



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

ADE

Facultad de Administración
y Dirección de Empresas /UPV

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Facultad de Administración y Dirección de Empresas

Nuevas tendencias en la industria alimentaria: Un análisis a
través de la huella digital de las empresas

Trabajo Fin de Grado

Grado en Administración y Dirección de Empresas

AUTOR/A: Chuliá Alfonso, José Ignacio

Tutor/a: Doménech i de Soria, Josep

CURSO ACADÉMICO: 2023/2024

Resumen

Con el cambio en los hábitos de consumo hacia opciones más sostenibles, saludables, vegetarianas y bajas en grasas, la industria alimentaria se enfrenta al reto de adaptarse continuamente a estas nuevas tendencias. Este Trabajo Final de Grado tiene como objetivo estudiar cómo la adaptación de las empresas se relaciona con la competitividad de estas. En primer lugar, se estudian los antecedentes en la industria agroalimentaria y se presentan diversas nuevas tendencias en el sector. A continuación, utilizando los indicadores digitales obtenidos de la base de datos SABI, se realiza el análisis mediante el lenguaje de programación R y técnicas de inteligencia de negocios. Para ello, tras usar dos técnicas previas (PCA y *clusters*) se aplican dos modelos estadísticos, clasificación y regresión. Para cada uno de estos dos modelos se aplica validación cruzada con tal de elegir el algoritmo que mejor resultado proporciona. Finalmente, se consigue concluir que las empresas que se adaptan a las nuevas tendencias son las que presentan mayor competitividad en el sector.

Palabras clave: Indicadores digitales, Industria agroalimentaria y Programación en R.

Abstract

With the shift in consumer habits towards more sustainable, healthy, vegetarian, and low-fat options, the food industry faces the challenge of continuously adapting to these new trends. This Final Degree Project aims to study how companies' adaptation relates to their competitiveness. First, the background of the agri-food industry is examined, and various new trends in the sector are presented. Then, using digital indicators obtained from the SABI database, the analysis is conducted using the R programming language and business intelligence techniques. For this purpose, after applying two preliminary techniques (PCA and clustering), two statistical models, classification and regression, are applied. For each of these models, cross-validation is performed to choose the algorithm that provides the best results. Finally, it is concluded that companies that adapt to new trends are the ones that show greater competitiveness in the sector.

Keywords: Digital indicators, Agri-food industry, and R programming.

Índice de contenidos

1. Introducción	6
1.1 Objetivos Trabajo Final de Grado	7
1.2 Estructura del Trabajo Final de Grado.....	7
2. Antecedentes y Marco contextual	9
2.1. El sector agroalimentario	9
2.1.2. Dimensión.....	9
2.1.2. Desarrollo	10
2.1.3 Cadena de suministro	12
2.2. Nuevas tendencias en el sector	13
2.2.1. Sostenibilidad y productos sostenibles.....	14
2.2.2. Productos alimenticios saludables	16
2.2.3. Productos vegetarianos y veganos.....	16
2.3. Páginas web de las empresas	17
2.3.1. Definición página web	17
2.3.2. Funciones de la página web	17
2.3.3 Indicadores digitales	18
3. Metodología	19
3.1 Herramienta de Análisis: R	19
3.2. Datos	19
3.2.1. Origen de datos	19
3.2.2 Descripción de variables y observaciones	20
3.2.3Preparación de datos.....	21
3.3. Modelos estadísticos	22
3.3.1 Análisis PCA.....	22
3.3.2 Análisis de clústeres	24
3.3.3 Clasificación	26
3.3.4 Regresión.....	27
4. Resultados	29
4.1. Análisis descriptivo univariante.....	29
4.2. Análisis descriptivo bivariante.....	36
4.2.1. Análisis de correlaciones	36
4.3. Análisis descriptivo multivariante	41
4.3.1 Resultados PCA.....	41
4.3.2 Resultados <i>clusters</i>	44

4.3.3 Resultado modelos de clasificación.....	50
4.3.4 Resultados regresión.....	53
5. Conclusiones.....	55
6. Bibliografía	57
Anexo I. Objetivos de Desarrollo Sostenible	60
Anexo II. Variables palabras clave	62

Índice de ilustraciones

Ilustración 1- Retos más urgentes	15
Ilustración 2: Boxplot datos anómalos.....	30
Ilustración 3: Boxplot datos anómalos corregidos.....	31
Ilustración 4: Grafico de correlaciones	36
Ilustración 5: Grafico del codo	41
Ilustración 6: Loading plot	42
Ilustración 7: Score plot.....	44
Ilustración 8: Gráfico coeficiente de silhouette k-medias.....	45
Ilustración 9: Gráfico intraclusters k-medias.....	45
Ilustración 10: Gráficos preliminar de scores	46
Ilustración 11: Gráfico silhouette k-medoides.....	46
Ilustración 12: Gráfico intraclusters K-medoides	47
Ilustración 13: Gráfico preliminar de scores k-medoides	47
Ilustración 14: Coeficiente de silhouette k-medias	48
Ilustración 15: Coeficiente silhouette k-medoides	48
Ilustración 16: Score plot.....	49
Ilustración 17: Gráfica perfil medio clusters	50
Ilustración 18: Gráfico curva Roc	51
Ilustración 19: Resultados modelo Random Forest clasificación	52
Ilustración 20: Resultado Random Forest regresión	54

Índice de tablas

Tabla 1: Descripción variables información	20
Tabla 2: Descripción variables económicas.....	21
Tabla 3: Resumen datos	29
Tabla 4: Análisis univariante de variables económicas	31
<i>Tabla 5: Análisis univariante de palabras clave</i>	<i>31</i>
Tabla 6: Análisis de correlaciones entre variables económicas y palabras clave	37
Tabla 7: Resultados clasificación	51
Tabla 8: Resultados regresión	53
Tabla 9. Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible.....	60

1. Introducción

En los últimos años la industria alimentaria ha tenido cambios importantes en la demanda de alimentos, impulsados por la preferencia de los consumidores y las nuevas tendencias. Dichos cambios han hecho que las empresas del sector deban adaptarse rápidamente para mejorar su competitividad en el mercado

Por eso, para atender a estas nuevas demandas de alimentos, muchas empresas del sector han adoptado herramientas digitales que les ayudarán a comprender mejor las necesidades de los consumidores a tiempo real. Por ello, la implementación de páginas web se ha convertido en una estrategia clave para promocionarse y dar mayor visibilidad a su gama de productos.

Además, el uso de los indicadores digitales ha permitido a las empresas estudiar y analizar los cambios en la industria alimentaria de manera más eficiente. Gracias a estos indicadores, las empresas pueden identificar tendencias, ajustar sus estrategias y mejorar su oferta de productos.

La digitalización no solo ha permitido a las empresas del sector agroalimentario competir de manera más eficiente, sino también innovar. La introducción de nuevos productos, ha generado gran interés en la población y ha contribuido positivamente en la economía española. Este sector, que incluye desde la pesca y la agricultura hasta la producción de alimentos procesados, ha visto mejoras en sus registros económicos gracias a estas innovaciones.

Para comprobar si las empresas se adaptan a las nuevas tendencias, se utilizará la herramienta de los indicadores digitales. Y poder ver si existe relación entre los beneficios obtenidos de las empresas y su adaptación a las nuevas tendencias.

1.1 Objetivos Trabajo Final de Grado

Los objetivos del presente Trabajo de Final de Grado consisten en analizar la relación que hay entre la adaptación de las empresas de la industria agroalimentaria a las tendencias de consumo saludable y su competitividad en el mercado, utilizando las herramientas de análisis de datos y técnicas de inteligencia de negocios.

A continuación, se detallan los objetivos para la realización de este trabajo de final de grado:

- 1- Identificar las principales tendencias de consumo saludable que están influyendo en el sector alimentario, con especial atención a la sostenibilidad y productos saludables.
- 2- Desarrollar una base de datos para evaluar la presencia digital de las empresas del sector alimentario, basado en datos extraídos de sus páginas web.
- 3- Analizar los datos económicos y financieros de las empresas alimentarias obtenidos de la base de datos SABI para determinar indicadores clave de competitividad.
- 4- Examinar la relación entre la adaptación a las tendencias de consumo saludable y la competitividad de las empresas, empleando técnicas estadísticas y de inteligencia de negocios en el lenguaje de programación R.

1.2 Estructura del Trabajo Final de Grado

El presente Trabajo Final de Grado (TFG) se estructura de la siguiente manera para proporcionar una visión clara y ordenada de los temas abordados:

El Capítulo 2, donde se habla de antecedentes y marco contextual, comienza con una descripción de la industria agroalimentaria, abordando las dimensiones y la cadena de suministro. También se proporciona información y se analizan las nuevas tendencias en el sector, como la sostenibilidad, los productos de alimentación saludable y productos vegetarianos y veganos, destacando su importancia creciente. Además, se examina la huella digital de las empresas e incluyendo la definición de los indicadores digitales, los cuales van a ayudar a poder ver si las empresas siguen las nuevas tendencias.

En el Capítulo 3, se va a describir el enfoque metodológico adaptado para la investigación. Se explica el origen de los datos que se han utilizado, la descripción de cada variable y observación utilizada para el estudio, y los pasos tomados para la preparación de los datos. Se van a explicar los modelos estadísticos aplicados, incluyendo el análisis de componentes principales (PCA), la clasificación, y la regresión. En esta sección se podrá ver el proceso de investigación y proporcionar una base sólida para interpretar los resultados.

En el Capítulo 4, que habla de resultados, se informará sobre los hallazgos principales del estudio. Se divide en tres partes: análisis descriptivo univariante, bivariante y multivariante.

Y por último, el Capítulo 5, que es el apartado de conclusiones, donde se resumirá los principales hallazgos conseguidos con el estudio, destacando la relevancia de los resultados e implicaciones para el sector agroalimentario.

2. Antecedentes y Marco contextual

En primer lugar, con tal de comprender con mayor profundidad el funcionamiento de la industria alimentaria, se va a dar una breve explicación de conceptos básicos.

2.1. El sector agroalimentario

El sector agroalimentario es todo aquello que proporciona el campo ya sea bienes o servicios con la finalidad de proporcionarle alimento a la población mundial. Es uno de los componentes más importantes en la economía global, ya que este agrupa todo el conjunto de actividades que tienen relación con la producción, procesamiento, distribución y comercialización de alimentos. Este sector se compone por actividades como la agricultura, la pesca y la ganadería e incluso hasta la elaboración de alimentos procesados y la distribución minorista.

Según Laajimi (1997), el consumo de alimentos es uno de los aspectos más importantes de la vida cotidiana, al ser la alimentación una actividad básica y necesaria del ser humano. Además, la forma de alimentarse es una de las manifestaciones socioculturales más importantes, aunque las modificaciones en los hábitos alimentarios son en general paulatinas.

Es muy importante comprender la importancia que tiene este sector, ya que es el encargado de proporcionarle alimentos a la población mundial, genera una gran cantidad de empleo en áreas rurales y contribuye significativamente al PIB de muchas naciones.

Es por ello por lo que surgió el interés de este estudio, para comprobar si las empresas más importantes del sector son las más competitivas gracias a tener en cuenta las tendencias del mercado.

2.1.2. Dimensión

Para poder saber cuántas empresas hay dentro de esta industria y poder entender mejor las conclusiones del estudio, se van a analizar las dimensiones, ya que es muy importante saber que dentro de este sector se encuentran tanto pequeñas como grandes empresas.

Según el Ministerio de agricultura, pesca y alimentación (2023), en Europa, la principal actividad de la industria manufacturera es la industria alimentaria, con un valor mayor a 1.093.000 millones de euros de cifra de negocios. Esto representa el 14,2% de la industria manufacturera, un porcentaje inferior al año 2020 debido a que en este año no se han contado con los datos de Reino Unido, ya que este no forma parte de Europa.

Este sector cuenta con 289.000 empresas que proporcionan empleo a 4,52 millones de personas. Además, las medianas y pequeñas empresas representan un 40,5% del total de la cifra de negocios del sector y el 58,4% del total de todos los puestos de trabajo que genera. El 95,4% de estas empresas son Pymes con menos de 50 empleados (pequeñas

empresas), y un 79,8% tienen menos de 10 trabajadores (microempresas). Por otro lado, cabe destacar también que 9 de cada 10 empresas son PYMES.

En España, la industria alimentaria se sitúa en el cuarto puesto en valor de cifra de negocios, representando un 10,9%. Los países que se sitúan por encima de España son, en primer lugar, Francia con un 19,5%, en segundo lugar, Alemania con un 16,9% y, por último, Italia con un 13,3%. En España, la primera rama manufacturera del sector industrial es la industria de bebidas y alimentación. Según los últimos datos de estadística estructural extraído del INE, esta rama tiene una cifra de negocios de 126.354,1 millones de euros, lo que representa un 25,4% del sector manufacturero. Esta rama representa el 2,5 % del PIB de España (en VAB) y asciende a 25.236 millones de euros (-2,2%).

La cantidad de empresas que hay en la industria de alimentación y bebidas ha ascendido a 30.260 según los últimos datos que ha registrado el INE (2020) en su página web, ese total de empresas representa un total de 15,7% de toda la industria manufacturera.

Actualmente, el número de trabajadores en esta industria asciende a 558.400 personas, con una tendencia de aumento progresivo durante el año.

Con el aumento de las necesidades de la población, es una realidad que la demanda de alimentos sigue creciendo, lo que hace que la producción agrícola sea crucial para la población.

La tecnología ha sido crucial para mejorar la eficiencia y la productividad en la cadena de suministros en la agricultura. La automatización de estos procesos ha permitido reducir tiempo y costos, además de incrementar la producción en las explotaciones agrícolas. Sin embargo, es importante mantener un equilibrio entre el uso de la tecnología y las prácticas agrícolas sostenibles, ya que es muy esencial proteger el medio ambiente y la salud de las cosechas a largo plazo.

En conclusión, el crecimiento y la modernización son factores claves en el sector para poder cubrir todas las necesidades alimentarias de la población mundial y adaptarse a las tendencias del mercado.

2.1.2. Desarrollo

El desarrollo es un tema crucial para entender la necesidad que tienen las empresas en implementar nuevas técnicas para ser más eficientes. Es por ello que, gracias al desarrollo que existe actualmente en este sector, las empresas pueden contar con herramientas suficientes para adaptarse a las nuevas tendencias rápidamente.

Según Cubero (2024), el sector agroalimentario ha experimentado nuevos desarrollos significativos que han influido en su evolución y transformación:

- 1- **Avances Tecnológicos:** Este aspecto es uno de los más importantes, ya que, cuanto mayor sean los avances tecnológicos en el sector, habrá una mayor reducción de tiempo en la recolección y procesamiento de alimentos, aumentará la productividad y mejorará la salud de los cultivos.

Según Sunedu (2020), los robots, drones y láseres son algunas de las innovaciones implementadas para mejorar la rentabilidad de las explotaciones y reducir su impacto ambiental.

- 2- **Sostenibilidad Ambiental:** Ha tenido un gran impacto en este sector debido a su enfoque en prácticas agrícolas sostenibles para reducir impactos negativos, como el uso de productos plásticos, la emisión de gases efecto invernadero, y la contaminación del agua. La agricultura regenerativa, el uso de cultivos de cobertura y la gestión de plagas son algunas de las técnicas adoptadas para mejorar la sostenibilidad en el sector.
- 3- **Cambio en las Preferencias del Consumidor:** Los consumidores están mostrando un mayor interés en productos orgánicos y sostenibles. La demanda de alimentos producidos de manera ética y respetando al medio ambiente ha aumentado. Actualmente, la sociedad ha experimentado un cambio drástico, con una gran parte de la población siguiendo una vida saludable con rutinas de alimentación basadas en productos bajos en grasas, elevados en proteínas y diuréticos, con el fin de mejorar su rutina de alimentación y poder tener un cuerpo saludable. Todo este cambio se está dando gracias a la visibilidad que proporcionan las redes sociales, y otros medios de *streaming*.
- 4- **Globalización y Comercio Internacional:** El comercio internacional del sector agroalimentario ha tenido un aumento debido al incremento de demanda de alimentos en los nuevos mercados y en aquellos que actualmente se encuentran en desarrollo.
Este punto tiene beneficios, ya que aumenta la facturación de las empresas. Sin embargo, por otro lado, los agricultores que no han podido expandirse tanto como las grandes empresas están sufriendo un gran impacto en su economía, ya que no puede competir con los precios que imponen las grandes empresas en el mercado. En los países que tienen una industrialización más elevada, la mano de obra es más barata.

En resumen, el sector agrario ha experimentado un desarrollo significativo en los últimos años, impulsado por avances tecnológicos, un enfoque creciente en la sostenibilidad, cambios en las preferencias del consumidor y la globalización del comercio agrícola. Estos desarrollos han llevado a una agricultura más eficiente, sostenible y orientada al mercado.

Gracias a todo este desarrollo que se está implementando en las empresas del sector, muchas de estas empresas se pueden adaptar más rápidamente a las nuevas tendencias, y conocer cuáles son las necesidades del consumidor. Es por ello por lo que se va a comprobar si las empresas que mejores resultados financieros tienen son aquellas que siguen las nuevas tendencias.

2.1.3 Cadena de suministro

Para entender el tiempo que se tarda en identificar cuál es la tendencia actual del mercado y el proceso que hay después para adaptarse a esa tendencia, es necesario conocer la cadena de suministro, ya que nos explica las etapas que atraviesa un producto hasta que llegue al consumidor.

Según Trazable (2022), la cadena de suministro en el sector agroalimentario es un proceso complejo que involucra múltiples etapas, desde la producción en el campo hasta la entrega de productos agrícolas frescos o procesados a los consumidores finales. Cabe destacar que no todos los procesos de la cadena entran dentro de la industria agroalimentaria.

Estas serían las etapas de la cadena de suministros:

- 1- **Producción Agrícola:** La cadena de suministro comienza en las explotaciones agrícolas, donde se cultivan productos agrícolas y se elaboran productos derivados. Esto puede incluir todo tipo de cultivos relacionados con el campo, el mar y otros tipos de explotaciones.
- 2- **Recolección y Postcosecha:** Una vez que los cultivos están maduros, se lleva a cabo la recolección. El proceso puede variar en duración dependiendo del tipo de cultivo o cosecha. Después de la recolección, los productos pasan por el proceso de postcosecha, que incluye la limpieza de los productos, su clasificación, su almacenamiento a una temperatura adecuada para su conservación y, dependiendo del alimento, su empaquetado para su transporte.
- 3- **Almacenamiento y Distribución:** La mayoría de los productos son perecederos, por lo que necesitan condiciones específicas para almacenamiento y para mantener su calidad y frescura. En esta fase, los alimentos se almacenan en instalaciones refrigeradas antes de ser distribuidos. Hay diferentes tipos de distribución; la más común dentro del país es la terrestre, pero también existen marítima y aéreo.
- 4- **Procesamiento y Envasado:** En algunos casos, determinados productos pasan por procesos de transformación antes de llegar al consumidor final. Esto puede incluir el lavado, corte, enlatado, deshidratado y muchos otros métodos que conocemos, los cuales prolongan la vida útil del alimento. En este caso, las empresas pertenecientes a esta etapa son las que entran en la definición del sector agroalimentario según Eurostat (2008).
- 5- **Venta al por Mayor y al por Menor:** Los productos pueden ser vendidos por minoristas, mayoristas o directamente a los consumidores.
- 6- **Consumo:** La cadena se culmina con el consumo del producto que previamente ha sido comprado por el consumidor final, ya sea en restaurantes, en el hogar u otros lugares.

Es importante tener en cuenta que no siempre es motivo de éxito adaptarse a todas las nuevas tendencias que surgen, ya que es un proceso costoso y es posible que el resultado final no salga rentable para las empresas. Por ello, el interés de este estudio es comprender si las empresas que mayor competitividad son aquellas que se adaptan a todas las nuevas tendencias o si, por lo contrario, no es necesario seguir las nuevas tendencias para ser una empresa líder del sector.

2.2. Nuevas tendencias en el sector

Es muy importante comprender cómo funcionan las tendencias dentro de un sector y saber que tendencias van a tener un impacto significativo y cuáles no. Por lo tanto, este apartado va a dar a conocer cuáles son las nuevas tendencias en este sector, para que posteriormente poder entender por qué las empresas siguen o no estas nuevas tendencias.

Según Arjona (2022), desde los principios de los años 70, las nuevas tendencias en el consumo de alimentos se caracterizan por una ruptura del consumo de alimentos básicos. Debido a el fin de la autarquía, se empezó a tener acceso a otros tipos de alimentos que antes no se podían conseguir, lo que generó un aumento de la productividad agrícola, además de un incremento en el poder adquisitivo debido al incremento de los salarios.

Tanto el sistema de aprovisionamiento alimentario como la política agraria se centran en la función de alimentación, prestando especial atención a las tendencias del consumo de alimentos y al comportamiento del consumidor.

Es una realidad que el sector agrícola es uno de los sectores que más cambios ha experimentado debido a las tendencias, ya que la tecnología ha incorporado muchos medios dentro de este sector, desde nuevas herramientas para hacer una producción más sostenible hasta nuevos productos más sanos. Además, gracias a las importaciones, se pueden adquirir productos que nos son posibles producir en España, lo que favorece a una dieta más variada.

Según Gracia (2001), una de las tendencias más notables actualmente son los cultivos genéticamente modificados (GM) y la biotecnología. Los avances en biotecnología han llevado al desarrollo de cultivos genéticamente modificados que son resistentes a plagas, enfermedades o condiciones ambientales adversas. Estos cultivos pueden aumentar la productividad y reducir la necesidad de pesticidas y herbicidas químicos. Entre los productos que más modificaciones genéticas han sufrido se encuentra la soja, el maíz, y el algodón.

El motivo de este estudio es analizar si las empresas que tienen mayor impacto económico en el sector están relacionadas con la incorporación de las nuevas tendencias. Por ello, vamos a conocer qué tendencias hay en el sector, ya que existe una gran variedad. Según Food Trends (2022), las tendencias para la industria agraria son:

- 1- **Plant-Based-Food:** La tendencia de consumir menos carne y más alimento de origen vegetal está creciendo debido a que los consumidores están concienciados

con el impacto medioambiental de la producción de carne, el bienestar animal y los beneficios para la salud al implementar nuevas dietas basadas en plantas.

- 2- **Alimentos saludables y sostenibles:** Los consumidores están cada vez más preocupados por su salud y el impacto medioambiental de su estilo de alimentación. Por ello, cada vez hay más demandas de alimentos sostenibles, saludables, nutritivos, libres de aditivos artificiales y respetuoso con el medio ambiente.
- 3- **Personalización:** Esto se refiere a que muchas personas tienen que adaptar sus dietas debido a intolerancias u otros aspectos. Como por ejemplo dietas sin gluten.
- 4- **Impresión 3D:** Es una medida que se está probando actualmente, ya que su finalidad sería la impresión en 3D de diferentes tipos de alimentos.
- 5- **Calidad:** Utilizar productos naturales los que contienen el 100% del producto y no buscar productos más baratos como los sucedáneos, ya que estos contienen trazas de alimento de peor calidad, según el artículo de Román (2003).

Una vez conocidas las tendencias del sector, se puede comprender mejor los retos a los que se tienen que enfrentar las empresas para poder adaptarse a estas tendencias. Gracias a este Trabajo Final de Grado, vamos a conocer si las empresas que mayor competitividad tienen en el mercado son aquellas que siguen las nuevas tendencias, incorporando aquellos productos veganos, vegetarianos, naturales y sostenibles.

2.2.1. Sostenibilidad y productos sostenibles

Ya que la sostenibilidad medioambiental es una tendencia muy significativa actualmente en el sector, como se ha podido ver en el apartado anterior, este punto va a portar información de por qué es tan importante para los consumidores y las empresas mejorar en este aspecto.

Según Gallopín (2003), En su sentido más general, la sostenibilidad de un sistema puede representarse mediante una función no decreciente de valuación de las salidas o productos del sistema analizado que son de interés. Es por ello que el sector agroalimentario debe tener muy en cuenta la sostenibilidad en todas las etapas de la cadena de producción. Si el cambio climático va en aumento, muchas cosechas se verán destruidas por tormentas imprevistas. Además, los cambios de temperatura drásticos reducirán la producción, por lo que generaría un caos tanto para el sector como para la humanidad, ya que no habría suficiente cantidad de alimentos para abastecer a la población.

Según Rodale Institute (2017), actualmente este sector está muy concienciado, por lo que utilizan muchas técnicas de sostenibilidad, como, por ejemplo:

1. La rotación de cultivos: sirve para mejorar la estructura del suelo y aumentar su fertilidad. También es muy útil para reducir plagas en las cosechas, maximizando

el rendimiento a largo plazo. Sin esta práctica, las plagas pueden afectar negativamente al cultivo y al equilibrio del ecosistema.

2. El compostaje: es otra práctica muy común en el sector agroalimentario, ya que es el proceso natural de la descomposición de materiales orgánicos. Este ayuda a reducir la cantidad de residuos orgánicos que van a los vertederos, disminuyendo así los gases de efecto invernadero y la producción de lixiviados.
3. La reducción de plástico: es una de las técnicas más utilizadas en la industria alimentaria. Muchos productos llegan a manos de los consumidores en envoltorios de plástico, cuya producción contribuye al cambio climático según The Nature Conservancy (2021). Los residuos de plástico contaminan vías fluviales y océanos, matan a cientos de animales y se filtra en nuestra cadena alimenticia. Por ello, muchas empresas se han renovado y han sustituido el plástico en los envoltorios de sus productos por otros materiales menos contaminantes.

Otras técnicas comunes que son ampliamente utilizadas incluyen el consumo responsable del agua, el uso de energías renovables y el reciclaje.

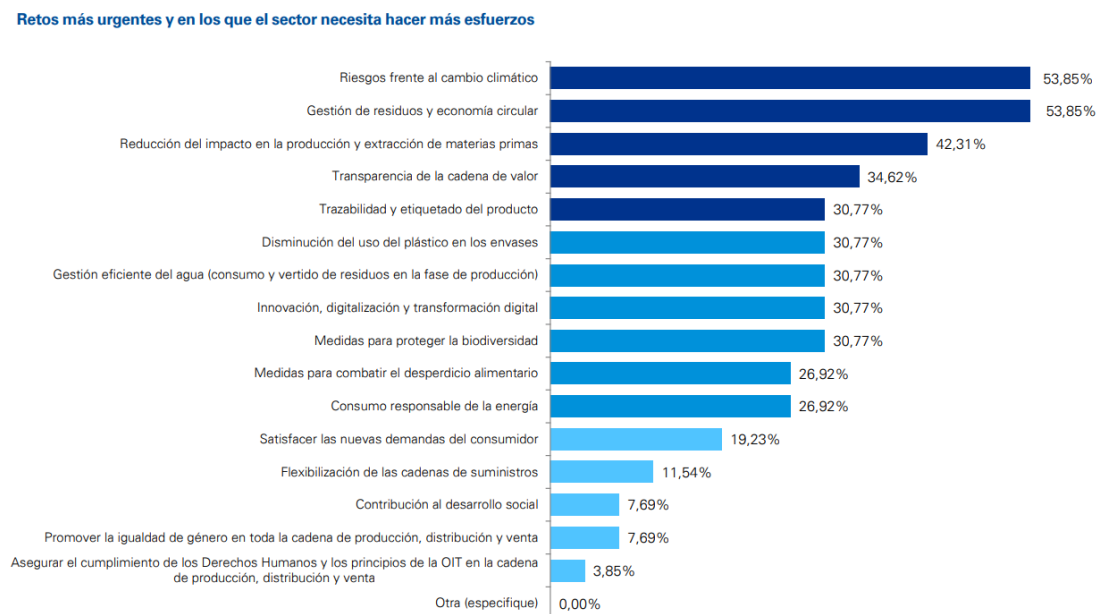


Ilustración 1- Retos más urgentes

Fuente: KPMG

Como se puede ver en la gráfica extraída de KPMG (2022), aún queda margen para mejorar. Sin embargo, según avanza la tecnología y los nuevos estudios que se realizan a diario, este sector irá introducción nuevas herramientas para aumentar la sostenibilidad y cumplir con más requisitos de la Agenda 2030.

Por lo tanto, sabiendo la importancia de ser sostenible, este estudio comprobará si las empresas que implementan medidas para que sus productos sean más sostenibles son las que reflejan mejores números y son las más importantes del sector.

2.2.2. Productos alimenticios saludables

La incorporación de productos alimenticios saludables en la industria es actualmente una tendencia creciente, ya que responde a la demanda del consumidor. Una gran parte de la población procura cuidar su alimentación para poder mejorar su salud a largo plazo y obtener un bienestar diario.

Los estudios de la OMS (2024) indican que una gran parte de la población sufre sobrepeso. En 2022, 1 de cada 8 personas en el mundo era obesa. Por lo tanto, la industria alimenticia ha optado por facilitar productos similares a los que ya existían, pero reduciendo en su gran medida grasas saturadas y otras sustancias que los hacían menos saludables, además de incorporar nuevas líneas de productos. Esta incorporación se está llevando a cabo de varias formas, como:

1. Desarrollo de nuevos productos.
2. Reformulación de productos existentes.
3. Etiquetado claro que proporcione toda la información nutricional del producto.
4. Mejora del marketing del producto, ya sea a través de envases más atractivos para el consumidor o mediante alianzas y colaboraciones con otras empresas para fomentar hábitos de alimentación saludables.

2.2.3. Productos vegetarianos y veganos

Uno de los temas que ha motivado la realización de este Trabajo Final de Grado, entre otros, es determinar si existe alguna relación entre las empresas que han introducido tendencias de alimentos en su gama de productos y su margen de competitividad. Muchas empresas han querido empatizar con un grupo muy vulnerable de personas que no consumen alimentos de origen animal ni derivados.

Esta nueva incorporación de productos en el mercado ha tenido un impacto muy significativo. Ante el crecimiento del mercado de productos vegetarianos y veganos, muchas empresas de alimentos convencionales están invirtiendo en el desarrollo de productos a base de plantas o adquiriendo empresas especializadas en este sector para expandir su oferta de productos y capitalizar esta tendencia en crecimiento.

Los productos vegetarianos y veganos están asociados con beneficios para la salud y el medio ambiente, lo que ha llevado a un aumento en la conciencia y la aceptación de estos productos entre los consumidores. Muchas personas están optando por dietas basadas en plantas por razones de salud, éticas y ambientales.

Según Lantern (2020), el 7,8% de la población española mayor de 18 años es vegana o vegetariana, y se esperaba que el mercado global de productos veganos y vegetarianos alcanzara un valor de 5.000 millones de dólares.

El crecimiento del mercado de productos vegetarianos y veganos ha sido notable en respuesta a la creciente demanda de consumidores que adoptan dietas basadas en plantas. Esto ha llevado a un aumento en la disponibilidad y la variedad de productos vegetarianos y veganos en supermercados, restaurantes y tiendas especializadas.

Una vez profundizado en las nuevas tendencias y su desarrollo, se intentará analizar con este Trabajo Final de Grado si las empresas que han implementado estas nuevas tendencias son las que han mejorado su competitividad, o si, por lo contrario, estas nuevas tendencias no han tenido una repercusión significativa en la mejora de competitividad de las empresas del sector alimentario.

2.3. Páginas web de las empresas

Este apartado explicará por qué las empresas han necesitado implementar nuevas técnicas para darse a conocer a más clientes, y cómo, con las herramientas adecuadas, como los indicadores digitales, pueden obtener la información necesaria para identificar cuáles son las tendencias en la industria.

Este apartado es fundamental para entender por qué en este Trabajo Final de Grado se utilizarán dichas herramientas e interpretar los resultados obtenidos para sacar conclusiones, ya que esta información provendrá de las páginas webs de las empresas.

Según Dans (2020), es muy importante que las empresas cuenten con páginas webs para tener presencia en el mercado y llegar más fácilmente al consumidor. Además, este punto ayudará a obtener información utilizando palabras claves que se emplean para recolectar datos y determinar si las empresas que se han adaptado a las nuevas tendencias son las empresas con mayor competitividad.

2.3.1. Definición página web

Según el libro de Hobbs (2015), una página web es un documento online que forma parte de un sitio web, este documento está diseñado para acceder a través de internet. Este archivo electrónico está compuesto principalmente de código HTML, CSS y JavaScript, que puede incluir texto, imágenes, videos, enlaces, formularios y otros elementos multimedia interactivos.

2.3.2. Funciones de la página web

Las páginas web son la unidad básica de contenido en la web y se utilizan para presentar información, compartir recursos, realizar transacciones en línea, comunicar ideas y promover productos y servicios. Cada página web tiene una URL (*Uniform Resource Locator*) única que permite a los usuarios acceder a ella a través de un navegador web.

Según Landois, (2021), la función de una página web es recopilar o difundir información a través de diferentes redes informáticas. Se podría decir que la función que ofrece una página web es equivalente a la de un folleto, revista, periódico o diario.

2.3.3 Indicadores digitales

Una vez entendido cómo funciona una página web, y por qué es tan importantes para las empresas contar con una, se va a explicar los indicadores digitales.

Según Llopart (2023), los indicadores digitales son herramientas que permiten hacer un seguimiento de una serie de acciones determinadas.

Estos pueden aportar diversos tipos de información, como el tráfico web, las páginas vistas, el tiempo en la página, la tasa de rebote, las conversaciones, la fuente de tráfico y las palabras clave.

En este Trabajo Final de Grado, esta herramienta es fundamental. El indicador que se va a utilizar para recopilar datos de las páginas web de las empresas serían las palabras claves.

Se ha generado una lista de palabras claves que reflejan las principales tendencias en este sector. Si estas palabras claves tienen mucha presencia en las páginas web de dichas empresas y esas empresas tienen una gran competitividad en el sector, se podría llegar al objetivo de este Trabajo Final de Grado: Determinar si la adaptación de las nuevas tendencias es un indicativo de la mejora de la competitividad de las empresas.

Conociendo el funcionamiento del sector y de las herramientas que se van a aplicar para obtener información, se puede empezar con el análisis para ver si las nuevas tendencias mejoran la competitividad de las empresas.

3. Metodología

En esta sección, se describirá la metodología empleada en el trabajo, explicando las variables, los métodos que se van a utilizar y los procesos de limpieza de datos necesarios para que la base de datos tenga sentido y se pueda trabajar con ella. El objetivo es determinar si la adaptación de las nuevas tendencias mejora la competitividad de las empresas.

3.1 Herramienta de Análisis: R

La herramienta utilizada para interpretar estos datos será R, un software cuyos creadores fueron Robert Gentleman y Ross Ihaka. R proporciona una amplia variedad de herramientas estadísticas y gráficas que ayudan a los usuarios a manipular sus datos y definir sus propias funciones. Según el artículo de Dataweek (2010), R es capaz de realizar modelos de regresión lineal y logísticos, análisis de series de tiempo, pruebas estadísticas clásicas, agrupamientos, *clustering* y muchas otras técnicas. R es uno de los softwares más utilizados por los usuarios debido a su facilidad de programación y su interfaz.

3.2. Datos

Los datos son una fuente de información crucial para entender sucesos específicos sobre un tema seleccionado. En este trabajo, se utilizará modelos de predicción y otras técnicas para analizar los datos y poder llegar a una conclusión.

Para la elaboración de este Trabajo Final de Grado, los datos proporcionarán información sobre empresas del sector agroalimentario. Se utilizarán variables económicas y otras relacionadas con la huella digital para observar el impacto de la implementación de nuevas líneas de productos (más saludables, veganos, etc.) en la competitividad de las empresas.

3.2.1. Origen de datos

Para obtener la información sobre el sector agroalimentario, se ha utilizado la herramienta Sistema de Análisis de Balances Ibéricos (SABI). Los datos recopilados corresponden al año 2022, ya que es el año más reciente con información completa disponible sobre el sector.

Según Van Dijk (2017), SABI es una base de datos que contiene información de más de 500000 empresas españolas y proporciona un exclusivo software de análisis financiero.

Los datos obtenidos para este estudio han pasado por una serie de filtros para garantizar su relevancia y actualidad. Antes de aplicarse cualquier filtro, la base de datos SABI contenía información sobre 26312 empresas de la industria agroalimentaria. Como había empresas que no eran útiles para el análisis se aplicaron los siguientes filtros:

1. Empresas con datos económicos actualizados al año 2022.
2. Empresas que tengan página web.

Estos filtros eran imprescindibles para el análisis. Tras aplicar estos filtros, el número de empresas se vio reducido a 5020 empresas. Estas empresas se consideraron necesarias para obtener información más actualizada y específica de las empresas

Es probable que el número de empresas se reduzca aún más durante los procesos de limpieza de datos.

3.2.2 Descripción de variables y observaciones

Es muy importante conocer los datos con los cuales se va a trabajar, por lo que este punto trata en una breve descripción de la base de datos.

La base de datos para el análisis de las empresas tiene un total de 220 variables y 5020 observaciones.

Una variable es una característica o atributo del cual queremos obtener información. Hay diferentes tipos de variables; pueden ser binarias, numéricas o de texto. Cada variable corresponde a una columna dentro de nuestra base de datos.

Con la base de datos que se ha obtenido, existen varios grupos:

1. **Información:** tres variables tipo texto, ya que aporta información de las empresas a nivel nombres, URL y otros. Estas variables son útiles al final de estudio para encontrar rápidamente las empresas que han cumplido con los parámetros solicitados.

Tabla 1: Descripción variables información

Variables	Descripción
Nombre	Nombre de la empresa
NIF	Identificador de la empresa
Web.site	Página web de la empresa

Fuente: Elaboración propia

2. **Económicos:** En este caso, hay cinco variables que son numéricas, ya que aporta información sobre beneficios, Roa y otras variables similares. Este tipo de variable es la que más información va a aportar al estudio, ya que los códigos de clúster se realizarán con estos datos.

Tabla 2: Descripción variables económicas

Variables	Descripción
IngExpl	Entradas de caja en la empresa como consecuencia de sus actividades. Esta medido en miles de euros.
ROA	Ratio financiero capaz de medir la capacidad de generar ganancias.
Emp	Cantidad de trabajadores en la empresa.
Va	Valor económico adicional de los productos. Esta medido en miles de euros.
Rdo	Saldo final obtenido de la empresa al final del cierre del periodo contable. Esta medido en miles de euros.

Fuente: Elaboración propia

Palabras clave: Son variables numéricas que aporta información sobre cuántas veces aparece una determinada palabra en la página web. Las columnas que empiezan por "keywords_" representan las frecuencias de aparición de las palabras con concordancia exacta (la palabra debe aparecer tal cual), mientras que las que empiezan por "kwstems_es_" son las frecuencias con concordancia amplia (puede aparecer la palabra tal cual o algún derivado). Todas las palabras claves se pueden consultar en el (Anexo II). Cabe recalcar que solamente se le va a dar uso a las palabras clave "kwstems_es_" ya que al ser menos restrictiva puede dar mejores resultados para el análisis.

Por otro lado, están las observaciones, que es un conjunto de valores de variables para una sola identidad. Las observaciones son las que van a aportar la información a cada variable, en este caso las observaciones representan las distintas empresas del sector.

3.2.3 Preparación de datos

El primer paso a seguir sería cargar los datos obtenidos de SABI, que previamente se exportaron a un archivo Excel y luego se leyeron con un programa especializado, en este caso R. Lo siguiente que se debe realizar es inspeccionar los datos mediante comandos como *summary*, para obtener un resumen estadístico o un *head* y *tail*, para ver la primera y última fila, respectivamente, y así entender como están organizados los datos. Es fundamental entender el contenido y la estructura de los datos.

Una vez conocida la estructura y contenido de los datos, se procederá a limpiar dichos datos, ya que existen valores faltantes, datos anómalos y duplicación de variables.

Por lo que, lo primero de todo ha sido, contar y tratar los valores faltantes en el conjunto de datos, y luego se ha aplicado un criterio específico de eliminar aquellas variables que su contenido era escaso.

Las variables eliminadas han sido aquellas que sus observaciones no contaban con:

1. Aquéllas para las que no se ha podido descargar la web o el contenido es escaso (columna *sizetext* < 100 o en blanco)
2. Aquéllas para las que no se cuenta con información económica
3. Una vez pasado por el filtro de limpieza de datos, la base está lista para trabajar con ella y toda aquella información garantiza calidad y precisión de los datos.

3.3. Modelos estadísticos

Para poder sacar conclusiones, será necesario aplicar modelos estadísticos a la base de datos. Estos modelos son herramientas muy útiles para entender los datos, hacer predicciones, clasificar observaciones, identificar relaciones entre variables, controlar procesos y tomar decisiones en base a la información recolectada.

Cabe recalcar que hay diferentes tipos de modelos, y dependiendo de la información de la base de datos, se aplicaran unos u otros. Algunos de los modelos que se podrían aplicar son el análisis PCA, el análisis de clústeres, y los modelos de clasificación y regresión, que posteriormente se explicarán en detalle.

Con los resultados de estos modelos, se podrán obtener conclusiones más precisas y útiles.

3.3.1 Análisis PCA

Uno de los modelos mencionados anteriormente es el análisis de componentes principales o PCA.

Según Tarazona (2024), el PCA es un método estadístico no supervisado que se utiliza para la reducción de la dimensionalidad. Esto significa que se utiliza para reducir el número de variables dentro del conjunto de datos, conservando la mayor cantidad posible de la variabilidad presente en los datos originales. Este método tiene mucha utilidad cuando se tiene un gran número de variables, muchas de las cuales pueden estar correlacionadas entre sí.

El PCA se aplica principalmente a variables numéricas, ya que implica operaciones matemáticas como la suma de cuadrados y multiplicaciones de matrices, que no son directamente aplicables a variables categóricas.

Las componentes principales se forman como combinaciones lineales de las variables originales y no están correlacionadas entre sí.

Si se tiene una matriz de datos X con I observaciones y J variables:

$$t_k = p_{k1}X_1 + p_{k2}X_2 + \cdots + p_{kJ}X_J = Xp_k$$

$$T = XP$$

Ecuación 1: Cálculo PCA

Fuente: (Tarazona, PCA, Análisis de componentes principales, 2024)

T= Matriz de scores.

P= Matriz de loadings

Como las componentes principales son ortogonales y si se exige que los vectores de loadings en cada componente sean unitarios (tenga norma 1):

$$PP^t = I \Rightarrow X = TP^t$$

Ecuación 2: Cálculo PCA

Fuente: (Tarazona, PCA, Análisis de componentes principales, 2024)

Se puede extraer tantas PC como $\min \{I-1, J\}$, aunque con pocas será suficiente para darle explicación a la mayor parte de la variabilidad de los datos (inercia). Si nos quedamos con K componentes principales:

$$X = TP^t + E$$

Ecuación 3: Cálculo PCA

Fuente: Libro inteligencia de negocios I

E = Matriz de residuos.

Por lo que el porcentaje de variabilidad de X que viene explicado por el PCAs se calculará como:

$$100\lambda_{\alpha} / \sum_{k=1}^K \lambda_k$$

Ecuación 4: Cálculo PCA

Fuente: (Tarazona, PCA, Análisis de componentes principales, 2024)

3.3.2 Análisis de clústeres

Otro modelo estadístico el cual es útil para poder obtener información sería el análisis de *clusters*.

Según Tarazona (2023), este modelo se trata de un método descriptivo no supervisado, el cual clasifica un conjunto heterogéneo de elementos en grupos, según las similitudes o diferencias que tengan entre ellos.

Existen varios tipos de clústeres los cuales pueden ser:

- **Cluster jerárquico:** Se basa en estructurar sus elementos en *clusters* en orden jerárquico, pudiendo así obtener un dendograma.
- **Métodos de partición:** Partición de los elementos en un número prefijo de *clusters*.
- **Métodos híbridos:** Se trata de una combinación de varios enfoques o algoritmos.

Es importante saber que hay una forma de comprobar si existen *clusters* dentro de tus datos y se podría saber aplicando el estadístico de Hopkins. Este funciona de tal manera que determina si un conjunto de datos está uniformemente distribuido o si contiene grupos significativos de puntos.

El estadístico de Hopkins H se calcula como la proporción de la suma de las distancias de los puntos aleatorios respecto a la suma total de distancias:

$$H = \frac{\sum_{i=1}^m y_i}{\sum_{i=1}^m x_i + \sum_{i=1}^m y_i}$$

Ecuación 5: Estadístico Hopkins

Fuente: Sonia Tarazona

En la ecuación x_i son las distancias de los puntos de la nuestra original y y_i son la distancia de los puntos aleatorios.

Por lo que la interpretación de los resultados del estadístico Hopkins según el manual de comandos de R (2021) sería:

- **H = 0,5**: Indica que los datos están distribuidos de forma aleatoria.
- **H = 1**: Indica fuerte tendencia de agrupación, es decir, existen *clusters*.
- **H = 0**: Indica que los datos están repartidos de una forma muy regular.

Por otra parte, una vez los datos estén preparados para continuar con el análisis, es importante seleccionar la medida de similitud o distancia.

La medida de distancia o similitud es una métrica que cuantifica la similitud o disimilitud entre dos puntos en un espacio de características. Estas medidas son fundamentales para determinar cómo se agrupan los puntos y cómo influyen a la formación de los *clusters*.

A continuación, se presentan algunas de las distancias que se podría utilizar, junto con sus respectivas ecuaciones:

Distancia euclídea: $d_{ii'} = \sqrt{\sum_{j=1}^J \frac{1}{s_j^2} (x_{ij} - x_{i'j})^2} = \sqrt{(\vec{x}_i - \vec{x}_{i'})' (\vec{x}_i - \vec{x}_{i'})}$

Distancia euclídea estandarizada: $d_{ii'} = \sqrt{\sum_{j=1}^J \frac{1}{s_j^2} (x_{ij} - x_{i'j})^2} = \sqrt{(\vec{x}_i - \vec{x}_{i'})' \bar{D}^{-1} (\vec{x}_i - \vec{x}_{i'})}$

Distancia de Mahalanobis: $d_{ii'} = (\vec{x}_i - \vec{x}_{i'})' \bar{S}^{-1} (\vec{x}_i - \vec{x}_{i'})$

Distancia Manhattan: $d_{ii'} = \sum_{j=1}^J |x_{ij} - x_{i'j}|$

Ecuación 6: Medidas de distancia

Fuente: Tarazona (2024)

Al obtener la distancia es necesario aplicar un algoritmo para la formación de los *clusters*. Dos de los algoritmos más comunes basados en el método de partición son el *k-medias* y el *k-medoides*.

El algoritmo *k-medias* selecciona *k* puntos como centros de los *clusters* iniciales, asignando aleatoriamente los elementos a los *clusters* y calculando los centroides de los *clusters* así formados. Luego, utiliza la distancia euclídea de cada elemento a los *k* centroides y los asigna al *cluster* cuyo centroide está más próximo.

Por otro lado, el algoritmo *k-medoides* (también conocido como PAM) utiliza uno de los puntos del conjunto de datos como el *medoide* para representar el centro del *cluster*. Este algoritmo es más robusto que el *k-medias* frente a datos con ruido y datos anómalos, además puede aplicarse a cualquier medida de distancia.

3.3.3 Clasificación

Otro tipo de modelo estadístico es el de clasificación. A diferencia de los modelos mencionados en los apartados anteriores, este modelo es un tipo de aprendizaje supervisado, en el cual se conoce la verdadera clase de cada elemento que se utilizó para construir el clasificador.

Según Debón, (2024), la función principal de este modelo es predecir una determinada clase (categórica) para un objeto. Partiendo de un conjunto de ejemplos ya clasificados, construye el modelo que permita clasificar casos nuevos.

Existen diversos tipos de modelos de clasificación los cuales son:

- 1- **Clasificación binaria:** Su función es asignar una observación a una de dos clases posibles.
- 2- **Clasificación multiclase:** Su principal función es asignar una observación a una de tres o más clases posibles.
- 3- **Clasificación multietiqueta:** Cada observación puede pertenecer a varias observaciones a la vez.

Cada uno de los modelos mencionados aplican determinados algoritmos comunes para poder desarrollar un resultado. Existen varios algoritmos los cuales se pueden aplicar que son:

- **Regresión logística:** Es un modelo lineal que se usa para la clasificación binaria. Su función es predecir la probabilidad que tiene esa observación a una clase usando una función *logit*.
- **Árboles de decisión o clasificación:** La entrada indica las observaciones caracterizables mediante variables o atributos. Se clasifican las observaciones desde la raíz hasta las hojas, además de que cada nodo especifica el test de algún atributo.
- **Bosque aleatorio:** Este algoritmo añade una nueva capa de aleatoriedad al proceso *bagging*, formando un conjunto de árboles de decisión, donde cada árbol se entrena mediante una muestra aleatoria seleccionada de un conjunto de datos.
- **Vecino más cercano:** La función de este algoritmo clasifica una observación nueva mediante las clases de sus K vecinos más cercanos en el espacio de características.
- **Naive Bayes:** Es un clasificador probabilístico basado en el teorema de Bayes.

$$P(C/X_1, X_2, \dots, X_p) = \frac{P(X_1, X_2, \dots, X_p/C)P(C)}{P(X_1, X_2, \dots, X_p)}$$

Fuente: (Debón, 2024)

Donde C es la clase y X_p son los valores predictores. La $P(C)$ es la probabilidad de una clase C . $P(X_1, \dots, X_p/C)$ es la verosimilitud dada la clase C . Y , por último, el denominador es la probabilidad de observar la evidencia observada.

- **Máquinas de soporte vectorial:** Son un conjunto de algoritmos de aprendizaje supervisado que se usan para la clasificación de datos. En resumen, un SVM busca un hiperplano que divide de forma más eficiente los puntos de una clase de la otra.

Existen diferencias técnicas para poder elegir el modelo adecuado. En este caso, la técnica utilizada es el *hold out* repetido, es una técnica de validación cruzada que consiste en dividir el conjunto de datos en dos subconjuntos: uno de prueba y otro de entrenamiento.

El objetivo principal es entrenar el modelo en el subconjunto de entrenamiento y luego evaluar el rendimiento en el subconjunto de prueba. Este proceso se repite varias veces, generando una división aleatoria en cada iteración.

Por lo tanto, esta técnica puede ser útil para escoger el modelo adecuado y poder seguir con el estudio.

3.3.4 Regresión

Los modelos de regresión, al igual que los de clasificación, son modelos de aprendizaje supervisado, ya que se conoce la predicción de una variable numérica de cada una de las observaciones que se usan para construir un modelo.

La funcionalidad de un modelo de regresión consta en predecir una variable numérica (ordinal o cardinal) para un objeto. Utilizando un conjunto de ejemplos en los que se conoce el valor de la variable la cual se desea predecir y con eso construir un modelo que permita predecir nuevos casos.

La principal diferencia entre clasificación y regresión es que la regresión trata de predecir una variable cuantitativa, y la clasificación trata de predecir una clase (variable categórica)

Los algoritmos utilizados en los modelos de regresión pueden ser los mismos mencionados en el apartado de clasificación, teniendo en cuenta que la función que desarrollan es similar a la de clasificación, aunque adaptada para predecir valores numéricos en lugar de clases categóricas.

Para realizar la evaluación de los modelos de regresión se van a usar dos medidas de evaluación de errores que son:

1. NMAE: El significado de sus siglas es Error Absoluto Medio Normalizado (*Normalized Mean Absolute Error*). Es una medida de la magnitud promedio de

los errores en una predicción. Un NMAE bajo indica que el modelo está produciendo predicciones precisas según *Rstudio* (2023). La ecuación del error según Hodson (2022) sería la siguiente:

$$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Ecuación 8: Error absoluto medio

Fuente: Hodson

Donde:

y_i son los valores observados.

\hat{y}_i son los valores predichos.

n son el número de observaciones.

2. RMSE: El significado de sus siglas es Raíz del Error Cuadrático Medio (*Root Mean Square Error*). Es una medida de la magnitud promedio de los errores al cuadrado en una predicción. Un RMSE bajo también indica que el modelo está produciendo predicciones precisas, pero penaliza más los errores grandes que el NMAE según *Rstudio* (2023). La ecuación del error según Hodson (2022) sería la siguiente:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Ecuación 9: Raíz del error cuadrático

Fuente: Hodson

Donde:

y_i son los valores observados.

\hat{y}_i son los valores predichos.

n son el número de observaciones.

4. Resultados

Una vez completado todo el código para el procesamiento de la base de datos, se procede a analizar los resultados obtenidos y a identificar las relaciones con los objetivos de este Trabajo Final de Grado. Este análisis es crucial para proporcionar una explicación y obtener información relevante sobre los objetivos planteados. Tras seguir todos los pasos descritos en la metodología, se procede a interpretar los resultados. los resultados.

4.1. Análisis descriptivo univariante

Es una etapa fundamental en la exploración de las variables para conocer la distribución, la tendencia y la dispersión de cada una de ellas

Lo primero que realizó fue una división de datos según la información proporcionada por las variables, tal como se mencionó en los apartados anteriores, ya que esto facilitó la elaboración del estudio.

Las variables se agrupan en cuatro categorías, siendo las variables económicas las que aportan mayor valor al estudio, debido a que proporcionan la información más relevante sobre la competitividad de las empresas

Tabla 3: Resumen datos

	IngExpl	ROA	Emp	Va	Rdo
Min	408	-154,46	1	-29387,8	-60280
1st Qu	1392	0,5258	9	356,4	7,38
median	3725	2,8825	19	819,2	57,01
Mean	26910	3,70	59,53	3561,9	476,51
3rd Qu	13374	7,5225	46	2320,9	289,35
Max	3547216	58,89	4116	518167	93500

Fuente: Elaboración propia

Dentro del grupo de variable económicas, se puede observar cinco variables que proporcionan información sobre la rentabilidad y del tamaño de empresas, lo que permite tener una idea del tipo de empresas presentes en la base de datos. Se observa que hay una empresa con más de 4000 empleados.

Además, es relevante notar que el ROA mínimo es negativo. Esto puede ocurrir debido a una disminución en las ventas, costos extraordinarios o un aumento en los costes operativos de la empresa.

Una vez conocida las variables económicas, se procederá a analizarlas gráficamente utilizando los gráficos de caja y bigotes (*boxplot*) para identificar posibles valores inconsistentes. Es importante destacar que, dado que cada variable numérica tiene un rango diferente, se realizará un escalado mediante normalización *Z-score* para estandarizar los datos. Esto permitirá comparar las variables de manera más efectiva y observar el rango en el que oscilan los resultados.

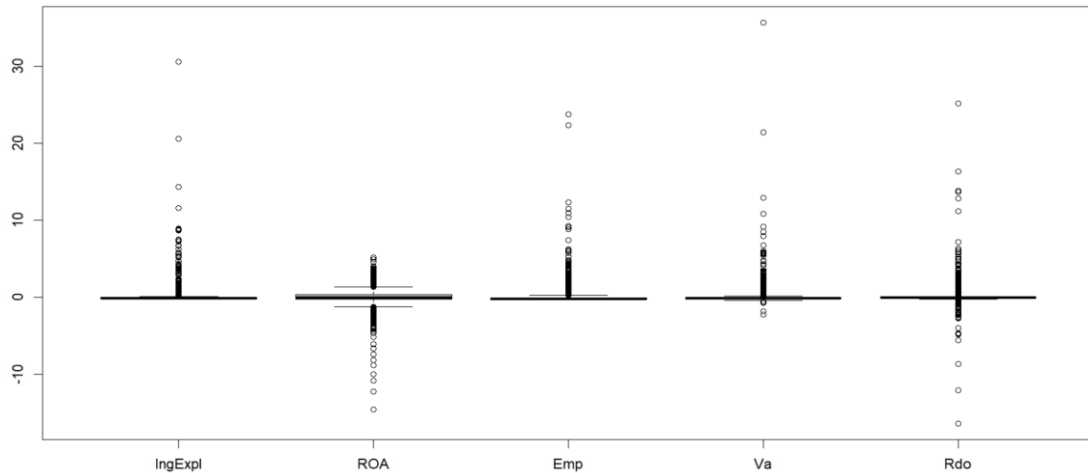


Ilustración 2: *Boxplot* datos anómalos

Fuente: *Elaboración propia*

Una vez obtenido el *boxplot*, se observaron la existencia de datos anómalos en todas las variables. Para determinar la naturaleza de estos datos, se verificó si podrían ser valores extremos en lugar de anómalos. Tras este análisis, se decidió eliminar los datos anómalos en lugar de transformarlos, ya que se identificó que muchas empresas, tanto grandes como pequeñas no aportaban información relevante al estudio.

En primer lugar, se utilizó el método del rango intercuartílico para la detección de anómalos. Finalmente, se optó por eliminar esos datos anómalos e imputar todos aquellos datos tuvieran menos de un 20% de valores faltantes, ya que así se preservaría una cantidad razonable de información. De esta manera, se asegura que el análisis se realice con datos más limpios y representativos.

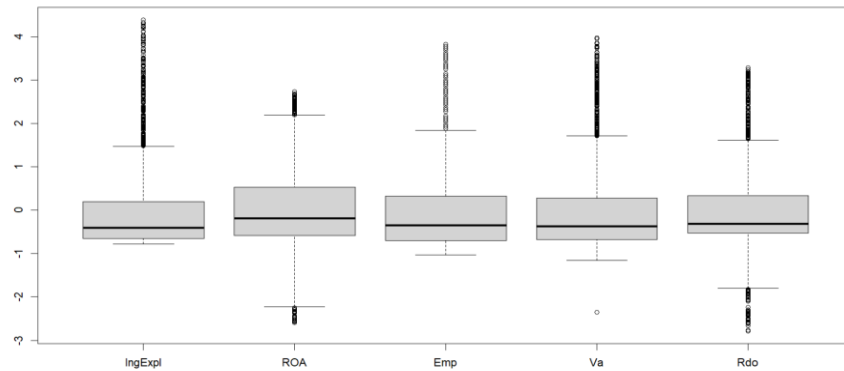


Ilustración 3: Boxplot datos anómalos corregidos

Fuente: Elaboración propia

Una vez eliminados los datos anómalos, se puede observar que la distribución de las variables se aproxima más a una distribución normal, lo que permite continuar con el estudio con los datos correctos y más fiables.

Después de eliminar los datos anómalos e imputar los datos faltantes utilizando la biblioteca MICE, se verificará el cambio experimentado en las variables para asegurar que la calidad de los datos sea adecuada para el análisis.

Tabla 4: Análisis univariante de variables económicas

	IngExpl	ROA	Emp	Va	Rdo
Min	408,2	-9,95	1	-1465,3	-408,82
1st Qu	1196,8	0,61	8	308,8	7,85
median	2614,1	2,64	15	633,4	45,02
Mean	5044,7	3,63	22,17	1032,5	104,37
3rd Qu	6222,3	6,41	29	1328,6	159,83
Max	31094,5	17,99	101	5266,9	711,11

Fuente: Elaboración propia

Una de las mayores diferencias observadas es el número de empleados, ya que antes el máximo era superior a 4000 empleados y ahora se ha reducido a 100. Esta reducción ayudará a que los gráficos no se vean distorsionados al extraer conclusiones. Además, se ha notado que los valores máximos de datos como el ROA y el VA también se ha reducido. Esto se debe a que, al eliminar los datos anómalos, se ha compactado la variabilidad de las empresas que conforman la base de datos.

Se observa también que los ingresos de explotación y los resultados han variado. Esto ocurre porque al eliminar los datos anómalos, se han eliminado las empresas con ingresos de explotación y resultados extremadamente bajos y altos, haciendo que las empresas restantes sean más similares entre sí.

Por otro lado, se ha hecho un análisis univariante de las palabras clave para recolectar más información sobre ellas.

Tabla 5: Análisis univariante de palabras clave

Variables	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
kwstems_es_adit	0	0	0	7,79	1	4134
kwstems_es_aeropon	0	0	0	0,00	0	0
kwstems_es_agav	0	0	0	1,55	0	1565
kwstems_es_agricultur_regener	0	0	0	0,24	0	503
kwstems_es_agricultur_vertical	0	0	0	0,00	0	11
kwstems_es_agu_alcalin	0	0	0	0,00	0	8
kwstems_es_aguacat	0	0	0	10,81	0	9672
kwstems_es_agu_de_mar	0	0	0	0,75	0	1854
kwstems_es_algas_marin	0	0	0	1,05	0	1206
kwstems_es_alimentacion_conscient	0	0	0	0,07	0	62
kwstems_es_aliment	0	0	2	112,15	23	45476
kwstems_es_alto_en_protein	0	0	0	0,31	0	393
kwstems_es_antioxid	0	0	0	38,89	4	37343
kwstems_es_ayun_intermitent	0	0	0	0,12	0	202
kwstems_es_azuc_de_coc	0	0	0	1,54	0	2819
kwstems_es_baj_en_calori	0	0	0	2,88	0	2662
kwstems_es_baj_en_gras	0	0	0	1,99	0	315
kwstems_es_baj_indic_glucem	0	0	0	0,15	0	55
kwstems_es_beb_funcional	0	0	0	0,02	0	24
kwstems_es_biodegrad	0	0	0	1,77	0	1294
kwstems_es_calid	0	7	35	254,69	17125	16973
kwstems_es_calori	0	0	0	9,30	0	5025
kwstems_es_carn	0	0	2	268,94	2725	59044
kwstems_es_comerci_just	0	0	0	0,49	0	406
kwstems_es_competit	0	0	0	12,86	2	4596
kwstems_es_compost	0	0	0	15,18	0	17773
kwstems_es_comunicacion	0	0	4	48,93	12	5307

kwstems_es_conservacion	0	0	2	25,46	10	2457
kwstems_es_conserv	0	1	9	251,75	48	56414
kwstems_es_crudivegan	0	0	0	0,01	0	31
kwstems_es_crud	0	0	0	43,36	2	24858
kwstems_es_cultiv_de_tej	0	0	0	0,00	0	1
kwstems_es_detox	0	0	0	10,56	0	9654
kwstems_es_ecolog	0	0	0	185,49	3	43810
kwstems_es_edulcor_natural	0	0	0	2,61	0	4449
kwstems_es_enriquec_con_mineral	0	0	0	0,00	0	7
kwstems_es_enriquec_con_vitamin	0	0	0	0,08	0	33
kwstems_es_envas	0	0	4	109,34	36	23090
kwstems_es_eritritol	0	0	0	0,85	0	2147
kwstems_es_espirulin	0	0	0	4,02	0	7853
kwstems_es_exportacion	0	0	0	11,73	0	9349
kwstems_es_ferment	0	0	0	32,83	0	33664
kwstems_es_fibr	0	0	0	52,02	4	41588
kwstems_es_gran_enter	0	0	0	0,62	0	428
kwstems_es_gras	0	0	2	133,85	41	47371
kwstems_es_harin_de_almendr	0	0	0	1,82	0	2007
kwstems_es_harin_de_coc	0	0	0	0,67	0	2036
kwstems_es_hidropon	0	0	0	0,00	0	1
kwstems_es_huell_de_carbon	0	0	0	2,95	0	1913
kwstems_es_importacion	0	0	0	2,61	0	3074
kwstems_es_integral	0	0	0	60,46	3	20543
kwstems_es_jug_verd	0	0	0	0,00	0	6
kwstems_es_kef	0	0	0	3,61	0	4128
kwstems_es_ket	0	0	0	39,73	0	64779
kwstems_es_kilometr_cer	0	0	0	0,06	0	59

kwstems_es_lech_de_almendr	0	0	0	1,37	0	3620
kwstems_es_lech_de_coc	0	0	0	4,04	0	5933
kwstems_es_lech_de_soj	0	0	0	0,10	0	63
kwstems_es_lech_vegetal	0	0	0	0,71	0	1614
kwstems_es_libr_de_transgen	0	0	0	0,02	0	12
kwstems_es_marc	0	2	9	154,86	46	37890
kwstems_es_match	0	0	0	26,95	0	22796
kwstems_es_miel_crud	0	0	0	0,94	0	2626
kwstems_es_natural	0	1	14	310,59	85	142301
kwstems_es_omeg_3	0	0	0	17,98	0	20338
kwstems_es_organ	0	0	2	38,29	10	13106
kwstems_es_pale	0	0	0	8,32	0	25900
kwstems_es_permacultur	0	0	0	0,01	0	13
kwstems_es_plant_bas	0	0	0	6,56	0	2087
kwstems_es_plastic	0	0	0	60,16	2	107313
kwstems_es_preci	0	0	2	444,76	56	284441
kwstems_es_precocin	0	0	0	14,65	0	11435
kwstems_es_probiot	0	0	0	12,44	0	22570
kwstems_es_produccion	0	0	4	52,18	20	13677
kwstems_es_product	0	45	245	1734,86	91125	435063
kwstems_es_product_de_tempor	0	0	0	1,46	0	949
kwstems_es_protein	0	0	0	114,14	29	57817
kwstems_es_quino	0	0	0	9,37	0	2799
kwstems_es_recet	0	0	1	405,43	34	226496
kwstems_es_rekurs	0	0	0	13,53	4	2952
kwstems_es_salud	0	0	2	194,58	26	176039
kwstems_es_seit	0	0	0	7,98	0	6751
kwstems_es_semill	0	0	0	83,85	2	88906

kwstems_es_sin_adit	0	0	0	4,64	0	2126
kwstems_es_sin_azuc	0	0	0	60,83	0	29355
kwstems_es_sin_colesterol	0	0	0	0,09	0	88
kwstems_es_sin_conserv	0	0	0	7,55	0	3097
kwstems_es_sin_glut	0	0	0	193,71	0	75135
kwstems_es_sin_lacte	0	0	0	1,16	0	3118
kwstems_es_sin_plastic	0	0	0	0,02	0	24
kwstems_es_sin_transgen	0	0	0	0,01	0	4
kwstems_es_smoothi	0	0	0	5,59	0	5609
kwstems_es_sosten	0	0	0	29,24	3	8910
kwstems_es_stevi	0	0	0	8,50	0	12106
kwstems_es_superaliment	0	0	0	7,15	0	16812
kwstems_es_superfood	0	0	0	0,73	0	1513
kwstems_es_suplement_vitamin	0	0	0	0,12	0	239
kwstems_es_tempeh	0	0	0	0,08	0	101
kwstems_es_tofu	0	0	0	3,89	0	3979
kwstems_es_vegan	0	0	0	95,28	0	128566
kwstems_es_vegetarian	0	0	0	19,67	0	10161
kwstems_es_xilitol	0	0	0	3,07	0	8642
kwstems_es_zer_wast	0	0	0	0,53	0	814

Fuente: Elaboración propia

Como se puede observar en la tabla, hay palabras que aparecen con más frecuencia que otras, como pueden ser “natural”, “precio”, “recetas”, “producto”, “salud” y “vegano”.

Una vez realizado la eliminación de los datos anómalos y la imputación de los datos faltantes, se procederá con el estudio de la base de datos, sabiendo que los datos tienen una distribución adecuada.

4.2. Análisis descriptivo bivariante

Para entender la relación que puede existir entre dos de las variables del estudio, es necesario realizar un análisis descriptivo bivariante. Este análisis ayuda a identificar si existe una relación significativa entre dos variables y a comprender los cambios que realizan juntas.

La manera más fácil de comprender la relación que existen entre dos variables es observándolas visualmente mediante un gráfico de correlaciones.

4.2.1. Análisis de correlaciones

Para generar dicho gráfico, se va a hacer primero un análisis de correlación para medir la dirección y la fuerza lineal entre dos variables cuantitativas.

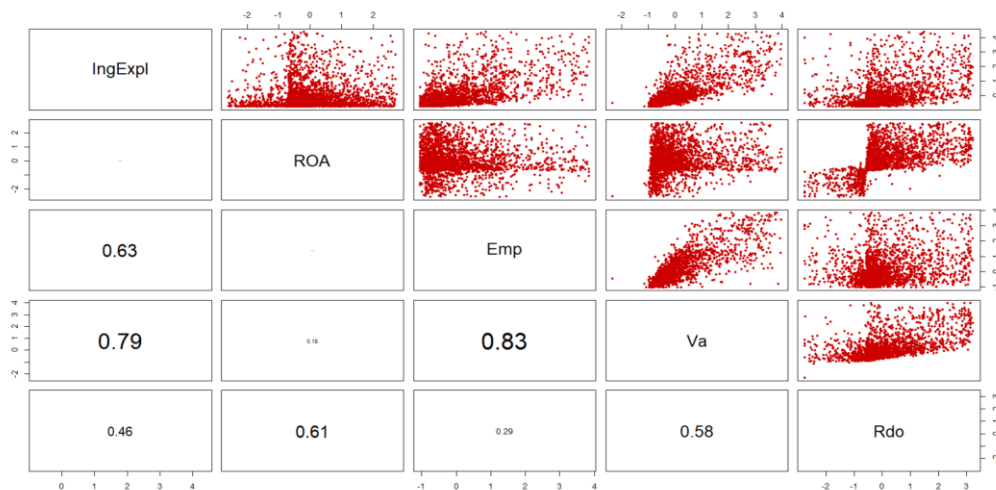


Ilustración 4: Gráfico de correlaciones

Fuente: Elaboración propia

En este gráfico se puede observar todas las relaciones posibles que existen entre dos variables.

La correlación más fuerte es la de empleados con valor añadido, lo cual puede deberse a que, por lo general, en las empresas del sector, cuanto más personal contratan, más producción generan. Por lo tanto, la relación entre estas variables tiene mucho sentido.

Por otro lado, la segunda correlación más fuerte es entre ingresos de explotación y valor añadido. Esta relación se explica porque los ingresos de explotación provienen de las actividades productivas que realizan las empresas, como la venta de productos o servicios, campañas de marketing, entre otras actividades del sector. Estas actividades contribuyen al valor añadido, por lo que, a mayor valor añadido, mayores serán los ingresos de explotación.

Las relaciones mencionadas hasta ahora son esperables, ya que todas informan sobre el tamaño de las empresas. Sin embargo, al prestar atención al gráfico, se puede observar que la variable ROA solo tiene relación con el resultado. Esto se debe a que el ROA es una variable que únicamente proporciona información sobre el rendimiento económico de las empresas del sector.

La relación entre el ROA y el resultado se debe a que el ROA depende de la variabilidad del resultado. En general, a mayor resultado, mayor será el ROA.

Por otro lado, se ha realizado un análisis bivariante entre las variables económicas y las palabras clave, ya que es fundamental para determinar el sentido de la relación, ya sea positivo o negativo.

Tabla 6: Análisis de correlaciones entre variables económicas y palabras clave

	IngExpl	ROA	Emp	Va	Rdo
kwstems_es_adit	0.0012	0.0007	0.0009	0.0038	0.0007
kwstems_es_agav	-0.0045	-0.0023	-0.0035	-0.0030	-0.0021
kwstems_es_agricultur_regener	0.4040	0.0038	0.4250	0.6452	0.4653
kwstems_es_agricultur_vertical	-0.0016	0.0029	-0.0013	-0.0012	-0.0005
kwstems_es_agu_alcalin	-0.0056	0.0132	-0.0067	-0.0056	-0.0034
kwstems_es_aguacat	0.0188	0.0097	0.0332	0.0350	0.0523
kwstems_es_agu_de_mar	-0.0008	0.0025	0.0028	0.0054	0.0088
kwstems_es_algas_marin	0.0162	0.0102	0.0127	0.0306	0.0604
kwstems_es_alimentacion_conscient	0.0203	-0.0013	0.0360	0.0430	0.0100
kwstems_es_aliment	0.0286	0.0097	0.0287	0.0290	0.0154
kwstems_es_alto_en_protein	0.0145	0.0294	0.0227	0.0286	0.0203
kwstems_es_antioxid	-0.0046	0.0085	0.0034	0.0059	0.0181
kwstems_es_ayun_intermitent	0.0045	0.0223	0.0188	0.0204	0.0333
kwstems_es_azuc_de_coc	-0.0051	-0.0059	-0.0055	-0.0048	-0.0036
kwstems_es_baj_en_calori	-0.0031	0.0026	-0.0007	0.0018	0.0102
kwstems_es_baj_en_gras	0.0894	0.0151	0.1292	0.1291	0.0974
kwstems_es_baj_indic_glucem	-0.0018	-0.0035	0.0039	0.0043	0.0162
kwstems_es_beb_funcional	0.0238	-0.0027	0.0168	0.0166	0.0108
kwstems_es_biodegrad	0.0158	-0.0089	0.0090	0.0130	0.0081

kwstems_es_calid	0.0389	0.0167	0.0584	0.0492	0.0212
kwstems_es_calori	0.0206	0.0108	0.0366	0.0403	0.0406
kwstems_es_carn	0.0157	0.0062	0.0119	0.0273	0.0512
kwstems_es_comerci_just	0.0020	-0.0068	0.0175	0.0153	-0.0042
kwstems_es_competit	0.0029	-0.0137	0.0076	0.0050	0.0004
kwstems_es_compost	-0.0022	0.0088	-0.0004	0.0022	0.0035
kwstems_es_comunicacion	0.0587	-0.0030	0.0616	0.0510	0.0197
kwstems_es_conservacion	-0.0071	0.0068	-0.0036	-0.0049	-0.0028
kwstems_es_conserv	-0.0057	-0.0017	0.0040	0.0033	0.0083
kwstems_es_crudivegan	-0.0058	-0.0363	-0.0054	-0.0059	-0.0072
kwstems_es_crud	-0.0005	0.0189	-0.0065	-0.0026	0.0069
kwstems_es_cultiv_de_tej	-0.0021	0.0180	0.0053	-0.0004	-0.0171
kwstems_es_detox	-0.0053	0.0568	0.0022	0.0154	0.0549
kwstems_es_ecolog	-0.0134	0.0008	-0.0148	-0.0130	-0.0062
kwstems_es_edulcor_natural	-0.0009	0.0002	0.0090	0.0096	0.0182
kwstems_es_enriquec_con_mineral	0.0438	-0.0060	0.0495	0.0798	0.0569
kwstems_es_enriquec_con_vitamin	0.0280	0.0027	0.0473	0.0571	0.0353
kwstems_es_envas	0.0168	0.0183	0.0308	0.0314	0.0154
kwstems_es_eritritol	-0.0045	0.0120	-0.0055	-0.0043	-0.0019
kwstems_es_espirulin	-0.0045	-0.0008	-0.0046	-0.0026	0.0013
kwstems_es_exportacion	0.0147	-0.0066	0.0222	0.0128	0.0032
kwstems_es_ferment	0.0288	-0.0092	0.0654	0.0742	0.0291
kwstems_es_fibr	0.0154	0.0057	0.0349	0.0424	0.0087
kwstems_es_gran_enter	0.0027	-0.0073	0.0127	0.0086	0.0043
kwstems_es_gras	0.0214	0.0034	0.0490	0.0369	0.0177
kwstems_es_harin_de_almendr	-0.0027	0.0028	-0.0008	-0.0030	-0.0030
kwstems_es_harin_de_coc	-0.0039	0.0066	-0.0051	-0.0037	-0.0012
kwstems_es_hidropon	0.0600	-0.0150	0.0061	-0.0071	-0.0598

kwstems_es_huell_de_carbon	0.0661	0.0024	0.0647	0.0474	0.0151
kwstems_es_importacion	0.0004	-0.0058	-0.0023	-0.0011	0.0018
kwstems_es_integral	0.0235	0.0111	0.0430	0.0410	0.0095
kwstems_es_jug_verd	-0.0030	0.0067	-0.0020	-0.0006	0.0009
kwstems_es_kef	0.0035	-0.0102	0.0083	0.0039	-0.0067
kwstems_es_ket	-0.0045	0.0383	-0.0065	-0.0042	-0.0005
kwstems_es_kilometr_cer	0.0073	0.0002	0.0275	0.0135	0.0015
kwstems_es_lech_de_almendr	-0.0030	-0.0003	-0.0036	-0.0022	-0.0001
kwstems_es_lech_de_coc	-0.0013	-0.0157	0.0049	-0.0005	-0.0029
kwstems_es_lech_de_soj	0.0263	0.0323	0.0205	0.0532	0.1116
kwstems_es_lech_vegetal	-0.0038	-0.0065	-0.0032	-0.0033	-0.0023
kwstems_es_libr_de_transgen	0.0029	-0.0180	0.0087	0.0029	-0.0245
kwstems_es_marc	0.0369	-0.0174	0.0571	0.0501	0.0160
kwstems_es_match	-0.0004	-0.0007	-0.0041	-0.0014	0.0007
kwstems_es_miel_crud	-0.0045	0.0130	-0.0057	-0.0043	-0.0020
kwstems_es_natural	0.0055	0.0357	0.0153	0.0252	0.0498
kwstems_es_omeg_3	0.0071	0.0172	0.0200	0.0233	0.0354
kwstems_es_organ	0.0090	0.0246	0.0123	0.0148	0.0259
kwstems_es_pale	-0.0041	0.0206	-0.0057	-0.0041	-0.0022
kwstems_es_permacultur	-0.0001	0.0036	-0.0021	0.0030	0.0067
kwstems_es_plant_bas	0.2507	0.0060	0.1613	0.2542	0.1850
kwstems_es_plastic	0.0096	-0.0006	0.0185	0.0194	0.0109
kwstems_es_preci	-0.0101	0.0076	-0.0078	-0.0074	-0.0033
kwstems_es_precocin	-0.0065	0.0145	0.0018	-0.0033	-0.0034
kwstems_es_probiot	-0.0025	0.0003	-0.0018	0.0008	0.0038
kwstems_es_produccion	0.0726	0.0070	0.0558	0.0585	0.0368
kwstems_es_product	0.0036	0.0115	0.0210	0.0231	0.0309
kwstems_es_product_de_tempor	-0.0034	-0.0675	0.0121	0.0029	-0.0034

kwstems_es_protein	0.0225	0.0338	0.0215	0.0264	0.0227
kwstems_es_quino	0.0390	0.0128	0.0483	0.0497	0.0493
kwstems_es_recet	0.0350	0.0199	0.0345	0.0635	0.0911
kwstems_es_rekurs	0.2438	0.0054	0.2999	0.3386	0.2147
kwstems_es_salud	0.0153	0.0036	0.0217	0.0262	0.0275
kwstems_es_seit	0.0345	0.0145	0.0440	0.0634	0.0456
kwstems_es_semill	-0.0050	0.0010	-0.0033	-0.0030	-0.0017
kwstems_es_sin_adit	-0.0047	-0.0029	0.0020	-0.0005	-0.0050
kwstems_es_sin_azuc	-0.0062	-0.0189	0.0073	0.0035	0.0074
kwstems_es_sin_colesterol	-0.0036	0.0211	-0.0054	-0.0035	0.0010
kwstems_es_sin_conserv	-0.0072	0.0081	0.0026	0.0015	0.0046
kwstems_es_sin_glut	0.0150	-0.0089	0.0491	0.0338	0.0102
kwstems_es_sin_lacte	0.0018	0.0893	0.0108	0.0271	0.0762
kwstems_es_sin_plastic	0.0011	-0.0041	0.0046	0.0029	0.0008
kwstems_es_sin_transgen	-0.0016	-0.0224	0.0011	-0.0028	-0.0165
kwstems_es_smoothi	0.0171	-0.0156	0.0291	0.0212	0.0055
kwstems_es_sosten	0.1294	0.0150	0.1367	0.1701	0.1127
kwstems_es_stevi	-0.0049	-0.0080	-0.0005	-0.0025	-0.0038
kwstems_es_superaliment	-0.0039	-0.0087	-0.0044	-0.0036	-0.0020
kwstems_es_superfood	-0.0029	-0.0094	-0.0027	-0.0017	-0.0027
kwstems_es_suplement_vitamin	0.0016	0.0004	-0.0010	0.0018	0.0024
kwstems_es_tempeh	0.0075	-0.0748	0.0105	0.0103	0.0038
kwstems_es_tofu	-0.0008	-0.0100	0.0000	0.0030	0.0073
kwstems_es_vegan	-0.0027	0.0397	-0.0004	-0.0002	0.0012
kwstems_es_vegetarian	0.0187	0.0070	0.0212	0.0366	0.0571
kwstems_es_xilitol	-0.0045	-0.0233	-0.0058	-0.0046	-0.0027
kwstems_es_zer_wast	0.0382	0.0150	0.0311	0.0420	0.0336

Fuente: Elaboración propia

Una vez observada la tabla, puede verse que las palabras que mayor correlación presentan con las variables económicas son “agricultura regenerativa”, “plant-based”, “recursos” y “sostenible”. En los cinco casos la relación con las variables económicas es positiva, indicando que, a mayor presencia de estas palabras, mayores ingresos de explotación, ROA, empleados, valor añadido y resultados. Cabe destacar que, para el caso del ROA, la relación lineal es más débil. Más adelante, en la sección 4.3.2 se explicará el porqué de esta diferencia.

4.3. Análisis descriptivo multivariante

Llegados a este punto, para obtener resultados más concretos y comprender cómo las variables interactúan entre sí y cómo afectan al resultado final que se desea obtener, se procederá a realizar un análisis multivariante.

Para llevar a cabo este análisis multivariante, se utilizarán técnicas explicadas en el apartado de metodología.

4.3.1 Resultados PCA

El primer método que se ha seguido para tener en cuenta todas las variables y poder encontrar la relación entre ellas es el análisis de PCA, como se explicó en el apartado 3.3.1. Cabe recordar que esta técnica reduce la dimensionalidad, disminuyendo el número de variables y extrayendo los componentes principales que explican la mayor parte de la variabilidad de los datos.

Lo primero que se debe hacer para comenzar con el análisis multivariante es identificar los componentes principales. Para obtener esa información, se va a realizar un *Scree Plot* o gráfico del codo, ya que este tipo de gráfico sirve para determinar el número óptimo de componentes principales en el modelo.

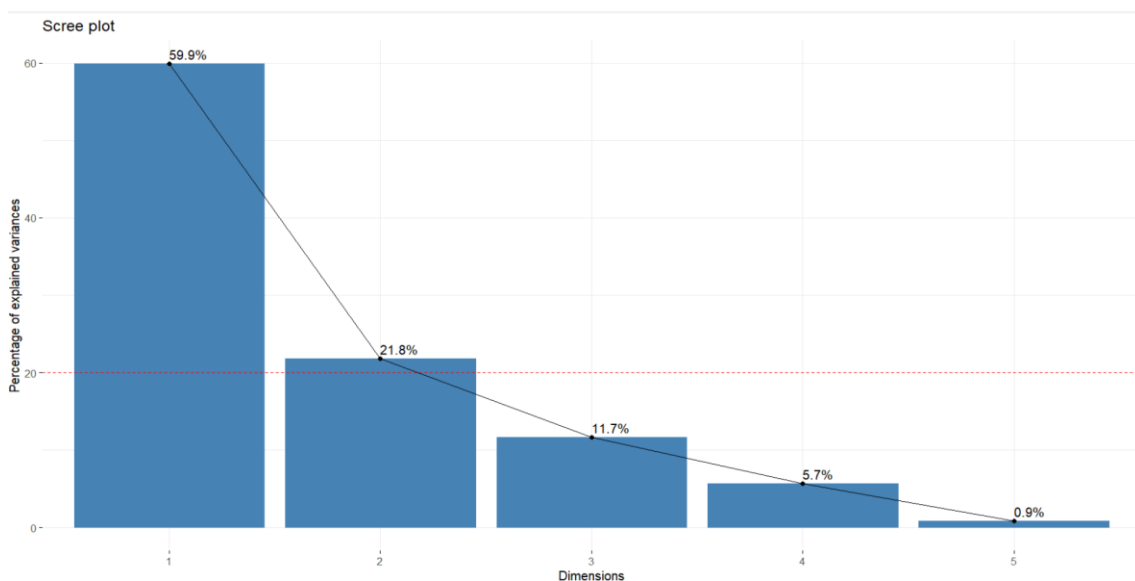


Ilustración 5: Gráfico del codo

Fuente: Elaboración propia

Este gráfico del codo muestra el eje X, que representa el número de componentes principales, y el eje Y, que representa el porcentaje de la varianza total explicado por cada componente. Por otra parte, la línea punteada roja representa el porcentaje de variabilidad de cada componente principal si todas explicaran lo mismo.

Aplicando tanto el método del codo como el de la variabilidad media, se puede observar en el gráfico que para el análisis deben seleccionarse las dos primeras componentes principales, ya que estas explican el 81,7% de la varianza total. Por lo tanto, el porcentaje total obtenido con las dos componentes será suficiente para continuar con el análisis.

A continuación, se realizará un *loading plot* de variables de PCA. Este gráfico proporcionará información sobre la contribución de cada variable a los componentes principales y la relación entre las variables. Este gráfico es muy útil para el análisis, ya que facilita una comprensión más clara de la relación entre las variables y entender la estructura de los datos.

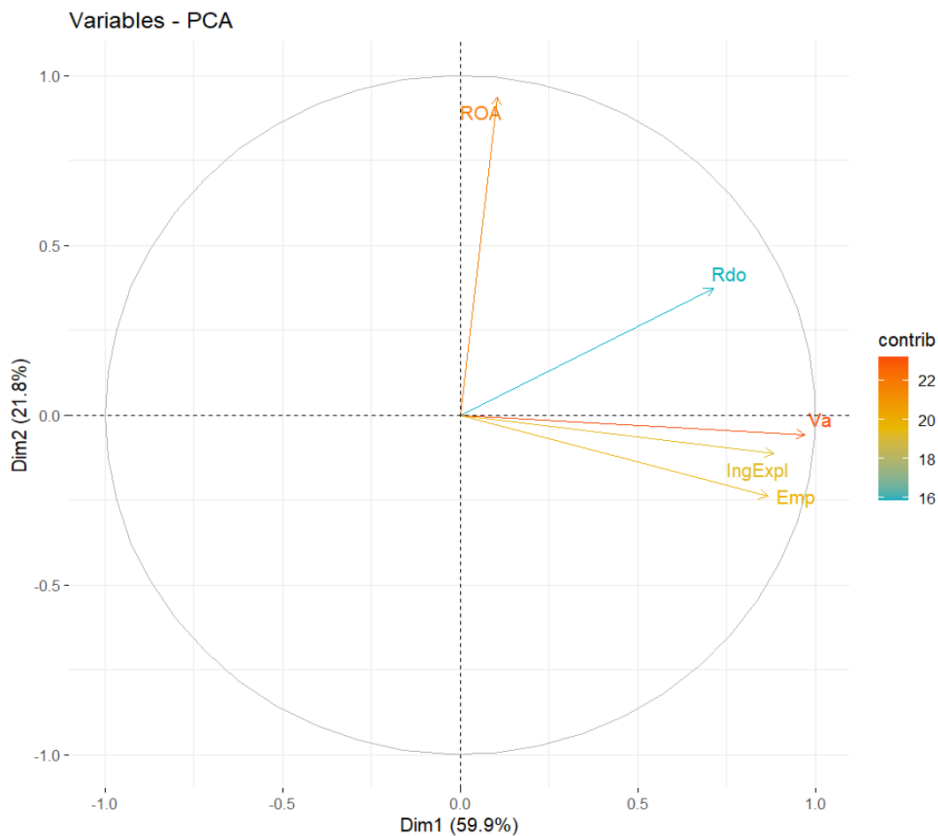


Ilustración 6: Loading plot

Fuente: Elaboración propia

Como se puede observar en el gráfico, cada flecha representa una variable de los datos. Además, el color y la distancia de cada flecha representa la contribución que tiene cada variable con las componentes principales, siendo azul poca contribución y rojo mucha contribución. Así que, como se observa en el gráfico, las variables que más contribuyen a la dimensión uno serían Valor añadido (VA) y empleados (Emp), y la variable que más contribuye a la dimensión dos sería ROA.

Por lo tanto, la dimensión uno recoge la mayor parte de la variabilidad, como ya se había observado en el gráfico del codo, permitiendo ver la relación existente entre las variables Va, IngExpl y Emp. Las variables más cercanas entre si están correlacionadas, y esta relación puede deberse a que todas las variables explican el tamaño de las empresas. Por lo general, las empresas del sector que cuentan con más trabajadores van a obtener mayores ingresos de explotación y valor añadido.

Por otro lado, las variables más alejadas, como el ROA con respecto a las variables mencionadas, tienen menos correlación. Esto se debe que el ROA aporta información sobre la rentabilidad de las empresas. Además, el ROA contribuye significativamente a la dimensión dos.

Por otro lado, se puede observar que la variable Rdo se encuentra con un tono azul entre Roa y el resto de las variables, lo que indica que existe poca relación con las demás variables y una baja contribución a las componentes principales.

Se puede observar que la variable Rdo tiene menor correlación con ROA que con el resto de las variables. Esto puede deberse a que las empresas del sector pueden estar obteniendo una ROA alta no por grandes resultados, sino por pequeño volumen de activos. En la industria agroalimentaria, muchas empresas no cuentan con producción propia, sino que actúan como intermediarios, por lo que su volumen de activo es más bajo.

Otro gráfico que puede ser útil para entender cómo se distribuyen las observaciones es un "Score plot". En este gráfico, la coordenada de cada punto corresponde a la puntuación que se le ha dado a la observación en las componentes principales seleccionadas, que en este caso son las dos primeras.

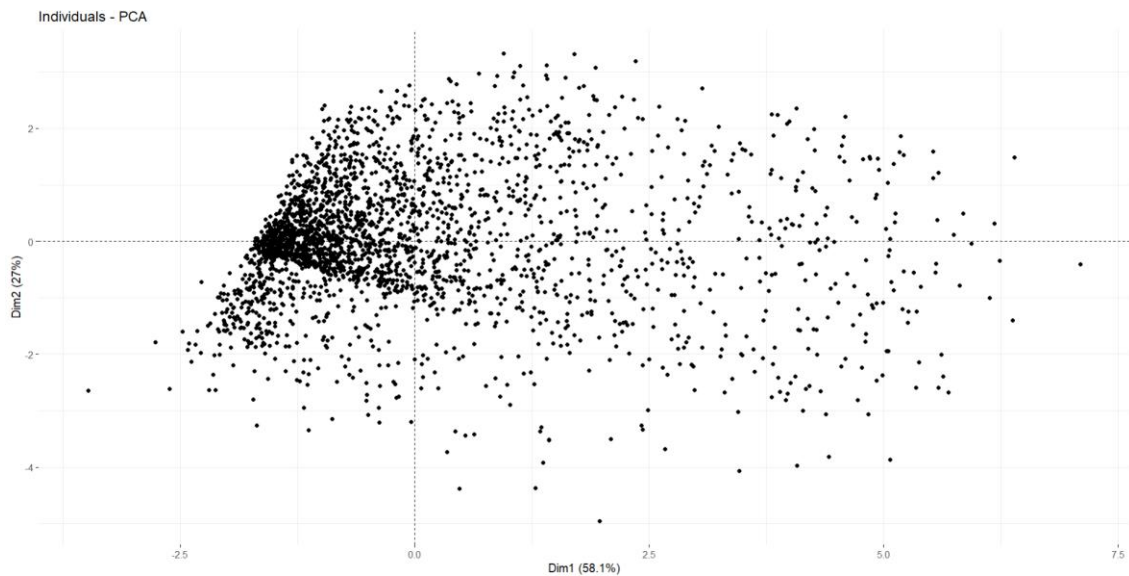


Ilustración 7: Score plot

Fuente: Elaboración propia

Como se puede observar en el gráfico, las empresas del sector agroalimentario se dividen en cuatro cuadrantes. Las que se encuentran en el cuadrante superior derecho son empresas con alta rentabilidad económica y un tamaño significativo dentro del sector.

4.3.2 Resultados *clusters*

El siguiente paso en el estudio multivariante es la aplicación del análisis de *clusters*. Como se explicó en el apartado 3.3.2, el análisis de *clusters* busca agrupar observaciones similares entre sí formando grupos o *cluster*.

Para llevar a cabo el análisis *cluster*, es necesario seleccionar los algoritmos de agrupamiento adecuados. En este caso, se han seleccionado los algoritmos *k-medias* y *k-medoides*, además de utilizar la distancia euclídea. El algoritmo *k-medias* como se explicó en otro apartado, requiere esta medida de distancia para su funcionamiento. Se utilizará para determinar el número óptimo de *clusters* y verificar cuántos grupos de datos similares se forman dentro de los datos económicos.

En primer lugar, se va a observar el gráfico de *k-medias* e interpretar los resultados que ofrece.

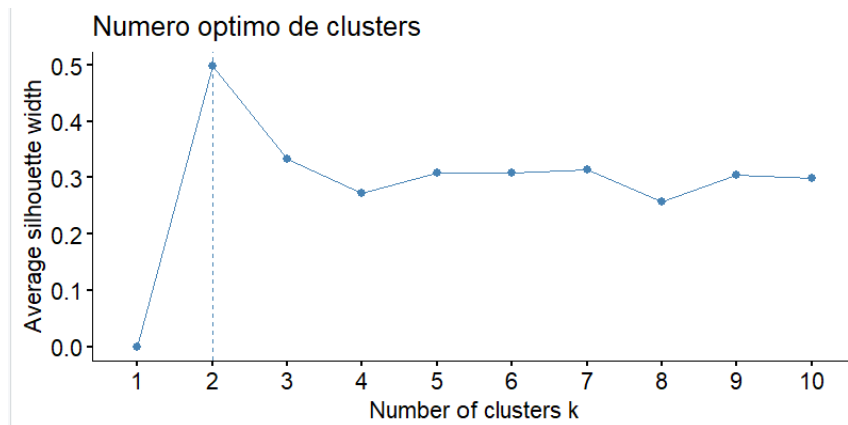


Ilustración 8: Gráfico coeficiente de silhouette k-medias

Fuente: Elaboración propia

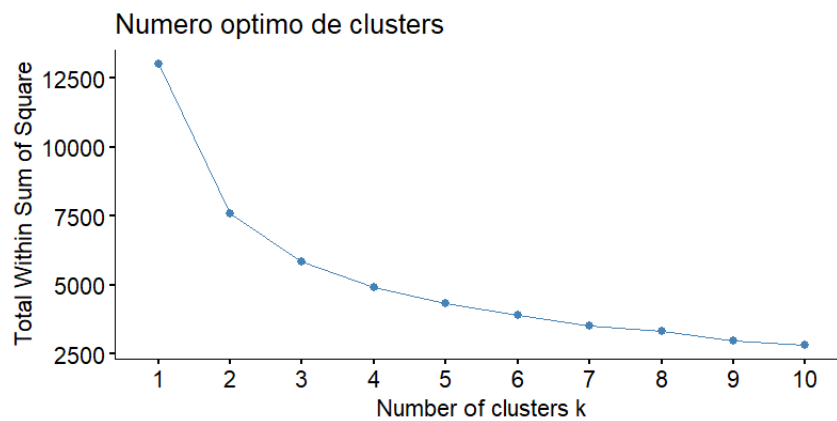


Ilustración 9: Gráfico intraclusters k-medias

Fuente: Elaboración propia

En el primer gráfico de *k-medias*, se puede observar que el método de *silhouette* sugiere que el número óptimo de *clusters* es 2. Sin embargo, en el segundo gráfico *k-medias* utilizando el método *intraclusters*, se observa que el número óptimo podría estar entre 2 y 3, dado que la varianza dentro de los *clusters* es suficientemente baja.

Por lo tanto, se ha seleccionado 2 *clusters* como el número óptimo para el análisis.

Finalmente, se ha procedido a realizar un gráfico preliminar de scores del PCA aplicado a los datos, en el que los puntos estarán coloreados según los *clusters* identificados.

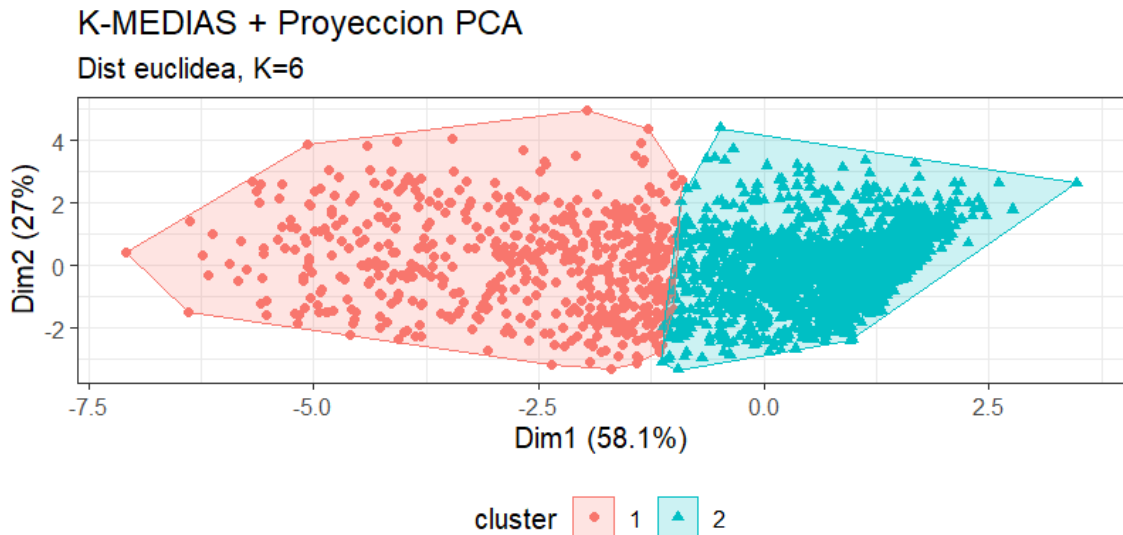


Ilustración 10: Gráficos preliminar de scores

Fuente: Elaboración propia

Como se puede observar en el gráfico, se identifican dos *clusters*. En el *cluster* verde, se nota una mayor concentración y una agrupación más compacta de empresas. Por otro lado, el *cluster* presenta una distribución más dispersa. Además, se observa que hay poco solapamiento entre los *clusters*, lo que sugiere un resultado favorable y bien definido.

A continuación, se procederá a analizar los gráficos generados por el algoritmo *k-medoides* para determinar el número óptimo de *clusters*.

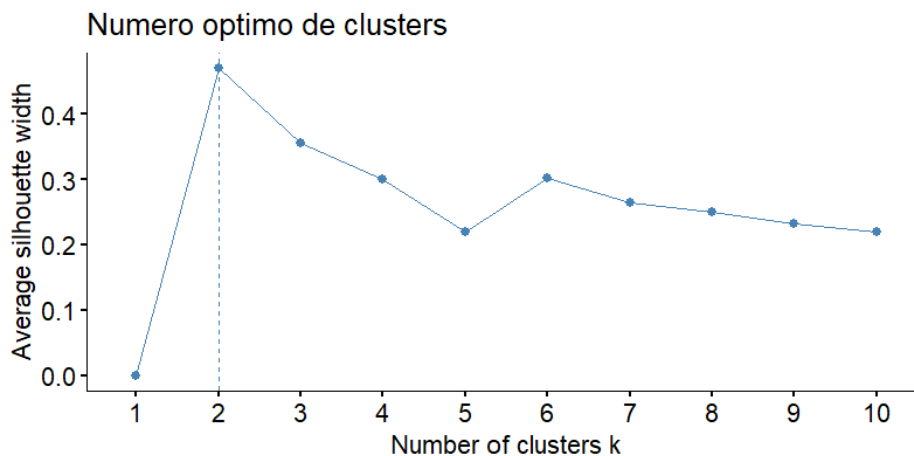


Ilustración 11: Gráfico silhouette *k-medoides*

Fuente: Elaboración propia

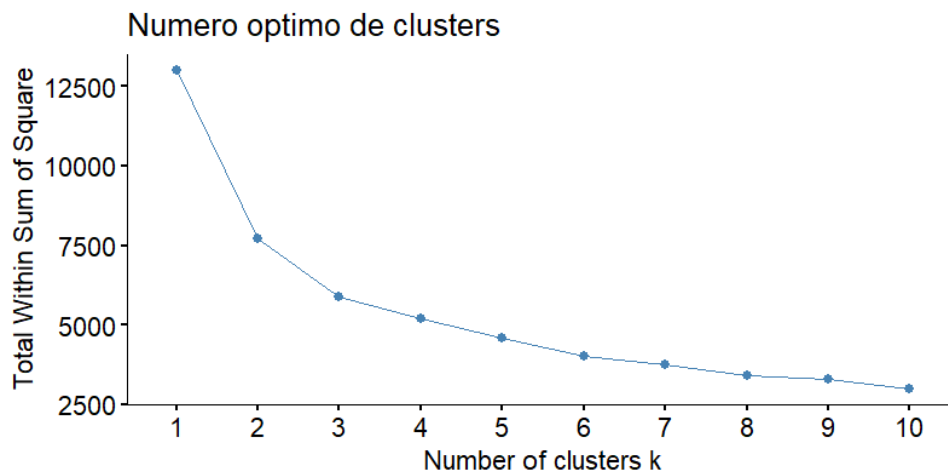


Ilustración 12: Gráfico intraclusters K-medoides

Fuente: Elaboración propia

Como se puede observar en el primer gráfico, el método de *silhouette* sugiere que el número óptimo de *clusters* es 2.

En el segundo gráfico, utilizando el método de *intracluster*, se confirma que el número óptimo de *clusters* también está entre 2 y 3, similar a los resultados obtenidos con el algoritmo K-medias.

Por lo tanto, para el algoritmo *k-medoides*, se selecciona 2 como el número óptimo de *clusters*.

Finalmente, se procederá a generar un gráfico preliminar de scores del PCA, aplicando el algoritmo *K-medoides* a los datos y coloreando según la existencia de *clusters*.

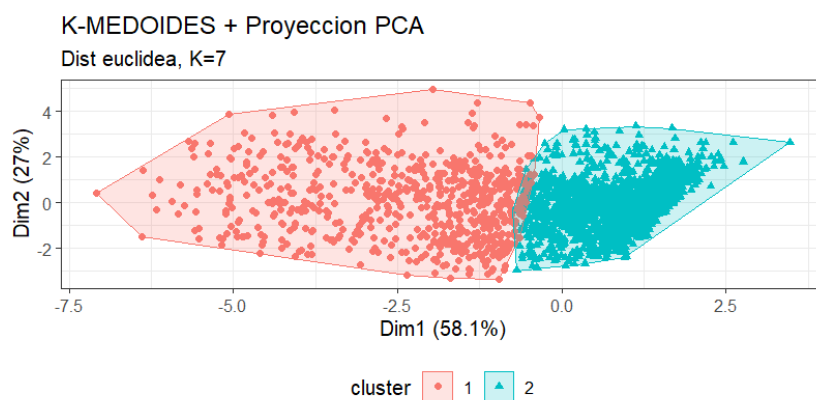


Ilustración 13: Gráfico preliminar de scores k-medoides

Fuente: Elaboración propia

Como se puede observar en la imagen, hay presencia de dos *clusters*, que fue el número óptimo seleccionado. El *cluster 2*, se observa una gran cantidad de observaciones muy agrupadas, mientras que en el *cluster 1* las observaciones están menos agrupadas. También se puede notar un poco de solapamiento en los *clusters*, lo cual podría ser un impedimento a la hora de hacer el análisis.

Con la información obtenida de las gráficas de los algoritmos de agrupación *k-medias* y *k-medoides*, se procederá a elegir el algoritmo más adecuado para continuar con el estudio, según el coeficiente de silhouette.

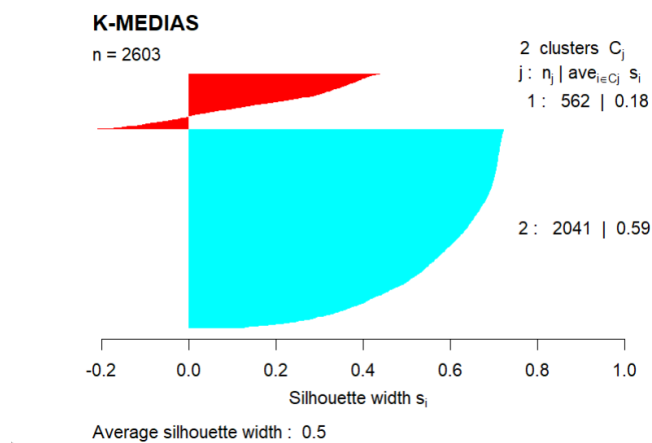


Ilustración 14: Coeficiente de silhouette *k-medias*

Fuente: Elaboración propia

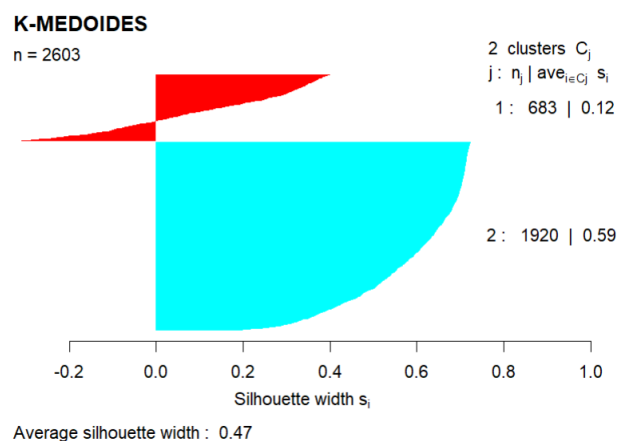


Ilustración 15: Coeficiente silhouette *k-medoides*

Fuente: Elaboración propia

A la vista de los coeficientes de silhouette, parece ser que el mejor es el algoritmo de *k-medias*, ya que muestra un coeficiente promedio ligeramente superior y presenta menos empresas del sector mal clasificadas, es decir, aquellas con un coeficiente

negativo. Esto sugiere que, aunque las diferencias son pequeñas, *k-medias* ofrece una mejor calidad de agrupamiento para los datos en cuestión, permitiendo una clasificación más precisa y coherente de las empresas del sector agroalimentario según sus características económicas

Por lo tanto, se observa cómo se han agrupado las observaciones según el *algoritmo k-medias* en el gráfico *Score Plot*.

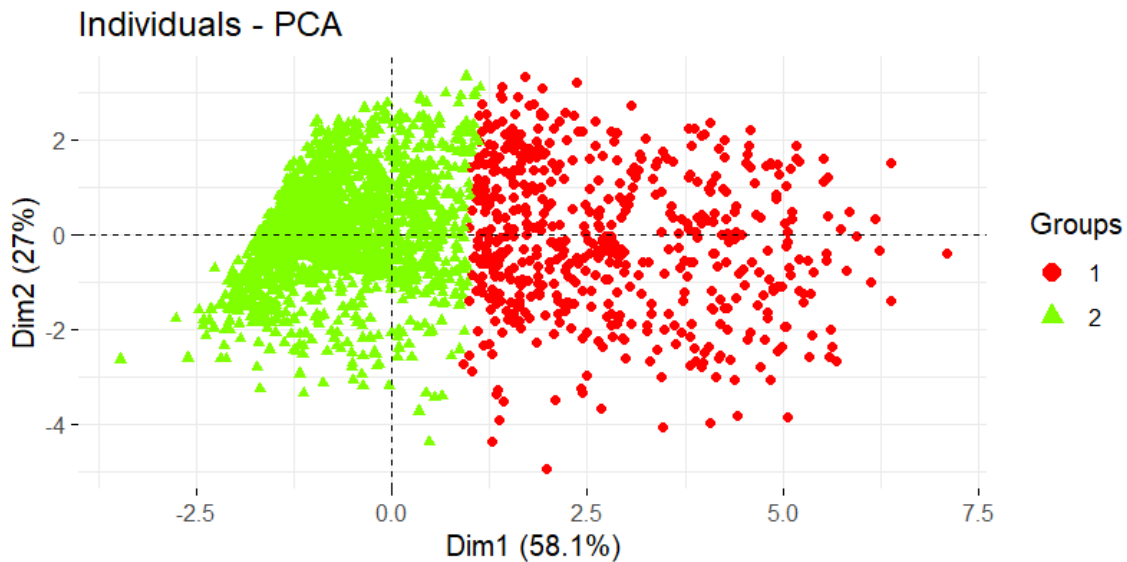


Ilustración 16: Score plot

Fuente: Elaboración propia

Como se puede observar en el gráfico, el cluster 1 (representado en verde) incluye principalmente empresas pequeñas y medianas, caracterizadas por un menor número de empleados, valor añadido e ingresos de explotación. Este grupo es predominante en el estudio, reflejando que la mayoría de las empresas analizadas son de menor tamaño.

Por otro lado, el cluster 2 (representado en rojo) agrupa a empresas más grandes, que, aunque son menos numerosas, destacan por tener mayores valores en las variables mencionadas, como empleados, valor añadido e ingresos de explotación.

El gráfico de perfil medio permitirá visualizar claramente estas diferencias, destacando cómo varían los valores medios de las principales características entre los dos clusters. Esto no solo ayuda a categorizar las empresas según su tamaño y capacidad económica, sino que también proporciona una perspectiva sobre las estructuras de las empresas del sector y las posibles estrategias competitivas que podrían estar siguiendo

A continuación, se procederá a analizar el gráfico de perfil medio de los clusters, lo cual permitirá identificar las diferencias en cada cluster y, así, profundizar en el entendimiento de las dinámicas presentes en el sector.

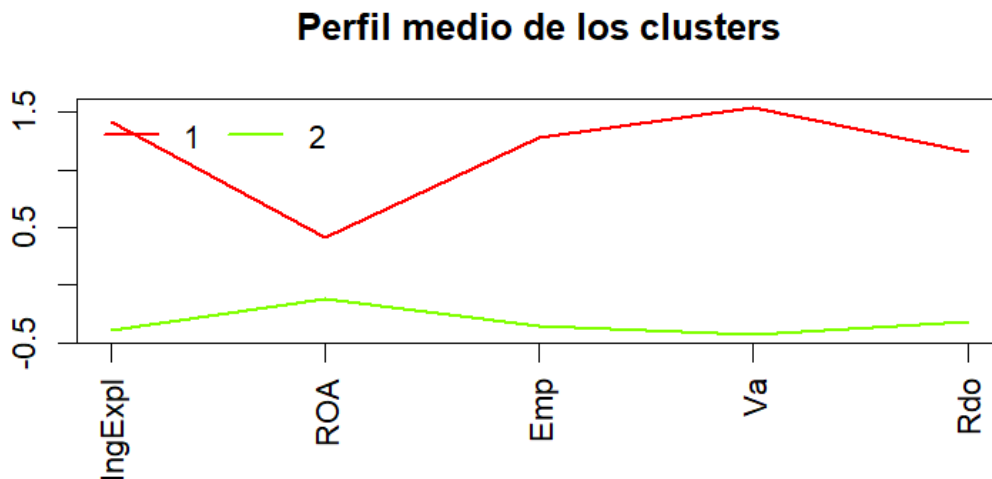


Ilustración 17: Gráfica perfil medio clusters

Fuente: Elaboración propia

El análisis de los perfiles medios de los *clusters* revela una tendencia significativa en la relación entre el tamaño de las empresas y su rendimiento financiero en el sector agroalimentario.

En el *cluster* 1 (grandes empresas), se puede observar que, a pesar de tener un mayor número de empleados, valor añadido, ingresos de explotación y resultados, estas empresas están experimentando un ROA bajo. Esto puede indicar problemas en la gestión de explotación, lo que resulta en una baja eficiencia operativa. Las grandes empresas podrían estar enfrentando desafíos como costos fijos elevados, ineficiencias en la producción, o una estructura organizativa pesada que afecta a su flexibilidad y adaptabilidad en el mercado.

En el *cluster* 2 (pequeñas empresas), aunque cuentan con menos recursos en términos de empleados y capacidad productiva presentan valores positivos y relativamente más altos de ROA. Esto sugiere que estas empresas están utilizando sus recursos de manera más eficiente, generando mayores retornos en relación con sus activos. Este comportamiento puede estar relacionado con una gestión ágil, menor carga de costos fijos, o una estrategia de nicho que les permite capitalizar oportunidades específicas en el mercado.

4.3.3 Resultado modelos de clasificación

Después de preparar los datos formando *clusters*, que se dividen en dos grupos que son pequeñas y grandes empresas del sector, se procederá a aplicar los modelos de clasificación. El objetivo es observar si estos modelos pueden predecir a qué grupo pertenece cada empresa, basándose en los indicadores digitales que presenta cada una, es decir, clasificar a una empresa en un grupo u otro según las palabras clave utilizadas.

En este caso, el grupo 1 o el grupo de interés será el de las empresas grandes, ya que presentan un mayor tamaño. Para evaluar los modelos de clasificación y seleccionar el modelo más adecuado para el estudio, se utilizará la técnica de validación cruzada *Hold out* repetido.

Una vez aplicada la técnica de *Hold out* repetido, se va a analizar los resultados y ver qué modelo es el más adecuado para seguir con el estudio.

Tabla 7: Resultados clasificación

Medias	Logit	Tree	Rf	NB	SVP	SVM	Vecino
Auc	0,53	0,58	0,68	0,51	0,59	0,45	0,55
Tasa de acierto	0,53	0,73	0,79	0,002	0,001	0,78	0,75

Fuente: Elaboración propia

Como se puede observar en la imagen, hay dos tipos de medidas de evaluación. Se puede notar que un modelo destaca sobre los demás, ya que sus puntuaciones en ambas medias son las más cercanas a uno. Esto nos lleva a concluir que el mejor modelo para poder continuar con el estudio, que nos permita clasificar adecuadamente las observaciones en los grupos de interés, es el modelo de *Random Forest* o Bosque aleatorio.

Este modelo tiene una tasa de acierto cercana al 80%, lo que indica que, de cada 10 observaciones, 8 son clasificadas correctamente. Esta precisión es más que aceptable para los propósitos del estudio

Para proporcionar una mejor comprensión visual, se presentará la curva ROC junto con la medida AUC.

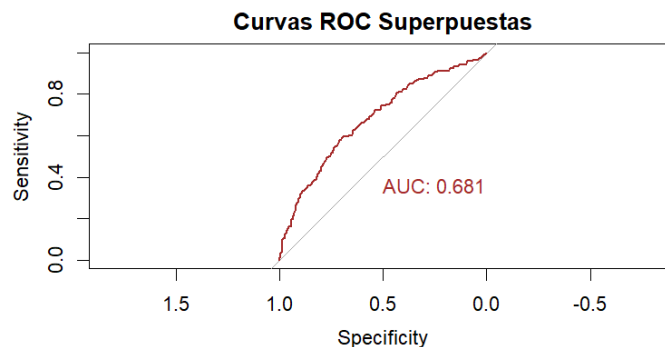


Ilustración 18: Gráfico curva Roc

Fuente: Elaboración propia

Al elegir el modelo *Random Forest*, se analizarán los resultados de clasificación obtenidos para identificar qué variables son las más relevantes para las empresas del grupo 1.

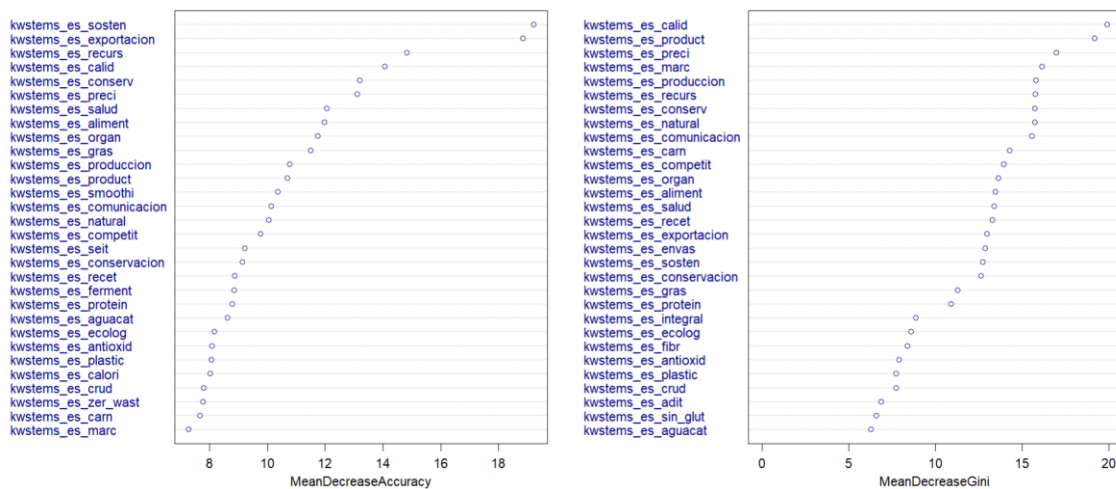


Ilustración 19: Resultados modelo Random Forest clasificación

Fuente: Elaboración propia

Se observan dos gráficos que muestran cuáles son las palabras más importantes para el algoritmo a la hora de clasificar una empresa como perteneciente al grupo 1, el cual es el más relevante para analizar la competitividad de las empresas del sector. Cada gráfico utiliza una medida diferente para representar la importancia de las palabras.

En el gráfico de la izquierda, se muestra la importancia de las variables en términos de la disminución de la precisión del modelo. En el gráfico de la derecha, se presenta la importancia de las variables en términos de la disminución del índice de Gini.

Según el gráfico de la izquierda, las palabras clave más importantes son “sostenibilidad” y “exportación”. Por otro lado, el gráfico de la derecha indica que las palabras con mayor importancia son “calidad” y “producto”. Se comprueba la relación con las nuevas tendencias en la industria.

Estos datos son muy importantes para la elaboración de estudios, ya que indican que las empresas grandes tienden a utilizar con frecuencia las palabras “sostenibilidad” y “calidad”. Se podría suponer que las empresas más competitivas en el mercado son aquellas que destacan en sus páginas web productos que son sostenibles y de alta calidad. Cabe destacar que se dispone de más de 200 variables de palabras, por lo que todas las palabras que aparecen en la tabla son términos que se utilizan con frecuencia en las páginas web de las empresas.

Al observar la imagen, se puede ver palabras como “proteína”, “saludable”, “conservantes” y “natural” que aparecen con frecuencia en las páginas web, indicando que las empresas que siguen las nuevas tendencias suelen ser las más competitivas.

4.3.4 Resultados regresión

En el apartado de clasificación, se ha usado el grupo 1, compuesto por las empresas grandes. Para la regresión, se empleará la segunda componente principal, que se centra en la rentabilidad económica de las empresas del sector. El objetivo, al igual que en clasificación, es examinar cómo los indicadores digitales están relacionadas con la segunda componente principal.

Se aplicará el método de *Hold out* repetido, al igual que en clasificación. Sin embargo, a diferencia de la clasificación, donde se mide a tasa de aciertos, en regresión no es posible utilizar esta medida. En su lugar, se considerarán las tasas de errores de los modelos, con el objetivo de encontrar el modelo que ofrezca las predicciones con la menor tasa de error.

En este caso, las dos unidades de medida de error serán NMAE y RMSE, explicadas previamente en la sección 3.3.4. Se han escogido estas unidades de medida porque ambas explican el error, aunque una es absoluta (RMSE) y la otra relativa (NMAE).

Tabla 8: Resultados regresión

	Regresión	Árbol	RF	Vecino	SVP	SVM
NMAE	2,023361	1,004535	1,000092	1,109373	1,012824	1,012931
RMSE	19,442213	1,158250	1,152825	1,249354	1,164015	1,177993

Fuente: Elaboración propia

Como se puede observar en la tabla, el modelo que menos tasa de error tiene en las dos medidas es el Random Forest o bosque aleatorio. El NMAE para el modelo de regresión es relativamente alto en comparación con los otros dos modelos. Esto indica que el error medio es aproximadamente 2 veces la media de los valores observados, lo cual no es lo ideal. Los otros modelos (Árboles, Rf, Vecino, SVP, SVM) tienen un NMAE muy cercano a 1, lo que sugiere que sus errores son comparables a la media de los valores observados, y por lo tanto, más aceptables.

Por otro lado, el RMSE para el modelo de regresión es significativamente mayor que los otros modelos. Esto indica que los errores en este modelo son mucho más grandes y dispersos y los otros modelos tienen un RMSE en torno a 1.15-1.25, lo que indica que los errores son bastante pequeños y los modelos tienen un rendimiento similar.

El modelo seleccionado ha sido el *Random Forest*, ha resultado ser el mismo que en clasificación, no por ello tiene que ser siempre así.

Esto indica que este es el modelo que mejor predice la rentabilidad de las empresas en función de los indicadores digitales. Ahora, se procederá a aplicar el modelo *Random Forest* para obtener los resultados finales del estudio.

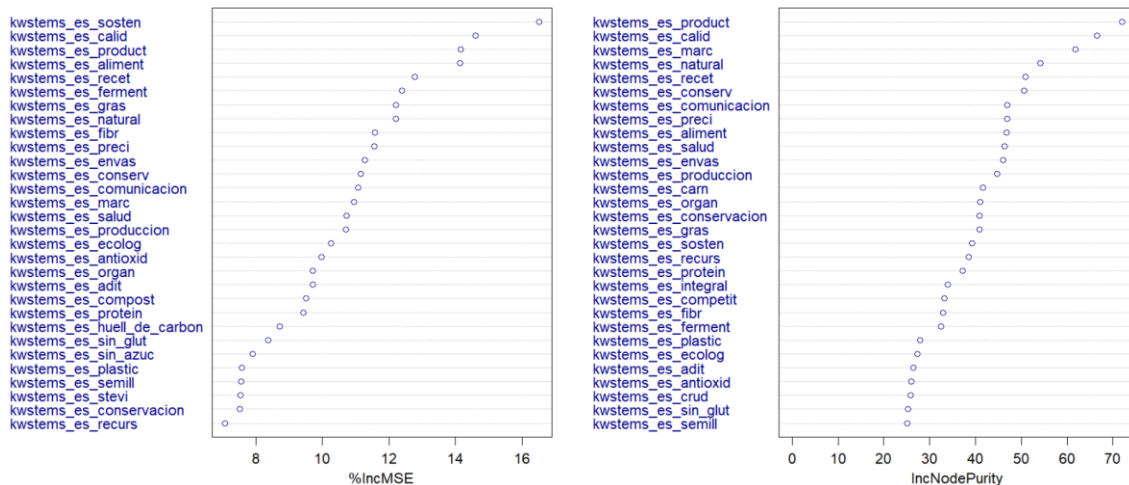


Ilustración 20: Resultado Random Forest regresión

Fuente: Elaboración propia

Como se observa en la gráfica, las palabras clave más importantes para la predicción del modelo siguen siendo “sostenible”, “producto” y “calidad”. Sin embargo, en este caso, las palabras “naturales”, “ecológico”, “sin gluten” y “receta”, cogen más importancia. Como se ha podido comprobar en el análisis de correlaciones, la contribución al rendimiento es positiva. Eso puede deberse a que estas tendencias son relativamente nuevas y aún no han sido adoptadas por todas las empresas, pero aquellas que sí lo han hecho están obteniendo un rendimiento económico satisfactorio.

En conclusión, las empresas que muestran un mayor rendimiento son aquellas que destacan en sostenibilidad, ofrecen una amplia variedad de productos atractivos para el mercado y mantienen una calidad que atrae a los consumidores. La sostenibilidad, en particular, se está consolidando como una tendencia crucial para mejorar la competitividad en el mercado, especialmente entre las empresas grandes.

Esto demuestra que la relación entre la adopción de nuevas tendencias y la competitividad es clara. Las empresas necesitan estar atentas a las necesidades y preferencias de los consumidores para posicionarse como líderes en el sector. Los resultados obtenidos son consistentes con los de la clasificación, sugiriendo que el uso de estas tendencias no solo aumenta la rentabilidad, sino que también contribuye al crecimiento del tamaño de la empresa. Así, se responde al objetivo del Trabajo Final de Grado.

5. Conclusiones

Al finalizar los resultados de este Trabajo Final de Grado y tras analizar toda la información del sector, se pueden extraer las siguientes conclusiones:

El estudio se inició con el objetivo de comprender si la competitividad de las empresas tenía relación con la incorporación de nuevas tendencias, tales como la incorporación de alimentos saludables, vegetarianos, veganos, naturales y todo lo relacionado con la mejora de la salud y la sostenibilidad medioambiental.

El objetivo principal de este trabajo se ha alcanzado mediante a la construcción de una base de datos a partir de información financiera extraída de la base de datos SABI y la implementación de herramientas de análisis de indicadores digitales. Estas herramientas han permitido identificar las palabras claves más relevantes, y se han encontrado que las empresas con mayor tamaño y rentabilidad financiera y por ende una alta competitividad, son aquellas empresas que han logrado adaptarse a las nuevas tendencias.

En particular, las empresas que destacan en el sector, ya sea por tamaño de empresa o rendimiento financiero, son aquellas que han puesto mayor énfasis en la sostenibilidad. Este parámetro resulta ser uno de los más importantes para alcanzar el liderazgo en el sector.

Además, no solo la sostenibilidad es crucial, sino también la calidad de los productos que ofrecen las empresas. La tendencia hacia productos saludables también es importante. El análisis revela que las empresas que se enfocan en ofrecer productos más saludables o naturales tienen una tendencia positiva hacia el crecimiento dentro del sector.

Aunque se esperaba que la incorporación de productos veganos y vegetarianos, que han emergido recientemente en el mercado, fuera un indicador de competitividad, los datos sugieren que su impacto aún es limitado. Es posible que, dado que los datos utilizados son de 2022, estos productos no hayan alcanzado aún una relevancia significativa.

Otro hallazgo relevante es que muchas empresas grandes presentan un bajo ROA, a pesar de tener un potencial para obtener un ROA más alto. Esto puede ser indicativo de una gestión ineficiente de sus activos. En contraste, las empresas más pequeñas, a pesar de su tamaño, parecen optimizar mejor sus recursos.

Finalmente, tanto en los resultados de clasificación como en los de regresión, se observa que no hay diferencias importantes en las palabras clave utilizadas por las grandes empresas y la más rentables. Dado que la diferencia en la rentabilidad entre grandes y pequeñas empresas es significativa, sugiere que factores ajenos al estudio podrían estar influyendo en esta variación. Por lo tanto, una vez analizados los resultados de clasificación y regresión, se puede concluir que se ha alcanzado el objetivo general de este Trabajo Final de Grado: las empresas que se adaptan a las nuevas tendencias son las que presentan mayor competitividad en el sector.

En conclusión, las empresas con mayor competitividad son aquellas que se han adaptado a las nuevas tendencias. Este estudio indica que para las empresas que desean crecer y aumentar su influencia en el sector, es crucial prestar atención a la adaptación a las nuevas tendencias, ya que esto es un indicio de mayor rentabilidad financiera y un crecimiento más significativo en el sector.

6. Bibliografía

- ABC. (13 de ENERO de 2020). *AGRONOMA*. Obtenido de <https://sevilla.abc.es/agronoma/noticias/agricultura/avances-tecnologicos-agricultura-2020/>
- Álvaro Martínez, Á. D. (s.f.). *AGRICULTURA 1.0*. Obtenido de https://www.mapa.gob.es/es/ministerio/servicios/informacion/agricultura-a-martinez_tcm30-563118.pdf
- Andrade Pinelo, A. (2011). Análisis del ROA, ROE y ROI. *Artículo+2+Noviembre+2011+Análisis+del+ROA+ROE+y+ROI.*, 4.
- Arjona Liesau, C. (2022). Las nuevas tendencias en innovación y sostenibilidad en el sector agroalimentario derivadas de los cambios en los hábitos de los consumidores. *Revista de Agricultura Sostenible*, 45-60.
- Bejar Gonzales, D., & Jove Castillo, E. (2020). *Estudio de las tecnologías 4.0 del sector de industrias alimentarias*. SUNEDU.
- Cerón, J. (27 de 03 de 2023). *Rpubs*. Obtenido de https://rpubs.com/stefens07/MAE_RMSE
- Club de emprendedores. (18 de febrero de 2018). *CEA Empresa Andaluza*. Obtenido de <https://masempresas.cea.es/el-sabi-una-herramienta-indispensable-para-el-inversor/>
- Cofece. (s.f.). *REPORTE SOBRE LAS CONDICIONES DE COMPETENCIA EN EL SECTOR AGROALIMENTARIO*.
- Cubero, V. (8 de marzo de 2024). *Plataforma Tierra*. Obtenido de <https://www.plataformatierra.es/innovacion/9-tendencias-de-la-industria-agroalimentaria-para-2024>
- Dans, E. (2020). La empresa y la web 2.0. *Harvard Deusto marketing & ventas*, 36-43.
- Dataweek. (2010). *Datamedia*. Obtenido de <https://datademias.es/blog/que-es-r>
- Debón, A. (2024). Aprendizaje supervisado: Clasificación.
- Economics, P. (31 de mayo de 2016). *Asociación de Biotecnología Vegetal Agrícola, Agro-Bio*. Obtenido de <https://agrobio.org/noticias/beneficios-economicos-de-cultivos-geneticamente-modificados-alcanzaron-150-mil-millones-de>
- Eurostat. (2008). *Statistical classification of economic activities in the European Community*. Luxembourg: EU Bookshop.
- Galiana, E., & Centeno, J. (2016). Reflexiones sobre responsabilidad social empresarial, responsabilidad pública y la sostenibilidad medioambiental. *Revista Galega de Economía*, 5-22.
- Gallopín, G. (2003). *Sostenibilidad y desarrollo sostenible*. CEPAL.

- Gracia, Á. (2001). Tendencias tecnológicas en el sector agroalimentario. *Economía industrial*, 39-46.
- Hobbs, L. (2015). *Diseñar su propia pagina web*. Marcombo, S.A.
- Hodson, T. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE). *Geoscientific Model Development*, 5481-5487.
- INE. (2020). *INSTITUTO NACIONAL DE ESTADISTICA*. Obtenido de https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176851&menu=ultiDatos&idp=1254735727106#:~:text=El%20n%C3%BAmero%20de%20explotaciones%20agr%C3%ADcolas,aumento%20del%207%2C4%25.
- institute, R. (2017). *Rodale institute*. Obtenido de <https://rodaleinstitute.org/es/porqu%C3%A9-org%C3%A1nico/pr%C3%A1cticas-de-agricultura-org%C3%A1nica/rotaciones-de-cultivos/#:~:text=La%20rotaci%C3%B3n%20de%20cultivos%20es,las%20plagas%20y%20las%20malezas>.
- Kassambara, A. (2021). *R-project.org*. Obtenido de https://search.r-project.org/CRAN/refmans/factoextra/html/get_clust_tendency.html
- KPMG. (mayo de 2022). Obtenido de <https://assets.kpmg.com/content/dam/kpmg/es/pdf/2022/05/transformacion-sostenible-sector-agroalimentario.pdf>
- Laajimi, A. &. (1997). *El Consumo de Alimentos en España. Cambios y nuevas tendencias*. Obtenido de https://www.researchgate.net/profile/Abderrouf-Laajimi/publication/44813354_El_consumo_de_alimentos_en_Espana_Cambios_y_nuevas_tendencias/links/54329cfe0cf225bddcc7bfc3/El-consumo-de-alimentos-en-Espana-Cambios-y-nuevas-tendencias.pdf.
- Landois. (3 de octubre de 2021). Obtenido de Landois.com: [https://blog.landois.com/paginas-web/que-es-una-pagina-web#:~:text=de%20%C3%A1rea%20local\),Funci%C3%B3n%20de%20una%20p%C3%A1gina%20web,trav%C3%A9s%20de%20diferentes%20redes%20inform%C3%A1ticas](https://blog.landois.com/paginas-web/que-es-una-pagina-web#:~:text=de%20%C3%A1rea%20local),Funci%C3%B3n%20de%20una%20p%C3%A1gina%20web,trav%C3%A9s%20de%20diferentes%20redes%20inform%C3%A1ticas).
- Lantern. (2020). Obtenido de Lantern.es: <https://www.lantern.es/papers/the-green-revolution-entendiendo-el-auge-del-mundo-veggie>
- Llopart Casanovas, A. (17 de febrero de 2023). *Iebschool*. Obtenido de <https://www.iebschool.com/blog/indicadores-estrategia-digital-redes-sociales/>
- Marcano, L., & Wilmer, F. (2013). *SciELO*. Obtenido de https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1315-01622013000200009#:~:text=Los%20valores%20an%C3%B3malos%20son%20oun,m%C3%A9todos%20estad%C3%ADsticos%20univariantes%20o%20multivariantes.

- México, G. d. (19 de enero de 2018). *Gobierno de México*. Obtenido de <https://www.gob.mx/asercia/articulos/conoces-el-origen-del-maiz?idiom=es#:~:text=Se%20considera%20que%20el%20ma%C3%ADz,a%20unos%20kil%C3%B3metros%20de%20Mitla>.
- Ministerio de agricultura, pesca y alimentación. (26 de enero de 2023). *Ministerio de agricultura, pesca y alimentación*. Obtenido de https://www.mapa.gob.es/es/alimentacion/temas/industria-agroalimentaria/20230126informeanualindustria2021-20224t22ok_tcm30-87450.pdf
- Organización Mundial de la Salud*. (1 de marzo de 2024). Obtenido de <https://www.who.int/es/news-room/fact-sheets/detail/obesity-and-overweight#:~:text=Datos%20y%20cifras,ellos%2C%20890%20millones%20er an%20obesos>.
- Román, J. (2003). La calidad en los productos del medio rural. *Acciones e Investigaciones Sociales*, págs. 191-195.
- Tarazona, S. (2023). Análisis Clustering.
- Tarazona, S. (2024). PCA, Análisis de componentes principales.
- The Nature Conservancy*. (19 de julio de 2021). Obtenido de <https://www.nature.org/es-us/que-hacemos/nuestras-prioridades/ciudades-saludables/detener-residuos-plasticos/#:~:text=La%20producci%C3%B3n%20e%20incineraci%C3%B3n%20de,infiltran%20en%20nuestra%20cadena%20alimentaria>.
- tierra, P. (4 de Enero de 2023). *Plataforma tierra*. Obtenido de <https://www.plataformatierra.es/actualidad/9-tendencias-de-la-industria-agroalimentaria-2023>
- Trazable. (30 de Mayo de 2022). *Trazable.io*. Obtenido de Trazable.io: <https://trazable.io/blog/que-es-la-cadena-de-suministro-en-la-industria-alimentaria/>
- Trends, F. (1 de 12 de 2022). *IA Coial*. Obtenido de <https://www.coial.org/las-principales-tendencias-que-marcaran-la-industria-alimentaria-en-2024/>
- Universidad de Valencia*. (s.f.). Obtenido de <https://www.uv.es/ceaces/multivari/cluster/CLUSTER2.htm>
- Universidad Europea*. (24 de junio de 2022). Obtenido de <https://universidadeuropea.com/blog/que-es-sostenibilidad-ambiental/>
- Van Dijk, B., & Informa, S. (2017). *SABI: Sistemas de Análisis de Balances Ibéricos*. Informa D&B.
- Vega, J. B. (s.f.). R para principiantes. En J. B. Vega.

Anexo I. Objetivos de Desarrollo Sostenible

Tabla 9. Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. Fin de la pobreza				x
ODS 2. Hambre cero	x			
ODS 3. Salud y bienestar	x			
ODS 4. Educación de calidad		x		
ODS 5. Igualdad de género				x
ODS 6. Agua limpia y saneamiento		x		
ODS 7. Energía asequible y no contaminante		x		
ODS 8. Trabajo decente y crecimiento económico	x			
ODS 9. Industria, innovación e infraestructuras	x			
ODS 10. Reducción de las desigualdades		x		
ODS 11. Ciudades y comunidades sostenibles				x
ODS 12. Producción y consumo responsables		x		
ODS 13. Acción por el clima	x			
ODS 14. Vida submarina			x	
ODS 15. Vida de ecosistemas terrestres	x			
ODS 16. Paz, justicia e instituciones sólidas				x
ODS 17. Alianzas para lograr objetivos			x	

Fuente: Elaboración propia.

Uno de los ODS con los que más se involucra este Trabajo Final de Grado es con el ODS 2 hambre cero, ya que las tendencias del sector agroalimentario hacen que pueda mejorar la eficacia y la productividad, lo cual hace que aumente la producción y el suministro de alimentos, haciendo que se pueda ver reducido el hambre en el mundo.

Por otro lado, el ODS 3 la salud y el bienestar. Este tiene una relación con la implementación saludable de los alimentos al mercado y seguros, ya que cada vez se aplica la practicas agrícolas sostenibles, lo cual hace que el producto final al ser consumido sea más saludable para las personas.

El ODS 8 trabajo decente y crecimiento económico. El sector agroalimentario es uno de los sectores que más empleo crea, ya que cada vez es más necesario gracias a la modernización.

El ODS 9 industria, innovación e infraestructuras. El uso de las nuevas tecnologías y la innovación del sector agroalimentario están siendo cruciales para el desarrollo de infraestructuras sostenibles y competitivas.

El ODS 13 Acción por el clima. Gracias a el desarrollo tecnológico las prácticas agroalimentarias sostenibles pueden reducir las emisiones de gases efecto invernadero pudiendo así reducir el impacto en el cambio climático.

Y por último el ODS 15 Vida de ecosistemas terrestres. Este punto es muy importante ya que es necesario tener un ecosistema terrestre sano para los cultivos, cosechas, pero gracias a las prácticas agrícolas sostenibles ayudan a preservar y restaurar los ecosistemas terrestres.

Esos son los ODS que más relación se ha podido tener con este Trabajo Final de grado, pero los demás quitando un par también tienen mínimamente relación ya que este sector es muy importante para todo ser humano, ya que la mayoría de los productos vienen de la naturaleza, la cual se intenta proteger gracias a la actuación de la agenda 2030.

Anexo II. Variables palabras clave

Todas estas palabras no tienen definición en sí ya que son el número de veces que aparece la palabra exacta (Keywords) o una palabra que se asemeja a la que se encuentra en la página web (kwstems). A continuación, mostramos unas tablas con dichas palabras:

Variables		
keywords_aditivos	keywords_alimentacion_consciente	keywords_bebidas_funcionales
keywords_aeroponico	keywords_alimentos	keywords_biodegradable
keywords_agave	keywords_alto_en_proteinas	keywords_calidad
keywords_agricultura_regenerativa	keywords_antioxidantes	keywords_calorias
keywords_agricultura_vertical	keywords_ayuno_intermitente	keywords_carne
keywords_agua_alcalina	keywords_azucar_de_coco	keywords_comercio_justo
keywords_aguacate	keywords_bajo_en_calorias	keywords_competitividad
keywords_agua_de_mar	keywords_bajo_en_grasas	keywords_compostable
keywords_algas_marinas	keywords_bajo_indice_glucemico	keywords_comunicacion
keywords_harina_de_almendra	keywords_harina_de_almendra	keywords_kilometro_cero
keywords_harina_de_coco	keywords_harina_de_coco	keywords_leche_de_almendras
keywords_hidroponico	keywords_hidroponico	keywords_leche_de_coco
keywords_huella_de_carbono	keywords_huella_de_carbono	keywords_leche_de_soja
keywords_importacion	keywords_importacion	keywords_leche_vegetal
keywords_integral	keywords_integral	keywords_libre_de_transgenicos
keywords_jugo_verde	keywords_jugo_verde	keywords_marca
keywords_kefir	keywords_kefir	keywords_matcha
keywords_keto	keywords_keto	keywords_miel_cruda
keywords_seitan	keywords_sin_transgenicos	keywords_vegano
keywords_semillas	keywords_smoothies	keywords_vegetariano
keywords_sin_aditivos	keywords_sostenible	keywords_xilitol
keywords_sin_azucar	keywords_stevia	keywords_zero_waste
keywords_sin_cholesterol	keywords_superalimento	keywords_conservacion
keywords_sin_conservantes	keywords_superfood	keywords_conservantes
keywords_sin_gluten	keywords_suplementos_vitaminicos	keywords_crudivegano
keywords_sin_lacteos	keywords_tempeh	keywords_crudo
keywords_sin_plasticos	keywords_tofu	keywords_cultivo_de_tejidos
keywords_detox	keywords_enriquecido_con_vitaminas	keywords_productos
keywords_ecologico	keywords_envases	keywords_productos_de_temporada
keywords_edulcorantes_naturales	keywords_eritritol	keywords_proteina
keywords_enriquecido_con_minerales	keywords_espirulina	keywords_quinoa
keywords_natural	keywords_exportacion	keywords_recetas
keywords_omega_3	keywords_fermentados	keywords_recursos
keywords_organico	keywords_fibra	keywords_saludable
keywords_paleo	keywords_grano_entero	keywords_precio
keywords_permacultura	keywords_grasas	keywords_precocinado
keywords_plant_based	keywords_probioticos	
keywords_plastico	keywords_produccion	

Variables		
kwstems_es_adit	kwstems_es_eritritol	kwstems_es_quino
kwstems_es_aeropon	kwstems_es_espirulin	kwstems_es_recet
kwstems_es_agav	kwstems_es_exportacion	kwstems_es_rekurs
kwstems_es_agricultur_regener	kwstems_es_ferment	kwstems_es_salud
kwstems_es_agricultur_vertical	kwstems_es_fibr	kwstems_es_seit
kwstems_es_agu_alcalin	kwstems_es_gran_enter	kwstems_es_semill
kwstems_es_aguacat	kwstems_es_gras	kwstems_es_sin_adit
kwstems_es_agu_de_mar	kwstems_es_harin_de_almendr	kwstems_es_sin_azuc
kwstems_es_algas_marin	kwstems_es_harin_de_coc	kwstems_es_sin_colesterol
kwstems_es_alimentacion_conscient	kwstems_es_hidropon	kwstems_es_sin_conserv
kwstems_es_aliment	kwstems_es_huell_de_carbon	kwstems_es_sin_glut
kwstems_es_alto_en_protein	kwstems_es_importacion	kwstems_es_sin_lacte
kwstems_es_antioxid	kwstems_es_integral	kwstems_es_sin_plastic
kwstems_es_ayun_intermitent	kwstems_es_jug_verd	kwstems_es_sin_transgen
kwstems_es_azuc_de_coc	kwstems_es_kef	kwstems_es_smoothi
kwstems_es_baj_en_calori	kwstems_es_ket	kwstems_es_sosten
kwstems_es_baj_en_gras	kwstems_es_kilometr_cer	kwstems_es_stevi
kwstems_es_baj_indic_glucem	kwstems_es_lech_de_almendr	kwstems_es_superaliment
kwstems_es_beb_funcional	kwstems_es_lech_de_coc	kwstems_es_superfood
kwstems_es_biodegrad	kwstems_es_lech_de_soj	kwstems_es_suplement_vitamin
kwstems_es_calid	kwstems_es_lech_vegetal	kwstems_es_tempeh
kwstems_es_calori	kwstems_es_libr_de_transgen	kwstems_es_tofu
kwstems_es_carn	kwstems_es_marc	kwstems_es_vegan
kwstems_es_comerci_just	kwstems_es_match	kwstems_es_vegetarian
kwstems_es_competit	kwstems_es_miel_crud	kwstems_es_xilitol
kwstems_es_compost	kwstems_es_natural	kwstems_es_zer_wast
kwstems_es_comunicacion	kwstems_es_omeg_3	kwstems_es_detox
kwstems_es_conservacion	kwstems_es_organ	kwstems_es_ecolog
kwstems_es_conserv	kwstems_es_pale	kwstems_es_edulcor_natural
kwstems_es_crudivegan	kwstems_es_permacultur	kwstems_es_enriquec_con_mineral
kwstems_es_crud	kwstems_es_plant_bas	kwstems_es_enriquec_con_vitamin
kwstems_es_cultiv_de_tej	kwstems_es_plastic	kwstems_es_envas
kwstems_es_product_de_tempor	kwstems_es_preci	kwstems_es_produccion
kwstems_es_protein	kwstems_es_precocin	kwstems_es_product
kwstems_es_probiot		