission of new changes, and (iii) unlike citations-per-article measures like Impact Factor, it is not skewed by a small number of individual, highly cited articles.

$$DoT_i = \sum Hindex_{journal} \tag{3}$$

The values of the interpretation are graphically expressed with different sizes for each dot (which represents a keyword) in both KEM and KIM maps. For the final consideration of terms related to weak signals, every *DoD* and *DoV* are multiplied by their *DoT*. This way, scientific journals have a higher weight in the detection of weak signals than the other sources.

2.6. Stage 6: Natural Language Processing (NLP): Multi-Word Expressions

The words in both maps are possible terms related to weak signals. As a result, all keywords not detected in both maps are discarded. However, it is hardly ever possible to extract valuable information with just a single word, considering that a single word can have several meanings, or at least, be connected to several sub-issues.

Natural language processing is widely used in controlled environments [54], but in this study, it is used as an additional stage to improve the quality of the information selected. As the system is only depending on the input dataset, NLP techniques are applied in no controlled environments, considering multi-word expressions.

In conclusion, the next step is a multi-word expression analysis, a natural language processing technique that will help obtain more accurate results, as is shown in previous studies [55,56]. The analysis is performed in the list of detected terms from the text mining analysis from the previous step. The study involves the first words immediately preceding and following the identified term in every appearance, but discarding all stopwords.

The result of the process is a network of expressions related to a keyword, ranked by their overall popularity.

2.7. Stage 7: Interpretation, Evaluation and Decision-Making

As a result of the whole system, experts and other stakeholders will have access to four outputs that will help them in the decision-making process:

1. A list of potential weak signals represented in the Keyword Issue Map, depending on their Degree of Diffusion and Degree of Transmission.
2. A list of potential weak signals represented in the Keyword Emergence Map, depending on their Degree of Visibility and Degree of Transmission.
3. A ranking of all the keywords present in both graphs, which are more likely to be connected to weak signals.
4. The results of the multi-word analysis, providing more accurate results to discard false signs.

3. Experimental Setup

3.1. Definition of the Experiment for Remote Sensing Sector

Once the parts of the proposed system have been described in the previous section, an experiment has been defined to test it. As previously stated, the system can be applied to any sector, because it is only dependent on the input dataset of documents. For this study, the sector that has been chosen is remote sensing. The main reason is because the global market for services related to remote sensing is in a huge expansion and will reach US $7 billion by 2024, due to applications that require the exploitation of satellite image data for companies and governments, such as disaster prevention, weather forecast or agriculture [57].

Remote sensing services are providing new applications in observation and analysis of information at remote locations by means of airborne vehicles, satellites, and on-ground equipment [58]. Remote sensing solutions have multifaceted social applications, such as the mapping of open water surfaces, soil moisture mapping, surface movements, data mosaic and satellite maps, geology and mineral resources, urban and rural development, disaster management support, and climate change studies, among others.

The remote sensing market is projected to witness a significant growth due to the increasing demand for remote sensing services in our society. This growth of remote sensing services is attributed to the effective and flexible data-gathering from remote locations without being physically there [59]. In conclusion, as there are currently many possibilities for social entrepreneurs and companies to work on remote sensing applications, the system has been applied to facilitate the prediction of the future impact of new technologies by the detection of weak signals.

To obtain the input data for the system, a search of the term “remote sensing” was performed in Science Direct and New York Times sources between 2007 and 2017. As recent appearances have more relevance in the weak signal analysis, documents prior to 2006 were not considered because their contribution to the results is negligible.

Although Science Direct also contains electronic books, only peer-reviewed scientific articles were considered in the study. The Python algorithm downloaded the required information from both websites and stored them in databases, as described in the previous section. After that, all tweets sharing all the detected documents from Science Direct and the New York Times were considered, together with all comments and responses from other users in the social network to those tweets.

As a result, more than 43,000 ScienceDirect scientific articles, 1800 New York Times newspaper articles and 59,000 Twitter tweets between 2007 and 2017 were extracted and divided into 11 groups of documents from every year for the analysis. Document distribution by type and year is shown in Figure 5.

Once the input dataset was created, the additional stages of the system were performed. First, a data warehouse was created to store information from every word of every document: source document, number of occurrences and year of publication of the source. As previously stated, all stopwords were discarded and a Porter’s process of stemming was carried out. After this, the steps of category assignation and text mining were developed. As a result, a list of keywords related to weak signals was obtained, and both Keyword Issue and Emergence Maps. The next step was to carry out the multi-word expressions analysis, with the detected words from the previous steps.

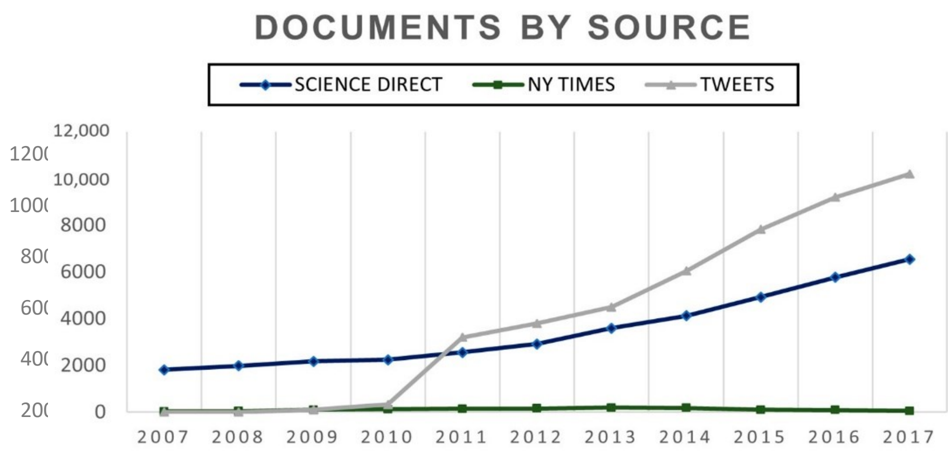


Figure 5. Dataset of documents by type of source.

3.2. Definition of the Evaluation Methods

In a quantitative analysis, words related to weak signals are expected to show a low absolute number of occurrences, but a high range of fluctuation [14]. The graph of *DoV* and *DoD* of a weak signal shows this behavior. A first analysis has been carried out to check that all detected weak signals follow this behavior.

In addition, two different methodologies have been used to evaluate the consistency of the algorithm. The first one consists of generating an additional dataset with the segment of documents for the years 2018 and 2019. Then, it is possible to compare which weak signals have become strong signals, knowing that no documents from these years are in the dataset of the experiment carried out to detect them.

The second method consists of consulting a group of experts to know if there is a match between their opinion and the obtained results [32]. A group of five experts in remote sensing from different institutions were interviewed to compare the results of the test with their predictions.

4. Results

In this section, the output results of the application of the implemented system applied to remote sensing will be described. Computing information about this experiment is shared in Appendix A.

4.1. Keyword Issue Map (KIM) for Remote Sensing

The two main factors that have been considered to configure the Keyword Issue Map are the number of documents where the term appears, and the geometric mean of an average time-weighted increasing rate of this frequency.

These potential weak signals have in common a low frequency of documents where the word can be found, but a high increasing rate. If a word has a high frequency of documents, the word is connected to a strong signal.

These two factors are used to measure the issue component of the sign from the semiotic model of the future sign.

A group of 248 words has been detected, that belong to the cluster of potential weak signals according to the criteria of their degree of diffusion, or *DoD*.

In addition, the interpretation dimension of the sign is measured using the impact factor or h-index of every word, giving a different size of every dot according to this value.

The clustering algorithm has determined a threshold of 104.54 of average document frequency.

Above this threshold, there are no words considered as weak signals.

Figure 6a shows the graph of the Keyword Issue Map that has been generated in the test of the system.

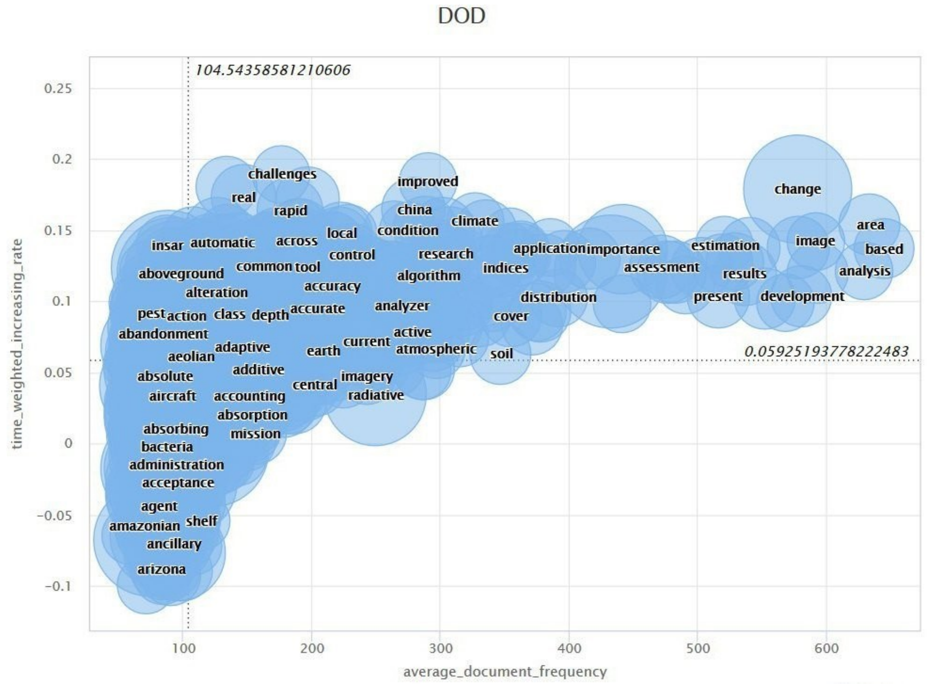


Figure 6a. Keyword Issue Map obtained for remote sensing test

4.2. *Keyword Emergence Map (KEM) for Remote Sensing*

The two main factors that have been considered to configure the Keyword Emergence Map are the number of appearances of a term, and the geometric mean of an average time-weighted increasing rate of this frequency.

These potential weak signals have in common a low frequency of occurrences of the word, but a high increasing rate. If a word has a high frequency of occurrences, the word is connected to a strong signal. These two factors are used to measure the signal component of the sign from the semiotic model of the future sign.

A group of 233 words has been detected, that belong to the cluster of potential weak signals, according to the criteria of their degree of visibility, or *DoV*.

In addition, the interpretation dimension of the sign is also measured using the Impact Factor or H index of every word, giving a different size of every dot according to this value.

The clustering algorithm has determined a threshold of 616.01 of average frequency of occurrences.

Above this threshold, there are no words considered as weak signals.

Figure 6b shows the graph of the Keyword Emergence Map that has been generated in the test of the system.

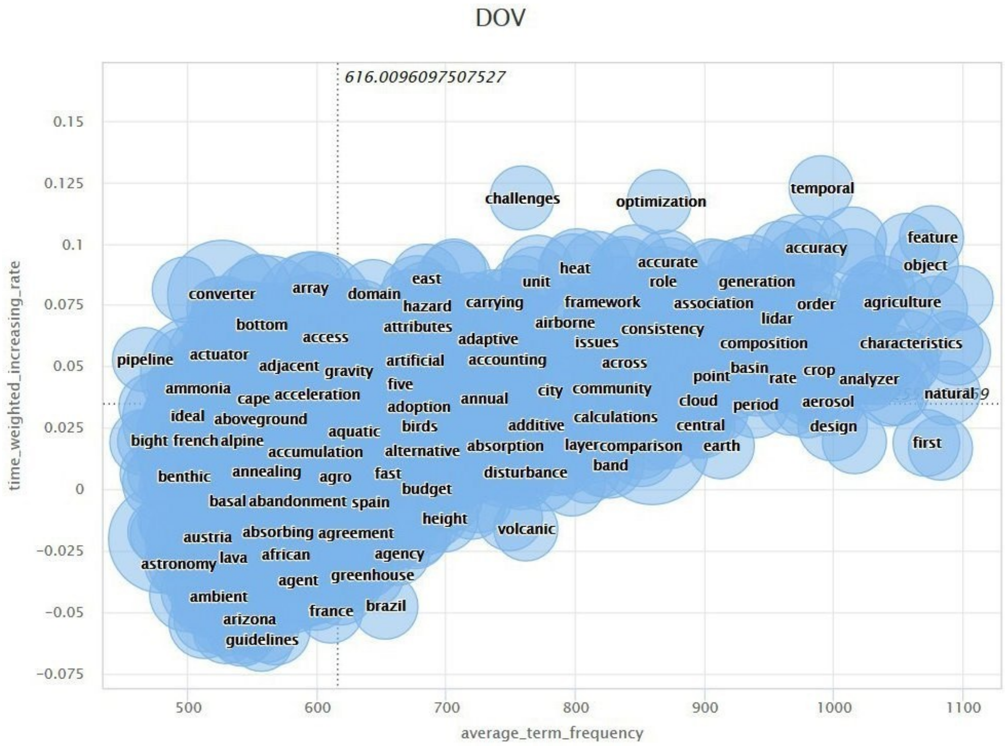


Figure 6b. Keyword Emergence Map obtained for remote sensing test.

4.3. Detected Terms as Potential Weak Signals

A list of 87 words from these two maps fulfill both requirements of low frequency of occurrences and low frequency of documents where the word can be found:

1. Keywords related to environmental, sustainability and geographical factors: Africa, alluvial, asteroids, attenuation, bedrock, Canadian, curvature, depression, desertification, disaster, diurnal, ENSO, extinction, foliar, forestry, Italy, Miocene, multitemporal, observatory, oceanography, pollen, rainforest, rangeland, southeast, sprawl, threat, topsoil, waste, weed and Wuhan.
2. Keywords related to business needs: adjacent, archival, breaking, care, check, consumption, diagnosis, forward, guidance, indirect, interior, intervention, invariant, kernel, maximization, mega, native, NOAA, physiological, plantation, preference, probabilistic, rational, residential, stakeholder, super, supervised, triggering, uptake, vibration and wild.
3. Keywords related to product/technological components: actuator, adaptative, array, bathymetry, cassini, clay, color, converter, endmember, excitation, gamma, hitran, inorganic, InSAR, oblique, passage, photometry, pigments, Rosetta, sounder, SRTM, stepwise, unmanned, UVSQ, volatile and voxel.

Some of the terms identified in both the Key Emergence Map and the Key Issue Map are shown in Table 2. This table shows a list of some of the keywords detected, divided into three different layers (business needs, environmental or geographical factors and product/technological components), their values of frequency of appearances (*DoD* and *DoV*) measured using Equations (1) and (2) *with tw=0* applied to the whole period of 11 years, their normalized *geometric mean of* time-weighted increase ratios in the whole period, and its degree of transmission (*DoT*) measured using Equation (3). The table also shows the categories that have been assigned automatically by the system, in the third stage of the system. Some detected categories for remote sensing are climate change, meteorology, geography, water research, space, or agriculture.

Table 2. Group of Keywords Related to Weak Signals Detected in Remote Sensing Test.

Keyword	DoD	Incr rate	DoV	Incr rate	DoT	Automatic category
BUSINESS NEEDS						
consumption	96.73	0.0975	579.36	0.0479	6.39	Agricultural and Forest Meteorology
diagnosis	91.18	0.1079	566.45	0.0472	6.39	Space Research, Water Research
kernel	96.64	0.07	540.09	0.0384	6.16	Space Research, Water Resources
NOAA	84.73	0.0839	531.27	0.0473	2.38	Climate Change, Space Research, Wind power
physiological	93.91	0.0718	576.73	0.0363	6.39	Radiology, Climate Change
residential	92.91	0.0813	536.64	0.0463	6.39	Climate Change, Applied Geography, Water Research
ENVIRONMENTAL/SUSTAINABILITY FACTORS						
asteroids	88.64	0.0796	582.64	0.0479	6.84	Space Research
bedrock	78.64	0.1007	568.55	0.0647	13.21	Space Research, Particle Physics
Africa	93.64	0.0699	528.45	0.0678	7.03	Climate Change, Water Research
Canadian	85.36	0.0781	545.09	0.0427	6.76	Space Research, Agriculture
desertification	93.91	0.0671	604.45	0.0357	6.39	Climate Change, Space Research
disaster	102.91	0.0712	559	0.045	6.39	Astronautics, Climate Change, Ecosystems
ENSO	86.91	0.0667	572.91	0.04	0.32	Agricultural and Forest Meteorology
extinction	83	0.0863	593.45	0.052	132.52	Space Research, Chemistry
PRODUCT/TECHNOLOGICAL COMPONENTS						
gamma	88.27	0.1081	528	0.0533	5.6	Sea Research, Space Research
HITRAN	95.73	0.0683	577.36	0.0684	6.39	Chemistry, Molecular Spectroscopy Research
InSAR	87	0.1393	87	0.1393	0.95	Space Research, Water Research
UVSQ	94.18	0.0831	555.64	0.0491	127.36	Aerospace Science, Aeronautics
SRTM	85.36	0.0617	527.73	0.0425	3.41	Wind power, Applied Geography, Biology

The pattern of the degrees of visibility and diffusion of the word "desertification", one of the terms identified with a weak signal in the field of remote sensing, is shown in Figure 7. All graphs of DoV and DoD from every detected word show an abnormal pattern with a low frequency.

(a) **(b)**

Figure 7. The fluctuant pattern in the *DoV* (**a**) and *DoD* (**b**) of the word "desertification"

4.4. Results of the Multi-Word Analysis

Some of the detected words from the text mining analysis were selected to perform a multi-word analysis with them, with the objective of obtaining more accurate information.

A few expressions identified as potential weak signals are "climate engineering", "biosatellite engineering", "spectral splitting", "dioxide splitting", "adaptative encoding", "spectroscopic voxel", "cooperative coevolution", "desertification reversion", "global desertification", "ground photometry", "urban sprawl", "West Africa" or "UVSQ-Sat". Figure 8 shows the words that have a higher correlation with the terms "desertification" and "photometry", two of the detected terms related to work signals. The figure below every term is the percentage of times that the word appears beside "desertification" or "photometry" in all the documents of the dataset.

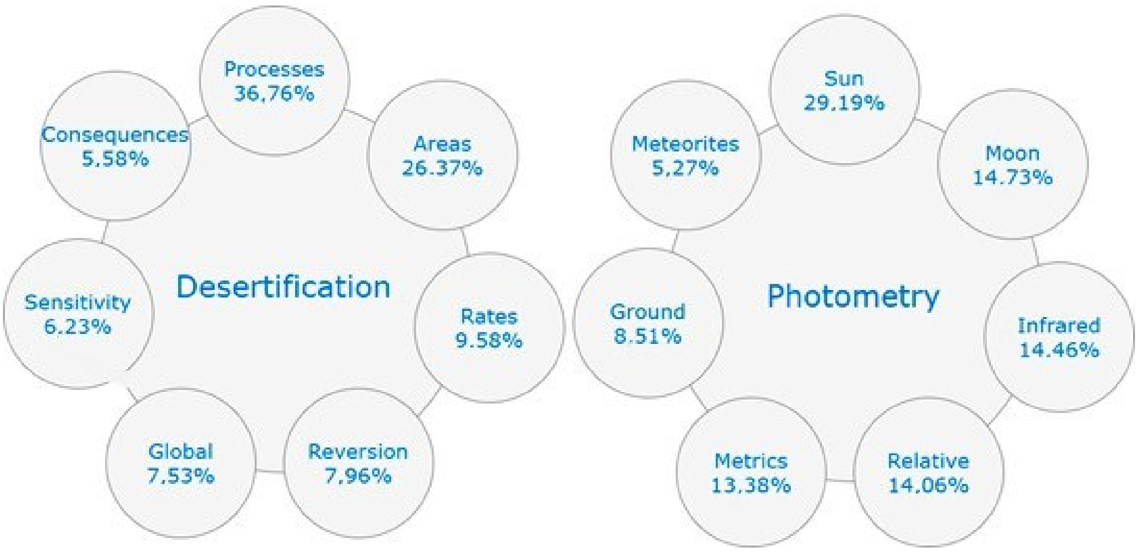


Figure 8. Multi-Word Expressions Related to the words "desertification" and "photometry".

4.5. Evaluation of the Results

In this experiment, some strong signals were detected in the period of 2018 and 2019, based on five scientific papers for 2018 and 2019 from the journal *Remote Sensing*, which showed the importance of West Africa [60–65], which was one of the expressions detected as a weak signal in the experiment related to the remote sensing sector. The expression "UVSQ-Sat", which means "UltraViolet and infrared Sensors at high Quantum efficiency onboard a small SATellite", detected as a candidate of weak signal and now a strong signal, appears for the first time in the title of a scientific paper of 2019 [66]. Finally, other detected keywords, such as "NOAA (National Oceanic and Atmospheric Administration)", "InSAR (Interferometric synthetic aperture radar)", "Rosetta" and "SRTM (Shuttle Radar Topography Mission) are also becoming strong signals in the last two years.

Regarding the group of experts, they indicated the following statements:

1. The growth of remote sensing services is attributed to the effective and flexible data-gathering, thanks to highest resolutions of the metrics, cloud computing software and machine learning techniques. Several terms, such as "adaptative encoding" or "voxel", were detected as related to weak signals.
2. Among the outstanding applications, agriculture and especially desertification, are areas in which remote sensors will be more relevant. Desertification and other terms related to agriculture are keywords that the algorithm identified is related to weak signals.
3. Interferometric synthetic aperture radar, abbreviated "InSAR", which is a radar technique used in geodesy and remote sensing, is becoming more and more important. InSAR is a keyword that the algorithm identified as related to weak signals
4. West Africa is becoming one of the most interesting areas in the world for remote sensing applications. "West Africa" is an expression that the algorithm identified as related to weak signals [60–65].

5. Discussion

The main findings of the obtained results and the main limitations of the methodologies will be described in this section.

5.1. Main Findings

The input dataset that was generated consists of multiple documents about the remote sensing sector that were extracted from three different types of sources. There are a small amount of newspaper articles compared with the amount of scientific papers, which means that newspapers filter the content of scientific journals, publishing only relevant results and their applications [67]. However, the large number of tweets found shows that society has special interest in the applications of this sector.

Previous studies such as Koivisto et al. [22], Yoon [33] and Griol-Barres et al. [49] have only detected weak signals related to terms within standard lists of keywords. The results show that several terms which rarely could be found on standard databases can be detected by this system, providing more reliable results. The results of the experiment show that the system can detect words such as "ENSO (El Niño–Southern Oscillation)", "NOAA (National Oceanic and Atmospheric Administration)" or the Chinese region of Wuhan, because every word in every document is processed as a potential keyword.

The keywords related to weak signals are isolated terms that do not provide enough information for a rich analysis. The automatic assignation of categories to the weak signals detected provides more useful information. For instance, the word “Africa” was detected as one related to weak signals. The Classification stage provided the categories of “climate change” and “water research” for the word “Africa”, giving additional information about applications and subsectors where this word is becoming more relevant. As previously stated, these categories are obtained automatically, considering the names, keywords, topics, and Special Issues of the scientific articles and journals where the word “Africa” is present.

A multi-word analysis was performed to detect expressions that provide more accurate information that can be useful for experts and other stakeholders in their decision-making processes. For this process, only the 87 terms detected were considered. For instance, “West Africa” was a popular expression detected, providing a bit more information about the weak signal. Another example is the term “desertification”, that refers to a problem that can be tackled by remote sensing applications, while the multi-word expression “desertification reversion” refers to new opportunities to solve the problem.

After obtaining the results, the next step is to evaluate and validate them. In standard applications of neuronal networks, the dataset is usually divided into subsets to train and test the network. However, this is not recommended in applications to detect signals of the future applications, due to the low frequency of occurrences of words related to weak signals. In a previous study related to weak signals [33], results were validated, with a deeper evaluation of the same documents of the dataset that was used for the test. This is also not recommended because results should be confirmed with external factors and not by using the same input dataset.

In another study [37], a machine learning technique known as transductive learning was used to train the system, exposing the algorithm to all the examples (i.e., all written texts), but using specific labels (i.e., actions selected) for a subset of these examples. The main difference with this study is that, in this proposal, labels or categories are built automatically to build a system than can be used in any sector.

Two evaluation methods have been performed. The first method of evaluation has been carried out in previous works [33,49], and is based on the detection of strong signals in a new input dataset, with more recent documents than the ones considered in the experiment. In this method, the presence of strong signals in recent documents that were detected as weak signals in earlier references is considered a confirmation of the success in weak signal detection. To confirm the experimental results of the experiment, a new test was conducted with a dataset of documents from 2018 and 2019 related to remote sensing; a group of the detected weak signals have become strong signals. The second method of evaluation was to leverage relevant indicators from a group of experts in the field. The results have confirmed that the weak signals detected in our experiments matched the indicators identified by a panel of experts in the field of remote sensing. In conclusion, the obtained results and the methodologies applied in the evaluation show interesting and promising points that could be useful to detect new opportunities to entrepreneurs and other stakeholders, but the study also shows several limitations which will be addressed in future work.

5.2. Limitations

Although the study results are interesting, the experiment has confirmed that single word expressions do not generally provide relevant information for a better understanding of future trends. Furthermore, although systems to detect weak signals have been evaluated in various sectors [68,69], new studies in other fields should be carried out, to continue testing whether the proposed system is portable to other domains.

The lists of detected words in both the Keyword Issue and Emergence Maps present promising results, but they can also contain false positives that are not really connected to weak signals. A cost-effective method to avoid this situation has been adopted, by considering the words that appear in both lists, but this approach has the limitation of discarding some weak signals.

The automatic assignation of categories and applying natural language processing techniques such as the multi-word expression analysis are providing more useful information to entrepreneurs and other stakeholders. However, to improve the system to provide clearer and definitive guidance for decision-making, more natural language processing techniques could be applied, such as bag of words recognition, regular expressions, or sentiment analysis [70], which will be tested in our future work.

The evaluation techniques applied to the system also present several limitations. The main problem of using a more recent input dataset to validate the results is that this can only be performed when a new dataset of documents is available. The conclusion is that this limitation can be assumed, as the study is concerned with the prediction of future trends, and consequently, the only fully reliable check method is to wait until that future becomes the present. On the other hand, using a group of experts to evaluate the results allows an immediate assessment, but has the limitation that they could discard some weak signals when they are out of their field of knowledge.

Despite these limitations, the performed experiment and its evaluation show that the implemented system is reaching good reliability, according to the opinion of the group of experts in remote sensing, and the dataset of documents from 2018 and 2019.

6. Conclusions

This study describes the design, implementation, and evaluation of a system to identify weak signals of the future. This system is designed to help experts, startups and other stakeholders in their decision-making processes, due to the difficulty of scanning complex environments full of textual information.

The system has been designed under the constraints of high efficiency, the use of multiple types of input sources, data sources of unstructured textual documents, its applicability in every field of study, and to extract conclusions based on quantitative results that are not dependent on the opinion of experts [71]. The experimental setup of the system includes a dataset of documents from 2007 to 2017 related to the field of remote sensing. Two methodologies to evaluate the experiment (new dataset with documents from 2018 and 2019 and a group of experts) were applied, with promising results that can be used by entrepreneurs and other organizations in their decision-making processes.

The system is only dependent of its input dataset of documents. If the system is tested in a different sector, it is only necessary to obtain a different input dataset with documents related to that sector, following the instructions explained in Section 2. In addition, although some studies use standard category lists [49], one of the main advantages of assigning categories automatically and dynamically is that they are only dependent on the input dataset, therefore being the most relevant ones for the field of study. In conclusion, the system can be used to perform the detection of weak signals of the future in multiple sectors. The outputs are obtained by a quantitative scoring methodology that does not depend on the opinion of experts.

Finally, in future research, the system will be tested in other sectors, and new natural language processing techniques will be applied.

Author Contributions: Conceptualization, I.G.-B., S.M. and J.M.; methodology, I.G.-B. and J.M.; software, I.G.-B. and S.M.; validation, I.G.-B., S.M., H.F. and J.M.; formal analysis, I.G.-B.; investigation, I.G.-B. and J.M.; resources, I.G.-B., S.M. and H.F.; data curation, I.G.-B. and S.M.; writing—original draft preparation, I.G.-B., writing—review and editing, I.G.-B., A.C., H.F. and J.M.; visualization, I.G.-B.; supervision, J.M.; project administration, I.G.-B. and J.M.; funding acquisition, I.G.-B., H.F. and J.M. All authors have read and agree to the published version of the manuscript.

Funding: This research is partially supported by EIT Climate-KIC of the European Institute of Technology (project EIT Climate-KIC Accelerator—TC_3.1.5_190607_P066-1A) and InnoCENS from Erasmus+ (573965-EPP-1-2016-1-SE-EPPKA2-CBHE-JP).

Acknowledgments: This research is also part of a PhD from the Department of Electronics Engineering of the Polytechnic University of Valencia.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

This appendix shows computing information about the experiment carried out in the study. The first consideration is that computer systems are moving from "central processing" performing on the CPU (central processing unit) to a "co-processing" performing distributed between the CPU and the GPU. To enable this new computational paradigm, there are new parallel hardware architectures that are used by a multitude of application developers. Multicore platforms are providing new computing capacity that enables the execution process of systems that analyze a vast number of documents.

The general-purpose GPU programming has become a standard in deep learning applications and keeps evolving. The Compute Unified Device Architecture (CUDA), which is the Nvidia framework for creating software to be executed in GPUs, is in fact one of the most reliable options for data engineers [72].

The tests have been executed on a computer with the following setups:

1. CPU: Intel core i7 7500U (Dual Core) - RAM: 16 GB DD4
2. GPU: NVidia Geforce GT 650M 1024 MB GDDR5—900MHz—384 CUDA Cores RAM: 16 GB DD4

The execution times show that the parallel GPU implementation is up to nine times faster compared with the standard sequential alternative using CPUs. The reason why the GPU rendering obtains this optimization is because GPUs have a large number of simple cores which allow parallel computing through thousands of threads computing at a time, and the system requires multiple operations related to every word in every document that can be easily executed in parallel.

Table A1 shows the execution times of the experiment.

Table A1. Execution times.

Stage	CPU	GPU
Data Warehouse creation	6257 min	688 min
Category assignation	68 min	9 min
Text mining	124 min	11 min
Multi-word expressions	43 min	6 min

These results show that, despite the large number of documents processed, the system has been designed in an efficient way so that the calculation can be performed on equipment with medium-level hardware, and, therefore, on an equipment that could be affordable for entrepreneurs and SMEs.

Finally, Table A2 shows a benchmark of operation performance for different database technologies [45]. After this study, the database technology selected to store the data extracted from documents was MongoDB [46], as shown in the definition and implementation of the system.

Table A2. Comparing operation performance of different DB technologies with ten thousand records (time in milliseconds).

Operation	Oracle	MySQL	MsSql	MongoDB	Redis	GraphQL	Cassandra
INSERT	0.076	0.093	0.093	0.005	0.009	0.008	0.011
SELECT	0.025	0.093	0.062	0.009	0.016	0.010	0.014

References

1. Eisenhardt, K.M.; Brown, S.L. Patching: Restitching business portfolios in dynamic markets. *Harv. Bus. Rev.* **1999**, *77*, 72–82. [PubMed]
2. Zahra, S.A.; Gedajlovic, E.; Neubaum, D.O.; Shulman, J.M. A typology of social entrepreneurs: Motives, search processes and ethical challenges. *J. Bus. Ventur.* **2009**, *24*, 519–532. [CrossRef]
3. Choo, C.W.; Auster, E. Environmental scanning: Acquisition and use of information by managers. *Ann. Rev. Inf. Sci. Technol.* **1993**, *28*, 279–314.
4. Ansoff, H.I.; McDonnell, E.J. *Implanting Strategic Management*; Prentice Hall: Cambridge, MA, USA, 1990; pp. 27–34.
5. Ansoff, H.I. Managing Strategic Surprise by Response to Weak Signals. *Calif. Manag. Rev.* **1975**, *18*, 21–33. [CrossRef]
6. Cooper, A.; Voigt, C.; Unterfrauner, E.; Kravcik, M.; Pawlowski, J.; Pirkkalainen, H. Report on Weak Signals Collection. TELMAP, European Commission Seventh Framework Project (IST-257822). 2011. Deliverable D4.1. pp. 6–7. Available online: https://cordis.europa.eu/docs/projects/cnect/2/257822/080/deliverables/001-D41Weaksignalscollectionfinal.doc (accessed on 25 April 2020).
7. Coffman, B. Part I. Introduction. In *Weak Signal Research*; MG Taylor Corporation:

Louisville, KY, USA, 1997.
8. Godet, M. *From Anticipation to Action, a Handbook of Strategic Prospective*; UNESCO Publishing: Paris, France, 1994; p. 59.
9. Molitor, G.T. Molitor Forecasting Model: Key Dimensions for Plotting the Patterns of Change. *J. Future Stud.* **2003**, *8*, 61–72.
10. Dator, J. Futures Studies as Applied Knowledge. In *New Thinking for a New Millennium*; Routledge: London, UK, 1996; pp. 66–74.
11. Dator, J. Universities without quality and quality without universities. *Horizon* **2005**, *13*, 199–215. [CrossRef]
12. Nikander, I.O. Early Warnings, a Phenomenon in Project Management. Ph.D. Thesis, Helsinki University of Technology, Helsinki, Finland, 2002.
13. Mannermaa, M. *Tulevaisuuden Hallinta Skenaariot Strategiatyoskentelyssa. (Managing the Future, Scenarios in Strategy Work)*; WSOY: Porvoo, Finland, 1999; p. 227.
14. Hiltunen, E. The future sign and its three dimensions. *Futures* **2007**, *40*, 247–260. [CrossRef]
15. Peirce, C.S. Some Consequences of Four Incapacities. *J. Specul. Philos.* **1868**, *2*, 140–157.
16. Han, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Morgan Kaufmann Publishers: Burlington, MA, USA, 2001.
17. Fischler, M.A.; Firschein, O. *Intelligence: The Eye, the Brain and the Computer*; Addison-Wesley: Boston, MA, USA, 1987; p. 221.
18. Hong, S.W.; Kim, Y.E.; Bae, K.J.; Park, Y.W.; Park, J.K. Development of analysis model for R&D environment change in search of the weak signal. *J. Korea Technol. Innov. Soc.* **2009**, *12*, 189–211.
19. Thorleuchter, D.; Scheja, T.; Van den Poel, D. Semantic weak signal tracing. *Expert Syst. Appl.* **2014**, *41*, 5009–5016. [CrossRef]
20. Julien, P.A.; Andriambeloson, E.; Ramangalahy, C. Networks, weak signals and technological innovations among SMEs in the land-based transportation equipment sector. *Entrep. Reg. Dev.* **2004**, *16*, 251–269. [CrossRef]
21. Wu, X.; Zhu, X.; Wu, G.Q.; Ding, W. Data mining with big data. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 97–107. [CrossRef]
22. Koivisto, R.; Kulmala, I.; Gotcheva, N. Weak signals and damage scenarios—Systematics to identify weak signals and their sources related to mass transport attacks. *Technol. Forecast. Soc. Chang.* **2016**, *104*, 180–190. [CrossRef]
23. Davis, J.; Groves, C. City/future in the making: Masterplanning London's Olympic legacy as anticipatory assemblage. *Futures* **2019**, *109*, 13–23. [CrossRef]
24. Irvine, N.; Nugent, C.; Zhang, S.; Wang, H.; Ng, W.W.Y. Neural Network Ensembles for Sensor-Based Human Activity Recognition Within Smart Environments. *Sensors* **2020**, *20*, 216. [CrossRef] [PubMed]
25. Huang, M.; Liu, Z. Research on Mechanical Fault Prediction Method Based on Multifeature Fusion of Vibration Sensing Data. *Sensors* **2020**, *20*, 6. [CrossRef] [PubMed]

26. Awan, F.M.; Saleem, Y.; Minerva, R.; Crespi, N. A Comparative Analysis of Machine/Deep Learning Models for Parking Space Availability Prediction. *Sensors* **2020**, *20*, 322. [CrossRef]
27. Baghmolaei, R.M.; Mozafari, N.; Hamzeh, A. Continuous states latency aware influence maximization in social networks. *AI Commun.* **2017**, *30*, 99–116.
28. McGrath, J.; Fischetti, J. What if compulsory schooling was a 21st century invention? Weak signals from a systematic review of the literature. *Int. J. Educ. Res.* **2019**, *95*, 212–226. [CrossRef]
29. Chao, W.; Jiang, X.; Luo, Z.; Hu, Y.; Ma, W. Interpretable Charge Prediction for Criminal Cases with Dynamic Rationale Attention. *J. Artif. Intell. Res.* **2019**, *66*, 743–764. [CrossRef]
30. Van Veen, B.L.; Ortt, R.; Badke-Schaub, P. Compensating for perceptual filters in weak signal assessments. *Futures* **2019**, *108*, 1–11. [CrossRef]
31. Thorleuchter, D.; Van den Poel, D. Idea mining for webbased weak signal detection. *Futures* **2015**, *66*, 25–34. [CrossRef]
32. Rowea, E.; Wrightb, G.; Derbyshirec, J. Enhancing horizon scanning by utilizing pre-developed scenarios: Analysis of current practice and specification of a process improvement to aid the identification of important 'weak signals'. *Technol. Forecast. Soc. Chang.* **2017**, *125*, 224–235. [CrossRef]
33. Yoon, J. Detecting weak signals for long-term business opportunities using text mining of Web news. *Expert Syst. Appl.* **2012**, *39*, 12543–12550. [CrossRef]
34. Yoo, S.H.; Won, D. Simulation of Weak Signals of Nanotechnology Innovation in Complex System.
Sustainability **2018**, *10*, 486.
35. Suh, J.H. Generating Future-Oriented Energy Policies and Technologies from the Multidisciplinary Group Discussions by Text-Mining-Based Identification of Topics and Experts. *Sustainability* **2018**, *10*, 3709. [CrossRef]
36. Kwon, L.-N.; Park, J.-H.; Moon, Y.-H.; Lee, B.; Shin, Y.; Kim, Y.-K. Weak signal detecting of industry convergence using information of products and services of global listed companies—Focusing on growth engine industry in South Korea. *J. Open Innov. Technol. Mark. Complex.* **2018**, *4*, 10. [CrossRef]
37. Ben-Porat, O.; Hirsch, S.; Kuchy, L.; Elad, G.; Reichart, R.; Tennenholtz, M. Predicting Strategic Behavior from Free Text. *J. Artif. Intell. Res.* **2020**, *68*, 413–445. [CrossRef]
38. Fink, L.; Yogev, N.; Even, A. Business intelligence and organizational learning: An empirical investigation of value creation processes. *Inf. Manag.* **2017**, *54*, 38–56. [CrossRef]
39. Ilmola, L.; Kuusi, O. Filters of weak signals hinder foresight: Monitoring weak signals efficiently in corporate decision-making. *Futures* **2006**, *38*, 908–924. [CrossRef]
40. Weng, J.; Bu-Sung, L. Event detection in twitter. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011.

41. Doulamis, N. Event detection in twitter microblogging. *IEEE Trans. Cybern.* **2015**, *46*, 2810–2824. [CrossRef] [PubMed]
42. Atefeh, F.; Wael, K. A survey of techniques for event detection in twitter. *Comput. Intell.* **2013**, *31*, 132–164. [CrossRef]
43. Mehmood, N.; Culmone, R.; Mostarda, L. Modeling temporal aspects of sensor data for MongoDB NoSql database. *J. Big Data* **2017**, *4*, 15. [CrossRef]
44. Bjeladinovic, S. A fresh approach for hybrid Sql/NoSql database design based on data structuredness. *Enterp. Inf. Syst.* **2018**, *12*, 1202–1220. [CrossRef]
45. Č erešňák, R.; Kvet, M. Comparison of query performance in relational a non-relation databases. *Transp. Res. Procedia* **2019**, *40*, 170–177. [CrossRef]
46. Yangui, R.; Nabli, A.; Gargouri, F. Automatic Transformation of Data Warehouse Schema to NoSQL Data Base: Comparative Study. *Procedia Comput. Sci.* **2016**, *96*, 255–264. [CrossRef]
47. Inmon, W.H. *Building the Data Warehouse*, 4th ed.; John Wiley and Sons: Hoboken, NJ, USA, 2005; p. 156.
48. Willett, P. The Porter stemming algorithm: Then and now. *Program* **2006**, *40*, 219–223. [CrossRef]
49. Griol-Barres, I.; Milla, S.; Millet, J. Implementación de un sistema de detección de señales débiles de futuro mediante técnicas de minería de textos. (Implementation of a weak signal detection system by text mining techniques). *Rev. Esp. Doc. Cient.* **2019**, *42*, 234. [CrossRef]
50. Kim, J.; Han, M.; Lee, Y.; Park, Y. Futuristic datadriven scenario building: Incorporating text mining and fuzzy association rule mining into fuzzy cognitive map. *Expert Syst. Appl.* **2016**, *57*, 311–323. [CrossRef]
51. Mendonca, S.; Cunha, M.P.; Kaivo-Oja, J.; Ruff, F. Wild Cards, Weak Signals and Organizational Improvisation. *Futures* **2004**, *36*, 201–218. [CrossRef]
52. Ishikiriyama, C.S.; Miro, D.; Gomesa, C.F.S. Text Mining Business Intelligence: A small sample of what words can say. *Procedia Comput. Sci.* **2015**, *55*, 261–267. [CrossRef]
53. Yuen, J. Comparison of Impact Factor, Eigenfactor Metrics, and SCImago Journal Rank Indicator and h-index for Neurosurgical and Spinal Surgical Journals. *World Neurosurg.* **2018**, *119*, e328–e337. [CrossRef] [PubMed]
54. Thomason, J.; Padmakumar, A.; Sinapov, J.; Walker, N.; Jiang, Y.; Yedidsion, H.; Hart, J.; Stone, P.; Mooney, R.J. Jointly Improving Parsing and Perception for Natural Language Commands through Human-Robot Dialog. *J. Artif. Intell. Res.* **2020**, *67*, 327–374. [CrossRef]
55. Guralnik, V.; Srivastava, J. Event detection from time series data. In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 15–18 August 1999; pp. 33–42.
56. Tseng, Y.-H.; Lin, C.-J.; Lin, Y.-I. Text mining techniques for patent analysis. *Inf. Process. Manag.* **2007**, *43*, 1216–1247. [CrossRef]
57. Wood, L. *Satellite Remote Sensing—Market Analysis, Trends, and Forecasts*; Global Industry Analysts Inc.: San Jose, CA, USA, 2019; p. 12.
58. Bindzarova-Gergelova, M.; Labant, S.; Kuzevic, S.; Kuzevicova, Z.; Pavolova, H.

Identification of Roof Surfaces from LiDAR Cloud Points by GIS Tools: A Case Study of Lučenec, Slovakia. *Sustainability* **2020**, *12*, 6847. [CrossRef]

59. Sugla, S.; Dhum, N. *Remote Sensing Services Market by Platform (Satellites, UAVs, Manned Aircraft, and Ground), End User (Defense and Commercial), Resolution (Spatial, Spectral, Radiometric, and Temporal), and Region—Global Forecast to 2022*; Markets and Markets: Dublin, Ireland, 2017; pp. 5–25.
60. Badmos, O.S.; Rienow, A.; Callo-Concha, D.; Greve, K.; Jürgens, C. Urban Development in West Africa—Monitoring and Intensity Analysis of Slum Growth in Lagos: Linking Pattern and Process. *Remote Sens.* **2018**, *10*, 1044. [CrossRef]
61. Thomson, E.R.; Malhi, Y.; Bartholomeus, H.; Oliveras, L.; Gvozdevaite, A. Mapping the Leaf Economic Spectrum across West African Tropical Forests Using UAV-Acquired Hyperspectral Imagery. *Remote Sens.* **2018**, *10*, 1532. [CrossRef]
62. Samasse, K.; Hanan, N.P.; Tappan, G.; Diallo, Y. Assessing Cropland Area in West Africa for Agricultural Yield Analysis. *Remote Sens.* **2018**, *10*, 1785. [CrossRef]
63. Anchang, J.Y.; Prihodko, L.; Kaptué, A.T.; Ross, C.W.; Ji, W.; Kumar, S.S. Trends in Woody and Herbaceous Vegetation in the Savannas of West Africa. *Remote Sens.* **2019**, *11*, 576. [CrossRef]
64. Jung, H.C.; Getirana, A.; Arsenault, K.R.; Holmes, T.R.H.; McNally, A. Uncertainties in Evapotranspiration Estimates over West Africa. *Remote Sens.* **2019**, *11*, 892. [CrossRef]
65. Mondal, P.; Liu, X.; Fatoyinbo, T.E.; Lagomasino, D. Evaluating Combinations of Sentinel-2 Data and Machine-Learning Algorithms for Mangrove Mapping in West Africa. *Remote Sens.* **2019**, *11*, 2928. [CrossRef]
66. Meftah, M.; Damé, L.; Keckhut, P.; Bekki, S.; Sarkissian, A.; Hauchecorne, A.; Bertran, E.; Carta, J.P. UVSQ-SAT, a Pathfinder CubeSat Mission for Observing Essential Climate Variables. *Remote Sens.* **2020**, *12*, 92. [CrossRef]
67. Zhang, W.; Yoshida, T.; Tang, X. Text classification based on multi-word with support vector machine.*Knowl. Based Syst.* **2008**, *21*, 879–886. [CrossRef]
68. Griol, I.; Milla, S.; Millet, J. Improving strategic decision making by the detection of weak signals in heterogeneous documents by text mining techniques. *AI Commun.* **2019**, *32*, 347–360. [CrossRef]
69. Griol, I.; Milla, S.; Millet, J. System Implementation for the Detection of Weak Signals of the Future in Heterogeneous Documents by Text Mining and Natural Language Processing Techniques. In Proceedings of the 11th International Conference on Agents and Artificial Intelligence, Prague, Czech Republic, 19–21 February 2019; pp. 631–638.
70. Dzedzickis, A.; Kaklauskas, A.; Bucinskas, V. Human Emotion Recognition: Review of Sensors and Methods. *Sensors* **2020**, *20*, 592. [CrossRef]
71. Haegeman, K.; Marinelli, E.; Scapolo, F.; Ricci, A.; Sokolov, A. Quantitative and qualitative approaches in future oriented technology analysis (FTA): From combination to integration. *Technol. Forecast. Soc. Chang.* **2013**, *80*, 386–397. [CrossRef]
72. Silva, V.O.; Martins, C.; Ekel, P. An Efficient Parallel Implementation of an Optimized Simplex Method in GPU-CUDA. *IEEE Lat. Am. Trans.* **2018**, *16*, 564–573. [CrossRef]

Article

Variational Quantum Circuits for Machine Learning. Application for the Detection of Weak Signals.

Israel Griol-Barres, Sergio Milla, Antonio Cebrián, Yashar Mansoori and José Millet.

Applied Sciences 2021, 11, 6427; doi: 10.3390/app11146427
Received: 30 May 2021; Accepted: 9 July 2021; Published: 12 July 2021

Authors	*Griol-Barres, I.; Milla, S.; Cebrián, A.; Mansoori, Y.; Millet, J.*
Title	*Variational quantum circuits for machine learning. An application for the detection of weak signals.*
Journal	*Applied Sciences*, **2021**, *11*(14), 6427
DOI	10.3390/app1114
Publisher	MDPI, ST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLAND
ISSN	2076-3417
eISSN	2076-3417
Impact Factor	2.679 (2020) – 2.736 (5 years)

JCR Category	Rank in Category	Quartile in Category
CHEMISTRY, MULTIDISCIPLINARY - SCIE	**101 of 179**	**Q3**
ENGINEERING, MULTIDISCIPLINARY - SCIE	**38 of 91**	**Q2**
MATERIALS SCIENCE, MULTIDISCIPLINARY - SCIE	**201 of 335**	**Q3**
PHYSICS, APPLIED - SCIE	**73 of 160**	**Q2**

CiteScore Highest percentile	71% 85/297 General Engineering
CiteScore 2020	3.0
SJR 2020	0.435 Q1
SNIP 2020	1.068
HIndex	52
Main contributions	This chapter is a published article that includes an analysis of all available quantum computing technologies. In addition, a variational circuit in quantum logic is proposed and applied as part of the system to detect weak signals in the field of medical imaging. This chapter addresses the Objective 3 and develops tasks 2, 7, 8 and 9.

Article
Variational Quantum Circuits for Machine Learning. Application for the Detection of Weak Signals.
Israel Griol-Barres, Sergio Milla, Antonio Cebrián, Yashar Mansoori and José Millet.
Applied Sciences 2021, 11, 6427; doi: 10.3390/app11146427
Received: 30 May 2021; Accepted: 9 July 2021; Published: 12 July 2021

4. Variational Quantum Circuits for Machine Learning. Application for the Detection of Weak Signals.

Featured Application: Quantum classifier to detect weak signals.

Abstract: Quantum computing is a new paradigm for a multitude of computing applications. This study presents the technologies that are currently available for the physical implementation of qubits and quantum gates, establishing their main advantages and disadvantages and the available frameworks for programming and implementing quantum circuits. One of the main applications for quantum computing is the development of new algorithms for machine learning. In this study, an implementation of a quantum circuit based on support vector machines (SVMs) is described for the resolution of classification problems. This circuit is specially designed for the noisy intermediate-scale quantum (NISQ) computers that are currently available. As an experiment, the circuit is tested on a real quantum computer based on superconducting qubits for an application to detect weak signals of the future. Weak signals are indicators of incipient changes that will have a future impact. Even for experts, the detection of these events is complicated since it is too early to predict this impact. The data obtained with the experiment shows promising results but also confirms that ongoing technological development is still required to take full advantage of quantum computing.

Keywords: quantum computing; variational quantum circuits; quantum support vector machines, weak signals of the future; machine learning.

1. Introduction

Approximately every two years the number of transistors in a classical microprocessor doubles [1]. As a result, transistors are becoming smaller and require lower voltages to operate. However, current classical computers are reaching the limit of computational capacity because when manipulated in circuits of such small size, electrons tend to act unpredictably. They can also pass through the walls of conduction channels in what is known as the 'tunnel effect' [2]. One of the alternatives to classical computing is the exploitation of the laws of quantum mechanics in computational environments.

Quantum computing is the development of computational technologies that are based on the laws of quantum mechanics. This type of computing enables the creation of systems to store, process, and transfer information encoded in quantum media. The first idea of a quantum computer was proposed in 1982 [3] and was based on the first classical computing machine defined by Alan Turing [4]. The Turing machine manipulated symbols on a strip according to a table of rules that simulated a computational algorithm. The first proposal for quantum computation replaced the strip with a sequence of two-state quantum systems. The machine evolved in steps. At the end of each step, the tape was always in a fundamental state of '1' or '0'. However, during each step, the machine could be in a superposition of spin states.

Soon after the first idea for a quantum computer, some scholars [5] introduced the 'speculative' idea that a computer based on the laws of quantum mechanics could perform complex calculations faster than conventional binary computers.

Conventional computers convert data into a series of binary digits called bits that constitute the basic units of information. These bits can only have the values of '0' and '1' and are encoded in integrated circuits according to electrical signals of different voltages. Technological advancements have enabled the creation of devices that act as bits. Today an integrated circuit contains millions and millions of transistors.

In contrast, data processed by quantum computers use 'qubits' that can have the value of '0' and '1' at the same time. Unlike bits, qubits can be in a superposition of both states [6].

The Bloch sphere [7] is a geometrical representation of all possible superposition states that a qubit can adopt, as can be seen in Figure 1a. Each point on the surface of the sphere corresponds to a pure state of the Hilbert space of complex dimension 2. The point of coordinates (0,0,1) corresponds to an eigenvector with a positive eigenvalue of the Pauli matrix and the point of coordinates (0,0,−1) corresponds to the eigenvector with negative eigenvalue. Both states are expressed with the notation |0> and |1> and correspond to spin up and spin down. The qubit can be represented as a linear combination of both states, as shown in the equation, where a and b are complex numbers, and where $|\alpha|^2 + |\beta|^2 = 1$.

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = r_\alpha e^{i\phi_\alpha}|0\rangle + r_\beta e^{i\phi_\beta}|1\rangle \quad (1)$$

The Bloch sphere makes it possible to visualize the action of different quantum gates graphically. For example, one of the most used quantum gates is the Hadamard gate that operates on a single qubit. This gate is equivalent to the combination of two rotations, one of pi about the z-axis followed by a rotation of pi/2 about the y-axis of the Bloch sphere as shown in Figure 1b.

The Hadamard gate is the 'quantum Fourier transform' performed on a qubit [8]. The Hadamard gate is usually the first stage in a quantum circuit because it places the qubit in a state of superposition. However, the Bloch sphere model, while conceptually determining the states that a qubit can have, has the major limitation of not being useful for showing the 'entanglement' of several qubits. One of the most important advantages of state superposition is that, while with n bits one can encode a single state out of 2^n possible states, with n qubits one could encode 2^n states simultaneously. This can be achieved by taking advantage of the possibility of rotation of the qubit values. Given this, quantum computers can dramatically increase the capacity of storage devices by storing huge amounts of data in a very little space [9].

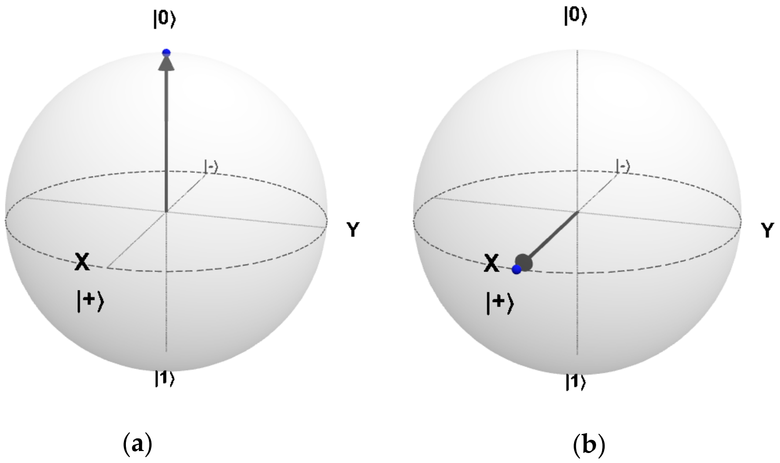


Figure 1. The Block sphere representing a qubit in the state of: (**a**) |0> and (**b**) after applying a Hadamard gate to (**a**).

In addition to superposition, qubits share another principle of quantum physics called entanglement [10]. If superposition is the qubit's characteristic of being in several states at the same time, entanglement is the correlation that exists between two qubits, even if they are not physically in the same place. When quantum entanglement occurs between two particles, their quantum states are determined as if they were a single particle. Therefore, any variation in the quantum state in one of them instantaneously causes the same variation in the other particle. As previously stated, the quantum state of n qubits implies the superposition of all possible configurations (2^n). The quantum entanglement between the electron spins of n qubits exponentially increases the Hilbert space from $2n$ to 2^n dimensions.

2. Quantum Technology Approaches

Although there is a technological challenge for the development of quantum computers, the theoretical framework on which they are based is solid and well defined. It allows all the possibilities that classical computing offers and far more. From a strictly mathematical point of view, the operations performed with qubits can be expressed with matrix algebra, tensor products and probability calculations, to name but a few concepts central to the discipline.

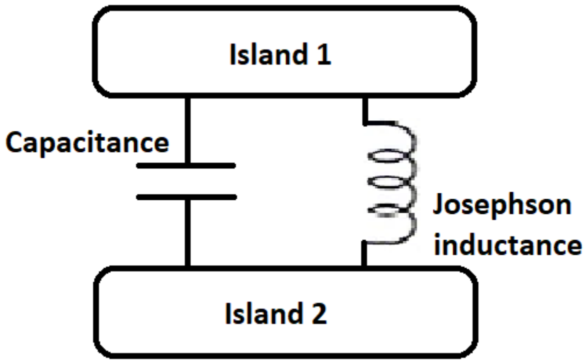
However, creating physical devices that behave in line with quantum mechanics has been a technological challenge; a challenge that has led several companies to invest heavily to achieve 'quantum supremacy' [11]. A variety of technologies exists to implement quantum devices.

2.1. Superconducting Circuits

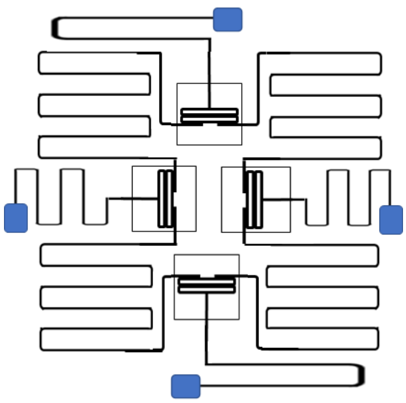
One type of technology used to generate qubits is based on superconduction. These circuits are characterized by the use of qubits that can be selectively introduced or removed. Superconduction circuits can take several different architectures. For instance, superconduction-based qubits can be designed by the interconnection of superconducting electrodes with Josephson junctions. The system is based on the capacitive contribution of the Cooper pairs as well as on the inductive contribution of the jump between two electrodes [12]. Some examples of qubits based on these concepts are the charge qubit, the flux qubit, the phase qubit, and the fluxonium. The transmon ("transmission line shunted plasma oscillation qubit" [12]) is a type of superconducting charge qubit that was developed in 2007 to reduce sensitivity to noise. This technology is based on a multilevel system, used as qubits if all their dynamics are confined to two quantum levels. The parameters of a qubit cannot be perfectly obtained during its fabrication process, despite knowing its Hamiltonian capacity [13]. At present there is still a high discrepancy between two fabricated qubits. Superconducting qubits typically operate roughly in the range between 4 GHz and 8 GHz, an energy range that gives rise to the states of 0 and 1. Despite the complexity of superconduction circuits, control of the qubit entanglement is achieved by changing the direction of the circuit current corresponding to the 0 and 1 states [14].

The transmon consists of two islands connected by a capacitor and a Josephson inductance. In other words, its operation is based on an LC oscillator. The use of a Josephson inductance provides inductive energy in the form of a cosine function. Consequently, the transition of the transmon from the first to the second excitement state is less than 300 MHz. In practice, this difference is sufficient to confine the dynamics to a subspace of two qubits using pulses of a duration of about 20 ns. For two-qubit quantum gates, two inductances in parallel are needed. Figure 2 shows the structure of a transmon and the implementation of a quantum superconducting device.

For this implementation, the quantum chip must be completely isolated and at a temperature of only a few tens of millikelvins above absolute zero. To achieve such low temperatures, cryogenic technologies are required. Therefore, a quantum computer requires the use of thermal insulators to prevent noise that would compromise the quality of the qubit. So far, due to the state-of-the-art technologies requiring a high investment, only a few companies such as Google, IBM, Rigetti, Intel, and Alibaba have ventured into developing quantum computers based on superconducting circuits.



(**a**)



(**b**)

Figure 2. Superconducting qubits with (**a**) diagram of a transmon, and (**b**) implementation of a quantum processor unit with 4 transmon qubits.

2.2. Ion Traps

Another technology is 'ion traps'. Ion traps are electromagnetic and optical fields used to trap electrically charged individual atoms, i.e., ions [15]. These ions are pushed by their vibrations to entangle their spin states. Ion traps also require operating temperatures close to absolute zero in order to achieve better control of the quantum states of the ion. This is done by decreasing the atom's vibrational energy, thus exciting the atom to increase its internal energy and thereby transferring the quantum superposition of the electronic states to vibrational modes. In short, qubits are stored in the internal states of ions. The resulting electron detraction causes the ions to interact with each other through electromagnetic fields [16].

For any quantum operation, however basic, sequences of several pulses are required to minimize perturbation or decay as much as possible. In addition, the applied magnetic fields and laser sources are required to be very stable [17]. Organizations such as IonQ and the Massachusetts Institute of Technology are advancing the development of quantum computers using this technology [18].

2.3. Majorana Quasiparticles

Fermions are quasiparticles that have been theoretically defined to form excitations in superconductors [19]. Majorana fermions are particles that correspond identically to their antiparticles. Indium antimonide nanowires can be used to create qubits in this way. The production of Majorana fermions in the laboratory is an arduous task that requires the combination of the latest advances in nanotechnology, superconductivity, and material sciences. First, a crystal nanowire consisting of a 100-nanometer-long column of atoms is constructed. The wire is then connected to a circuit capable of measuring single electrons passing through it. Theoretically, if all the electrons in the wire are magnetized, fermions emerge from both ends of the wire.

The bond that exists between each pair of Majorana particles could be exploited to store quantum information in two different intertwined locations. This duplicity and the greater theoretical stability of these fermions could result in qubits that are more stable and less sensitive to local ambient noise. While a study partly funded by Microsoft claimed to have detected Majorana fermions [20], its authors themselves retracted the claim. There are also several subsequent studies that cast doubt on the detection of these quasiparticles [21]. The main advantage of this technology with respect to superconductors that store data in a single object is the possibility of a qubit encoding a single bit of information in several fermions. The qubit remembers whether it has moved clockwise or counterclockwise.

2.4. Spin Qubits

Quantum computers based on spin qubit technology operate by controlling the spin of electrons and electron holes in semiconductors [22]. The Loss–DiVicenzo quantum computer is the first quantum computer built employing this technology [23].

The Loss–DiVicenzo quantum computer works in the half-spin degree of freedom of individual electrons confined in quantum dots. Quantum dots are semiconductor nanostructures that confine the motion of electrons in all three spatial directions and were discovered in 1981 [24].

The Fermi gas model, an ideal free fermion system in which electrons do not interact with each other, is used to operate the Loss–DiVicenzo quantum computer. Two-dimensional electron gases are introduced in semiconductors such as gallium arsenide [25], silicon [26], germanium [27] and graphene [28].

Spin qubits require slightly higher operating temperatures than superconducting qubits, namely 1 kelvin at 20 millikelvins, a factor than can potentially reduce system complexity. Another apparent advantage is that spin qubits can potentially remain coherent longer than in other technologies. Finally, spin qubits have a much smaller size than superconducting qubits, which would facilitate the scalability of quantum computing [29].

2.5. Photonic Chips

A photonic quantum computer uses particles of light (i.e., photons) unlike other technologies, which use electrons. Photons can take advantage of quantum phenomena in a very efficient way because they do not have significant coupling with the outside. Optical quanta have a wide freedom in terms of frequency, linear momentum or spin, on which qubits can be implemented in a relatively simple way. Thus, these systems contain a set of optical devices, such as light sources, splitters, mirrors and detectors, which process photons. Each time a photon interacts with the splitter, the photon takes two paths simultaneously. Consequently, a quantum superposition state is achieved [30]. Among the latest advances, a quantum device that can perform photon entanglement with very high levels of precision and control on a large scale shows the potential of this design [31]. This technology could make mass fabrication of future optical quantum computers possible.

According to research at Xanadu, since optical quantum chips operate at room temperature, the underlying technology is more scalable than other cooling-based technologies such as superconducting and can be integrated into optical-based telecommunications infrastructures [32].

2.6. Quantum Annealing

Qubits are the lowest energy states of the superconducting loops in QPU using this technology. These processors do not use quantum versions of logic gates but generate energy diagrams with two valleys known as double-well potential. Each of the valleys corresponds to states 0 and 1 respectively. The qubit collapses to one of these valleys at the end of the process. However, it is possible to control the probability that the qubit collapses to one state or the other. To do this, the qubit is programmed by applying subjecting it to an external magnetic field. The field tilts the energy diagram increasing the probability of a qubit collapsing into a particular valley, as can be seen in Figure 3 [33].

This indicates how these processors operate in superposition. In addition, these qubits can be interlocked together with devices called couplers. A coupler can cause two qubits to collapse into the same or opposite states. These two interlocked qubits can be interpreted as a single object with four possible states [34]. In short, although unlike other quantum technologies quantum annealing technology cannot be used for general purposes, the work of the D-Wave Systems company claims that it represents a very interesting approach to optimization problems [35].

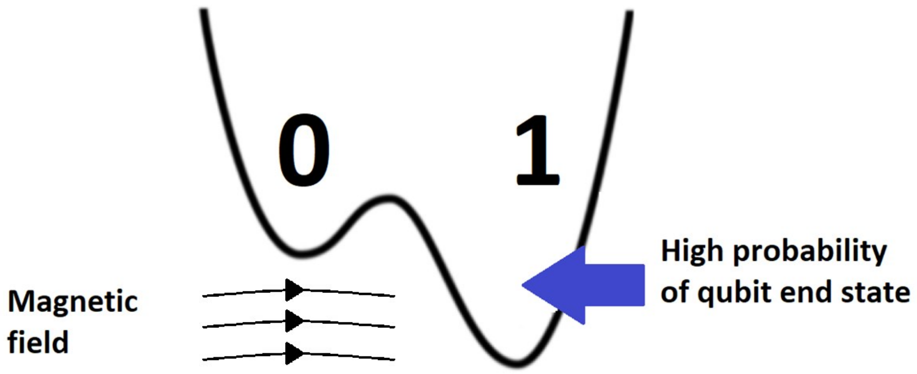


Figure 3. Quantum annealing allows the application of external magnetic fields to modify the probabilities of a qubit ending state.

Table 1 summarizes the current technologies related to quantum computing that have been addressed in this section.

Table 1. Available technologies for quantum computing.

Technology	Superconducting Circuits	Ion Traps	Majorana Quasiparticles	Spin Qubits	Photonic Quantum	Quantum Annealing
Description	Superconducting electrodes interconnected by Josephson junctions	Electromagnetic and optical fields to trap and control ions.	Majorana particles and their identical antiparticles act as qubits.	Control spin of electrons in semiconductors.	Optical devices to control photons.	Magnetic fields to alter energy diagrams.
Pros	Fast working. Build on the existing semiconductor industry.	Very stable. Highest achieved gate fidelities.	Very stable. Less sensitivity to noise.	Stable. Build on the existing semiconductor industry.	Operates at room temperature. Large scale integration. High precision.	Good performance in optimization applications.
Cons	Low coherence time. Must be kept cold.	Slow operation. Must be kept cold. Many lasers required.	Not technical evidence.	Only a few entangled. Must be kept cold.	High complexity.	Only a few applications.
Company support	IBM, Google, Quantum circuits, Rigetti, Intel, Alibaba, Amazon.	IonQ, Amazon.	Microsoft. Bell Labs.	Intel.	Xanadu.	D-Wave Systems.

2.7. Quantum Algorithms and Main Applications for Quantum Computing

Quantum algorithms are those that are executed on quantum computers and have computational advantages over algorithms executed on conventional computers.

The quantum paradigm does not consist of doing the same thing more quickly, but in executing quantum algorithms that allow certain operations to be performed in a totally different way, and, on many occasions, far more efficiently.

These advantages are usually in terms of speed and efficiency, especially in those problems that are not tractable on conventional computers. Quantum algorithms could present a substantial advantage when employing techniques that classical computation is unable to achieve, such as superposition or entanglement.

The computational complexity of an algorithm increases as the size of the algorithm increases in terms of variables or inputs. Complexity theory classifies problems into classes according to the relationship between the size of the problem and the resources required (mainly computational time and memory space) to reach a solution [36].

Problems are classified as (i) easy, those that use a constant number of operations or grow at a $\log^2(n)$ logarithmic rate; (ii) tractable, when their operations grow with a polynomial cost; (iii) intractable, when they grow at an exponential rate. If polynomial algorithms are found for problems that are currently classified as intractable, they undergo a change in their classification [37].

For example, if an algorithm must choose the optimal path among a million options, these paths must be encoded and analyzed in bits or qubits depending on whether it is run on a classical or quantum computer. A classical computer would need to analyze each of these paths one by one, requiring in the order of $N/2$ steps, i.e., 500,000 operations. However, a quantum computer can find an optimal path by taking advantage of quantum parallelism and finding the solution in *sqrt(N)* operations, i.e., 1000 tries.

In another class of algorithms, the quantum advantage is even greater. A conventional computer can perform the Fourier transform using $n2^n$ logic gates, while a quantum computer can solve a transform in n^2, which translates into exponential savings.

Thus, the advantage that quantum computing brings to the resolution of the transform is that for n qubits, the calculations are performed in 2^n amplitudes. This makes it possible to efficiently solve many problems that are not solvable on an ordinary computer.

Similarly, one of the possibilities of quantum computing is the application of quantum algorithms to artificial intelligence and machine learning applications [38]. Through machine learning, these algorithms can automatically learn to recognize recurring patterns from huge amounts of data. In these applications, there are a large number of combinations of analysis and computation, which in computational terms implies a very expensive optimization problem [39].

One of the main challenges of quantum computing are applications in physics and chemistry. The quantum characteristics of these devices give them the potential to solve molecular electronic structure simulation problems. These simulations have a wide range of applications such as the design of new materials or the discovery of new drugs [40].

Quantum computing can have a major impact on storage devices, permitting the storage of huge amounts of data in a very small space [9].

Many industries such as logistics, financial services or telecommunications require optimization algorithms. Quantum computing can be used to solve a wide variety of these problems, as in the application of Monte Carlo methods [35].

Another application for a quantum system is to increase the variety in *models* for pattern identification, recommendation systems, automatic classification and autonomous decision making [41].

An important role of quantum computing is also expected in the fields of cryptocurrencies and network security mechanisms. The computing power of quantum computers could render current encryption algorithms obsolete and open doors to new ways of protecting information and risk analysis [42,43].

3. Quantum Approaches for Machine Learning

In this section, several available approaches for quantum machine learning are presented.

3.1. Variational Quantum Circuits for Machine Learning

The quantum mode operation of the set of qubits is very fragile and unstable. In fact, the instability of the set of operating qubits of a quantum computer increases with the number of qubits and is related to the quality of the materials from which they are made. For this reason, the coherence time in which the qubits are stable is limited.

To efficiently solve problems such as unstructured database searches [44], the use of thousands of qubits with low error rates and high coherence times will be required. These quantum computers are not yet available, but noisy intermediate-scale quantum (NISQ) computers can already be used [45]. However, these computers are composed of numerous noisy qubits that remain stable for very few nanoseconds.

Variational quantum algorithms (VQAs) [46–49] use classical optimizers to train parameterized quantum circuits and represent an advancement in quantum computing running on NISQ computers [50].

Variational algorithms can be the basis for numerous applications, for example, in the design of a quantum classifier. The algorithm is similar in operation to *a* classical support-vector machine.

In machine learning, support-vector machines or SVMs are supervised learning models for solving classification and regression problems. Basically, SVMs build models capable of predicting whether a new point belongs to one category or another. Similarly, the quantum variational circuit performs hyperplane slices like a conventional SVM. On the other hand, kernel functions are used to solve nonlinear classification problems using linear classifiers. Classifiers based on quantum logic have no advantage over classical computers when the kernel function can be computed efficiently on a classical computer. *However,* on many occasions, when the feature space is very large and kernel functions are hard to find, quantum computing can be very useful [38].

The variational algorithm uses the Hilbert space in which the quantum processor operates to find the optimal hyperplane cut in the same way as a classical support-vector machine. The algorithm is run for training and for classification. The training stage uses an already classified dataset.

The algorithm is composed of three parts: (i) state preparation or feature map where the input data x are encoded; (ii) model circuit or optimization circuit with input parameters θ; and (iii) measurement stage z as shown in the following equation. A diagram with the three stages is shown in Figure 4.

$$f(x,\theta)=z \tag{2}$$

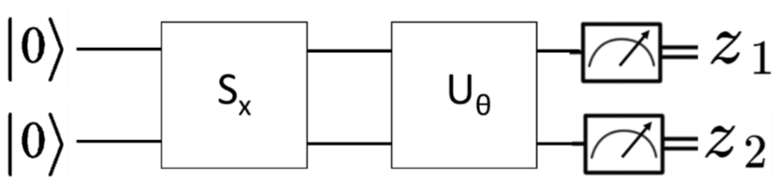


Figure 4. Variational quantum algorithm that applies the function $f(x, \theta) = z$, with a state preparation circuit S_x, where an input x is encoded, a model circuit U_θ where the parameters θ are applied, and the measurement stage.

3.2. Quantum Convolutional Neural Networks

Convolutional neural networks (CNNs) are a type of neural network in which neurons act as receptive agents, mimicking neurons in the primary visual cortex of a biological brain. The main applications of this type of network are related to visual computer tasks, such as, for example, image classification and segmentation, precision medicine or optical character recognition [51].

A CNN generally consists of three layers: a convolution layer, a pooling layer and fully connected layer. The convolution layer computes new pixel values from a linear combination of neighboring pixels with specific weights. The pooling layer aims to reduce the size of the feature map. Finally, the fully connected layer obtains the output through a linear combination of the remaining pixels with specific weights. The weights are calculated through training the neural network with an input dataset.

In the quantum version of this neural network, there is an additional first step of the convolution layer, encoding the image data into a quantum state.

The developed model of a quantum convolutional neural network (QCNN) in [52] performs well against noise in image recognition applications. In addition, the parameters it uses are independent of the size of the inputs, making it a suitable solution for currently available noisy quantum devices.

In study [53], a QCNN is used for quantum phase recognition (QPR) applications. In the experiment, the model discovers whether a quantum state belongs to a particular quantum phase. In addition, quantum error correction (QEC) optimization is performed, which consists of optimizing error correction in real experimental environments when the error model is unknown.

QCNNs require fewer parameters to obtain similar results compared to conventional convolutional neural networks. Therefore, both quantum and conventional large-scale neural networks can solve complex machine learning problems, but as the growth of Hilbert spaces is exponential, the quantum alternative is more efficient.

3.3. Quantum Distributed or Federated Machine Learning

One of the main limitations of quantum convolutional neural networks (QCNN) is that they are a centralized solution. The generation of deep learning models requires a huge amount of data. Some applications have public datasets for research, but in most advanced processes it is necessary to collect specific data in fields such as medical imaging or natural language processing (NLP). As sharing data from these applications can infringe on user rights, federated learning [54] can solve the problem if the model is trained directly at the user's source, and what is shared in the cloud are the parameters of the trained model. The components of a federated machine learning system are the central node and several client nodes that send the parameters calculated in the training to the central node.

Quantum distributed training consists of performing the training of a neural network through several quantum computers (that act as central and client nodes), with the objective of reducing the time needed to train the network. Hybrid quantum-classical neural networks can be used, although the concept is equally applicable to any type of machine learning model.

Some studies on network training [55,56] show that similar levels of precision can be obtained more quickly with quantum distributed machine learning. However, for the integration of quantum computers in future communication systems, it is necessary to create large-scale distributed quantum networks that surpass the quantum networks currently emerging [57].

3.4. Quantum Reinforcement Learning

In addition to supervised and unsupervised learning, there is a third area of machine learning known as reinforcement learning (RL). In reinforcement learning, intelligent agents make decisions to maximize cumulative reward. The scope of this approach is oriented towards exploring the unknown and exploiting available knowledge [58].

Conventional reinforcement learning systems have three main elements: (i) policy, (ii) reward function, and (iii) model of the environment. However, the implementation of quantum algorithms could be substantially different from traditional ones. Quantum machine learning has a potential acceleration of *O(sqrt(N))* for machine learning algorithms using quantum reinforcement learning [59].

In [60] quantum variational circuits are used to replace neural networks in reinforcement algorithms. Their conclusion is that QVC facilitates reaching a higher efficiency than traditional neural networks.

In [61] another experiment is presented in which the agent learning process is accelerated by using a quantum communication channel.

4. Materials and Methods

In this section, a variational quantum circuit is implemented. In addition, an experimental setup is defined for quantum simulation and execution applied to the detection of weak signals of the future.

4.1. Design of a Variational Quantum Algorithm for Quantum Machine Learning

As previously defined, a variational quantum algorithm is composed of a state preparation stage, a model circuit, and a measurement stage. The model circuit acts on a quantum state in which the input data x have been encoded.

The first step is to prepare the circuit by putting the qubits in a superposition state. To do this, Hadamard gates are introduced, which create a basic 50–50 superposition. This means that, when measuring the state, a result of '0' or '1' will be obtained with a 50% probability for each of the states. The Hadamard gate maps the Z-axis to the X-axis (and vice versa).

Once the quantum computer is in a superposition state, entanglement is used for state preparation Sx, which consists of embedding the classical data in a quantum state. The most common way to do this operation is amplitude encoding [62–65]. As a result, the coordinates of a vector are mapped into the amplitude values of a quantum state.

With this step, it is possible to map the input data in a nonlinear way into a quantum state, as can be seen in the following equation, which assumes a quantum feature map:

$$\Phi: \vec{x} \in \Omega \rightarrow |\Phi(\vec{x})\rangle\langle\Phi(\vec{x})| \tag{3}$$

Entanglement is achieved by combining Z quantum gates and two-qubit CNOTs as can be seen in Figure 5. This circuit acts on the initial state |00> and does not need to have a high depth to fulfill its function [38], and basically consists of Hadamard gates and a diagonal gate in the Pauli Z-basis [66].

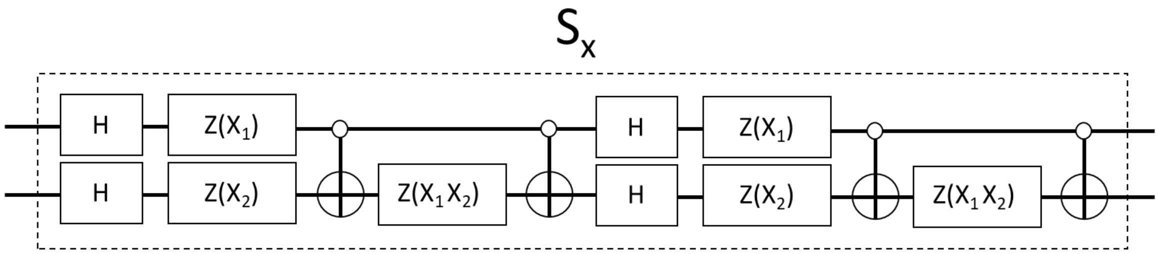


Figure 5. State preparation circuit Sx that maps the training and input data to a quantum feature map.

The Z-gate changes the sign of the amplitude when applied to the state of the qubit $|1\rangle$ and leaves it unchanged when operating with a qubit in the $|0\rangle$ state. Thus, the Z-gate shifts the qubit along the interval on the Bloch sphere over the surface of the sphere from $|0\rangle$ to $|1\rangle$, rotating values around the z-axis.

The CNOT or controlled NOT gate operates two-qubit registers. The CNOT inverts the second qubit if the first qubit has the state $|1\rangle$. However, the CNOT does not vary the state of the second qubit if the first qubit has the state $|0\rangle$.

Once the feature map has been created and the input data are encoded in qubit amplitudes, it is necessary to apply a model circuit U_θ, where θ is a set of classical parameters belonging to $R^{2n(l+1)}$. The parameters θ are optimized during the training phase.

The model circuit is composed of one-qubit quantum gates and quantum gates controlled by an additional single qubit. In this way, the algorithm retains a low depth [41]. The circuit is composed of *l* layers, being *l=5* in the experiment performed. By increasing the depth with larger *l*, higher success rates are expected [38].

The model circuit is shown in Figure 6. If a qubit with the state $\alpha|0\rangle + \beta|1\rangle$ is introduced at the input of a Y gate, the output will have the state $\beta i|0\rangle - \alpha i|1\rangle$, so it works as a combination of gates X and Z.

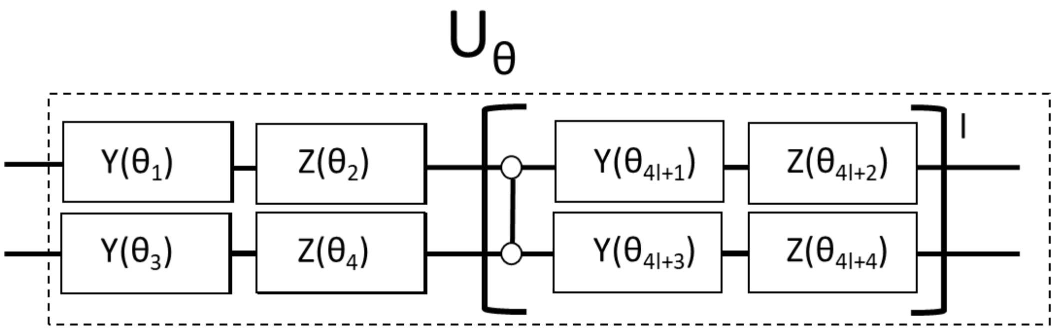


Figure 6. Model circuit UΘ that maps used for the classifier.

Once the model is executed, for a problem of classification 'y' with two labels, a binary measurement is performed on the output state of the qubits.

The probability of obtaining a specific outcome in the classification is described in the next equation:

$$py(x) = 2^{-1}(1 + y\langle\Phi(\vec{x})|U^T(\theta)\, f\, U(\theta)|\Phi(\vec{x})\rangle), \tag{4}$$

for a Boolean function $f{:}\{0,1\}^n$.

For decision making, the system is executed R times to obtain an empirical distribution.

An alternative to this circuit is to use a classical SVM to perform the classification instead of using a variational quantum circuit. In this case, the kernel is estimated directly within the quantum computer for a pair of input data [38].

4.2. Application to the Detection of Weak Signals of the Future

Weak signals are indicators of incipient changes that will have a future impact [67]. Even for experts, the detection of these events is complicated as their impact is as yet unpredictable.

The available studies on these types of signals are mostly qualitative [68], valid only for specific applications [69–71] and based on the opinion of experts and other stakeholders [72]. Even in cases where quantitative analysis is used, these studies are based only on structured data [73].

However, a system has been developed to assist experts in the detection of weak signals that are quantitative, multipurpose, and drawn from various types of unstructured data sources. This system has been successfully applied in various sectors such as solar panels [74], artificial intelligence [75] and remote sensing [44].

The data used by the system are all kinds of documents from a variety of sources, such as scientific articles, newspaper articles and social media posts. One of the most important phases of the system is text mining, in which all the words in these documents are classified into three different data classes as can be seen in Figure 7, where data are classified in three different classes: (i) possible weak signals, (ii) strong signals (those signals that have already reached the necessary impact to be known by a wide audience), and (iii) noise (words and expressions in which there has been no increase in usage). The two main factors are the average frequency of each word and a time weighted *increase* rate to give more value to the most recent occurrences.

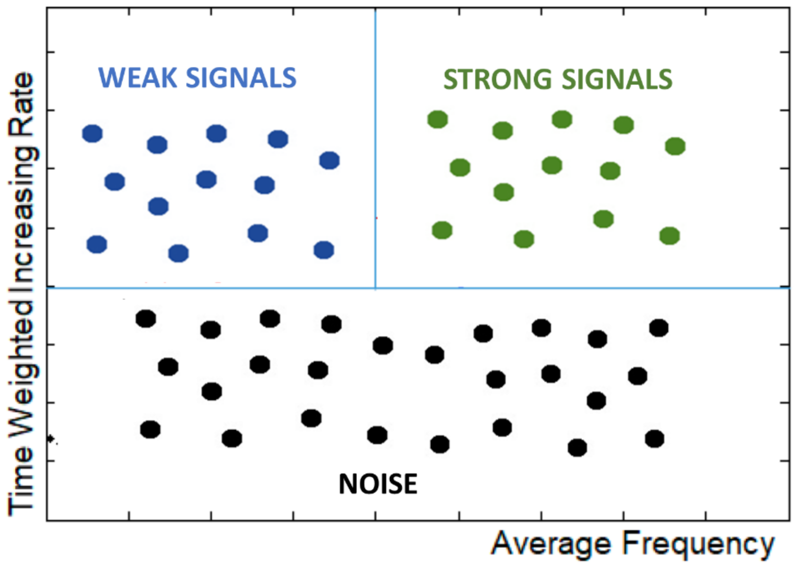


Figure 7. Keyword map showing the three different classes for an application for the detection of weak signals of the future.

Two different keyword maps are created with this structure. The first one is based on the 'degree of diffusion' or *DoD*, and the second one is based on the 'degree of visibility' or *DoV*. These two dimensions are part of the model of a weak signal by Hiltunen [76].

The degree of diffusion (*DoD*) uses the total number of documents where the keyword appears as an average frequency to define an increasing rate considering different periods of time. The degree of visibility (*DoV*) uses the total number of appearances of each keyword (independently of the number of documents) as an average frequency, to define an increasing rate considering again different periods of time. The keyword is considered a weak signal when in both keyword maps for *DoD* and *DoV* it is classified as weak signal.

In available studies, the opinion of experts [44] is used to define the thresholds that divide the word cloud into these three clusters. However, current results show that many of the expressions detected as weak signals may be false positives.

For this reason, a machine learning system to define the clusters in an automatic way may be more appropriate.

The process consists of creating a training set considering only those words classified as weak signals in both *DoV* and *DoD* maps. Then, the classifier is applied to classify the rest of the words to reduce the number of false positives detected.

In addition, the applications in which the system has been used has had an input dataset of about 90k documents, which implies about 150 million words. The processing advantage offered by quantum computers is a very interesting advance in the classification of these words.

In the next section, the system is analyzed by applying the quantum classifier described in this section to the medical imaging sector.

4.3. Quantum Circuit Simulation and Execution Framework

Qiskit [77] is a quantum simulation tool created by IBM for quantum software development using the Python programming language. Its functionalities are divided into four different components: (i) Terra, which provides tools for programming quantum circuits with levels close to machine code; (ii) Aqua, which provides already programmed quantum algorithms for the realization of applications in the fields of chemistry, AI, optimization and finance; (iii) Aer, *(a)* software simulator on small quantum devices; (iv) Ignis, a tool that characterizes noise in quantum devices to mitigate errors.

On the other hand, in IBM Quantum Experience [78] it is possible to access IBM quantum computers remotely to implement and run quantum algorithms in real environments. IBM's quantum processors are fabricated using superconducting transmon qubits that are at temperatures near absolute zero at IBM's New York research center.

For example, a test can be performed by testing a basic circuit as shown in Figure 8, consisting of a Hadamard gate applied only to one of the qubits and a CNOT gate controlled by this same qubit.

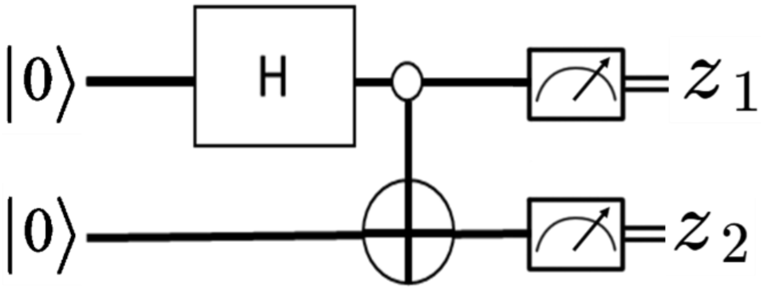


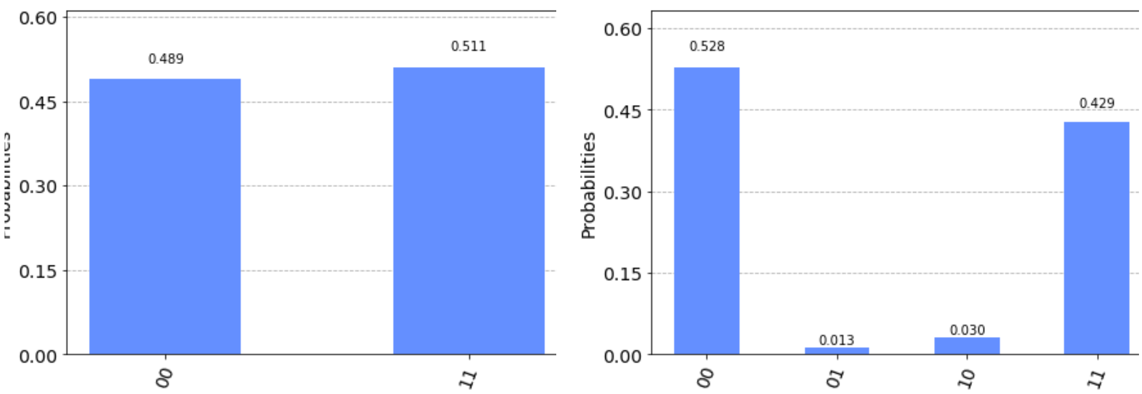
Figure 8. Basic circuit with a Hadamard gate, CNOT gate and binary measurements.

If the simulation of the circuit is performed using the Aer component of Qiskit, the result seen in Figure 9a is obtained, in which there is approximately a 50% probability of obtaining the state |00> and another 50% of obtaining |11>. However, when the same circuit is run on the 5-bit quantum computer 'ibmq_santiago' through IBM Quantum Experience, the result shown in Figure 9b is obtained, in which after running the circuit *for* 1024 shots, the states |01> and |10> are also found. This occurs because the simulator represents a perfect quantum device, but the real quantum device is noisy and susceptible to errors.

There are other frameworks for quantum programming such as Penny Lane [79], developed by Xanadu under the Apache 2.0 License. In addition, it allows interaction with real hardware through Amazon Braket [80] which allows access to several real quantum computers of various technologies.

5. Results

To perform the experiment, more than 40,000 ScienceDirect scientific articles, more than 1500 New York Times newspaper articles and more than 50,000 Twitter tweets between 2007 and 2017 were extracted and divided into 11 groups of documents from every year for the analysis. All these documents were related to the sector of medical imaging. A dataset of 937 elements has been generated to detect keywords related to weak signals of the future.



(**a**) (**b**)

Figure 9. Measurements obtained from the execution of the circuit from Figure 8 by (**a**) using the Aer component simulator of Qiskit, and (**b**) implementing and executing the circuit in a real quantum computer 'ibmq_santiago'.

Figure 10 shows the classification of these keywords by defining two thresholds to divide the word cloud as in [44] according to the *DoV* criteria. As previously stated, this simple methodology has the problem that many of the expressions detected as weak signals are false positives.

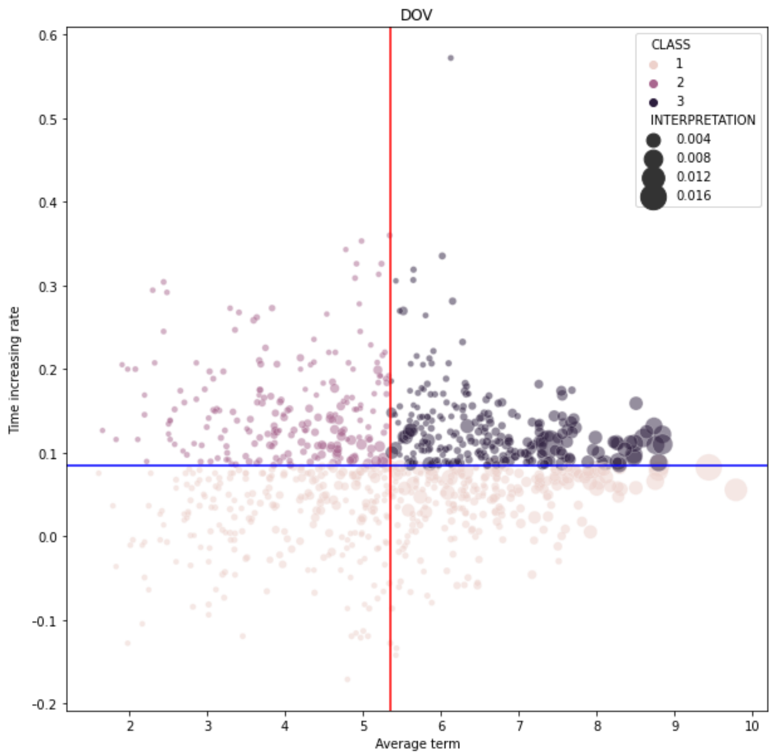


Figure 10. Definition of the three classes (1) weak signals, (2) strong signals and (3) noise, without using machine learning.

Table 2. shows some of the output results of the application of the system described in [44]. The full data set is available.

Keyword	*DoD*	Increasing Rate	*DoV*	Increasing Rate	WS
abdominal	59.250	0.120949	173.333	0.124932	1
ablation	629.167	−0.000349	106.667	0.040984	0
abuse	27.021	0.272509	4.029	0.121411	1
…					
window	300.833	0.171221	108.333	0.124527	1
young	39.750	0.039471	171.667	0.081619	0
zero	5.002	0.075291	24.167	0.075291	0

Table 2. Group of keywords from the weak signals test for medical imaging.

Figure 11 shows the result of applying support vector machines with different kernels in a conventional computer, including linear kernel, linear SVC, radial basis function (RBF) kernel and a polynomial kernel of degree 3.

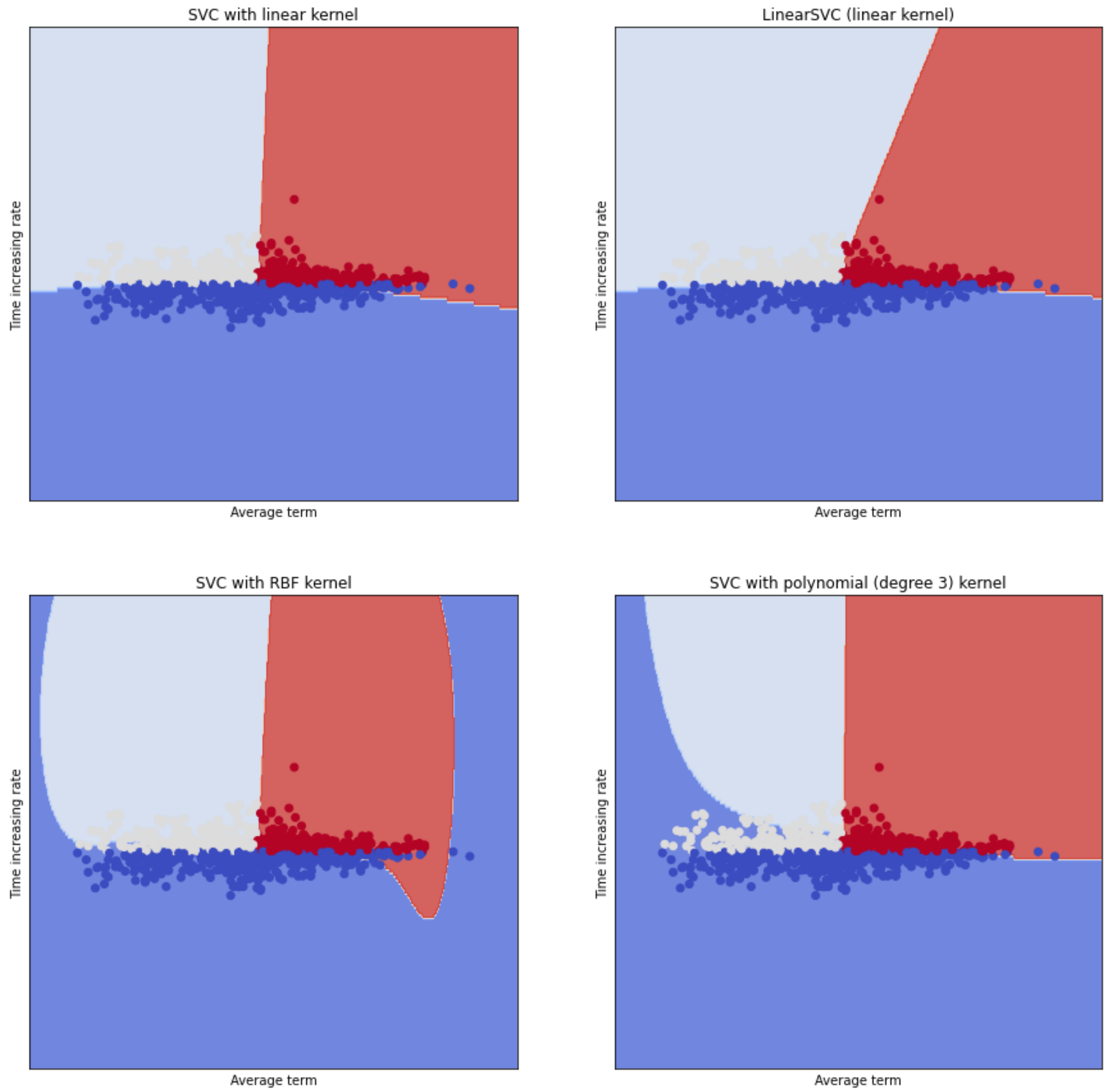


Figure 11. Result of applying support vector machines with different kernels.

To encode the data, the function "ZZFeatureMap" of qiskit was used, which encodes and implements the circuit shown in Figure 5. As a result, the input data was encoded in the amplitudes of a quantum state.

In the first quantum computation test, the complete dataset of 937 elements was used, divided in a training set of 749 elements and a test set of 188 elements. A total of 1,024 shots were programmed for each quantum measurement to obtain an accurate probabilistic distribution. However, when using the qiskit simulator and executing the quantum computer 'ibmq_santiago', the test did not obtain results after several days of uninterrupted execution. Because of this, it was decided to reduce the initial sample for a second test.

In the second test, a sample of 47 items was used. The training set was composed of 23 items classified as "noise", 12 items classified as "weak signals" and 12 items classified as "strong signals". When the system was used for a test set of 8 items classified as "noise", 3 items classified as "weak signals" and 1 item classified as "strong signals", a 70.21% success rate was obtained after its execution in 'ibmq_santiago'. A success rate of 78.72% was obtained after its execution using qiskit quantum simulator. Finally, a success rate of 61.70% was obtained after its execution using the classical algorithm. The algorithm 'SklearnSVM' is used to find the performance from classical computing.

These results are shown in Figure 12 where class A is 'noise', class B is 'weak signals' and class C is 'strong signals'.

```
Prediction from datapoints set:
  ground truth: ['A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A'
, 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'B', 'B'
, 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'C', 'C', 'C', 'C'
, 'C', 'C', 'C', 'C', 'C', 'C', 'C', 'C']
  prediction1:   ['A', 'A', 'A', 'A', 'A', 'A', 'A', 'B', 'A', 'A', 'B
', 'B', 'A', 'A', 'A', 'A', 'B', 'A', 'B', 'B', 'B', 'A', 'A', 'A', 'A
', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'C', 'A', 'A', 'A
', 'A', 'C', 'C', 'A', 'A', 'A', 'A', 'A']
  success rate: 61.702127659574465%
  prediction2:   ['C', 'A', 'A', 'A', 'A', 'A', 'A', 'B', 'A', 'B', 'B
', 'A', 'A', 'C', 'A', 'C', 'A', 'A', 'A', 'B', 'A', 'A', 'A', 'B', 'B
', 'A', 'B', 'B', 'B', 'B', 'C', 'B', 'B', 'B', 'B', 'C', 'C', 'C', 'C
', 'C', 'A', 'C', 'C', 'C', 'C', 'C', 'C']
  success rate: 78.72340425531915%
  prediction3:   ['C', 'A', 'C', 'A', 'A', 'A', 'A', 'C', 'A', 'B', 'C
', 'A', 'A', 'A', 'A', 'C', 'A', 'A', 'C', 'A', 'C', 'A', 'A', 'A', 'B
', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'C', 'C', 'C', 'C
', 'C', 'A', 'C', 'A', 'A', 'A', 'A', 'C']
  success rate: 70.2127659574468%
```

Figure 12. Result of applying quantum SVM algorithm to the input data for three classes where prediction1 is classical computer, prediction2 is quantum simulation and prediction3 is quantum real computer.

Another test was performed using a sample of 47 items but only two classes were considered: "weak signals" and "no weak signals". The latter includes both noise and strong signals. The training set was composed of 33 items classified as "no weak signals" and 14 items classified as "weak signals". When the system was used for a test set of 11 items classified as "noise" and 1 item classified as "weak signals", a 70.21% success rate was obtained after its execution in 'ibmq_santiago'. A success rate of 74.47% was obtained after its execution using qiskit quantum simulator. Finally, a success rate of 80.85% was obtained after its execution using the classical algorithm.

These results are shown in Figure 13 where class A is 'no weak signal' and class B is 'weak signals'.

The performance results of these tests are shown in Table 3.

Table 3. Prediction success rate for two and three classes using classical, quantum simulation and quantum computing.

Number of Classes	Classical	Quantum Simulation	Quantum Real Computer
2	80.75%	74.47%	70.21%
3	61.70%	78.72%	70.21%

```
Prediction from datapoints set:
  ground truth: ['A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A'
, 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A'
, 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'B', 'B', 'B', 'B'
, 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B']
  prediction1:   ['A', 'B', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'B
', 'A', 'A', 'A', 'B', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A
', 'A', 'A', 'A', 'A', 'B', 'A', 'A', 'A', 'A', 'A', 'A', 'B', 'A', 'A
', 'B', 'B', 'A', 'B', 'B', 'B', 'A', 'B']
  success rate: 80.85106382978724%
  prediction2:   ['A', 'A', 'A', 'A', 'B', 'A', 'B', 'A', 'A', 'B', 'B
', 'A', 'A', 'A', 'A', 'A', 'A', 'B', 'B', 'A', 'A', 'B', 'A', 'B', 'A
', 'A', 'A', 'B', 'B', 'B', 'A', 'A', 'A', 'A', 'A', 'B', 'B', 'B', 'B
', 'B', 'B', 'B', 'B', 'B', 'B', 'A', 'B']
  success rate: 74.46808510638297%
  prediction3:   ['A', 'A', 'B', 'A', 'B', 'B', 'A', 'A', 'A', 'A', 'A
', 'A', 'A', 'A', 'A', 'A', 'B', 'B', 'A', 'A', 'A', 'A', 'B', 'A', 'B
', 'B', 'A', 'B', 'B', 'B', 'A', 'A', 'A', 'B', 'B', 'B', 'B', 'B', 'B
', 'B', 'B', 'B', 'B', 'B', 'B', 'A', 'B']
  success rate: 70.2127659574468%
```

Figure 13. Result of applying quantum SVM algorithm to the input data for two classes where prediction1 is classical computer, prediction2 is quantum simulation and prediction3 is quantum real computer.

6. Discussion and Conclusions

There is a growing interest in the application of quantum computing to machine learning. For this reason, several quantum algorithms are emerging with the potential to be implemented on current quantum devices. However, because these devices are very noisy, these algorithms must be fast and robust against decoherence.

In addition, there are strong arguments showing that circuits of great simplicity are difficult to simulate on conventional computers [81,82]. Some studies such as [62] claim that there is an exponential improvement in this class of algorithms when the input data are in superposition. However, initially the data are encoded in classical logic and the operations required to encode the data in quantum logic considerably reduce this acceleration. [38].

In the experiment presented in the previous section, it has been proven that similar success rates can be obtained between classical and quantum computing.

In the first test, a comparative study has been performed for the detection of three different classes. Thus, a distinction has been made between 'weak signals', 'strong signals' and those expressions that have not led to an increase in their use, i.e., 'noise'. The results show that the quantum simulator obtained the highest level of success in its prediction, with 78.72%, being much higher than the performance of the classical computer. The quantum computer execution 'ibmq_santiago' has resulted in a lower prediction success but also higher than that obtained in the classical computer.

In the second test, the same comparative study was carried out, but considering only two classes. The 'weak signals' quadrant has been distinguished from the rest of the points that are classified as 'no weak signals' including both 'noise' and 'strong signals'. The results show that the classical computer obtained the highest success rate while the quantum simulator obtained a slightly lower performance. However, the use of the real IBM quantum computer 'ibmq_santiago', obtained a success rate of 70.21%, being lower than that obtained using the quantum simulator due to the noise resulting from the available quantum devices.

These results show an influence of the dataset on the performances, to which obtaining a suitable feature map is key. In addition, as stated in [38], available quantum classifiers do not offer a considerable advantage when the feature vector kernel can be efficiently obtained on a classical computer. In fact, this same source reveals that it is possible to achieve 100% performance if the optimal feature map is available. In future research, it is necessary to further study the potential for improvement if suitable feature maps are detected for other real datasets.

Although quantum computing seems to bring many advantages to machine learning algorithms, the results of this study show that with large datasets both the simulator and the real quantum computer have been unable to present results in less time than a classical computer. The need to encode the data in a quantum state and the need to run each circuit a large number of shots as a measure of robustness to noise confirm that current quantum technology cannot yet be considered an alternative to classical computing [59].

The four main problems that need to be solved are:

1. First, encoding the input data is especially complex when compared to data processing, and this complexity may come at a high cost to quantum algorithms, which may lose some of their fast-processing advantage.
2. Second, the output measurements are not easily interpretable, since qubit entanglement computes many solutions until the states collapse into a particular final state.
3. Third, quantum computation is thought to offer an advantage to solve large problems, but due to hardware limitations, this has not yet been demonstrated.
4. Finally, better algorithm evaluation systems are needed to appreciate the advantages of quantum algorithms over classical algorithms.

Some possible sources of improvement have been presented in Section 3. Quantum convolutional neural networks are giving good results to complex problems and have been shown to be more robust to noise than other quantum alternatives [52]. Quantum simulator success rates or similar are expected to be attained by the reduction of noise.

On the other hand, the integration of quantum variational circuits in reinforcement learning techniques allows higher efficiencies to be achieved [60].

Another problem that has been detected is that all these applications are centralized. Although there are some distributed solutions available, they usually combine mixed classical and quantum technologies. Although single quantum devices allow a high level of parallelism, federated machine learning improves the security and privacy of users by not sharing their information in the cloud. However, quantum computers are not yet available to the general public. In the future, distributed solutions may increase the speed of handling complex problems that require large amounts of data to be processed.

This paper has described the technologies available for the implementation of quantum devices. Some of these technologies require extreme conditions for their operation, such as temperatures close to absolute zero. In addition, all these technologies are still under development, so current quantum devices are very noisy and have different fabrication yields. However, several organizations are investing considerable resources in obtaining technologies that demonstrate the supremacy of quantum computing.

In addition, the latest advances related to the field of quantum machine learning have been presented in Section 3. Next, a possible application for quantum machine learning related to future weak signal detection is presented. In this application, a large number of words and documents are taken into account in the process of detecting future trends. The computational advantage that quantum computing possesses adds immense value to this field.

The experiments performed show promising results on technologies that are still under development. Some of the results show higher efficiencies using quantum computers compared to classical ones. In the future, new methods for the detection of suitable future maps, the combination with other technologies and machine learning algorithms, and the application of distributed solutions will allow major improvements in quantum machine learning applications.

Author Contributions: Conceptualization, I.G.-B., S.M., A.C., Y.M. and J.M.; methodology, I.G.-B., S.M. and J.M.; software, I.G.-B. and S.M.; validation, I.G.-B., S.M., A.C., Y.M. and J.M.; formal analysis, I.G.-B.; investigation, I.G.-B., S.M. and J.M.; resources, I.G.-B., S.M. and Y.M.; data curation, I.G.-B. and S.M.; writing—original draft preparation, I.G.-B., writing—review and editing, I.G.-B., A.C., Y.M. and J.M.; visualization, I.G.-B.; supervision, J.M.; project administration, I.G.-B. and J.M.; funding acquisition, I.G.-B. and J.M. All authors have read and agree to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors acknowledge Ismael Faro who is Chief Architect of Quantum Computing Cloud and Software at IBM for his help in this work. This research is also part of a PhD from the Department of Electronics Engineering of the Universitat Politècnica de València.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Moore, G.E. Cramming more components into integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. *IEEE Solid State Circuits Soc. Newsl.* **2006**, *11*, 33–35, doi:10.1109/N-SSC.2006.4785860.
2. Martín-Palma, R.J. Quantum tunneling in low-dimensional semiconductors mediated by virtual photons. *AIP Adv.* **2020**, *10*, 015145, doi:10.1063/1.5133039.
3. Benioff, P. Quantum mechanical models of Turing machines that dissipate no energy. *Phys. Rev. Lett.* **1982**, *48*, 1581–1585, doi:10.1103/PhysRevLett.48.1581.
4. Turing, A.M. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* **1937**, *42*, 230–265, doi:10.1112/plms/s2-42.1.230.
5. Trabesinger, A. Quantum simulation. *Nat. Phys.* **2012**, *8*, doi:10.1038/nphys2258.
6. Nielsen, M.A.; Chuang, I.L. *Quantum Computation and Quantum Information*; Cambridge University Press: Cambridge, USA, 2010.
7. Bloch, F. Nuclear induction. *Phys. Rev.* **1946**, *70*, 460–474, doi:10.1103/physrev.70.460.
8. Yorke, B.A.; Beddard, G.; Owen, R.L.; Pearson, A.R. Time-resolved crystallography using the Hadamard transform. *Nat. Methods* **2014**, *11*, 1131–1134, doi:10.1038/nmeth.3139.

9. Ruiz, I.E.K. Quatum Computing: Practical applications that classical computing cannot solve. Bachelor's Thesis, University Carlos III, Madrid, Spain, 2019.
10. Yamaguchi, K.; Watamura, N.; Hotta, M. Quantum information capsule and information delocalization by entanglement in multiple-qubit systems. *Phys. Lett. A* **2019**, *383*, 1255–1259, doi:10.1016/j.physleta.2019.01.036.
11. Gibney, E. Hello quantum world! Google publishes landmark quantum supremacy claim. *Nature* **2019**, *574*, 461–462, doi:10.1038/d41586-019-03213-z.
12. Koch, J.; Yu, T.M.; Gambetta, J.; Houck, A.A.; Schuster, D.I.; Majer, J.; Blais, A.; Devoret, M.H.; Girvin, S.M.; Schoelkopf, R.J. Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **2007**, *76*, 042319, doi:10.1103/PhysRevA.76.042319.
13. Lucas, A. Using formulations of many np problems. *Front. Phys.* **2014**, *2*, doi:10.3389/fphy.2014.00005.
14. Wallraff, A.; Schuster, D.I.; Blais, A.; Frunzio, L.; Huang, R.S.; Majer, J.; Schoelkopf, R.J. Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **2004**, *431*, 162–167, doi:10.1038/nature02851.
15. Cirac, J.I.; Zoller, P. A scalable quantum computer with ions in an array of microtraps. *Nature* **2000**, *404*, 579–581, doi:10.1038/35007021.
16. Seif, A.; Landsman, K.A.; Linke, N.M.; Figgatt, C.; Monroe, C.; Hafezi, M. Machine learning assisted readout of trapped ion qubits. *J. Phys. B At. Mol. Opt. Phys.* **2018**, *51*, doi:10.1088/1361-6455/aad62b.
17. Eschner, J. *Quantum Computation with Trapped Ions*; Institut de Ciències Fotòniques: Castelldefels, Spain, 2006.
18. Monroe, C.R.; Schoelkopf, R.J.; Lukin, M.D. Quantum connections. *Sci. Am.* **2016**, *314*, 50–57, doi:10.1038/scientificamerican0516-50.
19. Mourik, V.; Zuo, K.; Frolov, S.M.; Plissard, S.R.; Bakkers, E.; Kouwenhoven, L.P. Signatures of Majorana fermions in hybrid superconductor-semiconductor nanowire devices. *Science* **2012**, *336*, 1003–1007, doi:10.1126/science.1222360.
20. Zhang, H.; Liu, C.X.; Gazibegovic, S.; Xu, D.; Logan, J.A.; Wang, G.; van Loo, N.; Boomer, J.D.S.; de Moor, M.W.A.; Car, D.; et al. RETRACTED ARTICLE: Quantized Majorana conductance. *Nature* **2018**, *556*, 74–79, doi:10.1038/nature26142.
21. Frolov, S. Quantum computing's reproducibility crisis: Majorama fermions. *Nature* **2021**, *592*, 350–352, doi:10.1038/d41586-021-00954-8.
22. Vandersypen, L.M.K.; Eriksson, M.A. Quantum computing with semiconductor spins. *Phys. Today* **2019**, *72*, 38, doi:10.1063/PT.3.4270.
23. Loss, D.; DiVincenzo, D.P. Quantum computation with quantum dots. *Phys. Rev. A Am. Phys. Soc.* **1998**, *57*, 120–126, doi:10.1103/physreva.57.120.
24. Ekimov, A.I.; Efros, A.L.; Onushchenko, A.A. Quantum size effect in semiconductor microcrystals. *Solid State Commun.* **1985**, *56*, 921–924, doi:10.1016/S0038-1098(85)80025-9.
25. Bluhm, H.; Foletti, S.; Neder, I.; Rudner, M.; Mahalu, D.; Umansky, V.; Yacoby, A. Dephasing time of GaAs electron-spin qubits coupled to a nuclear bath exceeding 200 μs. *Nat. Phys.* **2010**, *7*, 109–113, doi:10.1038/nphys1856.
26. Wang, S.; Querner, C.; Dadosh, T.; Crouch, C.H.; Novikov, D.S.; Drndic, M. Collective fluorescence enhancement in nanoparticle clusters. *Nat. Commun.* **2011**, *2*, doi:10.1038/ncomms1357.
27. Watzinger, H.; Kukučka, J.; Vukušić, L.; Gao, F.; Wang, T.; Schäffler, F.; Zhang, J.J.; Katsaros, G. A germanium hole spin qubit. *Nat. Commun.* **2018**, *9*, 3902, doi:10.1038/s41467-018-06418-4.

28. Trauzettel, B.; Bulaev, D.V.; Loss, D.; Burkard, G. Spin qubits in graphene quantum dots. *Nat. Phys.* **2007**, *3*, 192–196, doi:10.1038/nphys544.
29. Guerreschi, G.G.; Park, J. Two-step approach to scheduling quantum circuits. *Quantum Sci. Technol.* **2018**, *3*, 0345003, doi:10.1088/2058-9565/aacf0b.
30. Brotons-Gisbert, M.; Proux, R.; Picard, R.; Andres-Penares, D.; Branny, A.; Molina-Sánchez, A.; Sánchez-Royo, J.F.; Gerardot, B.D. Out-of-plane orientation of luminescent excitons in two-dimensional indium selenide. *Nat. Commun.* **2019**, *10*, 3913, doi:10.1038/s41467-019-11920-4.
31. Wang, J.; Paesani, S.; Ding, Y.; Santagati, R.; Skrzypczyk, P.; Salavrakos, A.; Tura, J.; Augusiak, R.; Mancinska, L.; Bacco, D.; et al. Multidimensional quantum entanglement with large-scale integrated optics. *Science* **2018**, *360*, 285–291, doi:10.1126/science.aar7053.
32. Arrazola, J.M.; Bergholm, V.; Brádler, K.; Bromley, T.R.; Collins, M.J.; Dhand, I.; Fumagalli, A.; Gerrits, T.; Goussey, A.; Helt, L.G.; et al. Quantum circuits with many photons on a programmable nanophotonic chip. *Nature* **2021**, *591*, 54–60, doi:10.1038/s41586-021-03202-1.
33. Bian, Z.; Chudak, F.; Israel, R.B.; Lackey, B.; Macready, W.G.; Roy, A. Mapping constrained optimization problems to quantum annealing with application to fault diagnosis. *Front. ICT* **2016**, *3*, 14, doi:10.3389/fict.2016.00014.
34. Johnson, M.W.; Amin, M.H.S.; Gildert, S.; Lanting, T.; Hamze, F.; Dickson, N.; Harris, R.; Berkley, A.J.; Johansson, J.; Bunyk, P.; et al. Quantum annealing with manufactured spins. *Nature* **2011**, *473*, 194–198, doi:10.1038/nature10012.
35. King, A.D.; Raymond, J.; Lanting, T.; Isakov, S.V.; Mohseni, M.; Poulin-Lamarre, G.; Ejtemaee, S.; Bernoudy, W.; Ozfidan, I.; Smirnov, A.Y.; et al. Scaling advantage over path-integral Monte Carlo in quantum simulation of geometrically frustrated magnets. *Nat. Commun.* **2021**, *12*, doi:10.1038/s41467-021-20901-5.
36. Cobham, A. The intrinsic computational difficulty of functions. In *Logic, Methodology and Philosophy of Science, Proceedings of the 1964 International Congress*; North-Holland Publishing Company: Amsterdam, The Netherlands, 1965; pp. 24–30.
37. Goldreich, O. *Computational Complexity: A Conceptual Perspective*; Cambridge University Press: Cambridge, USA, 2008.
38. Havlíček, V.; Córcoles, A.D.; Temme, K.; Harrow, A.W.; Kandala, A.; Chow, J.M.; Gambetta, J.M. Supervised learning with quantum-enhanced feature spaces. *Nature* **2019**, *567*, 209–212, doi:10.1038/s41586-019-0980-2.
39. Barkoutsos, P.K.I.; Nannicini, G.; Robert, A.; Tavernilli, I.; Woerner, S. Improving variational quantum optimization using CVaR. *Quantum* **2020**, *4*, 256, doi:10.22331/q-2020-04-20-256.
40. Liu, H.; Hao Low, G.; Steiger, D.S.; Häner, T.; Reiher, M.; Troyer, M. Prospects of Quantum Computing for Molecular Sciences, *Quantum Phys.* **2021**, arXiv:2102.10081.
41. Schuld, M.; Bocharov, A.; Svore, K.; Wiebe, N. Circuit-centric quantum classifiers. *Phys. Rev. A* **2020**, *101*, 032308, doi:10.1103/PhysRevA.101.032308.
42. Woerner, S.; Egger, D.J. Quantum risk analysis. *Nat. Quantum Inf.* **2019**, *5*, 15, doi:10.1038/s41534-019-0130-6.
43. Eskandarpour, R.; Gokhale, P.; Khodaei, A.; Chong, F.T.; Passo, A.; Bahramirad, S. Quantum Computing for Enhancing Grid Security. *IEEE Trans. Power Syst.* **2020**, *35*, 4135–4137, doi:10.1109/TPWRS.2020.3004073.

44. Griol-Barres, I.; Milla, S.; Cebrián, A.; Fan, H.; Millet, J. Detecting weak signals of the future: A system implementation based on Text Mining and Natural Language Processing. *Sustainability* **2020**, *12*, 7848.
45. Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2018**, *2*, 79, doi:10.22331/q-2018-08-06-79.
46. Zhou, L.; Wang, S.-T.; Choi, S.; Pichler, H.; Lukin, M.D. Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices algorithm. *arXiv* **2019**, arXiv:1812.01041v2.
47. Farhi, E.; Goldstone, J.; Gutmann, S. A quantum approximate optimization algorithm. *arXiv* **2014**, arXiv:1411.4028.
48. McClean, J.R.; Romero, J.; Babbush, R.; Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.* **2016**, *18*, 023023, doi:10.1088/1367-2630/18/2/023023.
49. Romero, J.; Olson, J.P.; AspuruGuzik, A. Quantum autoencoders for efficient compression of quantum data. *Quantum Sci. Technol.* **2017**, *2*, 045001, doi:10.1088/2058-9565/aa8072.
50. Cerezo, M.; Arrasmith, A.; Babbush, R.; Benjamin, S.C.; Endo, S.; Fujii, K.; McClean, J.R.; Mitarai, K.; Yuan, X.; Cincio, L.; et al. Variational Quantum Algorithms. *arXiv* **2020**, arXiv:2012.09265.
51. Le Cun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444, doi:10.1038/nature14539.
52. Wei, S.; Chen, Y.; Zhou, Z.; Long, G. A Quantum Convolutional Neural Network on NISQ Devices. *arXiv* **2021**, arXiv:2104.06918.
53. Cong, I.; Choi, S.; Lukin, M.D. Quantum convolutional neural networks. *Nat. Phys.* **2019**, *15*, 1273–1278, doi:10.1038/s41567-019-0648-8.
54. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, Ft. Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.
55. Yen-Chi Chen, S.; Yoo, S. Computational Science Initiative. *arXiv* **2021**, arXiv:2103.12010.
56. Chehimi, M.; Saad, W. Quantum Federated Learning with Quantum Data. *arXiv* **2021**, arXiv:2106.00005.
57. Pirandola, S.; Braunstein, S.L. Physics: Unite to build a quantum internet. *Nature* **2016**, *532*, 169–171, doi:10.1038/532169a.
58. Sutton, R.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, USA, 1998.
59. Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum machine learning. *Nature* **2017**, *549*, 195–202, doi:10.1038/nature23474.
60. Lockwood, O.; Si, M. Reinforcement learning with quantum variational circuits. In Proceedings of the 16th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-20), Palo Alto, CA, USA, 19–23 October 2020; arXiv:2008.07524.
61. Saggio, V.; Asenbeck, B.E.; Hamann, A.; Strömberg, T.; Schiansky, P.; Dunjko, V.; Friis, N.; Harris, N.C.; Hockberg, M.; Englund, D.; et al. Experimental quantum speed-up in reinforcement learning agents. *Nature* **2021**, *591*, 229–233, doi:10.1038/s41586-021-03242-7.
62. Rebentrost, P.; Mohseni, M.; Lloyd, S. Quantum support vector machine for big data classification. *Phys. Rev. Lett.* **2014**, 113, 130503, doi:10.1103/PhysRevLett.113.130503.

63. Kerenedis, I.; Prakash, A. Quantum recommendation systems. *arXiv* **2016**, arXiv:1603.08675.
64. Wiebe, N.; Braun, D.; Lloyd, S. Quantum algorithm for data fitting. *Phys. Rev. Lett.* **2012**, *109*, 050505, doi:10.1103/PhysRevLett.109.050505.
65. Schuld, M.; Fingerhuth, M.; Petruccione, F. Implementing a distance-based classifier with a quantum interference circuit. *EPL* **2017**, *119*, 60002, doi:10.1209/0295-5075/119/60002.
66. Mitarai, K.; Negoro, M.; Kitagawa, M.; Fujii, K. Quantum circuit learning. *arXiv* **2018**, arXiv:1803.00745.
67. Ansoff, H.I. Managing Strategic Surprise by Response to Weak Signals. *Calif. Manag. Rev*. **1975**, *18*, 21–33, doi:10.2307/41164635.
68. McGrath, J.; Fischetti, J. What if compulsory schooling was a 21st century invention? Weak signals from a systematic review of the literature. *Int. J. Educ. Res*. **2019**, *95*, 212–226, doi:10.1016/j.ijer.2019.02.006.
69. Koivisto, R.; Kulmala, I.; Gotcheva, N. Weak signals and damage scenarios—Systematics to identify weak signals and their sources related to mass transport attacks. *Technol. Forecast. Soc. Chang*. **2016**, *104*, 180–190, doi:10.1016/j.techfore.2015.12.010.
70. Davis, J.; Groves, C. City/future in the making: Masterplanning London's Olympic legacy as anticipatory assemblage. *Futures* **2019**, *109*, 13–23, doi:10.1016/j.futures.2019.04.002.
71. Awan, F.M.; Saleem, Y.; Minerva, R.; Crespi, N. A Comparative Analysis of Machine/Deep Learning Models for Parking Space Availability Prediction. *Sensors* **2020**, *20*, 322, doi:10.3390/s20010322.
72. Van Veen, B.L.; Ortt, R.; Badke-Schaub, P. Compensating for perceptual filters in weak signal assessments. *Futures* **2019**, *108*, 1–11, doi:10.1016/j.futures.2019.02.018.
73. Thorleuchter, D.; Van den Poel, D. Idea mining for webbased weak signal detection. *Futures* **2015**, *66*, 25–34, doi:10.1016/j.futures.2014.12.007.
74. Griol-Barres, I.; Milla, S.; Millet, J. Implementación de un sistema de detección de señales débiles de futuro mediante técnicas de minería de textos. (Implementation of a weak signal detection system by text mining techniques). *Rev. EsDoc. Cient*. **2019**, *42*, 234, doi:10.3989/redc.2019.2.1599.
75. Griol, I.; Milla, S.; Millet, J. Improving strategic decision making by the detection of weak signals in heterogeneous documents by text mining techniques. *AI Commun*. **2019**, *32*, 347–360, doi:10.3233/AIC-190625.
76. Hiltunen, E. The future sign and its three dimensions. *Futures* **2007**, *40*, 247–260, doi:10.1016/j.futures.2007.08.021.
77. Qiskit. Available online: https://qiskit.org/ (accessed on 30 May 2021).
78. IBM Quantum Experience. Available online: https://quantum-computing.ibm.com/ (accessed on 30 May 2021).
79. Penny Lane. Available online: https://pennylane.ai/ (accessed on 30 May 2021).
80. Amazon Bracket. Available online: https://aws.amazon.com/braket/ (accessed on 30 May 2021).
81. Terhal, B.M.; DiVincenzo, D.P. Adaptive quantum computation, constant depth quantum circuits and arthur-merlin games. *Quantum Inf. Comput.* **2004**, *4*, 134–145.
82. Bremner, M.J.; Montanaro, A.; Shepherd, D.J. Achieving quantum supremacy with sparse and noisy commuting quantum computations. *Quantum* **2017**, *1*, 8, doi:10.22331/q-2017-04-25-8.

5. General discussion

This work provides the design and implementation of the architecture of a system to detect weak signals of the future in an automatic, multi-source, and quantitative way, that is also applicable to any field.

A complete study of the state-of-the-art research in the detection of weak signals of the future has been carried out in the two articles corresponding to chapters 2 and 3, corresponding to Task 1 of this work. In addition, Chapter 4 includes a complete discussion on available quantum technologies.

Current systems are models conceived to face specific problems, but they cannot be used for other applications beyond the one for which they have been designed [1-6]. Some solutions are based on quantitative systems but focus on the use of data from a single class [7] or are too dependent on expert opinion [8].

For the implementation of the system, several input datasets have been created on different topics, corresponding to Task 2: solar panels [9], artificial intelligence [10], remote sensing [11] and medical imaging [12].

In most data mining applications, the input dataset is divided into different sets to train, test and validate the system. Nevertheless, as weak signals present a low number of appearances, it is very difficult to assure they will be detected in a validation set. For this reason, new evaluation methodologies are required. Tasks 3, 4, 5 and 6 are completed with the design and implementation of the system.

The first version of the system was the simplest and was presented in [9]. This version of the system includes an information gathering and integration stage, an Extract-Transformation-Loading stage in the data warehouse, an information classification phase, a text mining phase and, finally, the interpretation and evaluation of the results obtained.

In this version of the system, a database of “keywords” was necessary. A list of standard factors provided by UNESCO was used. Therefore, only the words belonging to this list are considered.

The system was tested with a dataset on the field of "solar panels". A high correlation was observed between the results obtained using this architecture in comparison with another similar study that only used a dataset with documents of a single type [7]. In this study, only two of the three dimensions from the Hiltunen model were considered. As the implemented system defined in this thesis considers the three dimensions, the obtained results are more accurate.

The use of different types of sources also provides new possibilities for evaluating results. Another possibility is to assign different weights for every source depending on the field or application under study.

The second version of the system was presented in [10] (Chapter 2). In order to detect all potential weak signals, every term in every document of the dataset is considered as a keyword (except the stop words that are previously removed). Thanks to this action, the term under study does not need to be in a standard list of keywords to be detected, and new words or expressions that could become new trends can be identified. Another improvement of the system is the automatic assignment of categories for each of these terms. Categories are automatically created from the topics of the scientific articles where these words appear. Therefore, every keyword could be assigned to multiple categories. Some of the automated created categories are: “Applied Software Computing”, “Brain Research”, “Biomedical”, “Nuclear Energy”, “Expert Systems”, “Computer and Structures” among many others. Creating a dynamic and automatic list of categories depending on the sector and the real content of the dataset brings more accuracy and relevance to the study.

Furthermore, in this system the concept of degree of transmission (DoT) has been introduced, in order to include the third dimension of Hiltunen's future sign model [13], known as interpretation.

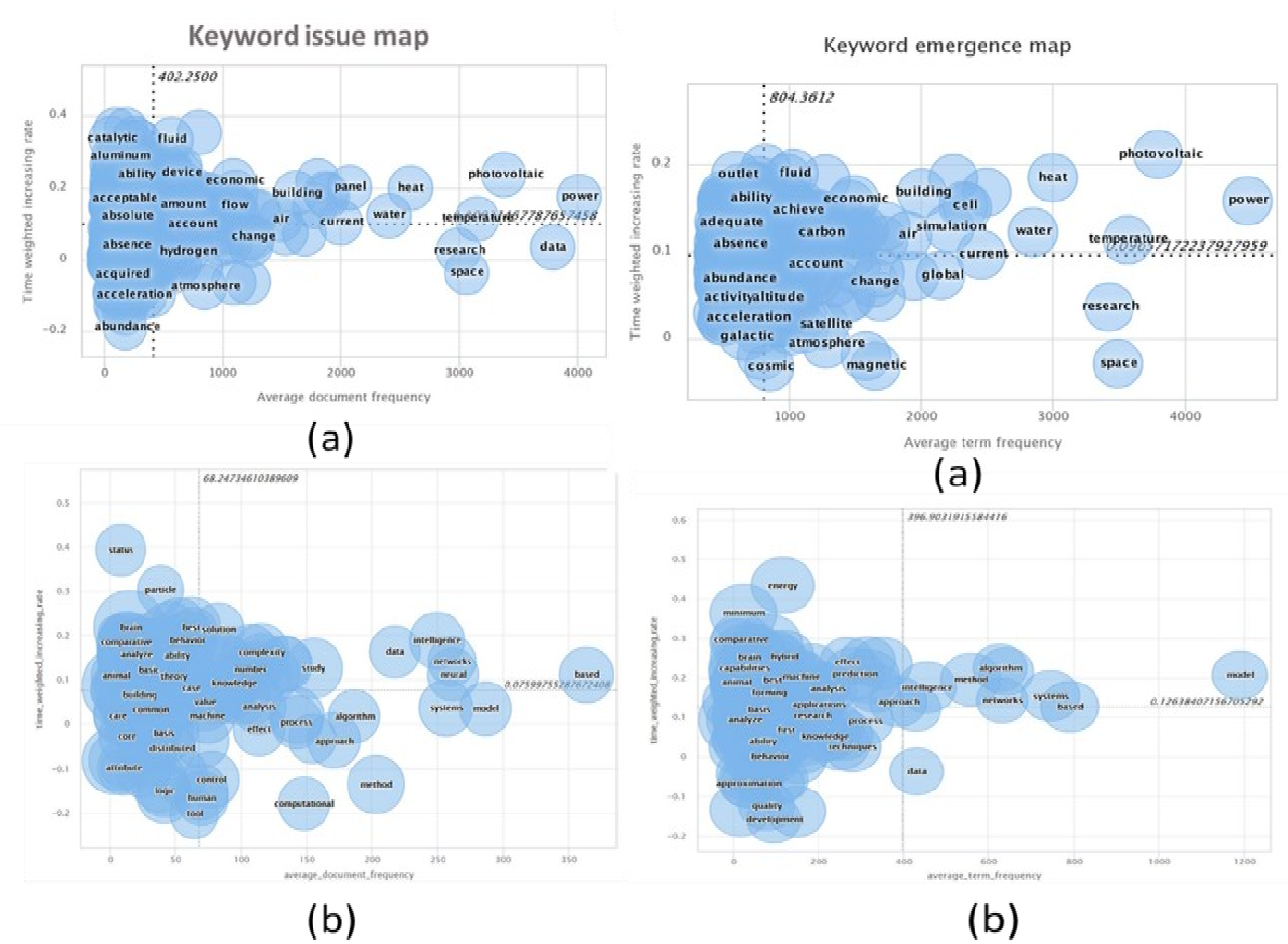


Figure 1. Keyword Issue Map (related to DoD) for (a) Solar Panels and (b) Artificial Intelligence and Keyword Emergence Map (related to DoV) for (a) Solar Panels and (b) Artificial Intelligence

If a word is included in both maps, it is more likely that it is related to a weak signal. However, the information provided by a single word is not sufficient to aid in the decision-making process. A word can have different meanings and can be related to different topics.

Disciplines such as Natural Language Processing (NLP) can be applied to provide more depth to the information obtained [14]. In this way, instead of considering only single words, this version of the system included a multi-word expressions analysis stage to obtain more complete and useful results as can be seen in Figure 2.

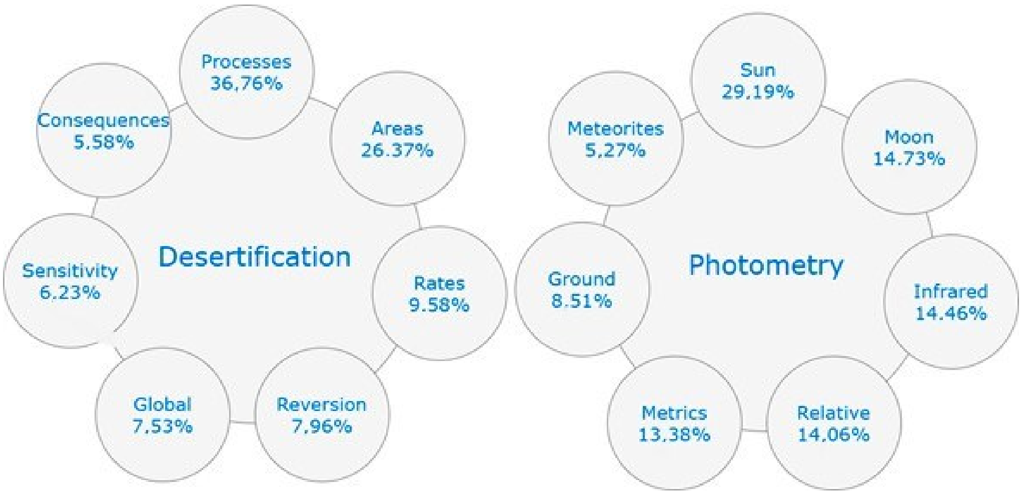


Figure 2. Multi-Word Expressions Related to the words: "desertification" and "photometry".

This second version of the system was tested with a dataset of documents related to "artificial intelligence".

The third version of the system is presented in [11] (Chapter 3). Thanks to the collaboration with experts from the Royal Institute of Technology (KTH) in Stockholm, it has been possible to use the system with a specific dataset on the subject of "remote sensing".

New improvements include two evaluation methods. The first method of evaluation is based on the detection of strong signals in a new input dataset, with more recent documents than the ones considered in the experiment.

This method is based on finding strong signals in recent documents, with terms that were identified as weak signals in previous documents. In this way, it is confirmed that the weak signals have become strong signals in a posterior period. A test was carried out with a dataset of remote sensing documents dated between 2018 and 2019 to confirm this method.

The second method of evaluation consists in comparing the results obtained with the vision of experts. A considerable number of terms identified as weak signals matched the list of terms identified by an independent panel of remote sensing experts.

The implementation of this system was carried out on both CPUs and parallel GPUs. The execution times show that the parallel GPU implementation is up to nine times faster compared with the standard sequential alternative using CPUs, completing Task 7.

Table 1. Execution times.

Stage	CPU	GPU	Improvement
Data Warehouse creation	6257 min	688 min	89,00%
Category assignation	68 min	9 min	86,76%
Text mining	124 min	11 min	91,13%
Multi-word expressions	43 min	6 min	86,05%
TOTAL	**6492 min**	**714 min**	**89,00%**

Although the results obtained are very useful for experts and decision makers, introducing new natural language processing techniques [15] is still necessary to improve the accuracy of the system.

As previously described in this work, quantum computing offers new possibilities for many applications including text mining and neural networks. Chapter 4 includes a description of the main technologies that are being developed to implement qubits and quantum computers. This study analyses their main pros and cons and the quantum computing platforms that are currently available.

One of the main problems of the quantum computers currently available is that data coherence is valid in a very limited time. For this reason, designers of new quantum algorithms must consider this reality of current quantum devices.

In all previous versions of the system for the detection of weak signals, the opinion of experts is used to define the thresholds that divide the word cloud into these three clusters: weak signals, strong signals, and noise. As a result, current results still rely on a decision carefully made by experts. To obtain more accurate results, a machine learning subsystem could be added to automatically define the thresholds for these three clusters.

The training set consists of all terms detected as weak signals in both keyword maps. The classifier assigns a class to all the other words.

The fourth version of the system is presented in [16] (Chapter 4) and includes a new quantum classifier that has been used with a new dataset of documents related to the "medical imaging" sector, as proposed in Task 8. This dataset consists of about 90k documents, which implies about 150 million words. The proposed quantum circuit is based on support vector machines (SVMs) and is applied to solve classification problems.

The processing advantage offered by quantum computers is a very interesting advance in the classification of these keywords. Some studies [17-18] show that circuits that are implemented with great simplicity in quantum logic require a great multitude of resources in conventional devices. Other studies [19] show that if the input data in algorithms related to machine learning are encoded in superposition, the benefit obtained is exponential. However, the required time to encode the input data in quantum logic reduces this improvement [20].

The results of the implemented quantum system show that it is possible to obtain success rates similar to those of classical computing. In the experiment carried out, the performance of the three-class system (weak signal, strong signal, noise) has been compared in (i) classical binary logic, (ii) quantum computing executing the algorithm in the IBM quantum simulator, and (iii) quantum computing by running the algorithm on a real IBM quantum computer. The results obtained with the quantum simulator show a success in the prediction of 78.72%, much higher than the performance obtained in the conventional computer (61.70%). Execution on the quantum computer 'ibmq_santiago' has also obtained results with greater success (70.21%) than the classical computer. Table 2 shows the confusion matrixes of this experiment.

However, the results obtained in the experiment with two classes (weak signal, non-weak signal) show a better performance in the classical computer. These results prove that there is a big influence of the encoding of the dataset on the performance of the system.

Table 2. Confusion matrixes in (i) classical binary logic, (ii) quantum computing executing the algorithm in the IBM quantum simulator, and (iii) quantum computing by running the algorithm on a real IBM quantum computer, for three classes.

	A (WS)	B (SS)	C (N)
A (WS)	**16**	2	9
B (SS)	7	**10**	0
C (N)	0	0	**3**

(i) Classical computer

	A (WS)	B (SS)	C (N)
A (WS)	**16**	1	1
B (SS)	4	**10**	0
C (N)	3	1	**11**

(ii) Quantum simulator

	A (WS)	B (SS)	C (N)
A (WS)	**15**	1	5
B (SS)	1	**11**	0
C (N)	7	0	**7**

(iii) Quantum real computer

Some of the most important results are summarized by fields in the following paragraphs:

1. Solar Panels

'Africa', 'Tunisia', 'Sahara' and 'Tunur' are terms that were detected as related to weak signals in the experiment related to the field of solar panels with articles prior to 2014. Some news articles confirmed the interest of the creation of future projects related to these locations, such as "Huge-scale Saharan solar energy project could provide Spain with electricity by 2015" [21], "Huge Tunisian solar park hopes to provide Saharan power to Europe" [22] or "Megaproject to bring solar energy from the Sahara to Europe" [23]. These articles confirm that the project called 'TuNur' is one of the most important ones at the moment.

In addition, 'zinc' was one of the materials that were detected as related to solar panels. According to NSW Department of Planning and Environment, some newer photovoltaic technology uses zinc oxide to achieve higher solar-cell efficiency via increased energy conversion [24].

2. Artificial intelligence

Some of the multiword expressions detected are becoming more and more important in the field of artificial intelligence now.

For example, 'desertification reversion' was only present in a few studies and now is becoming trendy in new available works such as [25] in which the impact of the desertification reversion of the eastern Hobq desert in China in greenhouse gasses emissions is studied.

Another example is "cooperative robotics" that is becoming trendy in artificial intelligence for health [26-27].

3. Remote Sensing

As previously reported, terms classified as strong signals in 2018-2019 were detected as weak signals in the test related to the field of remote sensing. The word "UVSQ-Sat" has become a strong signal and appears for the first time in the title of a published paper in 2019 [28]. Other words such as NOAA, InSAR, Rosetta and SRTM have also become strong signals.

The group of experts in this field reported the following findings:

- Remote sensing services are becoming more popular thanks to higher resolutions and better data collecting techniques. For this reason, experts expected several expressions detected as weak signals such as "adaptative encoding" and "voxel".
- Experts also expected new applications for remote sensors in agriculture and especially desertification. Some terms related to desertification and other agricultural factors were detected as weak signals.
- The experiment detected Interferometric synthetic aperture radar or "InSAR", which is a radar technique used in geodesy and remote sensing, as a weak signals. Experts reported that this term is becoming very relevant.
- According to experts, West Africa is becoming a very relevant location for remote sensing applications [29–34]. "West Africa" was detected as a weak signal multiword expression in the experiment.

Appendix 2 shows a complete list of the detected single terms present in both keyword maps.

4. Medical imaging

Some interesting multiword expressions were detected such as:

- 'clinical intelligence', a term to explore AI-based clinical decision support that is becoming more and more important [35].
- 'clinical reasoning', is a core component of clinical competency that is used in all patient encounters from simple to complex presentations, and it has become a high trend in the last years [36-38].
- 'brain computer', a term that is becoming trendy related to interfaces between human and artificial intelligences [39].
- 'abstract language', another term that is becoming important in neuroscience [40].

Figure 3 shows two maps of multiword expressions related to 'brain' and 'clinical'.

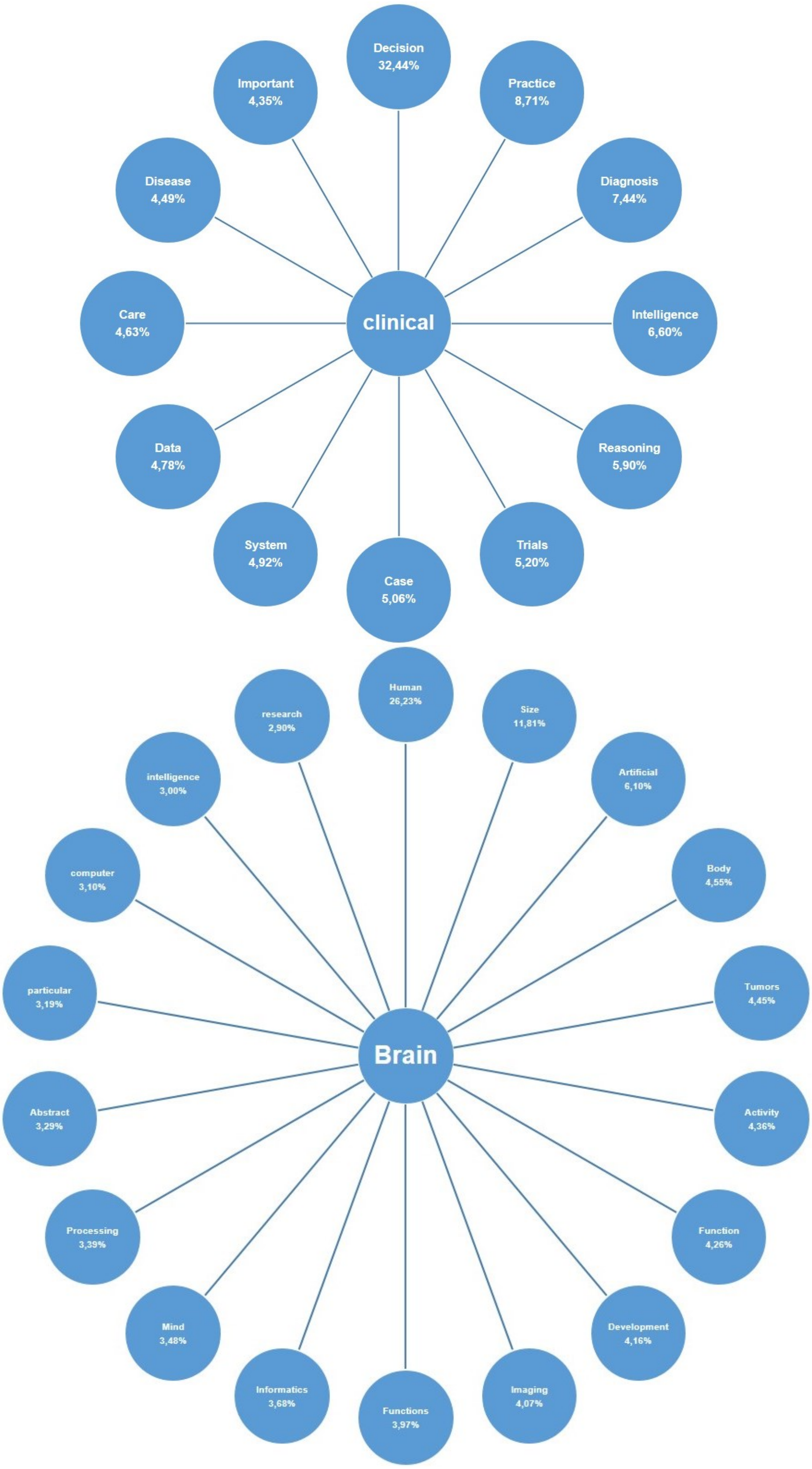


Figure 3. Multi-Word Expressions Related to the words "clinical" and "brain".

References

1. R. Koivisto, I. Kulmala and N. Gotcheva, Weak signals and damage scenarios – Systematics to identify weak signals and their sources related to mass transport attacks, *Finland Technological Forecasting and Social Change,* **2016,** 104, 180–190. doi:10.1016/j.techfore.2015.12.010.
2. O. Kuusi and E. Hiltunen, *The Signification Process of the Future Sign*, Finland Futures Research Centre ebook 4/2007, **2007**, p. 7.
3. S. Liu, M. Yamada, N. Collier and M. Sugiyama, Change-point detection in time-series data by relative density-ratio estimation, *Neural Networks,* **2013,** 43, 72–83. doi:10.1016/j.neunet. 2013.01.012.
4. M. Mannermaa, *Tulevaisuuden Hallinta Skenaariot Strate- giatyoskentelyssa (Managing the Future, Scenarios in Strategy Work)*, WSOY, Porvoo, **1999**, p. 227.
5. S. Mendonca, M.P. Cunha, J. Kaivo-Oja and F. Ruff, Wild cards, weak signals and organizational improvisation, FEUNL Working Paper No. 432, **2003**.
6. R. MohamadiBaghmolaei, N. Mozafari and A. Hamzeh, Continuous states latency aware influence maximization in social networks, *AI Communications*, **2017**, 30(2) (2017), 99–116. doi:10. 3233/AIC-170720.
7. J. Yoon, Detecting weak signals for long-term business opportunities using text mining of web news, Expert Systems *with Applications*, **2012**, 39 12543–12550. doi:10.1016/j. eswa.2012.04.059.
8. K. Haegeman, E. Marinelli, F. Scapolo, A. Ricci and A. Sokolov, Quantitative and qualitative approaches in future- oriented technology analysis (FTA): From combination to integration, *Technol. Forecast. Soc. Chang.* **2013**, 80, 386–397. doi:10.1016/j.techfore.2012.10.002.
9. Griol-Barres, I.; Milla, S.; Millet, J. Implementación de un sistema de detección de señales débiles de futuro mediante técnicas de minería de textos. (Implementation of a weak signal detection system by text mining techniques). *Rev. EsDoc. Cient*. **2019**, 42, 234, doi:10.3989/redc.2019.2.1599.
10. Griol, I.; Milla, S.; Millet, J. Improving strategic decision making by the detection of weak signals in heterogeneous documents by text mining techniques. *AI Commun*. **2019**, 32, 347–360, doi:10.3233/AIC-190625.
11. Griol-Barres, I.; Milla, S.; Cebrián, A.; Fan, H.; Millet, J. Detecting weak signals of the future: A system implementation based on Text Mining and Natural Language Processing. *Sustainability* **2020**, 12, 7848.
12. Griol-Barres, I; Milla, S.; Cebrián, A.; Mansoori, Y; Millet, J. Variational Quantum Circuits for Machine Learning. Application for the Detection of Weak Signals. *Applied Sciences* **2021**,11, 6427, doi: 10.3390/app11146427
13. Hiltunen, E. The future sign and its three dimensions. *Futures* **2007**, 40, 247–260, doi:10.1016/j.futures.2007.08.021.

14. Mendonca, S.; Cunha, M.P.; Kaivo-Oja, J.; Ruff, F. Wild Cards, Weak Signals and Organizational Improvisation. *Futures* **2004**, *36*, 201–218.
15. Zhang, W.; Yoshida, T.; Tang, X. Text classification based on multi-word with support vector machine.*Knowl. Based Syst.* **2008**, *21*, 879–886.
16. Griol-Barres, I.; Milla, S.; Cebrián, A.; Mansoori, Y.; Millet, J. Variational Quantum Circuits for Machine Learning. Application for the Detection of Weak Signals. *Applied Sciences*, **2021**, 11, 6427, doi: 10.3390/app11146427.
17. Terhal, B.M.; DiVincenzo, D.P. Adaptive quantum computation, constant depth quantum circuits and arthur-merlin games. *Quantum Inf. Comput.* **2004**, 4, 134–145.
18. Bremner, M.J.; Montanaro, A.; Shepherd, D.J. Achieving quantum supremacy with sparse and noisy commuting quantum computations. *Quantum* **2017**, 1, 8, doi:10.22331/q-2017-04-25-8.
19. Rebentrost, P.; Mohseni, M.; Lloyd, S. Quantum support vector machine for big data classification. *Phys. Rev. Lett.* **2014,** 113, 130503, doi:10.1103/PhysRevLett.113.130503.
20. Havlíček, V.; Córcoles, A.D.; Temme, K.; Harrow, A.W.; Kandala, A.; Chow, J.M.; Gambetta, J.M. Supervised learning with quantum-enhanced feature spaces. *Nature* **2019**, 567, 209–212, doi:10.1038/s41586-019-0980-2.
21. Huge-scale Saharan solar energy project could provide Spain with electricity by 2015, https://english.elpais.com/elpais/2011/11/17/inenglish/1321510846_850210.html, *El País*, **2011**.
22. Huge Tunisian solar park hopes to provide Saharan power to Europe, https://www.theguardian.com/environment/2017/sep/06/huge-tunisian-solar-park-hopes-to-provide-saharan-power-to-europe, *The Guardian*, **2017**.
23. Megaproyecto para llevar energía solar desde el Sáhara hasta Europa, https://www.lavanguardia.com/natural/20141022/54417391167/megaproyecto-tunur-energia-solar-electricidad-sahara-europa.html, La Vanguardia, **2014**.
24. Top Four metals and minerals that help solar panels work, https://www.planning.nsw.gov.au/News/2019/Top-four-metals-and-minerals-that-help-solar-panels-work, NSW Department of Planning and Environment, **2019**.
25. Wang, B.; Duan, Y.; Wang, W. Desertification reversion alters soil greenhouse gas emissions in the eastern Hobq Desert, China. *Environ Sci Pollut Res*, **2020,** 27**,** 15624–15634, doi:10.1007/s11356-020-08117-4.
26. Zhai, A.; Wang, J.; Zhang, H.; Lu, G.; Li H. Adaptive robust synchronized control for cooperative robotic manipulators with uncertain base coordinate system. *ISA Trans.* **2021**, 26:S0019-0578(21)00403-1. doi: 10.1016/j.isatra.2021.07.036.
27. Schleer, P.; Drobinsky, S.; de la Fuente, M.; Radermacher, K. Toward versatile cooperative surgical robotics: a review and future challenges. *Int J Comput Assist*

Radiol Surg. **2019**, 14(10):1673-1686. doi: 10.1007/s11548-019-01927-z.
28. Meftah, M.; Damé, L.; Keckhut, P.; Bekki, S.; Sarkissian, A.; Hauchecorne, A.; Bertran, E.; Carta, J.P. UVSQ-SAT, a Pathfinder CubeSat Mission for Observing Essential Climate Variables. *Remote Sens.* **2020**, *12*, 92.
29. Badmos, O.S.; Rienow, A.; Callo-Concha, D.; Greve, K.; Jürgens, C. Urban Development in West Africa—Monitoring and Intensity Analysis of Slum Growth in Lagos: Linking Pattern and Process. *Remote Sens.* **2018**, *10*, 1044.
30. Thomson, E.R.; Malhi, Y.; Bartholomeus, H.; Oliveras, L.; Gvozdevaite, A. Mapping the Leaf Economic Spectrum across West African Tropical Forests Using UAV-Acquired Hyperspectral Imagery. *Remote Sens.* **2018**, *10*, 1532.
31. Samasse, K.; Hanan, N.P.; Tappan, G.; Diallo, Y. Assessing Cropland Area in West Africa for Agricultural Yield Analysis. *Remote Sens.* **2018**, *10*, 1785.
32. Anchang, J.Y.; Prihodko, L.; Kaptué, A.T.; Ross, C.W.; Ji, W.; Kumar, S.S. Trends in Woody and Herbaceous Vegetation in the Savannas of West Africa. *Remote Sens.* **2019**, *11*, 576.
33. Jung, H.C.; Getirana, A.; Arsenault, K.R.; Holmes, T.R.H.; McNally, A. Uncertainties in Evapotranspiration Estimates over West Africa. *Remote Sens.* **2019**, *11*, 892.
34. Mondal, P.; Liu, X.; Fatoyinbo, T.E.; Lagomasino, D. Evaluating Combinations of Sentinel-2 Data and Machine-Learning Algorithms for Mangrove Mapping in West Africa. *Remote Sens.* **2019**, *11*, 2928.
35. Romero-Brufau, S.; Wyatt, K.D.; Boyum, P.; Mickelson, M.; Moore, M.; Cognetta-Rieke, C. A lesson in implementation: A pre-post study of providers' experience with artificial intelligence-based clinical decision support. *Int J Med Inform*. **2020**, 137:104072. doi: 10.1016/j.ijmedinf.2019.104072.
36. Huhn, K.; Gilliland, S.J.; Black, L.L.; Wainwright, S.F.; Christensen, N. Clinical Reasoning in Physical Therapy: A Concept Analysis. *Phys Ther*. **2019,** 99(4):440-456. doi: 10.1093/ptj/pzy148.
37. Richards, J.B.; Hayes, M.M.; Schwartzstein, R.M. Teaching Clinical Reasoning and Critical Thinking: From Cognitive Theory to Practical Application. *Chest*. **2020**, 158(4):1617-1628. doi: 10.1016/j.chest.2020.05.525.
38. Thampy, H.; Willert, E.; Ramani, S. Assessing Clinical Reasoning: Targeting the Higher Levels of the Pyramid. *J Gen Intern Med*. **2019**, 34(8):1631-1636. doi: 10.1007/s11606-019-04953-4.
39. Paulmurugan, K.; Vijayaragavan, V.; Ghosh, S.; Padmanabhan, P.; Gulyás, B. Brain-Computer Interfacing Using Functional Near-Infrared Spectroscopy (fNIRS). *Biosensors (Basel)*. **2021,**11(10):389. doi: 10.3390/bios11100389.
40. Spunt, R.P.; Adolphs, R. The neuroscience of understanding the emotions of others. Neurosci Lett. **2019,** 693:44-48. doi: 10.1016/j.neulet.2017.06.018.

6. General conclusions and future work

6.1 General conclusions

In this doctoral thesis, a system for the detection of weak signals of the future is conceptualized, implemented, and optimized, with analysis and validation in four different fields.

The main objective of this doctoral thesis has been achieved by the conceptualization, implementation, and hardware optimization of this system. The main characteristics of the system are described in the following table compared to most available studies.

Table 1. Comparing implemented system with available studies.

Implemented system	Available previous works
Quantitative analysis of the input data.	Mainly qualitative analysis.
Multiple fields: Model only dependent on the input dataset.	Specific model for a specific topic.
Unstructured input data. Text mining and Natural Language Processing.	Mainly structured data.
All words are considered keywords.	Pre-determined keywords.
Input data from multiple sources.	Mainly one source type.
All 3-dimensions from Hiltunen are considered.	Mainly one or two dimensions are considered.
Hardware optimization by quantum computing.	Mainly theoretical models or conventional software.

Specific Objective 1 is described and achieved in Chapter 2 in which the complete system for the detection of weak signals is defined and implemented. To test the system, an input dataset of unstructured documents in the field of artificial intelligence is created.

Specific Objective 2 is described and achieved in Chapter 2, 3 and 4, but main validation techniques have been defined in Chapter 3. The system has been tested in four different sectors: solar panels, artificial intelligence (Chapter 2), remote sensing (Chapter 3) and medical imaging (Chapter 4). Table 2 shows a list of multiword expressions related to the four fields that were considered.

Several methodologies to evaluate the experiment have been defined and applied in Chapter 3. Both methods have provided interesting results. The first one is based on studying if terms and expressions detected as weak signals have become strong signals with a new analysis with an input dataset with more recent documents in the same sector. The main limitation of this method is that new recent documents are required for this validation. However, the test that has been carried out in Chapter 3 proves that the system has predicted some terms that became strong signals in the period dated between 2018 and 2019.

The second evaluation method is based on the comparation of a list of detected weak signals with the opinion of experts in the field. A considerable number of words and expressions matched the list of terms identified by a group of experts on remote sensing. Consulting a group of experts allows an immediate evaluation but has also the limitation that they can only agree on what it is within their field of knowledge.

Despite these limitations, the performance of the system shows interesting results that prove that the designed system can provide useful information for decision-makers.

Table 2. Some multiword expressions detected as weak signals in the experiments related to the four fields under study.

Solar Panels				Medical Imaging			
Engineering		**Splitting**		**Encoding**		**Voxel**	
Climate	21,43%	Water	83,21%	Image	20,79%	Grid	23,60%
Thermal	20,85%	Spectral	4,66%	Transformation	18,35%	Image	18,81%
Application	19,79%	Technology	3,63%	System	18,12%	Volume	13,43%
Mechanical	18,06%	Cycles	3,06%	Reduce	12,66%	Classification	13,08%
System	12,81%	Efficient	2,58%	Pattern	10,69%	Coding	11,92%
Greenhouse	4,76%	Reactor	1,44%	Adaptative	9,76%	Spectroscopic	10,16%
Biosatellite	2,30%	Dioxide	1,41%	Slice	9,64%	Similarity	9,00%

Artificial Intelligence				Remote Sensing			
Brain		**Cooperative**		**Desertification**		**Photometry**	
Human	26,23%	Control	18,54%	Processes	36,76%	Sun	29,19%
Size	11,81%	Learning	14,42%	Areas	26,37%	Moon	14,73%
Artificial	6,10%	Robotics	10,98%	Rates	9,58%	Infrared	14,46%
Body	4,55%	Mobile	9,38%	Reversion	7,96%	Relative	14,46%
Tumors	4,45%	Intelligent	8,92%	Global	7,53%	Measurements	13,38%
Activity	4,36%	Coevolution	8,01%	Sensitivity	6,23%	Ground	8,51%
Computer	3,10%	Hybrid	3,89%	Consequences	5,58%	Meteorites	5,27%

Specific Objective 3 is described and achieved in Chapter 4 in which a new subsystem based on quantum computing is created, implemented, and tested for a better hardware optimization. Chapter 3 also includes the implementation and execution of the system in CPUs and parallel GPUs. The hardware optimization show that parallel GPU implementation provides nine times acceleration over using CPUs.

The implementation of the text mining subsystem on a quantum computer based on a support vector machine algorithm has brought a new paradigm in the detection of weak signals of the future. Thanks to the IBM Quantum Experience programme, the algorithm has been executed in a real available quantum computer by IBM. The obtained results show that a quantum simulator and the execution in the real quantum computer has provided better success performances than in classical digital devices. However, the requirements to encode a large input dataset into a quantum state, and the need to run each experiment many times as a measure of robustness to decoherence has also proved that quantum computing is still not able to obtain results in less time than a classical computer.

6.2 Challenges and future work

In order to obtain more accurate results, the system will be improved by adding new natural language processing tools and the generation of new validation techniques. Some natural language processing tools have been already applied such as stemming, eliminating stop words or multi word analysis. However, new techniques should be implemented in the future as for example, parts-of-speech tagging, sentiment analysis, bag of words and regular expressions.

Some challenges in the field of quantum computing should be faced in order to improve the performance and robustness of the system. First, encoding the input data into a quantum state is complex in terms of computation. Therefore, new encoding algorithms are required to take advantage of the fast time of processing that quantum computing brings. A second challenge is that interpreting quantum states is not easy since qubit entanglement computes many solutions until they collapse to a final state. New algorithms are required to take the maximum advantage of quantum entanglement. A third challenge is the hardware limitation that current quantum devices present. Decoherence makes it impossible to solve large problems with current technologies. Finally, better algorithm evaluation methodologies are required to obtain accurate and trustable results.

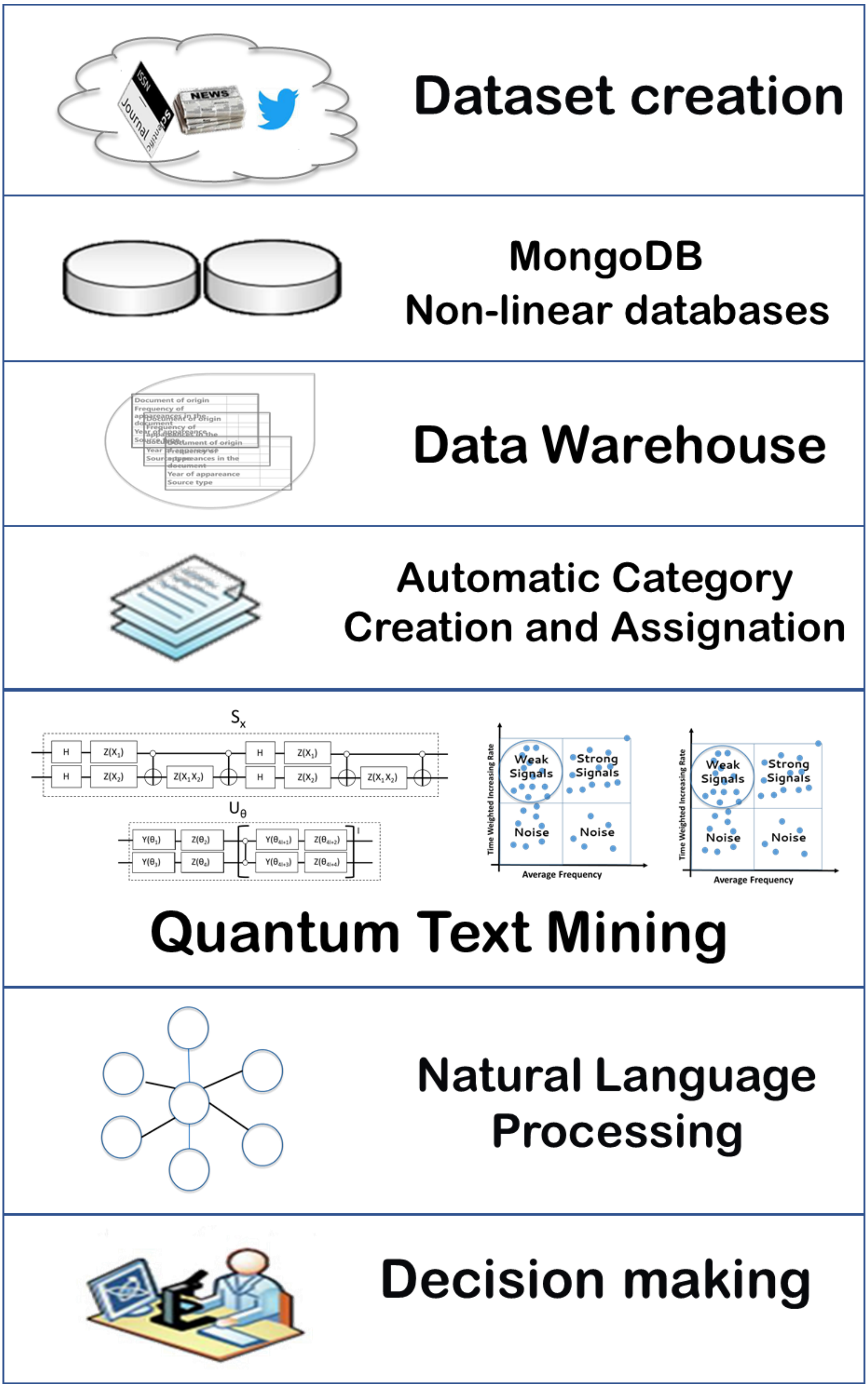
Some future improvements include quantum convolutional neural networks that are showing more robustness to noise, or integrating quantum variational circuits in reinforcement learning techniques to obtain a higher efficiency.

A possible solution to decoherence for applications with many documents to process is distributed quantum computing. Distributed solutions allow parallel quantum computing that will increase the speed when facing complex problems with a big number of terms to process. Among all available solutions, federated machine learning improves the security and privacy of users sharing their information in the cloud.

Chapter 4 also includes a description of the main available technologies for the hardware implementation of quantum computers. The quantum experiments that have been performed in this doctoral thesis have been implemented in a quantum simulator and in a quantum computer by IBM in the technology of superconducting qubits. Other technologies include ion traps, majorana quasiparticles, spin qubits, photonic quantum computing and quantum annealing. Not all these technologies are currently available but could bring new efficiency improvements in the future. A large number of resources from companies and other prestigious organizations are being dedicated to show the supremacy of quantum computing. Further research is required in the detection of suitable future maps, and the combination of different quantum technologies, including distributed solutions and combining quantum devices with classical ones.

In conclusion, the results that have been obtained with the conceptualization and implementation of this system and its hardware optimization in quantum computing prove that it is feasible to detect weak signals of the future by quantifying factors from an input dataset of unstructured information. These results have allowed the detection of new trends at a very early stage that have become more and more important today.

Appendix 1. Flow chart of the complete system



Appendix 2. List of keywords for DoV-DoD study in remote sensing

1. Keywords related to environmental, sustainability and geographical factors: Africa, alluvial, asteroids, attenuation, bedrock, Canadian, curvature, depression, desertification, disaster, diurnal, ENSO, extinction, foliar, forestry, Italy, Miocene, multitemporal, observatory, oceanography, pollen, rainforest, rangeland, southeast, sprawl, threat, topsoil, waste, weed and Wuhan.
2. Keywords related to business needs: adjacent, archival, breaking, care, check, consumption, diagnosis, forward, guidance, indirect, interior, intervention, invariant, kernel, maximization, mega, native, NOAA, physiological, plantation, preference, probabilistic, rational, residential, stakeholder, super, supervised, triggering, uptake, vibration and wild.
3. Keywords related to product/technological components: actuator, adaptative, array, bathymetry, cassini, clay, color, converter, endmember, excitation, gamma, hitran, inorganic, InSAR, oblique, passage, photometry, pigments, Rosetta, sounder, SRTM, stepwise, unmanned, UVSQ, volatile and voxel.

Group: ENVIRONMENTAL, SUSTAINABILITY AND GEOGRAPHICAL FACTORS

Keyword	DoD	Incr rate	DoV	Incr rate	DoT	Automatic Categories
alluvial	97,18	0,0747	581,91	0,0445	6,39	Advances in Engineering Software, Aeolian Research, Agricultural Water Management, Applied Geography, CATENA
asteroids	88,64	0,0796	582,64	0,0479	6,84	Astronautica, Advances in Space Research, Aerospace Science and Technology, Astronomy and Computing, Chemie der Erde
attenuation	98,36	0,1033	524	0,0576	6,39	Advances in Space Research, Agricultural and Forest Meteorology, Atmospheric Environment, Atmospheric Pollution Research, Atmospheric Research
bedrock	78,64	0,1007	568,55	0,0647	13,21	Advances in Space Research, Astroparticle Physics, CATENA, Computers & Geosciences, Continental Shelf Research
Africa	93,64	0,0699	528,45	0,0678	7,03	Advances in Climate Change Research, Advances in Water Resources, Agricultural Water Management, Applied Geography, Applied Ocean Research
canadian	85,36	0,0781	545,09	0,0427	6,76	Astronautica, Advances in Space Research, Agricultural and Forest Meteorology, Agriculture and

						Agricultural Science Procedia, Applied Geography
curvature	81	0,1143	541,55	0,0621	6,39	Astronautica, Agricultural and Forest Meteorology, Atmospheric Environment, Automation in Construction, CATENA
depression	83,45	0,1088	553,73	0,0652	6,39	Aeolian Research, Atmospheric Research, Biological Conservation, CATENA, Computers & Geosciences
desertification	93,91	0,0671	604,45	0,0357	6,39	Ecologica Sinica, Advances in Space Research, Advances in Water Resources, Aeolian Research, Agricultural and Forest Meteorology
disaster	102,91	0,0712	559	0,045	6,39	Astronautica, Ecologica Sinica, Advances in Climate Change Research, Advances in Space Research, Aeolian Research
diurnal	101,64	0,0629	593,91	0,0371	6,5	Astronautica, Advances in Space Research, Agricultural and Forest Meteorology, Agriculture, Ecosystems & Environment, Atmospheric Environment
enso	86,91	0,0667	572,91	0,04	0,32	Advances in Space Research, Agricultural and Forest Meteorology, Applied Geography, Aquaculture, Atmospheric Environment
extinction	83	0,0863	593,45	0,052	132,52	Advances in Space Research, APCBEE Procedia, Atmospheric Environment, Atmospheric Pollution Research, Atmospheric Research
foliar	85,09	0,0718	571,09	0,0503	6,09	Agricultural and Forest Meteorology, Agriculture, Ecosystems & Environment, Atmospheric Environment, Biosystems Engineering, Chemosphere
forestry	97,73	0,0617	587,91	0,0387	6,39	Ecologica Sinica, Advances in Space Research, Agricultural and Forest Meteorology, Applied Energy, Applied Geography
Italy	89,27	0,0835	538,82	0,0504	6,39	Astronautica, Oecologica, Advances in Space Research, Advances in Water Resources, Agricultural and Forest Meteorology
miocene	80,73	0,0741	527,91	0,0736	6,39	CATENA, Comptes Rendus Geoscience, Egyptian Journal of Petroleum, Geomorphology, Geothermics
multitemporal	93,27	0,0905	589,82	0,0397	6,39	Advances in Space Research, Agricultural and Forest Meteorology, Applied Geography, Applied

						Mathematical Modelling, Applied Soft Computing
observatory	82,73	0,0637	575,64	0,0532	6,39	Astronautica, Advances in Space Research, Astronomy and Computing, Atmospheric Environment, Atmospheric Pollution Research
oceanography	102,09	0,0658	557,18	0,054	6,39	Advances in Space Research, Aquatic Procedia, Continental Shelf Research, Deep Sea Research Part I: Oceanographic Research Papers, Deep Sea Research
pollen	81,73	0,0776	535,91	0,0401	20,85	Oecologica, Atmospheric Environment, CATENA, Chemosphere, Comptes Rendus Biologies
rainforest	89,36	0,0688	610,55	0,0385	1,33	Advances in Space Research, Agricultural and Forest Meteorology, Anthropocene, Applied Geography, Atmospheric Environment
rangeland	76,64	0,1007	568	0,0438	9,97	Oecologica, Aeolian Research, Agricultural and Forest Meteorology, Agricultural Systems, Applied Geography
southeast	102,64	0,1015	579,18	0,0438	6,52	Ecologica Sinica, Tropica, Advances in Water Resources, Aeolian Research, Agricultural and Forest Meteorology
sprawl	101,27	0,0773	556,64	0,0482	6,62	APCBEE Procedia, Applied Geography, Atmospheric Environment, Atmospheric Research, CATENA
threat	95,64	0,1192	575,91	0,0632	28,24	Astronautica, Ecologica Sinica, Advances in Space Research, Advances in Water Resources, Agricultural and Forest Meteorology
topsoil	87,45	0,1065	525,27	0,0443	6,09	Agricultural and Forest Meteorology, Agricultural Water Management, CATENA, Ecological Indicators, Geoderma
waste	83,91	0,0607	563,45	0,0494	6,39	Astronautica, Ecologica Sinica, Advances in Space Research, the ICRP, Applied Energy
weed	77,64	0,0905	567,73	0,0458	4,2	Agricultural and Forest Meteorology, Agricultural Systems, Agricultural Water Management, Agriculture, Ecosystems & Environment, Applied Soft Computing
Wuhan	86,55	0,0603	527,73	0,0425	196,43	Applied Geography, Cities, Computers, Environment and Urban

						Systems, Ecological Modelling, Energy

Group: BUSINESS NEEDS

Keyword	DoD	Incr rate	DoV	Incr rate	DoT	Automatic Categories
adjacent	95	0,0816	578,18	0,05	6,39	Advances in Space Research, Advances in Water Resources, Aeolian Research, AEU, Agricultural and Forest Meteorology
archival	86	0,1088	549,73	0,045	6,39	Astronautica, Advances in Space Research, Aeolian Research, Applied Geography, Astronomy and Computing
care	76	0,0676	533,27	0,0584	9,78	Astronautica, Tropica, Advances in Climate Change Research, Advances in Space Research, Advances in Water Resources
check	79,18	0,0785	545,64	0,0576	4,62	Astronautica, Advances in Space Research, Agricultural Systems, Agricultural Water Management, Aquatic Procedia
consumption	96,73	0,0975	579,36	0,0479	6,39	Astronautica, Networks, Agricultural and Forest Meteorology, Agricultural Water Management, Agriculture, Ecosystems & Environment
diagnosis	91,18	0,1079	566,45	0,0472	6,39	Tropica, Advances in Space Research, Advances in Water Resources, Agricultural and Forest Meteorology, Agricultural Sciences in China
forward	88,64	0,0778	523,82	0,0581	6,07	Astronautica, Ecologica Sinica, Advances in Space Research, Advances in Water Resources, Agricultural Systems
guidance	83,27	0,0939	554,64	0,0826	6,39	Astronautica, Tropica, Aerospace Science and Technology, Agriculture and Agricultural Science Procedia, Anthropocene
indirect	93,09	0,0743	584,55	0,0357	19,36	Ecologica Sinica, Tropica, Advances in Space Research, Agricultural and Forest Meteorology, Agricultural Sciences in China
interior	80,09	0,0729	533,45	0,0417	2,32	Astronautica, Advances in Space Research, Agricultural and Forest Meteorology, Applied Geography, Atmospheric Research
intervention	84,09	0,0665	578,55	0,0519	6,39	Astronautica, Oecologica, Tropica, Advances in Engineering Software, Agricultural and Forest Meteorology

invariant	83,55	0,0623	599,55	0,0751	6,39	AASRI Procedia, AEU, Agricultural and Forest Meteorology, Applied Mathematics and Computation, Chaos, Solitons & Fractals
kernel	96,64	0,07	540,09	0,0384	6,16	Advances in Space Research, Advances in Water Resources, AEU, Analytica Chimica , Applied and Computational Harmonic Analysis
maximization	93,36	0,0779	600,91	0,0436	6,39	Astronautica, Mathematica Scientia, Advances in Space Research, Advances in Water Resources, Agricultural and Forest Meteorology
mega	95,18	0,0899	580,82	0,0572	4,55	Agriculture, Ecosystems & Environment, Applied Geography, Atmospheric Environment, Atmospheric Research, CATENA
native	89,45	0,0603	559,36	0,0384	6,39	Ecologica Sinica, Oecologica, Agricultural and Forest Meteorology, Agricultural Systems, Agricultural Water Management
noaa	84,73	0,0839	531,27	0,0473	2,38	Ecologica Sinica, Advances in Space Research, Aeolian Research, Agricultural and Forest Meteorology, Agricultural Water Management
physiological	93,91	0,0718	576,73	0,0363	6,39	Academic Radiology, Astronautica, Ecologica Sinica, Oecologica, Agricultural and Forest Meteorology
plantation	86,73	0,0619	558,27	0,0621	6,39	Agricultural and Forest Meteorology, Agricultural Sciences in China, Agricultural Systems, Agricultural Water Management, Agriculture, Ecosystems & Environment
preference	93,73	0,0698	572,73	0,0494	155,22	Ecologica Sinica, Oecologica, Advances in Space Research, Advances in Water Resources, Agricultural and Forest Meteorology
probabilistic	92,36	0,1067	558,55	0,0589	6,39	Advances in Space Research, Advances in Water Resources, Agricultural and Forest Meteorology, Agriculture, Ecosystems & Environment, Applied Energy
rational	79	0,0786	555,09	0,0723	63,24	Agronomica Sinica, Agricultural Water Management, Applied Geography, Applied Soft Computing, CATENA
residential	92,91	0,0813	536,64	0,0463	6,39	Ecologica Sinica, Applied Geography, Aquatic Procedia, Atmospheric Environment, Building and Environment
social	90,73	0,0842	580,73	0,0403	9,47	Astronautica, Advances in Climate Change Research, Advances in Space Research, Advances in Water

						Resources, Agricultural Sciences in China
stakeholder	78,45	0,0698	526,18	0,0377	6,39	Astronautica, Tropica, Agricultural Systems, Anthropocene, Applied Geography
super	91,91	0,1271	552,18	0,0599	4,88	Astronautica, Advances in Space Research, Analytica Chimica , Applied and Computational Harmonic Analysis, Atmospheric Environment
supervised	89	0,1022	574,73	0,0679	6,39	Ecologica Sinica, Advances in Space Research, Agricultural Water Management, Agriculture, Ecosystems & Environment, Applied Geography
triggering	91,45	0,1199	597,18	0,0723	192,87	Advances in Space Research, Agricultural and Forest Meteorology, Agricultural Systems, Agriculture, Ecosystems & Environment, Applied Geography
uptake	91,18	0,0845	578,09	0,0691	6,39	Agronomica Sinica, Advances in Space Research, Agricultural and Forest Meteorology, Agriculture, Ecosystems & Environment, Aquacultural Engineering
vibration	88,27	0,0893	577,18	0,0424	6,39	Astronautica, Advances in Space Research, Aerospace Science and Technology, Agricultural and Forest Meteorology, Atmospheric Environment
wild	78,64	0,1021	552,45	0,04	8,29	Ecologica Sinica, Tropica, Advances in Space Research, Aerospace Science and Technology, Animal Behaviour

Group: PRODUCT/TECHNOLOGICAL COMPONENTS

Keyword	**DoD**	**Incr rate**	**DoV**	**Incr rate**	**DoT**	**Automatic Categories**
actuator	77,73	0,0917	523,27	0,0547	6,39	Astronautica, Aerospace Science and Technology, Automation in Construction, Computer Standards & Interfaces, Energy Procedia
adaptative	86,73	0,0575	509,36	0,0459	6,39	Astronautica, Networks, Agricultural and Forest Meteorology, Agricultural Water Management, Agriculture, Ecosystems & Environment
array	93	0,0882	594,55	0,082	14,76	Astronautica, Networks, Advances in Space Research, Aeolian Research, AEU
bathymetry	72,55	0,0625	530,91	0,043	6,39	Advances in Water Resources, Aquatic Procedia, Biological

						Conservation, CATENA, Coastal Engineering
cassini	102,09	0,0659	581,82	0,067	10,24	Astronautica, Geomorphology, Icarus, Infrared Physics & Technology, Journal of Quantitative Spectroscopy and Radiative Transfer
clay	88,55	0,124	543,91	0,0593	210,97	Advances in Space Research, Aeolian Research, Agricultural and Forest Meteorology, Agricultural Water Management, Ain Shams Engineering Journal
colour	95,55	0,0987	564,82	0,04	3,61	Advances in Space Research, Agricultural and Forest Meteorology, Analytica Chimica , Applied Geography, Aquatic Procedia
converter	91	0,0747	526	0,0792	143,52	Astronautica, Advances in Space Research, Agricultural and Forest Meteorology, Agriculture, Ecosystems & Environment, Applied Geography
endmember	78,82	0,1092	533,64	0,0454	6,39	Advances in Space Research, Analytica Chimica , Computers, Environment and Urban Systems, Icarus, Information Sciences
excitation	81,18	0,1007	587,64	0,0435	6,39	Astronautica, Advances in Space Research, Analytica Chimica , Applied Surface Science, Ceramics International
gamma	88,27	0,1081	528	0,0533	5,6	Astronautica, Oecologica, Advances in Space Research, AEU, Agricultural and Forest Meteorology
hitran	95,73	0,0683	577,36	0,0684	6,39	Chemical Physics Letters, Journal of Molecular Spectroscopy, Journal of Quantitative Spectroscopy and Radiative Transfer, Optics & Laser Technology, Transactions of Nonferrous Metals Society of China
insar	87	0,1393	87	0,1393	0,95	Advances in Space Research, Advances in Water Resources, Agriculture and Agricultural Science Procedia, Applied Geography, Computers & Geosciences
inorganic	81,36	0,0935	523,91	0,0729	6,39	Advances in Space Research, Agriculture, Ecosystems & Environment, Aquaculture, Aquatic Procedia, Arabian Journal of Chemistry
oblique	86,55	0,0857	584,55	0,0397	6,39	Advances in Space Research, AEU, Applied Geography, Aquatic Botany, Coastal Engineering

passage	82,73	0,0708	529,64	0,062	6,39	Advances in Space Research, Aeolian Research, Atmospheric Environment, Basic and Applied Ecology, Cold Regions Science and Technology
photometry	83,64	0,0596	552	0,0514	6,39	Advances in Space Research, Atmospheric Environment, Icarus, Infrared Physics & Technology, ISPRS Journal of Photogrammetry and Remote Sensing
pigments	95,27	0,0669	561,36	0,0707	189,9	Ecologica Sinica, Agricultural and Forest Meteorology, Analytica Chimica , Applied Surface Science, Biochemical Systematics and Ecology
rosetta	82,27	0,1019	496,55	0,035	8,21	Astronautica, Advances in Space Research, Astronomy and Computing, Geomorphology, Icarus
UVSQ	94,18	0,0831	555,64	0,0491	127,36	Aerospace Science and Technology, Chinese Journal of Aeronautics, Cold Regions Science and Technology, Computerized Medical Imaging and Graphics, Computers & Geosciences
sounder	68,27	0,096	542,18	0,0462	8,39	Astronautica, Tropica, Advances in Space Research, Atmospheric Environment, Atmospheric Pollution Research
srtm	85,36	0,0617	527,73	0,0425	3,41	Tropica, Aeolian Research, Applied Geography, Biosystems Engineering, CATENA
stepwise	83,09	0,0597	516,09	0,0472	6,39	Tropica, Atmospheric Environment, Atmospheric Research, Chemical Geology, Computers and Electronics in Agriculture
unmanned	90,91	0,0909	557,73	0,0664	6,39	AASRI Procedia, Astronautica, Automatica Sinica, Advanced Engineering Informatics, Advances in Space Research
volatile	95,55	0,0907	565,55	0,0713	6,39	Astronautica, Advances in Space Research, Analytica Chimica , Applied Geography, Atmospheric Environment
voxel	86,73	0,0575	509,36	0,0459	6,39	Astronautica, Networks, Agricultural and Forest Meteorology, Agricultural Water Management, Agriculture, Ecosystems & Environment

Appendix 3. Impact Factors of the Journals

ARTICLE 1

Authors	*Griol-Barres, I.; Milla, S.; Cebrián, A.; Mansoori, Y.; Millet, J.*
Title	*Variational quantum circuits for machine learning. An application for the detection of weak signals.*
Journal	*Applied Sciences*, **2021**, *11*(14), 6427
DOI	10.3390/app1114
Publisher	MDPI, ST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLAND
ISSN	2076-3417
eISSN	2076-3417
Impact Factor	2.679 (2020) – 2.736 (5 years)

JCR Category	Rank in Category	Quartile in Category
CHEMISTRY, MULTIDISCIPLINARY - SCIE	**101 of 179**	**Q3**
ENGINEERING, MULTIDISCIPLINARY - SCIE	**38 of 91**	**Q2**
MATERIALS SCIENCE, MULTIDISCIPLINARY - SCIE	**201 of 335**	**Q3**
PHYSICS, APPLIED - SCIE	**73 of 160**	**Q2**

CiteScore Highest percentile	71% 85/297 General Engineering
CiteScore 2020	3.0
SJR 2020	0.435 Q1
SNIP 2020	1.068
HIndex	52

ARTICLE 2

Authors	*Griol-Barres, I.; Milla, S.; Cebrián, A.; Fan, H.; Millet, J.*
Title	*Detecting Weak Signals of The Future: A System Implementation to detect new business opportunities based on Text Mining and Natural Language Processing*
Journal	*Sustainability,* **2020**, *12*(19), 7848
DOI	*10.3390/su12197848*
Publisher	MDPI, ST ALBAN-ANLAGE 66, CH-4052 BASEL, SWITZERLAND
ISSN	2071-1050
eISSN	2071-1050
Impact Factor	3.251 (2020), 2.576 (2019) – 3.473 (5 years)

JCR Category	Rank in Category	Quartile in Category
ENVIRONMENTAL SCIENCES SCIE	**124 of 274**	**Q2**
ENVIRONMENTAL STUDIES SSCI	**60 of 125**	**Q2**
GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY SSCI	**6 of 9**	**Q3**
GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY SCIE	**30 of 44**	**Q3**

CiteScore Highest percentile	84% 110/704 Geography, Planning and Development
CiteScore 2020	3.9 (4.0 2021)
SJR 2020	0.612 Q1
SNIP 2020	1.242
HIndex	85

ARTICLE 3

Authors	*Griol-Barres, I.; Milla, S.; Millet, J.*
Title	*Improving strategic decision making by the detection of weak signals in heterogeneous documents by text mining techniques.*
Journal	*AI Communications, **2019,** vol. 32, no. 5-6, pp. 347-360.*
DOI	*10.3233/AIC-190625*
Publisher	IOS PRESS, NIEUWE HEMWEG 6B, 1013 BG AMSTERDAM, NETHERLANDS
ISSN	0921-7126
eISSN	1875-8452
Impact Factor	1.0 (2019) – 1.072 (5 years)

JCR Category	Rank in Category	Quartile in Category
COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	**118 of 136**	**Q4**

CiteScore Highest percentile	54% 103/227 Artificial Intelligence
CiteScore 2020	3.7
SJR 2020	0.337
SNIP 2020	1.281
HIndex	38

ARTICLE 4

Authors	*Griol-Barres, I.; Milla, S.; Millet, J.*
Title	*Implementación de un sistema de detección de señales débiles de futuro mediante técnicas de minería de textos.*
	(Implementation of a weak signal detection system by text mining techniques).
Journal	*Revista Española de Documentación Científica, **2019,** vol. 42, no. 2.* Article number e234
DOI	10.3989/redc.2019.2.1599
Publisher	CONSEJO SUPERIOR INVESTIGACIONES CIENTIFICAS-CSIC, VITRUVIO 8, 28006 MADRID, SPAIN
ISSN	0210-0614
eISSN	1988-4621
Impact Factor	1.295 (2019) – 1.259 (5 years)

JCR Category	Rank in Category	Quartile in Category
INFORMATION SCIENCE & LIBRARY SCIENCE	**60 of 86**	**Q3**

CiteScore Highest percentile	72% 64/235 Library and Information Sciences
CiteScore 2020	2.0
SJR 2019	0.497 Q2 LIBRARY AND INFORMATION SCIENCES
SNIP 2019	1.215
HIndex	18

ARTICLE 5

Authors	*Griol-Barres, I.; Milla, S.; Millet, J.*
Title	*System Implementation for the Detection of Weak Signals of the Future in Heterogeneous Documents by Text Mining and Natural Language Processing Techniques.*
Journal	*Proceedings of the 11th International Conference on Agents and Artificial Intelligence – (Volume 2), ScitePress, 2019, pp. 631-638.*
DOI	10.5220/0007380306310638
Publisher	SciTePress
ISBN	978-989758350-6
Quality indicator	IEEE as technical sponsor and reviewer