BMT22
4th International Conference
Business Meets Technology
Ansbach, 7th – 9th July 2022

A REVIEW OF THE USE OF MACHINE LEARNING TECHNIQUES IN ECO-INNOVATION RESEARCH

Diez Martinez, Inés [a] and Peiró Signés, Ángel [b]

[a] *Universitat Politècnica de València, Spain (indiemar@alumni.upv.es)*
[b] *Universitat Politècnica de València, Spain (anpeisig@omp.upv.es)*

ABSTRACT: Machine learning is a powerful tool used across research all over the world. Machine learning algorithms are a form of artificial intelligence that allows more accurate predictions of causal conditions of all kinds, being able to analyze complex data samples beyond what a human could do. Machine learning mimics human reasoning by creating a neural network, and this has proven to be a useful technique to solve complex problems.

The thread of climate change is one of the most complex problems that humanity is currently facing. On one hand, we need the industries and the market to continue to function to guarantee covering the needs of the population, and its continued development. On the other hand, this development must guarantee the conservation of the planet and its habitability conditions, which are essential for the continued existence of a world to be left to future generations. Reducing the harmful effects that business-related activities have on the natural environment is key to guarantee a sustainable future, and this done, among other elements, through eco-innovation techniques.

Therefore, both machine learning and eco-innovation are striving topics across researchers nowadays, but: Are these two topics linked to each other? Is machine learning used as a tool to support a better understanding of eco-innovation (i.e., environmental innovation)?

This review aims to understand what is the role that machine learning has in the context of eco-innovation.

Results show that machine learning is not a widely used technique in the field of eco-innovation research and that there is a wide spectrum of research in which machine learning could be used in the future alongside the increasing research linked to eco-innovation.

Keywords: *Machine learning; Sustainability; Innovation; Eco-innovation; Environmental innovation*

How to cite: Diez Martinez, Inés and Peiró Signés, Ángel. 2022. A review of the use of machine learning techniques in eco-innovation research. In Proc.: 4th International Conference Business Meets Technology. Ansbach, 7th – 9th July 2022. Doi: https://doi.org/10.4995/BMT2022.2022.15550

2022, Editorial Universitat Politècnica de València

1. PURPOSE OF THE PAPER

The purpose of this paper is to better understand how machine learning is used in the context of research linked to eco-innovation through an analysis of bibliometric results.

Eco-innovation and environmental innovation practices are those that meet the demand of quality and sustainable products and services, using innovations, and while protecting the environment (Martínez-Pérez, García-Villaverde, and Elche, 2015).

Eco-innovation (i.e., environmental innovation) is key to be able to reach the sustainable development goals set out by global leaders. Nevertheless, sustainability comes along in many dimensions and through the entire value chain of processes and products. Depending on factors such as the economic sector, the geographical location, the characteristics of the infrastructure, the availability of resources or the R&D budget environmental orientation may be approached differently by each specific firm. As the saying of Ortega y Gasset goes, "I am myself and my circumstances": Each company has a specific situation and its orientation towards environmental practices may defer depending on the company's external and internal factors. Environmental innovation of companies comes through many perspectives and may be done differently for each firm, presenting complex datasets where many characteristics and potential causal conditions to trigger or block eco-innovation practices are present, with a high difficulty to analyse the firms' characteristics, external and internal factors.

On the other hand, machine learning is defined as leveraging on artificial intelligence and statistics among other topics, is designed to emulate learning from evidence and "train" algorithms so that they learn from the dataset automatically (Jordan and Mitchell 2015, El Naqa and Murphy, 2015). It is therefore useful to analyse bid datasets with numerous factors, which would be impossible to analyse manually by humans.

In consonance with this, we propose to research whether machine learning is used as a tool to support a better understanding of eco-innovation (i.e., environmental innovation). In principle, the high complexity of eco-innovation research including a wide range of external and internal factors, and the complex datasets, suggests that machine learning may be a viable tool to address eco-innovation research.

This paper aims at understanding what is the current use of machine learning to promote a better understanding of the triggers, barriers, and characteristics of eco-innovation. The paper will provide clarity regarding the who researchers may use machine learning algorithms in the context of environmental innovation and eco-innovation research.

2. RELATED WORK

Machine learning is one of the key methodologies used across wide scope of sectors and economic activities. Machine learning are computational methods, where past information is analyzed to make accurate predictions or improve performance (Mohri, Rostamizadeh and Talwalkar, 2018; Wang et al., 2016) and a data sample is studied to develop a set of algorithms to make accurate predictions of the future. In order words, machine learning can also be defined as an "evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment" (El Naqa and Murphy, 2015). As discussed by Jordan and Mitchell (2015), overall, machine learning allows computers to improve by-themselves, without requiring human intervention, through their own experience. It aims at improving decision making across industries (e.g., manufacturing, healthcare etc.) based on the analysis of evidence.

As described by Mohri, Rostamizadeh and Talwalkar (2018), the larger the sample, the easier the task is. It is key to understand the complexity of the topic and to collect the information accordingly. Without having sufficient data, the learner cannot properly analyze the solutions and would make an arbitrary choice among several options as they all seem like good solutions. On the other hand, it is also possible that over complex models may lead to complicated boundaries, entangling the analysis for the learner (Wang et al., 2016).

One use case for analysis employed in machine-learning is "basket analysis", which is, for example, finding associations between products bought by customers. As described by Alpaydin (2020), machine learning can be used to study the conditional probability of the form P(X|Y) where Y is the product, for example in the case of a supermarket, and X are the products that the customers have already purchased (e.g., how many of the customers who buy wine also buy cheese). Machine learning can be used in a similar way to study the characteristics of a company with X characteristics to understand what the conditional probability for Y is, for example in this case, integration of environmental considerations into innovation. Also, it would allow the study of the association between product (good or service), process, organizational or marketing innovation and their related environmental characteristics. Additionally, the relationship between the factors driving environmental innovation and the types of innovation.

Also following the guidelines presented by Alpaydin (2020), it is possible to introduce a third element to showcase the characteristics of the company (i.e., company size, sector etc.), to integrate company's attributes into the association analysis. In this case, the form would P(X|Y, D), where D is the set of company attributes.

3. METHODOLOGY

The database Scopus was used to explore the application of machine learning techniques to study eco-innovation.

Scopus is the world's largest abstract and citation database, and it can be used to monitor and measure research since it contains a wide scope of peer-reviewed literature such as papers of scientific journals, conference proceedings, and scientific books. (Baas, Schotten, Plume, Côté, and Karimi, 2020; Schotten, Meester, Steiginga, and Ross, 2017).

Several queries were performed to understand the role of machine learning in the context of eco-innovation research based on the publications' combinations of keywords. The method used is presented in Figure 1.

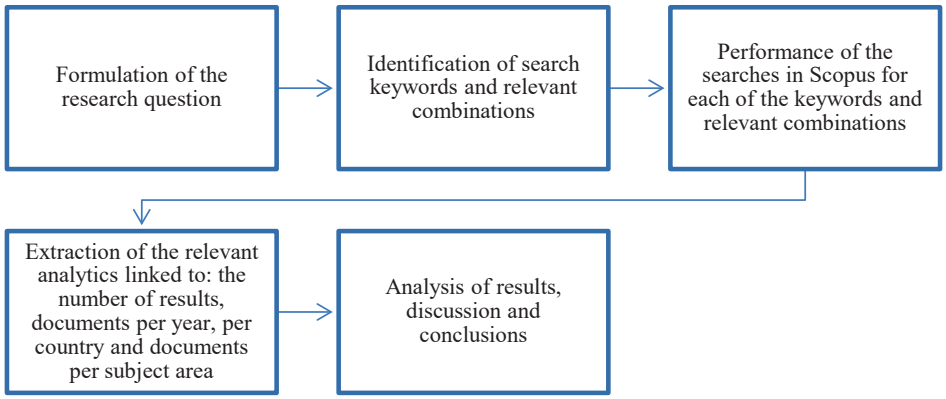

Figure 1. Research methodology

Regarding the formulation of the research question, the key element of this study is to understand whether machine learning is used in the context of eco-innovation research. Since, as described in the previous sections, eco-innovation research having often a complex set of parameters and large scope of elements to analyse, at first glance, make machine learning an ideal tool for this type of research.

Within this context, the research question of this paper is: Is machine learning used as a tool to support a better understanding of eco-innovation (i.e., environmental innovation)?

The key words selected for the search aimed at understanding the existing literature where one or several topics were present.

Table 1. Keywords used for the search

Keyword(s)	Rationale
Machine learning	Understand the total volume of literature linked to "machine learning", where "machine learning" is material to the published research as it is included as a keyword. Please note, terms such as "artificial intelligence" or "AI" were not included in the search as "machine learning" is a type of AI (Allegra et al., 2022) and therefore all article linked to machine learning is by definition also linked to artificial intelligence.
Innovation	Understand the total volume of literature linked to "innovation", where "innovation" is material to the published research as it is included as a keyword.
(Environmental innovation) OR Eco-innovation	Understand the total volume of literature linked to "eco-innovation" or "environmental innovation", where "eco-innovation" or "environmental innovation" is material to the published research as it is included as a keyword and are both terms used interchangeably referring to innovation with environmental considerations
Environmental innovation	Understand the total volume of literature linked to "environmental innovation", where "environmental innovation" is material to the published research as it is included as a keyword.
Eco-innovation	Understand the total volume of literature linked to "eco-innovation", where "eco-innovation" is material to the published research as it is included as a keyword.
Innovation AND Machine learning	Understand the total volume of literature linked to "innovation", and "machine learning", where both "innovation" and "machine learning" are material to the published research as they are both included as a keyword.
Environmental innovation AND Eco-innovation	Understand the total volume of literature linked to "environmental innovation", and "eco-innovation" to understand the use of these terms across literature.
(Environmental innovation OR Eco-innovation) AND Machine learning	Understand the total volume of literature linked to "environmental innovation" or "eco-innovation", and "machine learning", where "environmental innovation" or "eco-innovation", and "machine learning", are material to the published research as they are included as a keyword.

Environmental innovation AND Machine learning	Understand the total volume of literature linked to "environmental innovation" and "machine learning", where both "environmental innovation" and "machine learning" are material to the published research as they are both included as a keyword.
Eco-innovation AND Machine learning	Understand the total volume of literature linked to "eco-innovation" and "machine learning", where both "eco-innovation" and "machine learning" are material to the published research as they are both included as a keyword.

This allows the understanding of the current use of machine learning through the analysis of bibliometric figures.

4. FINDINDS

Machine learning is definitely a key research topic, with almost 300,000 publications having "machine learning" as a keyword. The highest volume of research regarding machine learning is published in United States, China, Indica, United Kingdom, Germany and Canada.

Results displayed in Table 2 show that machine learning, even though it is widely used for other topics (showing more almost 300,000 results in Scopus) is not a popular technique, rarely appearing for eco-innovation (i.e., environmental innovation) research.

Results of the search by keywords (without any filter per year, document type, source type or topic) show that among 1,037 results of publications containing the keyword "eco-innovation", only 2 contain as well the keyword "machine learning", representing less than 0.2% of results, with one article published in Spain (Peiró-Signes, Segarra-Oña, Trull-Domínguez and Sánchez-Planelles, 2022) and other in South Korea (Moon, Park and Woo, 2021). The latter included the term "machine learning" as part of the keyword "unsupervised machine learning approach", specifying that research leveraged on an unsupervised learning technique, K-means clustering method. Regarding the keywords "environmental innovation" and "machine learning" appearing together, a total of 21 results appear, whereas environmental innovation by itself has a total of 9,786 results. Thus, machine learning is used in ~0.21% of the total of research linked to environmental innovation.

Table 2. Number of results

Keyword(s)	Number of results
Machine learning	299,960
Innovation	206,469
(Environmental innovation) OR Eco-innovation	10,350
Environmental innovation	9,786
Eco-innovation	1,037
Innovation AND Machine learning	778
Environmental innovation AND Eco-innovation	473
(Environmental innovation OR Eco-innovation) AND Machine	22
Environmental innovation AND Machine learning	21
Eco-innovation AND Machine learning	2

The total of 22 results linked to machine learning and eco-innovation (i.e., environmental innovation) showed the following characteristics (please note the paper published by Moon, Park and Woo (2021) is duplicated appearing in both results for "Environmental innovation AND Machine learning" and for "Eco-innovation AND Machine learning"). Overall, 50% of documents are articles, 36.4% are conference papers and 13.6% reviews, where the different authors have only one publication on the topic.

Regarding their publication year, Figure 2, shows the breakdown of documents per year, with 2021 and 2020 being the two most relevant years so far in terms of publications, and a similar proportion will be followed in 2022 considering that the search was performed in May 2022. This shows that there is a growing interest in the use of machine learning techniques for eco-innovation research.

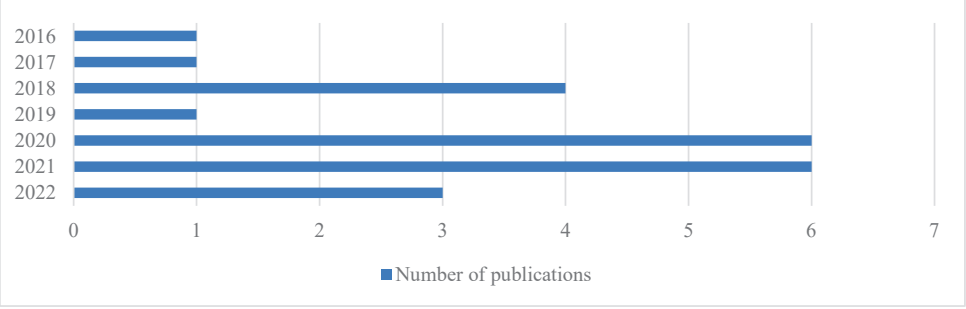

Figure 2. Documents per year

Figure 3 shows the breakdown of documents per country, with the highest being China, United States and the European countries. Results show that the research concerning both eco-innovation and machine learning is spread across regions and that there is a limited number of publications across the different countries, finding a maximum of 3 papers per country, which is a surprisingly low amount considering that both eco-innovation and machine learning are both popular terms striving today in terms of research interest.

Figure 3. Documents per country

Concerning the split per subject area, Computer Science, Engineering an Environmental Science and the three predominant ones as shown in Figure 4. This is aligned with the fact that machine learning, as it is artificial intelligence and based on algorithms, has a strong link with areas such as computer science, mathematics and engineering, and of course, more specifically, linked to decision sciences. Other subject areas such as environmental science, earth and planetary sciences are linked to the environmental aspects of eco-innovation. Additionally, other research areas appear linked to innovation having a strong link to business, such as business, management and accounting or economics.

Figure 4. Documents per subject area

5. DISCUSSION AND CONCLUSIONS

Approximately 1 in 20 papers (~5.0%) dedicated to innovation are linked to environmental innovation (or eco-innovation), whereas only 1 in 35 papers (~2.8%) dedicated to innovation and machine learning, are linked to environmental innovation (eco-innovation). This shows that the overall tendency to use machine learning in eco-innovation studies is lower than in the case of general innovation. Machine learning is linked to innovation research almost twice as often than it is linked to eco-innovation research. This suggests a lower maturity in the analysis of causal conditions and use of machine learning techniques for environmental innovation than for general innovation. Overall, results also suggest that eco-innovation is studied following other tools and that machine learning is not a widespread approach in the context of eco-innovation.

Overall, the number of publications using machine learning for eco-innovation related research is 3.14 documents per year, which is quite limited considering the high volume of machine learning publications and the high volume of eco-innovation research.

Also, research on this topic is relatively recent. The first research including machine learning and environmental innovation was published in 2016. This year is aligned with the overall main-stream use of machine learning techniques, considering that the 2010s came with the deep learning interest (Molnar, Casalicchio, Bischl, and 2020), but yet still this yearly average of 3.14 publications per year suggests that there is a research area still to be further developed and knowledge niche for the time being.

Regarding the countries, looking into Scopus publications on machine learning, also United States and China are the two key country publishers. Nevertheless, no country is publishing more than 3 papers per year addressing environmental innovation or eco-innovation and machine learning.

Also, these numbers do not reflect the global trends in the use of machine learning. In the overall number of publications of machine learning, India is in the top 3 of key publishers with over 29,000 documents, whereas there is no research published linked to eco-innovation and machine learning by any Indian research institution. This suggests that maybe the eco-innovation topic is less developed in this country, which also confirm when looking into the number of publications linked to eco-innovation and environmental innovation in the area, where India does not make the cut for the top 10, and China, United States, United Kingdom, Italy, and Spain are the leading countries in eco-innovation / environmental innovation research. In other words, supports the idea that the development of machine learning in other fields does not necessarily represent a development of machine learning techniques linked to eco-innovation research. And the development of eco-innovation studies does not necessarily trigger an interest in a machine learning approach.

Overall, the fact that machine learning present only in 0.2% of eco-innovation publications suggests that machine learning is not a widely used tool in eco-innovation research. This may suggest that the field of understanding causal conditions linked to eco-innovation may be a subject still to be developed, on which for the time being the analysis has not required powerful tools such as machine learning. Potentially eco-innovation use cases and data may be feasible to be processed manually or through simpler methods than the machine learning algorithms, at least for the time being. This also suggests an overall low maturity in the topic, as machine learning is one of the preferred alternatives when an exhaustive analysis identifying a wide scope of characteristics needs to be analysed.

6. RESEARCH LIMINATIONS/IMPLICATIONS

The key limitation of this research is to better understand what the alternatives to machine learning are and to what extent there are other tools to study causal conditions for the integration or preventing the integration of eco-innovation practices in firms.

7. ORIGINALITY/VALUE OF THE PAPER

This paper providers and overview of the use of machine learning in eco-innovation (i.e., environmental innovation) research, which, as per the results of the analysis, seems to be a field yet to explore by researchers in the eco-innovation field.

REFERENCES

Allegra, A., Tonacci, A., Sciaccotta, R., Genovese, S., Musolino, C., Pioggia, G., & Gangemi, S. (2022). Machine Learning and Deep Learning Applications in Multiple Myeloma Diagnosis, Prognosis, and Treatment Selection. Cancers, 14(3), 606.

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Baas, J. ; Schotten, M. ; Plume, A. ; Côté, G. ; Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. Quantitative Science Studies, 1(1), pp. 377-386.

El Naqa I.; Murphy M.J. (2015) What Is Machine Learning? In: El Naqa I., Li R., Murphy M. (eds) *Machine Learning in Radiation Oncology*. Springer, Cham. https://doi.org/10.1007/978-3-319-18305-3_1

Jordan, M.I.; Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260. DOI: 10.1126/science.aaa8415

Martínez-Pérez, A.; García-Villaverde, P; Elche, D. (2015) Eco-innovation antecedents in cultural tourism clusters: external relationships and explorative knowledge, Innovation: Management, Policy & Pract., 17, pp. 41-57, https://doi.org/10.1080/14479338.2015.1011058

Mohri, M.; Rostamizadeh, A.; Talwalkar, A. (2018). Foundations of machine learning. MIT press.

Molnar, C.; Casalicchio, G.; Bischl, B. (2020). Interpretable machine learning–a brief history, state-of-the-art and challenges. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 417-431). Springer, Cham.

Moon, H.; Park, S.; Woo, J., (2021). Staying on convention or leapfrogging to eco-innovation?: Identifying early adopters of hydrogen-powered vehicles. Technological Forecasting and Social Change, 171, p.120995. https://doi.org/10.1016/j.techfore.2021.120995

Peiró-Signes, Á.; Segarra-Oña, M.; Trull-Domínguez, Ó.; Sánchez-Planelles, J. (2022). Exposing the ideal combination of endogenous–exogenous drivers for companies' ecoinnovative orientation: Results from machine-learning methods. Socio-Economic Planning Sciences, 79, p.101145, https://doi.org/10.1016/j.seps.2021.101145

Schotten, M.; Meester, W. J.; Steiginga, S.; Ross, C. A. (2017). A brief history of Scopus: The world's largest abstract and citation database of scientific literature. In Research analytics (pp. 31-58). Auerbach Publications.

Wang, H. ; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. (2016). Machine learning basics. Deep learning, pp. 98-164.