

UNIVERSIDAD DE CANTABRIA
FACULTAD DE CIENCIAS ECONÓMICAS Y
EMPRESARIALES



MÁSTER OFICIAL EN EMPRESA Y TECNOLOGÍAS DE
LA INFORMACIÓN

CURSO ACADÉMICO 2024-2025

TRABAJO FIN DE MÁSTER

Transformación de la Agricultura con Ciencia de Datos

Autor

D. Santiago Bravo Moreno

Director

Dr. D. Pedro Solana González

Santander, 16 de septiembre de 2025

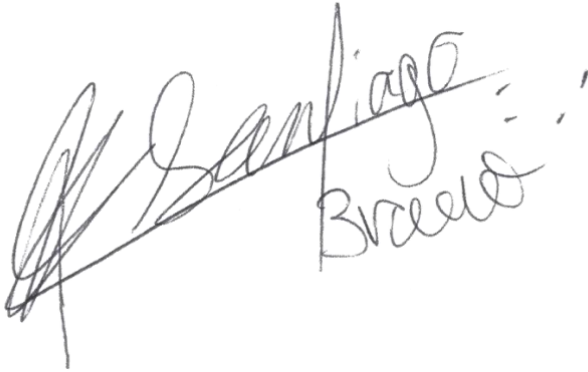
DECLARACIÓN RESPONSABLE

El autor Santiago Bravo Moreno es el único responsable del contenido del Trabajo Fin de Máster que se presenta. La Universidad de Cantabria, así como los profesores directores del mismo, no son responsables del contenido último de este Trabajo.

En tal sentido, el/la autor/a se hace responsable:

- 1. De la AUTORÍA Y ORIGINALIDAD del trabajo que se presenta.*
- 2. De que los DATOS y PUBLICACIONES en los que se basa la información contenida en el trabajo, o que han tenido una influencia relevante en el mismo, han sido citados en el texto y en la lista de referencias bibliográficas.*

El/La autor/a declara que el Trabajo Fin de Máster tiene una extensión de entre 10.000 y 15.000 palabras, excluidas tablas, gráficos y bibliografía.

A handwritten signature in black ink, reading 'Santiago Bravo'. The signature is written in a cursive, flowing style. The word 'Santiago' is written above 'Bravo', and there are some additional scribbles and dots to the right of the name.

Fdo. Santiago Bravo Moreno

Transformación de la Agricultura con Ciencia de Datos

Resumen:

Este trabajo analiza cómo la ciencia de datos puede cambiar la forma en que entendemos y trabajamos la agricultura, centrándose en cómo ayuda a mejorar la producción y hacerla más eficiente. Primero, se repasa la evolución del sector, desde métodos tradicionales hasta la agricultura 4.0, mostrando que cada avance ha resuelto problemas, pero también ha enfrentado retos, como el uso eficiente de recursos, la sostenibilidad y el impacto del clima. De ahí surge la idea central: aprovechar los datos para gestionar mejor todo el proceso agrícola.

Se explican las oportunidades que ofrece la ciencia de datos y cómo puede aplicarse en la agricultura. Se analiza el papel del *Big data*, IoT, minería de datos y otras tecnologías que permiten recopilar y procesar información para tomar decisiones más rápidas y acertadas. También se aborda el proceso ETL, base para convertir datos en información útil. Este flujo, común en distintas industrias, se adapta a la agricultura para que los agricultores puedan aprovecharlo.

Asimismo, se desarrolla un caso práctico en la región de La Rioja, utilizando datos históricos del clima (temperatura, lluvias, humedad) para estudiar su relación y efecto con la producción de uva. Se aplican modelos de ciencia de datos como árboles de decisión y clustering con la ayuda de la herramienta Weka para identificar cómo influyen estas variables en el rendimiento. El objetivo es demostrar que los datos no solo sirven para analizar lo que pasó, sino también para anticipar resultados y mejorar la gestión de la cosecha.

Los resultados muestran que, con estas herramientas, es posible prever escenarios y hacer ajustes que aumenten la eficiencia y la calidad del cultivo. Sin embargo, también se señalan retos como la falta de acceso a tecnología para pequeños productores, la dependencia digital y la protección de datos. En conclusión, la integración de la ciencia de datos en la agricultura no es solo una ventaja competitiva, sino una necesidad para avanzar hacia una producción más sostenible e inteligente.

Palabras clave: agricultura inteligente; eficiencia agrícola; sostenibilidad alimentaria; agricultura 4.0; ciencia de datos; análisis predictivo; aprendizaje automático

Transforming Agriculture through Data Science

Abstract:

This paper analyzes how data science can change the way we understand and work in agriculture, focusing on how it helps improve production and make it more efficient. First, it reviews the evolution of the sector, from traditional methods to Agriculture 4.0, demonstrating that each advancement has solved problems but also left challenges, such as the efficient use of resources, sustainability, and the impact of climate change. From this arises the central idea: leveraging data to better manage the entire agricultural process.

The main section explains the opportunities offered by data science and how it can be applied in agriculture. It examines the role of Big Data, IoT, data mining, and other technologies that enable the collection and processing of information to make faster and more accurate decisions. It also addresses the ETL process, which forms the foundation for converting data into useful information. This workflow, common in various industries, is adapted to agriculture so that farmers can take advantage of it.

Then, a practical case is developed in the region of La Rioja, using historical climate data (temperature, rainfall, humidity) to study its relationship and effect with grape production. Models such as decision trees and clustering are applied with the help of Weka to identify how these variables influence yield. The objective is to demonstrate that data is not only useful for analyzing what happened but also for anticipating results and improving crop management.

The results show that, with these tools, it is possible to predict scenarios and make adjustments that increase efficiency and crop quality. However, challenges such as a lack of access to technology for small producers, digital dependency, and data protection are also highlighted. In conclusion, integrating data science into agriculture is not just a competitive advantage but a necessity to move toward more sustainable and intelligent production.

Keywords: smart agriculture; agricultural efficiency; food sustainability; agriculture 4.0; data science; predictive analytics; machine learning

ÍNDICE

1. MARCO GENERAL	7
1.1 INTRODUCCIÓN.....	7
1.2 JUSTIFICACIÓN DEL TEMA	9
1.3 OBJETIVOS DEL TRABAJO.....	9
1.3.1 Objetivo general	9
1.3.2 Objetivos específicos	9
2. MARCO TEÓRICO	10
2.1 CIENCIA DE DATOS	10
2.2 AGRICULTURA 4.0.....	11
2.3 EFICIENCIA, SOSTENIBILIDAD Y SEGURIDAD ALIMENTARIA	13
3. EVOLUCIÓN DE LA AGRICULTURA CON LA TECNOLOGÍA.....	14
3.1 AGRICULTURA TRADICIONAL Y SUS LIMITACIONES.....	14
3.2 REVOLUCIÓN VERDE Y MECANIZACIÓN	15
3.3 AGRICULTURA DE PRECISIÓN.....	18
3.4 AGRICULTURA 4.0.....	19
3.4.1 Internet de las cosas.....	19
3.4.2 Big data	20
3.4.3 Inteligencia artificial	21
3.4.4 Cambio del rol del agricultor	21
3.4.5 Retos actuales y barreras tecnológicas	22
4. LA CIENCIA DE DATOS Y SU APLICABILIDAD EN LA AGRICULTURA..	23
4.1 ESTUDIOS PREVIOS APLICANDO CIENCIA DE DATOS EN LA AGRICULTURA	23
4.2 COMPARACIÓN DE ENFOQUES Y HERRAMIENTAS	24
4.3 OPORTUNIDADES Y BENEFICIOS	24
4.4 TECNOLOGÍAS CLAVE QUE PERMITEN APLICAR CIENCIA DE DATOS EN LA AGRICULTURA.....	25
4.4.1 Base de datos	26
4.4.2 Lenguajes de programación	27
4.4.3 Machine learning	29
4.4.4 Deep learning.....	30
4.4.5 Herramientas de visualización.....	30
5. CASO PRÁCTICO: SEGUIMIENTO DE LA PRODUCCIÓN DE UVA PARA LA FABRICACIÓN DE VINO EN LA RIOJA.....	32
5.1 PLANTEAMIENTO	32
5.2 METODOLOGÍA	34

5.3	HERRAMIENTAS Y RECURSOS UTILIZADAS	34
5.4	DESARROLLO	35
5.4.1	Data set.....	35
5.4.2	Preparación de los datos.....	36
5.4.3	Selección de algoritmos	37
5.4.4	Configuración en Weka	37
5.5	RESULTADOS Y DISCUSIÓN	40
5.5.1	Evaluación de atributos	40
5.5.2	J48	41
5.5.3	Random forest.....	42
5.5.4	Simple K-Means.....	45
6.	CONCLUSIÓN	47
7.	REFERENCIAS.....	49

ÍNDICE DE FIGURAS

<i>Figura 1. Trabajadores cosechando fresas en EE. UU.</i>	14
<i>Figura 2. Cosechadoras cultivando maíz de manera industrial.</i>	16
<i>Figura 3. Concentraciones medias anuales de nitratos en las aguas subterráneas para el periodo 2015-2019.</i>	17
<i>Figura 4. Infografía de la Comisión Europea con un resumen de los objetivos de la Directiva de Nitratos.</i>	17
<i>Figura 5. Integración de equipos con Internet de las cosas.</i>	20
<i>Figura 6. Analítica de datos masivos en tiempo real con un sistema de gestión agrícola.</i>	20
<i>Figura 7. Cambio de rol del agricultor tomado decisiones críticas con su software de gestión agrícola.</i>	22
<i>Figura 8. Proceso ETL y el viaje del dato adentro de una empresa.</i>	25
<i>Figura 9. Tabla de relaciones de una base de datos.</i>	26
<i>Figura 10. Tabla de relaciones de una base de datos.</i>	27
<i>Figura 11. Machine learning aplicado en la producción de uva.</i>	29
<i>Figura 12. Ejemplo de una granja inteligente con una tablet para visualizar los datos.</i>	31
<i>Figura 13. Ejemplo de un dashboard interactivo creado en PowerBi.</i>	31
<i>Figura 14. Cifra de Negocio por Categoría de la Industria Alimentaria en Millones de euros.</i>	32
<i>Figura 15. Bodega Marqués de Riscal en La Rioja Alavesa.</i>	33
<i>Figura 16. Data set de rendimientos de la producción del vino y condiciones climáticas de la rioja entre los años 2005-2023.</i>	36
<i>Figura 17. Aplicación del filtro de normalización de los datos.</i>	38
<i>Figura 18. Pestaña de ajustes del algoritmo J48 en Weka.</i>	39
<i>Figura 19. Ranking de atributos siendo de más influyentes a menos al modelo.</i>	40
<i>Figura 20. Árbol de decisión con las top 6 atributos.</i>	41
<i>Figura 21. Árbol de decisión con la totalidad de las variables.</i>	42
<i>Figura 22. Random Forest con totalidad de las variables.</i>	43
<i>Figura 23. Random Forest con las 6 variables principales.</i>	44
<i>Figura 24. Prueba 1 Simple K-Means con totalidad de las variables.</i>	45
<i>Figura 25. Prueba 2 Simple K-Means con las 6 variables principales.</i>	46

ÍNDICE DE TABLAS

<i>Tabla 1: Estudios previos</i>	23
----------------------------------	----

1. MARCO GENERAL

1.1 INTRODUCCIÓN

La agricultura, a lo largo de la historia, ha sido una pieza clave en el desarrollo de la sociedad. “La agricultura es comparable al bipedalismo y al dominio de fuego; transformó completamente nuestra interacción con el entorno y entre nosotros mismos” (Carey, 2023). Este avance permitió al ser humano dejar de ser nómada y asentarse en territorios permanentes gracias a la domesticación de animales y al cultivo de diversas especies vegetales. A partir de ello, se abrió la puerta a la expansión de civilizaciones, facilitando el comercio internacional con otros reinos y territorios lejanos. Incluso puede considerarse como el inicio del concepto de economía, con la invención del trueque de bienes (por ejemplo, 1 kg de patatas por 3 piezas de ropa) o con la aparición de formas primitivas de moneda como la sal, el cacao o los cereales, utilizados como medios de transacción.

Sin embargo, ese fue solo el comienzo de la humanidad y de su evolución. “El cambio de la recolección a la agricultura, que comenzó hace unos 12.000 años, lo cambió todo” (Larson, 2023). A partir del asentamiento en territorios fijos, la vida comenzó a volverse más estable, lo que favoreció el sedentarismo y contribuyó significativamente a la formación de la especie humana tal como se conoce en la actualidad. Este proceso sentó las bases para que, en los siglos posteriores, se produjera una sucesión de transformaciones: conflictos bélicos, nuevos descubrimientos, invenciones, la formación de imperios y naciones, entre otros acontecimientos fundamentales en la historia.

A medida que la práctica agrícola se fue perfeccionando y permitió alimentar a un mayor número de personas, también impulsó el desarrollo del conocimiento, las ideas, las innovaciones y los movimientos en distintas áreas como la ciencia, la física, la geografía, la literatura y el arte. La agricultura fue uno de los principales motores que propiciaron el mayor movimiento de la historia moderna: la revolución industrial. “La revolución industrial fue en realidad primero y sobre todo una revolución agrícola” (Bairoch, 1979).

Se produjo un efecto dominó. Al mejorar los métodos de cultivo mediante la introducción de nuevas técnicas y maquinaria, aumentaron tanto la productividad como la disponibilidad de alimentos. Como consecuencia, se favoreció el crecimiento de la población, lo que a su vez impulsó el desarrollo del comercio y de la economía. Posteriormente, se inició un nuevo ciclo, la mecanización del campo, que liberó gran parte de la mano de obra, lo que llevó a muchos trabajadores a buscar empleo en las fábricas y a desplazarse hacia las ciudades, promoviendo así el auge de una nueva industria.

Desde la revolución industrial, la sociedad ha experimentado transformaciones profundas: dos guerras mundiales, recesiones y depresiones económicas, innovaciones tecnológicas, regímenes autoritarios, cambios políticos y la globalización, entre otros. “También ha sido alimento para el alma y motor de nuevos cambios personales, sociales y ecológicos” (Santandreu y Rea, 2014). La agricultura ha demostrado ser un motor

fundamental de cambio social, económico y político, con presencia constante en los principales acontecimientos históricos de la humanidad. Su relevancia ha alcanzado tal magnitud que, en la actualidad, constituye una variable clave que influye directamente en la vida cotidiana.

Un ejemplo reciente lo demuestra con claridad; desde el inicio del conflicto entre Rusia y Ucrania, Europa y otros territorios se han visto afectados por la reducción en la producción y exportación de cereales. Según el Consejo Europeo (2025), Ucrania es uno de los principales exportadores mundiales de productos como trigo, maíz, cebada y aceite de girasol. La disminución de su actividad agrícola impactó directamente en el comercio internacional y en los mercados globales, contribuyendo al incremento de la inflación y a la escasez de estos bienes a nivel mundial.

Aunque este caso pueda parecer excepcional, situaciones como una mala cosecha o una sequía prolongada como las que han afectado recientemente a España generan consecuencias similares. En ambos escenarios, las variables económicas, sociales y logísticas se ven alteradas, provocando efectos negativos en cadena.

En este contexto, se hace evidente que la agricultura constituye un pilar fundamental para el funcionamiento de la sociedad. La disponibilidad de alimentos está directamente ligada al bienestar colectivo, lo que subraya la necesidad de proteger e invertir en esta industria estratégica. Actualmente, la agricultura atraviesa una nueva fase de transformación, caracterizada por una orientación más tecnológica y moderna, con el fin de hacer frente a los retos contemporáneos: (1) el cambio climático, (2) la creciente frecuencia de desastres naturales, (3) el aumento de la demanda alimentaria, (4) la presión por prácticas sostenibles tanto por parte de gobiernos como de consumidores, y (5) la necesidad de garantizar la seguridad alimentaria.

Para responder a estos desafíos, el sector agrícola (al igual que muchas otras industrias) ha adoptado nuevas técnicas, métodos, sistemas y tecnologías. Hoy en día, herramientas como sensores de humedad, drones, sistemas de riego inteligente, así como métodos de cultivo innovadores como la hidroponía o la agricultura vertical, ya forman parte del funcionamiento cotidiano de numerosas explotaciones. Todo ello conforma una nueva etapa de transformación en la que la tecnología desempeña un papel cada vez más relevante. De acuerdo con Cravero y Sepúlveda (2021), para la agricultura, hemos encontrado varios estudios que explican las técnicas de aprendizaje automático para abordar diversos problemas, como el control de enfermedades, la predicción del clima, y la mejora de la producción y la calidad.

Dentro de esta transformación, una herramienta en particular ha comenzado a destacar por su potencial y por la accesibilidad que ha alcanzado en los últimos años: la ciencia de datos. Si bien anteriormente parecía un recurso exclusivo de grandes corporaciones o centros de investigación, en la actualidad es más asequible, fácil de aprender e implementar, y útil para una amplia gama de productores. Desde pequeñas explotaciones familiares hasta grandes unidades agrícolas industriales, la ciencia de datos facilita la toma de decisiones fundamentadas, la predicción de resultados, la optimización de recursos y la mejora del rendimiento. Además, su aplicación no se limita únicamente a la agricultura, sino que se extiende a otras ramas del sector primario,

como la ganadería, la silvicultura o la acuicultura, generando beneficios concretos y significativos en cada una de ellas.

A partir de esta introducción, el trabajo se estructura en varias secciones que permiten entender con mayor profundidad la relación entre la ciencia de datos y la transformación del sector agrícola. En primer lugar, se presentan la justificación del tema y los objetivos que guían la investigación. A continuación, el marco teórico aborda los conceptos fundamentales de ciencia de datos y Agricultura 4.0, así como su impacto en la eficiencia, sostenibilidad y seguridad alimentaria. La tercera sección analiza la evolución histórica de la agricultura desde sus formas más tradicionales hasta el surgimiento de la agricultura inteligente, destacando los avances tecnológicos clave. Posteriormente, se profundiza en el papel de la ciencia de datos dentro de esta evolución, explorando estudios previos, resultados alcanzados, tecnologías asociadas y retos éticos. Para ilustrar su aplicabilidad, se desarrolla un caso práctico basado en el análisis de la producción de uva en la región de La Rioja, evaluando cómo los factores climatológicos influyen en la producción de uva para la elaboración de vino. Finalmente, se presentan los resultados, una discusión crítica de los mismos, y una reflexión final con las conclusiones, las limitaciones del estudio y posibles líneas de investigación futuras.

1.2 JUSTIFICACIÓN DEL TEMA

El uso de tecnologías y metodologías avanzadas en la agricultura se presenta como un campo de investigación relevante para enfrentar retos como la sostenibilidad, la variabilidad climática y el aumento de la demanda alimentaria. La transición hacia modelos más tecnificados permite mejorar la eficiencia y garantizar la seguridad alimentaria en un entorno global complejo.

La integración de análisis de datos, logística y dirección de operaciones en este sector favorece la optimización de procesos, la gestión eficiente de recursos y la reducción de costes; en este contexto, la ciencia de datos se posiciona como una herramienta estratégica para comprender relaciones causa-efecto, extraer conocimiento y generar modelos predictivos que contribuyan a una agricultura más sostenible e inteligente.

1.3 OBJETIVOS DEL TRABAJO

1.3.1 Objetivo general

Explorar cómo la ciencia de datos puede aplicarse en el sector vitivinícola para analizar el impacto de factores climatológicos en la producción de uva, identificando patrones y relaciones que permitan mejorar la planificación y el rendimiento mediante técnicas analíticas y modelos predictivos.

1.3.2 Objetivos específicos

- Objetivo 1: Investigar las aplicaciones actuales de la ciencia de datos en el sector agrícola, especialmente en sistemas de cosecha.

- Objetivo 2: Identificar los retos más comunes en los procesos de cosecha relacionados con la eficiencia, eficacia y seguridad alimentaria.
- Objetivo 3: Evaluar herramientas, métodos y tecnologías basadas en ciencia de datos que puedan aplicarse para mejorar dichos procesos.

2. MARCO TEÓRICO

A continuación, se presenta una breve introducción sobre el marco teórico y los temas tratados en este trabajo, con el objetivo de facilitar la comprensión del contenido.

2.1 CIENCIA DE DATOS

Una pregunta clave que puede surgir es: ¿Qué es la ciencia de datos? Existen varias definiciones, pero según (Microsoft, 2023), “La ciencia de datos es el estudio científico de los datos para obtener conocimientos. Este campo combina varias disciplinas para extraer conclusiones de conjuntos de datos masivos con el fin de tomar decisiones y realizar predicciones de manera informada”. La ciencia de datos es una herramienta utilizada por profesionales de perfiles y sectores muy diversos, pero con un mismo objetivo: tomar decisiones más informadas. Para ello, combina conocimientos de matemáticas, estadística, programación, analítica y el uso de herramientas informáticas.

El estudio y análisis de datos sigue un proceso conocido como ETL (extraer, transformar y cargar). Este procedimiento permite trabajar de forma organizada. Consiste en extraer datos desde diversas fuentes, transformarlos y modificarlos según las necesidades del analista, y luego cargarlos en una base de datos para su posterior consulta. El analista adapta los datos al área de investigación deseada, facilitando así la obtención de conclusiones útiles y relevantes.

Existen diversas herramientas informáticas y lenguajes de programación que facilitan enormemente la tarea del análisis de datos.

Los lenguajes de programación más utilizados en ciencia de datos son:

- **Python:** muy utilizado para análisis de datos, aprendizaje automático (*machine learning*) y desarrollo de software.
- **R:** especialmente útil en análisis estadístico y visualización de datos.
- **SQL:** lenguaje de consulta utilizado para trabajar con bases de datos estructuradas.

Entre los programas y plataformas más comunes en ciencia de datos destacan:

- **Google Colab:** una herramienta en la nube que permite crear y compartir cuadernos de trabajo, ejecutar código Python y realizar análisis de datos sin necesidad de instalar software localmente.

- **Jupyter Notebook:** aplicación de código abierto que permite combinar texto, código, visualizaciones y ecuaciones en un mismo documento interactivo.
- **Kaggle:** plataforma web que ofrece conjuntos de datos, competencias y *notebooks* para trabajar colaborativamente en ciencia de datos.
- **Weka:** software de código abierto que permite realizar minería de datos y aplicar técnicas de aprendizaje automático.
- **Power BI:** herramienta de Microsoft para la visualización interactiva de datos y generación de informes dinámicos.

Las principales técnicas utilizadas en ciencia de datos incluyen:

- **Machine learning (aprendizaje automático):** desarrollo de algoritmos que permiten a las máquinas aprender de los datos y realizar predicciones.
- **Modelos supervisados:** algoritmos que aprenden a partir de datos estructurados y etiquetados, es decir, con información conocida sobre las categorías o resultados. Algunos ejemplos son los algoritmos de clasificación como los árboles de decisión o los k-vecinos más cercanos, y los de regresión, como la regresión lineal o logística.
- **Modelos no supervisados:** algoritmos que trabajan con datos no etiquetados y buscan identificar patrones o estructuras ocultas. Ejemplos comunes son los algoritmos de agrupamiento (*clustering*) o de asociación.

El campo de la ciencia de datos es muy amplio, y en cada etapa del proceso puede intervenir un especialista diferente. Sin embargo, actualmente, el acceso al aprendizaje y la aplicación de la ciencia de datos se ha ampliado, permitiendo su implementación en diversos ámbitos, tanto académicos, científicos e industriales. Su utilidad radica en su capacidad para mejorar procesos, aumentar la eficiencia, reducir costes, apoyar investigaciones científicas y, sobre todo, permitir una toma de decisiones más informada y precisa.

Con un amplio abanico de posibilidades, la ciencia de datos encuentra un terreno lleno de oportunidades en sectores que, aunque tradicionalmente alejados de la tecnología, están en plena transformación, como es el caso de la agricultura.

2.2 AGRICULTURA 4.0

La transformación tecnológica que ha impactado a numerosas industrias también tiene el potencial de modificar profundamente el sector primario, especialmente en el ámbito agrícola. En este contexto, surge el concepto de agricultura 4.0, también denominada agricultura inteligente, la cual “Aplica innovaciones tecnológicas inteligentes en la agricultura mediante tecnologías avanzadas de información y comunicación” (Tzounis, 2017).

La agricultura inteligente se basa en la implementación de tecnologías como sensores, software, ordenadores, servidores, sistemas en la nube, Big Data e IoT. Estas herramientas permiten la recolección de datos que, posteriormente, son analizados con el objetivo de optimizar distintos aspectos del sistema agrícola, tales como la predicción de cosechas, la organización de la producción, la mejora en la calidad del producto y el control de plagas, entre otros procesos clave.

Un componente esencial que posibilita el desarrollo de la agricultura inteligente es el IoT, tecnología que permite la interconexión entre dispositivos electrónicos, facilitando el intercambio de información en tiempo real. Mediante la conexión a bases de datos y sistemas de gestión agrícola, es posible obtener información precisa sobre el estado de las explotaciones, lo cual contribuye a un control más eficiente de los cultivos, una mejor gestión de los recursos y una planificación más precisa, aumentando así la productividad. Todo ello se logra gracias a la posibilidad de tomar decisiones más informadas basadas en los datos recogidos por este sistema.

Un ejemplo representativo de la aplicación del IoT en el sector agrícola es el caso de SAT Marigómez, una explotación ubicada en Huelva dedicada al cultivo de mandarinas. “Esta organización utiliza la tecnología de Vodafone e IoT para recopilar y analizar datos relacionados con el suelo, el clima, las plagas y el consumo de recursos, lo que ha permitido mejorar significativamente su eficiencia y productividad” (Empresas, 2022). El sistema empleado incluye sensores que registran variables como temperatura, humedad, horas de sol, y consumo de agua. El objetivo principal es optimizar el uso del riego, especialmente en un contexto climático caracterizado por altas temperaturas y escasez de agua. Gracias a estos sensores, se puede determinar de forma precisa qué zonas requieren riego, en qué cantidad y con qué frecuencia.

La agricultura 4.0 no solo representa una evolución tecnológica, sino que también incorpora la ciencia de datos como elemento central en la toma de decisiones. La integración de sensores, dispositivos conectados y sistemas informáticos genera grandes volúmenes de información que, al ser adecuadamente analizados, permiten transformar los datos en decisiones estratégicas. De este modo, el análisis de datos se convierte en el motor de mejora continua dentro del sector agrícola.

La adopción de un modelo de gestión basado en datos permite identificar patrones, anticiparse a problemas y optimizar el uso de recursos como el agua, la energía o los fertilizantes. Casos como el de SAT Marigómez demuestran que, mediante el estudio constante de variables como el clima, la humedad o la presencia de plagas, es posible aumentar la eficiencia y reducir el desperdicio en las explotaciones agrícolas. En este marco, la ciencia de datos actúa como un vínculo fundamental entre la tecnología y el conocimiento agronómico aplicado, facilitando la toma de decisiones informadas.

Este nuevo modelo agrícola no solo mejora la productividad, sino que también favorece la adopción de prácticas más sostenibles, responsables y seguras, tanto desde una perspectiva ambiental como social.

2.3 EFICIENCIA, SOSTENIBILIDAD Y SEGURIDAD ALIMENTARIA

La evolución reciente del sector agrícola, impulsada por la adopción de tecnologías emergentes y el uso estratégico de la ciencia de datos, ha facilitado la transición hacia explotaciones caracterizadas por una mayor eficiencia, sostenibilidad y seguridad. Estas herramientas tecnológicas permiten optimizar los procesos productivos, mitigar el impacto ambiental y fortalecer la seguridad alimentaria. En este sentido, la integración tecnológica se consolida como un elemento clave para alcanzar los pilares fundamentales de una agricultura moderna y responsable.

Las definiciones operativas de estos pilares son las siguientes:

- **Eficiencia agrícola:** Hace referencia al aprovechamiento óptimo de los recursos naturales, junto con la aplicación de métodos y sistemas que permiten maximizar la producción, minimizando los tiempos de trabajo, las pérdidas y los desperdicios generados en el proceso productivo.
- **Sostenibilidad:** Implica el uso racional y consciente de los recursos disponibles, fomentando prácticas agrícolas que reduzcan el impacto ambiental. Entre estas prácticas destacan la reutilización de residuos orgánicos, el uso de abonos compostados y la implementación de sistemas como la agricultura hidropónica.
- **Seguridad alimentaria:** Se centra en garantizar la inocuidad y trazabilidad de los alimentos, mediante el control preventivo de plagas, enfermedades y contaminantes desde la fase de cultivo hasta la llegada del producto al consumidor final. Este objetivo se logra a través de planes de control de calidad en origen, certificaciones oficiales y sistemas de monitoreo y seguimiento.

En un entorno global cada vez más digitalizado, donde los datos constituyen un recurso estratégico, las organizaciones líderes en distintos sectores son aquellas que poseen mayor acceso a información relevante, y que, además, son capaces de interpretarla y actuar en consecuencia. Este fenómeno también se manifiesta en la agricultura, donde la toma de decisiones requiere el respaldo de datos precisos y actualizados, así como del conocimiento técnico necesario para evaluar alternativas y anticipar riesgos.

La agricultura moderna enfrenta desafíos significativos derivados del crecimiento poblacional, el cambio climático y la escasez de recursos naturales. “Según un informe de las Naciones Unidas, se proyecta que la población mundial alcanzará los 8.500 millones de personas en 2030 y superará los 9.700 millones en 2050. Esta evolución demográfica intensificará la demanda mundial de alimentos, al mismo tiempo que se verá limitada por factores como la reducción de tierras cultivables, la urbanización acelerada, la escasez hídrica y las alteraciones climáticas” (Siregar, Wahjuni, Seminar, y Santosa, 2022). En este contexto, la investigación científica y tecnológica en el ámbito agrícola adquiere una importancia estratégica.

La ciencia de datos se presenta, por tanto, como una herramienta fundamental para afrontar estos retos. Su aplicación permite realizar análisis en tiempo real, mejorar la eficiencia operativa y garantizar decisiones basadas en evidencia. Desde la predicción de rendimientos hasta la gestión del riego y el control de plagas, los datos posibilitan

una agricultura más precisa y adaptativa. El análisis de grandes volúmenes de información, facilitado por sensores, sistemas de monitoreo e inteligencia artificial, mejora no solo la productividad, sino también la sostenibilidad y la seguridad del sistema agroalimentario.

A medida que avanza la presente investigación, se profundizará en el papel que desempeña la ciencia de datos en la transformación del sector agrícola, y en su capacidad para responder de manera sostenible y eficaz a las demandas alimentarias del futuro.

3. EVOLUCIÓN DE LA AGRICULTURA CON LA TECNOLOGÍA

3.1 AGRICULTURA TRADICIONAL Y SUS LIMITACIONES

La transición de un modelo de subsistencia basado en la caza y la recolección hacia un sistema agrícola marcó un hito en la evolución de las sociedades humanas. “La agricultura tradicional es el resultado de experiencias proporcionadas por prácticas agrícolas locales a lo largo de miles de años” (Singh y Singh, 2017). Esta primera fase histórica constituye la base del desarrollo agrícola posterior y continúa siendo una práctica vigente en diversas regiones del mundo.

Dicha modalidad se caracteriza por la utilización de métodos manuales y rutinarios, una alta dependencia de las condiciones climáticas y un escaso nivel de mecanización. Predomina el uso de herramientas básicas y una elevada demanda de mano de obra, con un conocimiento técnico transmitido de forma empírica. Las explotaciones agrícolas suelen presentar una estructura de pequeña escala, con recursos limitados, gestión familiar y una orientación hacia el autoconsumo o los mercados locales.

Figura 1. Trabajadores cosechando fresas en EE. UU.



Fuente: Grupo (2022)

Este modelo, aunque históricamente efectivo para satisfacer las necesidades locales, presenta diversas limitaciones frente a los desafíos actuales. Los métodos tradicionales han quedado obsoletos ante retos cada vez más complejos frente a las exigencias actuales del sector agrícola: el aumento de la demanda alimentaria por parte de una población creciente, la baja capacidad de adaptación a fenómenos climatológicos, una productividad reducida e ineficiente por el limitado acceso a insumos modernos, el incremento de plagas y enfermedades, el uso de técnicas de riego poco eficientes y semillas tradicionales menos resistentes.

Ante las crecientes limitaciones de la agricultura tradicional, surgió la necesidad de transformar los métodos de producción agrícola para hacer frente a una población en aumento y a una demanda alimentaria cada vez mayor. Fue en este contexto cuando comenzaron a introducirse avances científicos, tecnológicos y mecánicos que marcaron un punto de inflexión en la evolución del sector agrícola. Así dio comienzo la *revolución verde y mecanización* una etapa clave en la modernización de la agricultura.

3.2 REVOLUCIÓN VERDE Y MECANIZACIÓN

A raíz de los desafíos que presentaba la agricultura tradicional, surgió una nueva etapa conocida como la (revolución verde) también llamada revolución agrícola, que marcó un punto de inflexión en la historia del sector. Esta fase se caracterizó por un notable aumento en la producción de alimentos, gracias a la incorporación de maquinaria agrícola, fertilizantes, pesticidas, nuevas variedades de cultivos y prácticas agrícolas más tecnificadas, aplicadas a gran escala.

La maquinaria introducida al campo durante esta etapa incluyó tractores, sembradoras y cosechadoras, que optimizaron tareas rutinarias y esenciales para el cultivo. Estas máquinas permitieron preparar el suelo de manera más eficiente, sembrar y plantar con mayor precisión, aplicar fertilizantes y abonos, así como fumigar y controlar plagas y enfermedades de forma más eficiente y amplia. Además, la mecanización también contribuyó al transporte dentro de las explotaciones y, en general, permitió reducir tiempos, disminuir la carga de trabajo manual y optimizar tanto la producción como el uso de los recursos disponibles.

Una de las grandes transformaciones que fue introducida en la agricultura en esta fase fueron los insumos modernos: fertilizantes químicos, pesticidas sintéticos y semillas genéticamente modificadas o mejoradas.

Los fertilizantes permitieron aportar nutrientes esenciales a los agricultores directamente al suelo como “el nitrógeno promueve el crecimiento vegetativo, el fósforo mejora el desarrollo de raíces y la floración, y el potasio fortalece la resistencia de las plantas” (Zambrano, Rodríguez, y Pire, 2002). Para acelerar el crecimiento y mejorar el rendimiento de los cultivos, especialmente en terrenos con suelos empobrecidos. Por otro lado, las pesticidas sintéticas ayudaron a combatir plagas, hongos y enfermedades que antes reducían considerablemente los rendimientos, aplicándose de forma preventiva o correctiva sobre las plantas. Finalmente, al paso del tiempo se desarrollaron semillas mejoradas genéticamente diseñadas para resistir condiciones climáticas adversas, enfermedades comunes y ciertos herbicidas.

Figura 2. Cosechadoras cultivando maíz de manera industrial.



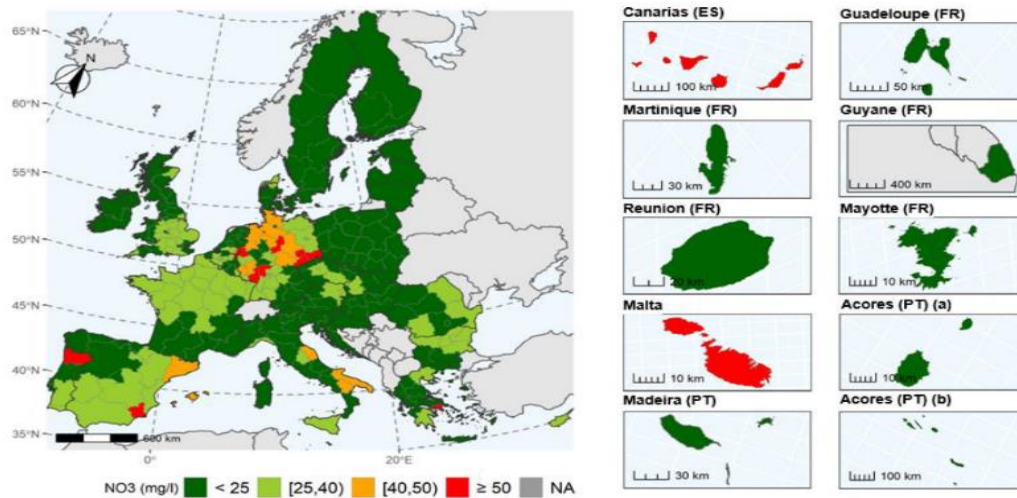
Fuente: Group (2022)

Gracias a estas innovaciones, los agricultores pudieron contar con cultivos más resistentes, estables y con mayores posibilidades de tener una buena cosecha. En conjunto, todo esto no solo ayudó a aumentar la productividad, sino que también permitió reducir muchas de las pérdidas que antes eran comunes. Esta etapa marcó un antes y un después en la historia de la agricultura moderna.

Estos nuevos avances en la mejora de los métodos, recursos y herramientas utilizadas por los agricultores en la agricultura moderna se perciben comúnmente como avances positivos en términos de eficiencia y productividad: permitieron tener plantas más resistentes y produjeron mucho más para una población que no deja de crecer. Sin embargo, “A diferencia de los métodos de reproducción anteriores, estas técnicas plantearon dudas sobre su seguridad para los consumidores y el medio ambiente” (Harlander, 2002). Durante un tiempo se consideraron una solución eficaz, pero con el paso de los años han ido apareciendo nuevos problemas. Por ejemplo, se ha empezado a observar el impacto ambiental que pueden generar estas prácticas “reducir los impactos ambientales, por ejemplo, en la calidad del agua, en la integridad funcional del suelo, en la pesca afectada por la escorrentía agrícola o, más ampliamente, en la biodiversidad nativa” (Thrall et al., 2010).

Un caso ampliamente documentado es el uso excesivo de fertilizantes en los cultivos, que, al mezclarse con la lluvia, provocan escorrentías agrícolas que terminan en lagos, ríos, mares y océanos. Lo que genera un desequilibrio ecológico al alterar las condiciones naturales del ecosistema acuático, disminuyendo la calidad del agua.

Figura 3. Concentraciones medias anuales de nitratos en las aguas subterráneas para el periodo 2015-2019.



Fuente: Demográfico (2025)

La Figura 3 muestra la concentración de nitratos en aguas subterráneas en distintas regiones de Europa. Se observa que las zonas con mayor densidad de explotaciones agrícolas coinciden con aquellas que registran los niveles más elevados de nitratos (expresados en mg/l).

Figura 4. Infografía de la Comisión Europea con un resumen de los objetivos de la Directiva de Nitratos.



Fuente: Comisión Europea (2025)

En la Figura 4 se presenta el marco de control directivo establecido por la Comisión Europea, junto con una serie de soluciones y objetivos concretos dirigidos a que los agricultores puedan hacer frente a las problemáticas actuales. Estas medidas tienen

como finalidad evitar la contaminación de los recursos hídricos, proteger los ecosistemas locales y mitigar el deterioro en la calidad del agua.

Adicionalmente, los fertilizantes no constituyen el único factor asociado a los problemas emergentes en la agricultura moderna; el uso de pesticidas también desempeña un papel relevante. Los pesticidas son sustancias químicas empleadas para prevenir, eliminar, repeler o controlar organismos nocivos que afectan a los cultivos, con el propósito de garantizar su protección.

Según diversos estudios, se utilizan 5.600 millones de libras de pesticidas en todo el mundo. En muchos países en desarrollo, los programas para controlar la exposición son limitados o inexistentes. En consecuencia, se estima que hasta 25 millones de trabajadores agrícolas en todo el mundo sufren intoxicaciones accidentales por pesticidas cada año (Alavanja, 2009; Donaldson, Kiely, y Grube, 1999; Jeyaratnam, 1990).

El uso intensivo y, en ocasiones, inadecuado de pesticidas ha contribuido a la degradación de la biodiversidad, así como a la pérdida de calidad de los recursos del suelo y del agua. Este fenómeno genera un ciclo retroalimentado en el que el deterioro de las condiciones naturales del cultivo incentiva el uso de fertilizantes adicionales, lo cual puede conllevar una aplicación excesiva de pesticidas, incluso en ausencia de amenazas reales.

Esta dependencia de insumos químicos ha sido señalada por diversos organismos internacionales como un factor que compromete la sostenibilidad de los sistemas agrícolas.

La creciente evidencia científica sobre los efectos negativos asociados a estas prácticas ha impulsado un debate global sobre la necesidad de incorporar enfoques agrícolas más sostenibles y ambientalmente responsables.

3.3 AGRICULTURA DE PRECISIÓN

Esta fase se originó como consecuencia del uso intensivo e insostenible de insumos agrícolas. “Puso en peligro la integridad ecológica de los agroecosistemas debido al uso intensivo de combustibles fósiles, recursos naturales, agroquímicos y maquinaria. Además, amenazó las prácticas agrícolas tradicionales milenarias” (Singh y Singh, 2017). Esta nueva etapa se conoce como agricultura de precisión, y se caracteriza por la toma de decisiones fundamentadas en datos e información obtenida mediante tecnologías avanzadas.

A diferencia de fases anteriores, la agricultura de precisión utiliza herramientas digitales, sensores, sistemas GPS, imágenes satelitales y análisis de datos para aplicar de forma exacta y localizada insumos como fertilizantes, pesticidas o agua. El objetivo es reducir el uso innecesario de agroquímicos, minimizar el impacto ambiental, proteger la biodiversidad y optimizar los recursos empleados en cada cultivo, logrando así una producción más eficiente, sostenible y controlada.

La agricultura de precisión marca un punto de inflexión al ser la primera etapa en la que se incorpora de forma consciente la sostenibilidad ambiental como parte esencial del modelo productivo. A diferencia de la Revolución Verde, que priorizó el aumento de la productividad sin considerar del todo las consecuencias ecológicas, esta nueva fase tiene como finalidad el equilibrio entre la eficiencia agrícola y la protección del entorno natural. Según Singh y Singh (2017) la producción sostenible de alimentos es uno de los principales desafíos del siglo XXI en la era de los problemas ambientales globales, como el cambio climático, el aumento de la población y la degradación de los recursos naturales, incluyendo la degradación del suelo y la pérdida de biodiversidad. Además, ha puesto en peligro las prácticas agrícolas tradicionales milenarias. En este contexto, la agricultura de precisión surge como una respuesta a esos desafíos, promueve prácticas agronómicas fundamentales en criterios de sostenibilidad ambiental.

3.4 AGRICULTURA 4.0

La fase más reciente en la evolución agrícola es conocida como agricultura 4.0 o agricultura inteligente. Esta etapa, mencionada previamente en el marco teórico, se caracteriza por diferencias significativas respecto a fases anteriores, tanto en su enfoque como en sus aplicaciones actuales.

La agricultura 4.0 constituye una evolución de la agricultura de precisión. Mientras que esta última se basa en la aplicación localizada de insumos agrícolas mediante el uso de tecnologías avanzadas, la agricultura inteligente amplía dicho enfoque, permitiendo una gestión integral de toda la explotación agrícola. Para ello, se emplean tecnologías similares integradas en sistemas de gestión avanzados que incorporan inteligencia artificial, big data e IoT.

Con la ayuda de estas herramientas, es posible tomar decisiones basadas en datos en tiempo real, optimizando el uso de recursos y mejorando tanto la sostenibilidad como la productividad del sistema agrícola.

A continuación, se presentan las principales tecnologías que están siendo implementadas y que están transformando el sector agrícola contemporáneo:

3.4.1 Internet de las cosas

Una de las tecnologías fundamentales en el contexto de la agricultura 4.0 es el IoT. Esta tecnología se basa en la interconexión de dispositivos a través de internet con el fin de recopilar datos y comunicarse entre sí sin necesidad de intervención humana constante. En el entorno agrícola, esto se aplica mediante sensores, estaciones meteorológicas, drones o sistemas de riego automatizados, los cuales permiten el monitoreo en tiempo real de variables como las condiciones del suelo, la humedad, la temperatura ambiental y el estado fisiológico de los cultivos. De acuerdo con Khanna y Kaur (2019), IoT es una técnica que combina recursos existentes en Internet para obtener control sobre los dispositivos.

Otra tecnología que está transformando la agricultura moderna es el Big Data. Según

vez más datos a un coste de adquisición cada vez menor. Esta herramienta se basa en la recopilación y el análisis de grandes volúmenes de datos provenientes de múltiples fuentes, como sensores instalados en el campo, satélites, drones, maquinaria agrícola, bases de datos climáticas y registros históricos de producción.

El análisis de estos datos permite identificar patrones, prever comportamientos y tomar decisiones estratégicas fundamentadas. De este modo, la toma de decisiones deja de depender únicamente de la intuición o de la experiencia empírica, y se apoya en información precisa y en tiempo real. Esto facilita la optimización de procesos agrícolas, como la planificación de la siembra, la rotación de cultivos, la gestión del riego o la aplicación de insumos en zonas específicas del terreno, contribuyendo así a una producción más eficiente y sostenible.

3.4.3 Inteligencia artificial

La inteligencia artificial (IA) se ha convertido en uno de los componentes centrales del sistema agrícola moderno. Según Cravero y Sepúlveda (2021), sus aplicaciones abarcan la predicción del rendimiento, la detección de enfermedades y malezas, la evaluación de la calidad de los cultivos, el reconocimiento de especies, así como el monitoreo del bienestar y la producción animal.

En este contexto, la IA se integra con otras tecnologías como el IoT, encargado de recopilar datos en tiempo real, y el *Big data*, que permite organizar y analizar grandes volúmenes de información. La inteligencia artificial, a través de algoritmos de aprendizaje automático (*machine learning*), interpreta estos datos para identificar patrones, anticipar eventos y optimizar procesos agrícolas. El *machine learning* es una rama de la IA que permite que los sistemas aprendan automáticamente a partir de datos sin ser programados de forma explícita. Una de sus técnicas más utilizadas en la agricultura 4.0 son las redes neuronales artificiales, que imitan la forma en que el cerebro humano procesa información y son especialmente efectivas para reconocer imágenes, clasificar información y predecir comportamientos complejos en sistemas agrícolas.

Entre sus aplicaciones más destacadas se encuentra la detección temprana de enfermedades en los cultivos, la recomendación de dosis óptimas de insumos según las condiciones del terreno y el historial de producción, así como la gestión eficiente de recursos naturales. Este enfoque contribuye significativamente a una agricultura más sostenible, eficiente y basada en la toma de decisiones fundamentadas en datos.

3.4.4 Cambio del rol del agricultor

El rol del agricultor ha experimentado una transformación significativa en muchos contextos, impulsada por la integración de tecnologías avanzadas, análisis de datos y toma de decisiones basada en información en tiempo real. No obstante, la adopción de estas innovaciones no se produce de forma homogénea, ya que varía según los recursos disponibles, la experiencia previa y las condiciones particulares de cada explotación.

Figura 7. Cambio de rol del agricultor tomado decisiones críticas con su software de gestión agrícola.



Fuente: Group (2022)

Este nuevo enfoque ha incrementado la complejidad técnica del trabajo agrícola, exigiendo conocimientos no solo agronómicos, sino también digitales. El agricultor contemporáneo debe ser capaz de interpretar datos, aplicar insumos con precisión y ajustar sus prácticas productivas en función de parámetros técnicos, lo que ha elevado el perfil profesional hacia una figura más técnica y especializada.

Sin embargo, esta transformación también plantea desafíos. La creciente dependencia de tecnologías digitales podría contribuir a una progresiva desvalorización del conocimiento tradicional, acumulado a lo largo de generaciones, lo cual supone una posible pérdida de prácticas sostenibles y adaptadas a contextos específicos.

3.4.5 Retos actuales y barreras tecnológicas

La agricultura 4.0 representa un avance significativo en comparación con las etapas iniciales de desarrollo del sector. A pesar de los beneficios que ofrece este nuevo enfoque de cultivo, basado en el uso de tecnologías avanzadas para mejorar la sostenibilidad y productividad del sistema agrícola, todavía enfrenta importantes retos, limitaciones y desafíos estructurales.

Aunque esta fase representa un avance tecnológico considerable respecto a las anteriores, todavía existen barreras tanto tecnológicas como económicas. El primer desafío evidente es que la implementación de un sistema avanzado, basado en tecnologías interconectadas y automatizadas, no resulta ni económica ni técnicamente accesible para todos. En segundo lugar, muchas personas dedicadas a la agricultura carecen de formación específica en estas herramientas, lo que las obliga a depender de proveedores externos para la instalación, gestión y mantenimiento de los sistemas, reduciendo así su autonomía. Finalmente, no todas las explotaciones agrícolas se

encuentran en zonas con infraestructura adecuada, enfrentando problemas como la falta de redes eléctricas fiables o una conectividad a internet estable.

Otro reto que enfrenta esta fase es la resistencia al cambio por parte de ciertos sectores del ámbito agrícola. La adopción de nuevas formas de cultivo basadas en tecnología genera preocupación por diversos motivos, entre ellos, la posible pérdida de conocimientos tradicionales, la alta dependencia de sistemas tecnológicos, y el desconocimiento o desconfianza hacia estas nuevas herramientas.

Entre los principales retos destacan la seguridad de los datos y el riesgo de ciberataques, agravados por la alta dependencia tecnológica y la falta de regulación. Además, el control de estas herramientas por grandes corporaciones puede excluir a pequeños productores y acentuar desigualdades en el sector agrícola.

4. LA CIENCIA DE DATOS Y SU APLICABILIDAD EN LA AGRICULTURA

4.1 ESTUDIOS PREVIOS APLICANDO CIENCIA DE DATOS EN LA AGRICULTURA

Tabla 1: Estudios previos

Estudios previos aplicando ciencia de datos en la agricultura

Autor	Título	Objetivo	Resultados
(Kamilaris, 2018)	Deep learning in agriculture: A survey	Explorar cómo el Deep learning puede aplicarse en la agricultura.	Aplicando en reconocimiento de cultivos, plagas y estimación de rendimiento.
(Tace et al., 2022)	Smart irrigation system based on IoT and machine learning	Desarrollar un sistema de riego inteligente basado en datos en tiempo real.	Se logró una reducción del 30% en el uso de agua sin afectar el rendimiento del cultivo
(Singla et al., 2024)	Exploration of machine learning approaches for automated crop disease detection	Predecir enfermedades en cultivos mediante clasificación de imágenes.	Se alcanzó una precisión del 92% usando modelos de una red neuronal convolucional.
(Luyckx et al., 2022)	The Future of Farming: The (Non)-Sense of Big Data Predictive Tools for Sustainable EU Agriculture	Analizar los desafíos del uso de Big Data en agricultura sostenible dentro de la UE.	Se cuestiona la capacidad del Big Data para captar la complejidad agrícola.
(Carbonell, 2016)	The ethics of big data in big agriculture	Explorar las implicaciones éticas y sociales del Big Data en la agricultura.	Riesgo de pérdida de control de datos por parte del agricultor y

4.2 COMPARACIÓN DE ENFOQUES Y HERRAMIENTAS

Analizando los trabajos citados en la tabla anterior, se observa que los autores han adoptado distintos enfoques para explorar la aplicabilidad y funcionalidad de la ciencia de datos en la agricultura. Algunos estudios se centran en la reducción de insumos durante la producción con el objetivo de mejorar la eficiencia en costes y rendimiento; otros abordan el uso de técnicas de detección para identificar plagas, malas hierbas y enfermedades comunes en los cultivos; mientras que algunos cuestionan la utilidad, accesibilidad y seguridad de estas tecnologías, planteando dudas sobre su verdadero impacto y equidad en el sector agrícola.

Estudios como los de Kamilaris (2018), Tace et al. (2022) y Singla et al. (2024), se orientan hacia un enfoque técnico, buscando optimizar procesos agrícolas mediante herramientas de ciencia de datos. Por otro lado, investigaciones como las de Carbonell (2016) y Luyckx et al. (2022), adoptan una perspectiva más crítica, ética y social, analizando cómo estas tecnologías afectan la distribución del poder, la soberanía de los agricultores y la brecha digital entre grandes y pequeñas explotaciones.

En cuanto a las técnicas y herramientas utilizadas, se identifica una amplia variedad de enfoques dentro del campo de la ciencia de datos. Kamilaris destaca la aplicación del *deep learning*, especialmente mediante redes neuronales convolucionales (CNN), para el reconocimiento automático de cultivos, detección de plagas y estimación de rendimiento. Singla et al. (2024), también emplea CNN para predecir enfermedades en cultivos a partir de imágenes. Por su parte, Tace et al. (2022), combinan el IoT con algoritmos de *machine learning* para crear un sistema de riego inteligente capaz de reducir el uso de agua sin afectar la productividad. En contraste, los estudios de Luyckx et al. (2022) y Carbonell (2016), no se basan en modelos computacionales, sino que analizan críticamente el uso de herramientas predictivas basadas en *Big data* y su implicación en la gobernanza agrícola y la ética de los datos. Cada uno de estos trabajos, desde su enfoque particular, contribuye a una visión más amplia sobre las oportunidades y límites de aplicar ciencia de datos en el contexto agrícola.

4.3 OPORTUNIDADES Y BENEFICIOS

En el ámbito agrícola, contar con información detallada y la capacidad de predecir el comportamiento de los cultivos representa una ventaja significativa, ya que incrementa las probabilidades de obtener un producto final exitoso. La principal fortaleza de incorporar nuevas tecnologías y enfoques basados en ciencia de datos radica en su impacto directo en la productividad del agricultor, al permitir una detección temprana de los insumos necesarios, los riesgos potenciales y los recursos que requieren los cultivos. Como señalan algunos autores, “las técnicas basadas en datos representan una gran oportunidad para sectores como la agrícola, al ofrecer información procesable que permite mejorar las prácticas de cultivo, aumentando así la eficiencia y el rendimiento.” (Taneja et al., 2020)

La integración de ciencia de datos en la agricultura ha revelado múltiples oportunidades de mejora en la eficiencia y sostenibilidad del sector. Entre las más destacadas se encuentran: la optimización del uso de recursos como el agua y fertilizantes (Tace et al.,

2022), la detección temprana de enfermedades y plagas mediante visión por computador (Singla et al., 2024), la predicción del rendimiento agrícola (Kamilaris, 2018), y el apoyo a la toma de decisiones con base en datos en tiempo real.

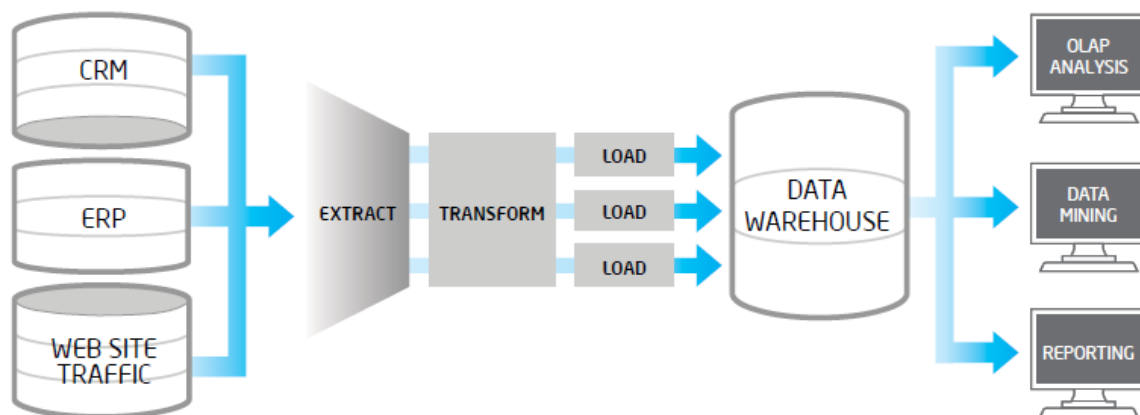
Estos son solo algunos ejemplos de los autores mencionados anteriormente, aunque se han realizado numerosos experimentos aplicando nuevas tecnologías y ciencia de datos en diversas industrias, especialmente en la agricultura. Existen múltiples oportunidades adicionales, como el desarrollo de sistemas de trazabilidad y seguridad alimentaria, la predicción y preparación ante condiciones meteorológicas, el monitoreo general del estado de los cultivos, así como la optimización logística y de la cadena de suministro.

En conjunto, estas aplicaciones reflejan el principal beneficio de la ciencia de datos en la agricultura: convertirla en un proceso más preciso y con un control integral de la producción. Esto no solo permite que los agricultores sean más eficientes y eficaces en sus operaciones, sino que también promueve una gestión más sostenible de los recursos.

4.4 TECNOLOGÍAS CLAVE QUE PERMITEN APLICAR CIENCIA DE DATOS EN LA AGRICULTURA

La integración de tecnologías avanzadas se ha convertido en un pilar esencial para la agricultura de precisión, al permitir una gestión más eficiente y sostenible de la producción. Tal como señala Rehman et al. (2017), la tecnología nueva o moderna en el sector agrícola puede mejorar sustancialmente la producción agrícola y su sostenibilidad, lo que respalda la necesidad de adoptar sistemas inteligentes interconectados que generen un flujo constante de datos. Estos datos deben ser almacenados de manera segura en bases de datos, para posteriormente pasar por un proceso estructurado propio de la ciencia de datos que incluye las fases de recolección, almacenamiento, limpieza, análisis, visualización y toma de decisiones.

Figura 8. Proceso ETL y el viaje del dato adentro de una empresa.



Fuente: Vidal (2014)

Este flujo de trabajo, conocido como proceso ETL, constituye un procedimiento estandarizado aplicable a cualquier industria, cuyo objetivo es transformar los datos en información valiosa para optimizar procesos y respaldar decisiones estratégicas. Cada fase del ETL requiere conocimientos especializados y el uso de herramientas tecnológicas específicas para garantizar que la información obtenida sea confiable, interpretable y útil para la gestión agrícola. A continuación, se detallan las principales tecnologías que permiten aplicar la ciencia de datos de manera efectiva en este sector.

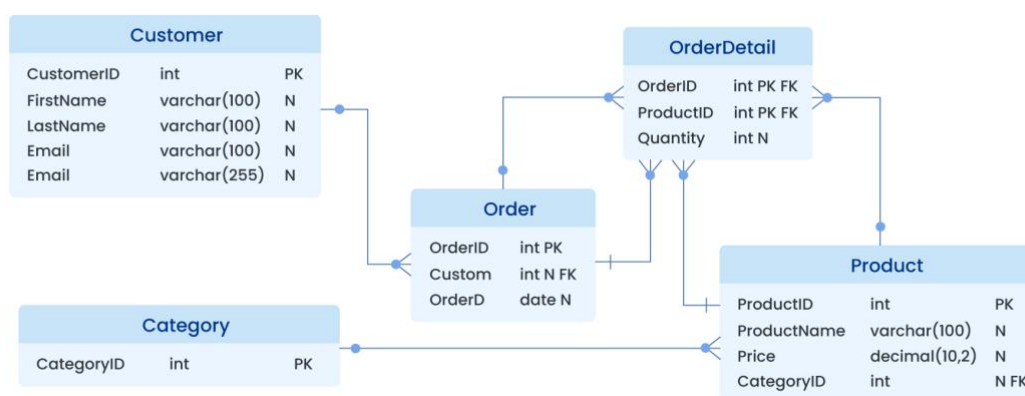
4.4.1 Base de datos

En cualquier sistema de ciencia de datos y analítica es imprescindible contar con un mecanismo de almacenamiento que permita resguardar la información para su posterior exploración, análisis y extracción de valor. Sin embargo, antes de llegar a la fase de almacenamiento, es necesario disponer de un sistema que facilite la captura y extracción de los datos.

Tomemos como ejemplo una explotación agrícola dedicada al cultivo de uvas, es decir, una explotación vitivinícola. En el campo, el agricultor dispone de sensores de diversos tipos distribuidos estratégicamente para medir variables como el nivel de agua en los cultivos, el pH y la concentración de nutrientes en el agua, la temperatura, la humedad, las horas de luz, las condiciones meteorológicas, entre otros factores. Todos estos dispositivos están interconectados mediante el IoT, lo que les permite comunicarse y compartir datos entre sí.

Toda esta información se almacena y gestiona en una base de datos local, de modo que el agricultor pueda interpretarla y, con base en ella, decidir cómo gestionar y cuidar el viñedo en su día a día para obtener el mejor rendimiento y calidad posible.

Figura 9. Tabla de relaciones de una base de datos.



Fuente: Vidal (2014)

Todos los datos que se capturan no pueden almacenarse sin control ni un sistema establecido, ya que, sería un caos interpretarlos y utilizarlos posteriormente. Es necesario darles estructura y significado, pues un dato por sí solo carece de valor. Por

ello, dentro de las bases de datos se lleva a cabo un proceso de limpieza y organización que permite que la información sea útil y coherente.

Dentro de una base de datos, los datos no se almacenan de forma aleatoria. Cada valor se clasifica y etiqueta según su tipo de dato (número entero, decimal, texto, fecha, booleano, etc.), lo que garantiza que cada información ocupe su lugar correspondiente, como si fuera un cajón dentro de un sistema de archivo bien organizado. Esta categorización facilita que, al crecer el volumen de información, la base de datos pueda mantenerse ordenada y optimizada para consultas rápidas y precisas.

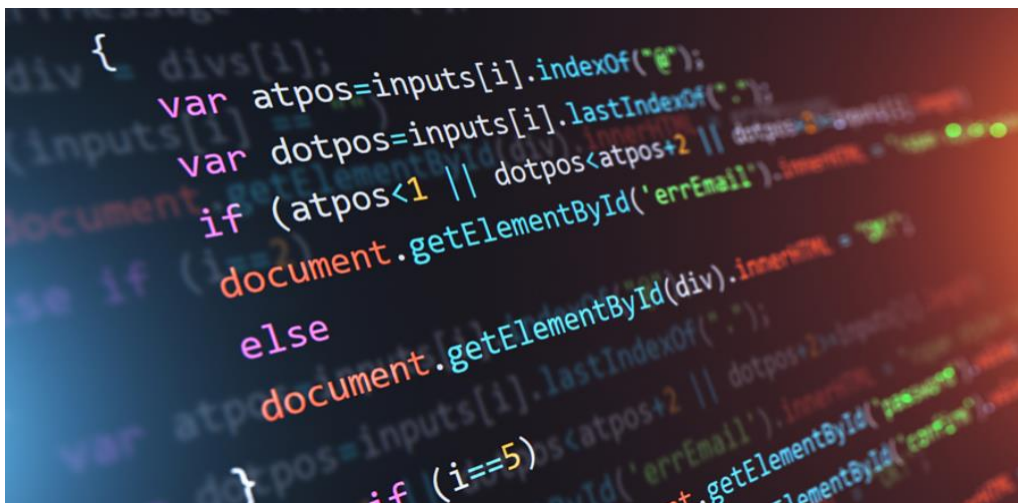
Entre las tecnologías más utilizadas se encuentra SQL (*Structured Query Language*), un lenguaje de programación diseñado específicamente para gestionar y manipular datos almacenados en bases de datos relacionales. Dentro de este ecosistema, *MySQL* es uno de los sistemas de gestión de bases de datos relacionales.

Para mantener un orden y gestionar adentro la base de datos se utiliza un el modelo entidad-relación, una metodología que representa gráficamente cómo se estructuran y conectan los datos dentro de la base, como se muestra arriba en la Figura 9. Las entidades corresponden a los elementos principales que se desean almacenar (por ejemplo: parcela, cosecha, cultivo) mientras que las relaciones describen cómo interactúan entre sí las entidades y las describe (por ejemplo: ID cosecha, fecha de recolección, variedad, n.º de kilogramos obtenidos, condición y mucho más). A medida que la base de datos crece, la información se organiza automáticamente con este modelo.

4.4.2 Lenguajes de programación

Una vez que los datos han sido recopilados y almacenados de manera estructurada, el siguiente paso consiste en su análisis y estudio con el fin de extraer información valiosa. Para ello, es necesario utilizar lenguajes de programación que permitan procesar y analizar la información.

Figura 10. Tabla de relaciones de una base de datos.



Fuente: Networks (2021)

Los lenguajes de programación pueden entenderse como un sistema de comunicación estructurado entre el ser humano y el ordenador, diseñado para dar instrucciones precisas a las máquinas. En el ámbito de la ciencia de datos, destacan varios lenguajes especializados: *Python*, ampliamente utilizado para análisis de datos y desarrollo de algoritmos de *machine learning*; *R*, orientado principalmente a la estadística y el análisis avanzado; y *SQL*, empleado para la gestión y consulta de bases de datos. En algunos casos, también se recurre a *JavaScript* para el tratamiento y visualización interactiva de datos.

La programación puede realizarse en diversos entornos y plataformas, cada uno con características específicas que facilitan la escritura, depuración y ejecución del código.

Entre los principales se encuentran:

1. Entornos de desarrollo integrados (IDE), que permiten escribir, depurar y ejecutar código dentro de una misma aplicación, como es el caso de *RStudio* para el lenguaje *R*, ampliamente utilizado en cálculos y análisis estadísticos.
2. *Notebooks* interactivos y colaborativos, que ofrecen la posibilidad de ejecutar código en un documento abierto y compartido, como *Jupyter Notebook* o *Google Colab*.
3. Consolas y terminales propias del sistema operativo, que permiten la ejecución directa de instrucciones y *scripts*.
4. Plataformas en la nube, como *AWS SageMaker* o *Azure Machine Learning*, que facilitan el desarrollo, entrenamiento y despliegue de modelos de forma escalable y remota.

Cada lenguaje de programación dispone de un ecosistema propio de paquetes o bibliotecas que facilitan, ampliar y optimizar sus capacidades, permitiendo incorporar funciones avanzadas para realizar distintos tipos de análisis y operaciones. En el caso de *Python*, uno de los lenguajes más utilizados en ciencia de datos, destacan las siguientes bibliotecas: *Pandas*, que facilita la manipulación y el procesamiento de datos; *NumPy*, orientada al cálculo numérico de alto rendimiento; *Matplotlib*, que permite la visualización de datos en forma de gráficos y diagramas; *Scikit-learn*, especializada en la implementación de algoritmos de machine learning; y *TensorFlow*, enfocada en el desarrollo de modelos de *deep learning*.

Como ejemplo, puede considerarse el caso de un agricultor que cultiva uvas de la variedad Tempranillo y registra datos diarios de su viñedo, incluyendo la fecha de recolección, los kilogramos cosechados, el nivel de azúcar en el mosto y la temperatura media del día. Utilizando el lenguaje *Python* en un entorno como *Google Colab* o *Jupyter Notebook*, y con bibliotecas como *Pandas* y *Matplotlib*, es posible importar y organizar los datos, calcular promedios y representar gráficamente la relación entre temperatura y cantidad cosechada. Este análisis permite identificar las condiciones óptimas para maximizar el rendimiento y la calidad, demostrando cómo la programación facilita la transformación de datos en información útil para la toma de decisiones agrícolas.

4.4.3 Machine learning

Dentro del ámbito de la analítica y el estudio de datos, existen diversos tipos de análisis, cada uno orientado a obtener conclusiones específicas y a extraer diferentes aprendizajes. Los principales enfoques son: el análisis descriptivo, cuyo objetivo es comprender qué ha ocurrido en el pasado; el análisis diagnóstico, que busca explicar por qué sucedió un determinado hecho; el análisis predictivo, enfocado en anticipar escenarios futuros; el análisis prescriptivo, que tiene como propósito recomendar acciones para alcanzar un resultado distinto y optimizado; y, por último, el análisis exploratorio, destinado a identificar patrones, relaciones y anomalías presentes en los datos sin partir de hipótesis previas.

El *machine learning* es una rama de la ciencia de datos que involucra el uso de inteligencia artificial, en la cual los ordenadores aprenden de forma autónoma las relaciones y patrones presentes en los datos. Esto les permite tomar decisiones o realizar predicciones sin estar programados específicamente para cada situación, sino aprendiendo mediante procesos iterativos de prueba y error. La inteligencia artificial busca imitar la inteligencia humana, replicando en los sistemas informáticos la capacidad de aprender a partir de la experiencia y de extraer conclusiones de forma similar a como lo hace una persona.

Figura 11. Machine learning aplicado en la producción de uva.



Fuente: Marr (2025)

Numerosos estudios de ciencia de datos aplicados al ámbito agrícola hacen uso del *machine learning* y lo citan como herramienta central. De los seis tipos de análisis que pueden realizarse en ciencia de datos, el *machine learning* se enfoca principalmente en tres:

- **Análisis predictivo:** orientado a anticipar resultados y situaciones futuras (por ejemplo, estimar la producción de uva en función del clima y las condiciones del cultivo).

- **Análisis prescriptivo:** destinado a recomendar y aplicar ajustes para optimizar el resultado final (por ejemplo, modificar los niveles de riego y fertilización para maximizar la calidad del producto).
- **Análisis exploratorio:** enfocado en identificar patrones y relaciones entre los datos, mediante técnicas como el clustering.

Un ejemplo de aplicación del machine learning en un viñedo consiste en utilizar datos históricos sobre clima, horas de luz, riego, fertilización y producción de uva para entrenar un modelo predictivo. Una vez entrenado, el modelo puede anticipar cuántos kilogramos de uva se cosecharán y su calidad estimada según las condiciones actuales. Además, puede recomendar ajustes como aumentar el riego en días de alta temperatura o modificar la dosis de nutrientes para maximizar la calidad del fruto.

4.4.4 Deep learning

El *deep learning* es una subrama del machine learning que emplea redes neuronales artificiales con múltiples capas de procesamiento para aprender patrones complejos a partir de grandes volúmenes de datos. Esta técnica posee la capacidad de identificar relaciones y estructuras en los datos sin intervención manual, a diferencia de otros métodos de machine learning que requieren una mayor participación humana en la selección de características.

En la agricultura, el *deep learning* se aplica, por ejemplo, para analizar imágenes de drones y satélites y así detectar plagas, enfermedades o deficiencias nutricionales en cultivos. También puede clasificar automáticamente la calidad de la fruta en una cinta de producción, reduciendo errores y aumentando la eficiencia.

4.4.5 Herramientas de visualización

La transformación de los datos en herramientas visuales permite convertir información compleja en un formato comprensible para facilitar la toma de decisiones. Un buen gráfico o una tabla dinámica pueden resumir meses de información y favorecer su interpretación, lo que contribuye a una mejor toma de decisiones. En esta fase, se establece un vínculo entre el agricultor y los datos, permitiéndole actuar de forma rápida y precisa, además de obtener una visión global del proceso.

Existen diversas herramientas y programas que permiten transformar los datos en representaciones visuales, algunas más avanzadas que otras. Entre ellas destacan: *Power BI* y *Tableau*, ideales para paneles interactivos; *Excel*, útil para gráficos y tablas dinámicas, además de ser compatible con *Power BI*; *QGIS*, especializado en mapas y análisis geoespacial; y, por último, librerías de *Python* que facilitan la expansión y visualización de los datos como *Matplotlib* y *Seaborn*.

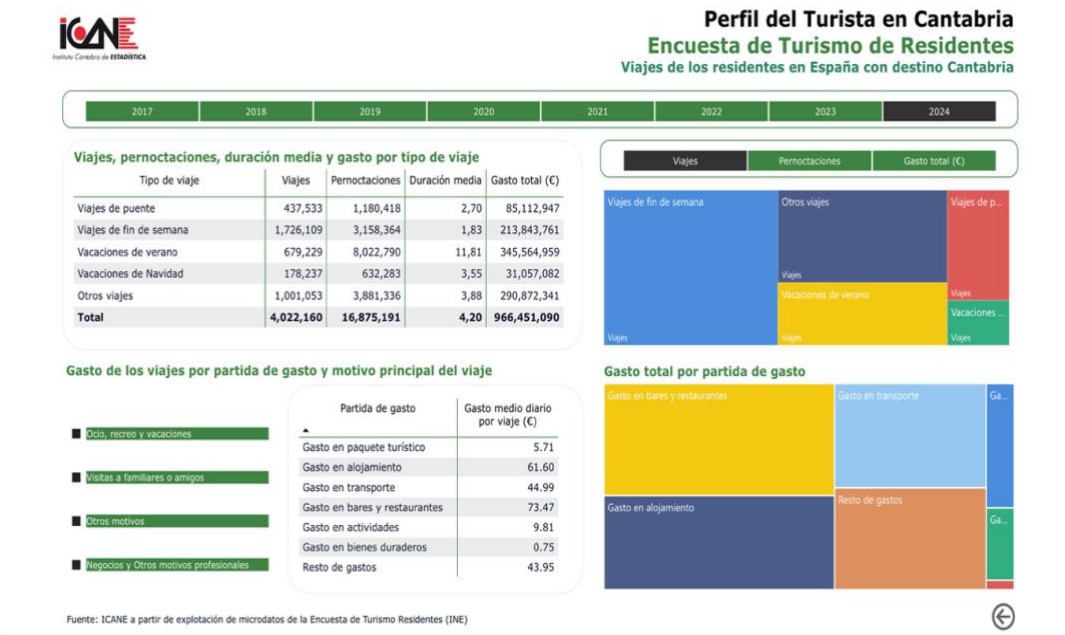
Las herramientas de visualización en el ámbito agrícola pueden mostrar tendencias en la producción y las cosechas, generar mapas de calor que permitan un análisis detallado del terreno para identificar las zonas con mayor rendimiento, y ofrecer paneles en tiempo real que se actualizan automáticamente para reflejar el estado actual de los cultivos.

Figura 12. Ejemplo de una granja inteligente con una tablet para visualizar los datos.



Fuente: Synox (2025)

Figura 13. Ejemplo de un dashboard interactivo creado en PowerBi.



Fuente: ICANE (2024)

Con el ejemplo del viñedo, las herramientas de visualización pueden mostrar en un único panel interactivo datos como la evolución semanal de la producción de uvas, el consumo de agua por parcela y la temperatura media registrada. Un mapa de calor podría resaltar las zonas con mayor rendimiento, mientras que un gráfico de líneas permitiría comparar la producción de este año con la de temporadas anteriores. Además, un panel en tiempo real podría actualizarse automáticamente con datos de sensores en campo, permitiendo al agricultor detectar de inmediato cualquier cambio significativo y tomar decisiones rápidas para optimizar la cosecha.

5. CASO PRÁCTICO: SEGUIMIENTO DE LA PRODUCCIÓN DE UVA PARA LA FABRICACIÓN DE VINO EN LA RIOJA

5.1 PLANTEAMIENTO

España destaca por diversos aspectos, y uno de los más relevantes es su gastronomía. “La industria alimentaria constituye una de las principales y más influyentes del país, situándose como la rama de actividad con mayor aportación a la facturación industrial en 2022, representando el 15,5 % del total del PIB” (INE, 2022).

Figura 14. Cifra de Negocio por Categoría de la Industria Alimentaria en Millones de euros.



Fuente: Ministerio de Agricultura (2023)

La gastronomía española se encuentra entre las más reconocidas y apreciadas a nivel mundial, destacando por su diversidad y calidad. Entre sus productos más emblemáticos se incluyen el jamón y otros embutidos, los quesos, las carnes, los pescados, los mariscos y, por supuesto, el vino. Según el Ministerio de Agricultura, Pesca y Alimentación, la fabricación de bebidas es la segunda categoría más relevante dentro de la industria alimentaria, solo por detrás del sector cárnico. Dentro de esta categoría se encuentra la producción de vinos, entre otros productos.

Los principales actores de la industria del vino a nivel mundial son Francia, Italia y España. “Este último se posiciona como el tercer mayor exportador de vino por volumen”, según (Statista, 2024).

La cultura, el estudio y la producción de vino tienen una gran relevancia en España, donde prácticamente todas las comunidades autónomas participan en su elaboración.

Entre las más destacadas se encuentran Castilla-La Mancha, Castilla y León, La Rioja, Galicia, Comunidad Valenciana y Andalucía. A nivel internacional, los vinos más reconocidos de España provienen principalmente de dos regiones: La Rioja y Ribera del Duero.

Figura 15. Bodega Marqués de Riscal en La Rioja Alavesa.



Fuente: Elaboración propia

Para el desarrollo del caso práctico, se ha seleccionado la región de La Rioja como objeto de estudio, debido a sus condiciones climáticas y a su ubicación geográfica, que la convierten en un entorno especialmente interesante para analizar la relación entre los datos y la producción agrícola.

La Rioja constituye un territorio privilegiado para la producción de uvas y la elaboración de vinos. Su localización geográfica reúne características similares a otras regiones vitivinícolas del mundo, ya que combina condiciones climáticas y factores geográficos que generan un microclima estable y favorable para la viticultura. El río Ebro contribuye al riego natural de los suelos, mientras que la presencia de montañas y cordilleras regula las temperaturas y protege frente al exceso de lluvias. A ello se suma la calidad nutritiva de la tierra y su posición en el interior de la península, que evita una influencia excesiva de la humedad costera. En conjunto, estos elementos conforman un entorno ideal que ha favorecido el florecimiento de la industria vitivinícola en la región.

El objetivo de este caso de estudio es analizar cómo diferentes factores climáticos como la temperatura, las precipitaciones, la humedad o las horas de sol influyen directamente en la producción de uva y vino en la región. Con este enfoque se busca identificar patrones y relaciones de causa-efecto que permitan comprender mejor el papel del clima en el rendimiento de las cosechas vitivinícolas.

5.2 METODOLOGÍA

Se emplea un diseño de observación anual que integra las condiciones climáticas (humedad, temperaturas, olas de calor, heladas, horas de sol y precipitaciones) junto con los rendimientos de producción de uva para la elaboración de vino en La Rioja durante los últimos años. Los datos se han obtenido de fuentes abiertas proporcionadas por el Gobierno de La Rioja, la AEMET, el Ministerio de Agricultura, Pesca y Alimentación, así como de organizaciones vitivinícolas de la región.

Una vez recopilados los *datasets*, se aplica el proceso ETL de la ciencia de datos. En esta fase se realiza una limpieza y reorganización de la información, conservando únicamente las variables relevantes y con valor analítico, y eliminando aquellas con alta correlación para evitar redundancias y confusión. Asimismo, los datos se normalizan para llevarlos a una escala común y facilitar la obtención de conclusiones más precisas.

Posteriormente, se utiliza el software de minería de datos *Weka*, aplicando algoritmos como *Random Forest*, *J48* y *Simple K-Means* con el objetivo de identificar posibles relaciones entre las condiciones climáticas y la producción vitivinícola. Los resultados permiten determinar qué variables ejercen un mayor efecto sobre la producción y cómo esta información se aprovecha en las fincas de uva y bodegas para mejorar la toma de decisiones, anticiparse a escenarios futuros y optimizar tanto la gestión en el campo como en el proceso productivo.

5.3 HERRAMIENTAS Y RECURSOS UTILIZADAS

Para el desarrollo del caso práctico se empleó el programa *Weka*, un software de minería de datos de código abierto ampliamente utilizado en el ámbito académico y de investigación. Este programa permite aplicar algoritmos de clasificación, regresión y clustering de manera sencilla e intuitiva, facilitando la interpretación de relaciones entre variables. La documentación oficial del programa se encuentra disponible en su propio wiki (Weka Wiki, 2025), que sirvió como referencia para la consulta de funciones y procesos durante el análisis. Además, se recurrió a material audiovisual de formación impartido por uno de los principales desarrolladores y difusores de *Weka* de la Universidad de Waikato en Nueva Zelanda, el profesor Ian H. Witten, a través del canal de *YouTube* WekaMOOC, donde se explican ejemplos prácticos de aplicación de los distintos algoritmos.

En cuanto a los datos, la información utilizada provino de recursos abiertos del Gobierno de La Rioja. En el portal oficial de información agroclimática se seleccionaron registros anuales comprendidos entre 2005 y 2023, recopilados y publicados gracias al AMET (Gobierno de La Rioja, 2025). Estos datos incluyeron variables climáticas como temperatura, precipitaciones, humedad, radiación y otras condiciones ambientales. Posteriormente, en la sección de estadísticas agrarias del mismo portal se consultaron los datos históricos de producción, superficie, rendimiento y elaboración de vino, seleccionando específicamente la categoría correspondiente al viñedo (Gobierno de La Rioja, 2025).

Finalmente, el programa Microsoft Excel se utilizó como herramienta de apoyo para la limpieza, transformación y organización de los datasets. En esta fase se eliminaron inconsistencias, se homogeneizaron los valores y se ajustaron las variables de interés, convirtiendo los archivos a formato CSV para su posterior procesamiento en *Weka*.

5.4 DESARROLLO

5.4.1 Data set

La información empleada en este trabajo procede de recursos abiertos del Gobierno de La Rioja. A través del portal oficial de información agroclimática se recopilaban registros históricos comprendidos entre 2005 y 2023, elaborados y publicados por el AMET (Gobierno de La Rioja, 2025). Dichos registros incluyen indicadores climáticos como temperatura, precipitaciones, humedad, radiación y otros parámetros ambientales. A su vez, en la sección de estadísticas agrarias del mismo portal se consultaron los datos de producción, superficie cultivada, rendimientos y elaboración de vino, seleccionando en todo momento la categoría específica del viñedo (Gobierno de La Rioja, 2025).

Con base en esta información se construyó un dataset propio que inicialmente estuvo compuesto por 37 variables y 19 instancias, correspondientes a los años 2005 a 2023. En su versión inicial, el dataset incluía tanto variables productivas y económicas como climáticas. No obstante, tras un análisis preliminar se decidió reducir el número de variables a 17, eliminando aquellas que mostraban una alta correlación con otras o que no aportaban valor directo al objetivo de este estudio. Entre las variables eliminadas se encuentran las económicas como el valor del mercado en euros y los precios del vino por tipo (blanco, rosado y tinto), así como las de producción desglosada por tipo de vino en hectolitros y las superficies de cosecha diferenciadas en regadío y secano.

El dataset final quedó conformado por las siguientes variables y sus respectivas unidades de medida:

- **Año** (numérica, en años calendario).
- **Cosecha** (categórica, buena o mala, variable clase creada a partir del rendimiento).
- **Rendimiento total** (kg/ha).
- **Rendimiento secano** (kg/ha).
- **Rendimiento regadío** (kg/ha).
- **Temperatura media** (°C).
- **Temperatura máxima** (°C).
- **Temperatura mínima** (°C).
- **Olas de calor** (binaria: 0 = no, 1 = sí).

- **Heladas** (binaria: 0 = no, 1 = sí).
- **Humedad media** (%).
- **Velocidad media del viento** (km/h).
- **Lluvias acumuladas** (l/m²).
- **Evapotranspiración calculada** (mm/día).
- **Temperatura máxima del suelo** (°C).
- **Temperatura mínima del suelo** (°C).
- **Radiación acumulada** (MJ/m²).

Figura 16. Data set de rendimientos de la producción del vino y condiciones climáticas de la rioja entre los años 2005-2023.

Año	Cosecha	Rendimiento Total kg entre ha	Rendimiento o Secano Total kg entre ha	Rendimiento o Regadio Total kg entre ha	Temp. Media.C	Temp. Max.C	Temp. Min.C	Ola de Calor	Heladas	Humedad Med Porcentaje	Velocidad Med Viento km entre hr	Lluvias acumuladas l entre m2	Evapotranspiración calculada mm entre día	Temp. Max. Suelo. C	Temp. Min. Suelo. C	Radiación Acumulada MJ entre m2
2005	Buena	16881	7389	9492	12.71	27.78	-0.44	1	1	67.22	9.01	23.84	113.98	20.39	9.7	498537
2006	Buena	15430	6674	8756	14.58	28.19	2.2	0	0	68.22	8.26	36.86	105.3	22.72	12.11	545668
2007	Buena	15235	6775	8460	11.3	24.99	-1.04	0	0	69.08	8.39	39.03	86.37	18.36	8.66	452545
2008	Mala	13701	6331	7370	11.37	23.85	-0.16	0	0	70.55	8.38	56.5	82.75	17.52	9.3	412428
2009	Mala	13823	6423	7400	12.06	26.33	-0.08	1	0	74.25	8.8	34.33	89.11	18.42	10.39	412227
2010	Mala	13590	6357	7233	10.94	25.69	-1.98	0	1	75.33	9.18	36.84	80.63	17.8	7.8	423363
2011	Mala	13094	6038	7056	12.29	26.63	0.53	1	0	76.08	8.93	26.64	84.7	17.93	10.73	463425
2012	Mala	12531	5388	7143	11.69	26.32	0.36	1	0	74.58	8.84	35.45	87.1	19.07	9.71	409249
2013	Mala	12910	5695	7215	11.37	24.74	0.66	0	0	77	9.17	46.28	79.62	18.57	9.23	382759
2014	Buena	14565	6712	7853	12.77	25.34	1.61	0	0	74.92	9.33	42.96	87.65	18.04	10.46	441229
2015	Buena	14917	6906	8011	12.48	26.23	1.66	1	0	75.58	9.13	35.53	89.96	17.9	10.63	467396
2016	Buena	15322	7114	8208	12.25	26.23	0.76	0	0	77.42	8.73	41.61	86.1	18.07	10.55	436579
2017	Mala	11214	4884	6330	12.53	26.88	0.21	0	1	71.92	8.96	32.42	94.63	18.85	10.4	461935
2018	Buena	15667	7260	8407	12.37	24.78	1.67	0	0	76.92	8.83	51.33	84.12	19.99	9.33	462119
2019	Mala	12072	5609	6463	12.53	27.11	0.79	1	1	73.75	9.41	42.77	94.1	19.94	9.74	468147
2020	Mala	13192	6083	7109	12.64	26.26	0.58	0	0	77.17	8.94	43.27	85.09	19.58	9.31	435327
2021	Mala	13324	6114	7210	12.06	26.06	0.8	0	0	76.83	9.45	41.47	87.77	20.91	9.6	475625
2022	Mala	13760	5790	7970	13.23	27.03	1.2	1	1	72.5	9.23	25.28	96.73	20.56	9.35	455867
2023	Mala	12722	5414	7308	13.43	27.63	1.53	0	1	71.42	9.63	29.8	98.77	21.67	10.14	485266

Fuente: Elaboración propia

5.4.2 Preparación de los datos

El proceso de preparación del dataset siguió las fases clásicas de Extracción, Transformación y Carga, adaptadas a las necesidades de este trabajo. En primer lugar, se depuraron las variables redundantes y aquellas con poca relevancia para los objetivos de investigación, como las de carácter económico o las medidas repetitivas de producción, quedando el conjunto final de 17 variables descritas anteriormente.

Posteriormente, se definió la variable objetivo para los modelos de clasificación: la Cosecha, construida como una variable categórica (buena o mala) a partir de los valores de rendimiento total (kg/ha). Este paso permitió transformar un valor numérico continuo

en una etiqueta clasificable, indispensable para el análisis con algoritmos supervisados como *J48* y *Random Forest*.

Asimismo, las variables categóricas originales olas de calor y heladas fueron transformadas en formato binario (0/1) con el fin de facilitar su procesamiento por parte de los algoritmos. Finalmente, se aplicó un proceso de normalización únicamente en el caso del algoritmo *Simple K-Means*, dado que este requiere que todas las variables estén en una misma escala para evitar sesgos en la formación de clústeres. Para los algoritmos de árboles de decisión y bosques aleatorios no se aplicó normalización, ya que estos no son sensibles a las diferencias de escala entre variables.

5.4.3 Selección de algoritmos

Para el análisis se seleccionaron tres algoritmos disponibles en el software Weka: *J48*, *Random Forest* y *Simple K-Means*. La elección de estos métodos respondió tanto a la naturaleza del dataset como a la intención de aplicar técnicas de clasificación y exploración de datos que permitieran identificar posibles relaciones entre las condiciones climáticas y el rendimiento vitivinícola.

El algoritmo *J48*, basado en árboles de decisión, se empleó por su capacidad de representar gráficamente las relaciones entre variables y de mostrar de manera intuitiva qué factores influyen más directamente en la clasificación de una cosecha como buena o mala.

Por su parte, el *Random Forest* se utilizó como un método más robusto de clasificación, ya que combina múltiples árboles de decisión y permite obtener una estimación más estable y precisa. Además, este algoritmo facilita identificar el conjunto de variables con mayor influencia en el modelo, lo cual resulta clave para determinar los factores climáticos más relevantes en la producción.

Finalmente, el algoritmo *Simple K-Means* se aplicó como técnica no supervisada con el objetivo de explorar patrones y agrupar los años en función de las similitudes entre sus condiciones climáticas y productivas. De esta manera, se buscó observar si era posible diferenciar grupos de cosechas (buenas o malas) a partir de la agrupación natural de las variables.

Dado el tamaño reducido del dataset (19 instancias y 17 variables), se realizaron ajustes en los parámetros de cada algoritmo para adaptarlos al contexto del estudio. Estos incluyeron, entre otros, la configuración de la profundidad máxima en los árboles, el número de iteraciones en los bosques aleatorios y la variación en el número de clústeres en K-Means, con el fin de evaluar la estabilidad y consistencia de los resultados.

5.4.4 Configuración en Weka

Con el conjunto de datos preparado, se procedió a implementar los experimentos en *Weka*. Desde el inicio se tomó la decisión de excluir las variables de rendimiento, ya que la clase objetivo (cosecha buena o mala) había sido construida a partir de dichos

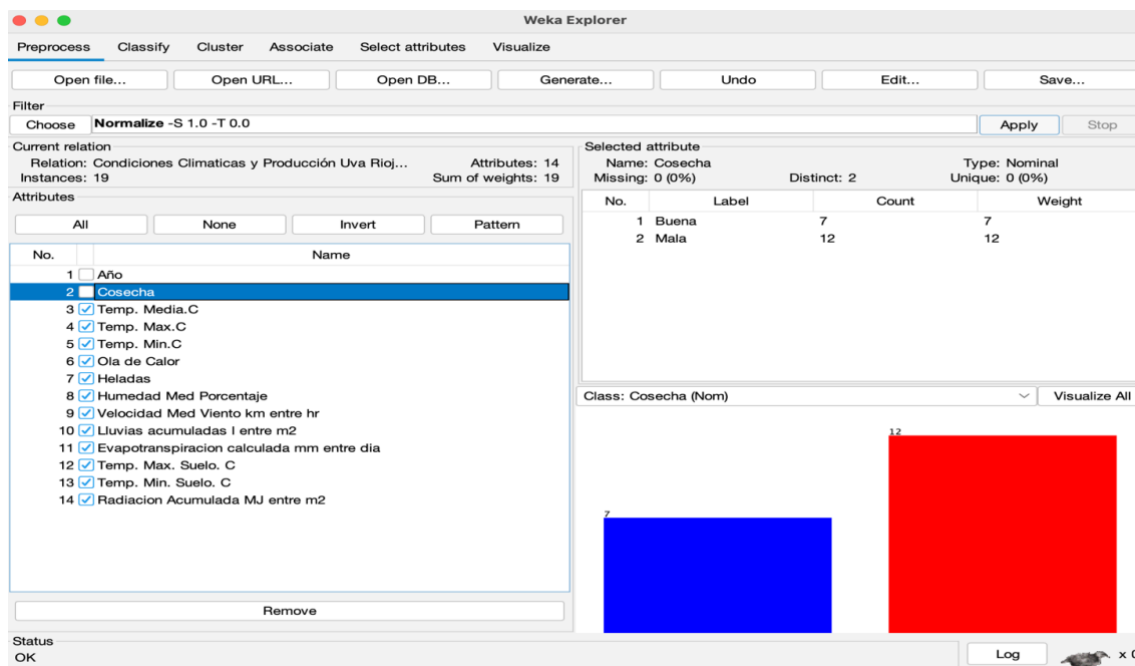
valores. Mantenerlas en el modelo suponía una relación directa que podía sesgar los resultados y dificultar el análisis del efecto real de las variables climáticas.

A partir de ahí, se plantearon dos escenarios de análisis para cada algoritmo:

1. **Uso de todas las variables disponibles**, ya sin incluir las de rendimiento.
2. **Uso de únicamente las seis variables más influyentes** según el evaluador *InfoGainAttributeEval* combinado con *Ranker*. Este segundo escenario buscaba comprobar si un número reducido de atributos era suficiente para mantener la capacidad predictiva y, al mismo tiempo, facilitar una interpretación más clara de los modelos.

Para la evaluación se utilizó la opción *Cross Validation* con 19 *folds*, lo que equivale a un procedimiento *leave-one-out*. En este enfoque, cada instancia del conjunto de datos actúa una vez como prueba mientras los 18 restantes se emplean para entrenamiento, garantizando así un mayor aprovechamiento de los datos en un dataset pequeño como el de este estudio.

Figura 17. Aplicación del filtro de normalización de los datos.



Fuente: Elaboración propia

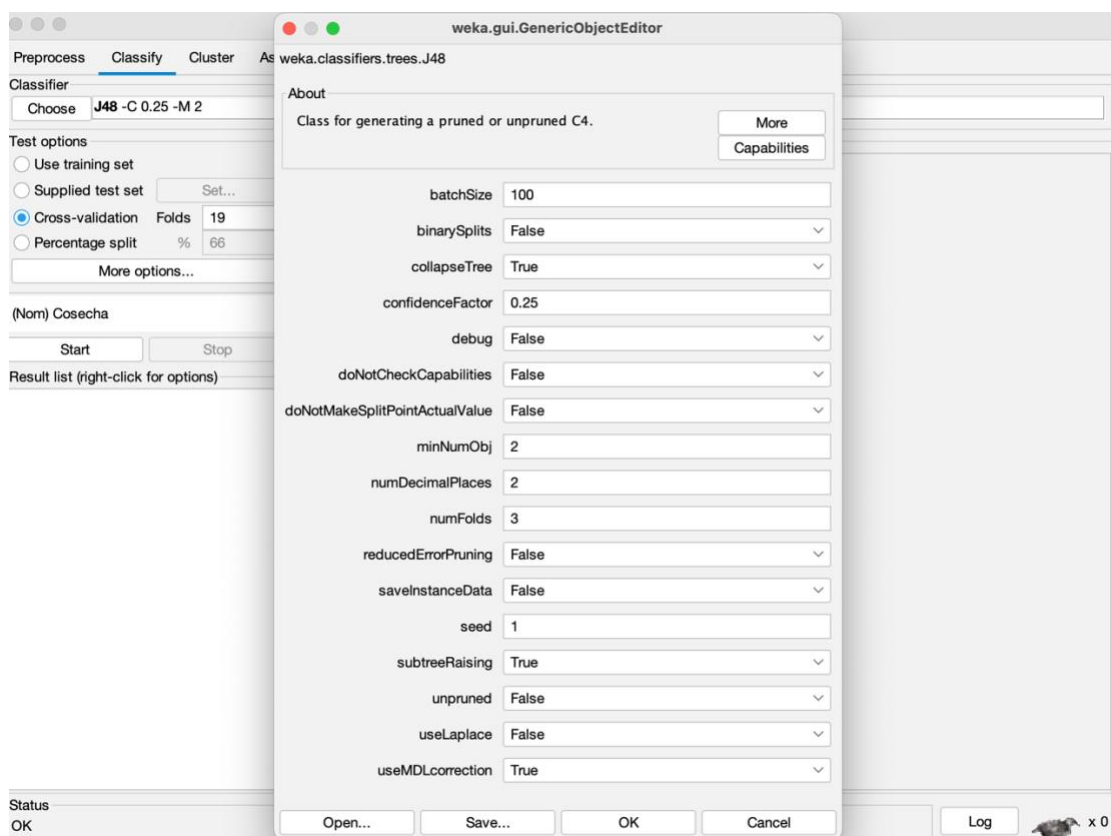
En cuanto a los parámetros específicos de cada algoritmo:

- **J48 (árbol de decisión)**: se ajustó el parámetro *confidence factor* con valores de 0.25 (por defecto), 0.15 y 0.10, con el objetivo de observar cómo la poda influía en la complejidad y profundidad del árbol. También se modificó el valor de *minNumObj* y se contrastó la opción *unpruned* en *true* y *false*, lo que permitió comparar entre árboles más simplificados y árboles sin poda.

- **Random Forest:** se probaron diferentes configuraciones en el número de árboles generados, la profundidad máxima y el número de atributos considerados en cada partición. Estos ajustes permitieron analizar el impacto de la complejidad y diversidad de los árboles en la capacidad del modelo para predecir la clase objetivo.
- **Simple K-Means:** se aplicó la normalización de los datos mediante el filtro *Normalize* en la pestaña *Preprocess* de Weka, excluyendo de la normalización las variables *Año* y *Cosecha*. Se ensayaron agrupaciones con $k = 2$ y $k = 3$, para explorar qué número de clústeres ofrecía una mejor representación de los patrones presentes en los datos. La variable “Cosecha” fue excluida de los atributos de entrada mediante la opción *ignore attributes*, y se empleó la evaluación *Classes to Clusters* para comparar la agrupación obtenida con la clasificación real de buena o mala cosecha.

De esta manera, el proceso de implementación en *Weka* no se limitó a aplicar directamente los algoritmos, sino que incluyó un conjunto de pruebas y ajustes técnicos orientados a mejorar la robustez del análisis y a maximizar la extracción de conocimiento a partir de un dataset reducido.

Figura 18. Pestaña de ajustes del algoritmo J48 en Weka.



Fuente: Elaboración propia

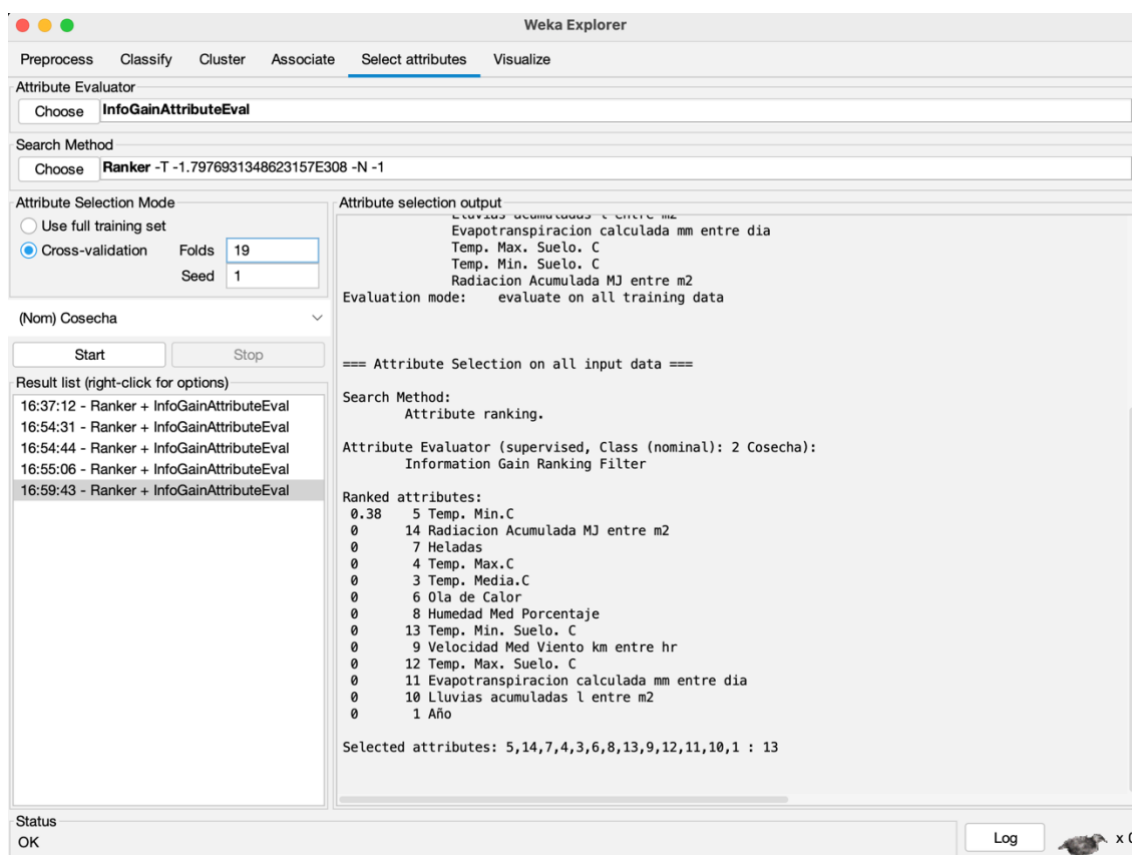
5.5 RESULTADOS Y DISCUSIÓN

5.5.1 Evaluación de atributos

La evaluación de atributos mediante *InfoGainAttributeEval* combinado con *Ranker* permitió identificar las variables que presentan una mayor influencia sobre la clasificación de la cosecha como buena o mala. Entre las variables más relevantes se encontraron, por orden de importancia: temperatura mínima, radiación acumulada, heladas, temperatura máxima, temperatura media y olas de calor, siendo la temperatura mínima la que mostró un impacto más significativo en el modelo.

Estos resultados sugieren que las condiciones de temperatura extrema, la radiación solar acumulada y la incidencia de heladas son factores determinantes en la producción de uva, mientras que variables como la temperatura media u otras condiciones climáticas presentan una influencia secundaria. Este análisis preliminar de atributos proporcionó una base sólida para la selección de variables en los modelos de clasificación y clustering, permitiendo enfocarse en los factores climáticos más relevantes y facilitando la interpretación de los resultados de los algoritmos aplicados posteriormente.

Figura 19. Ranking de atributos siendo de más influyentes a menos al modelo.



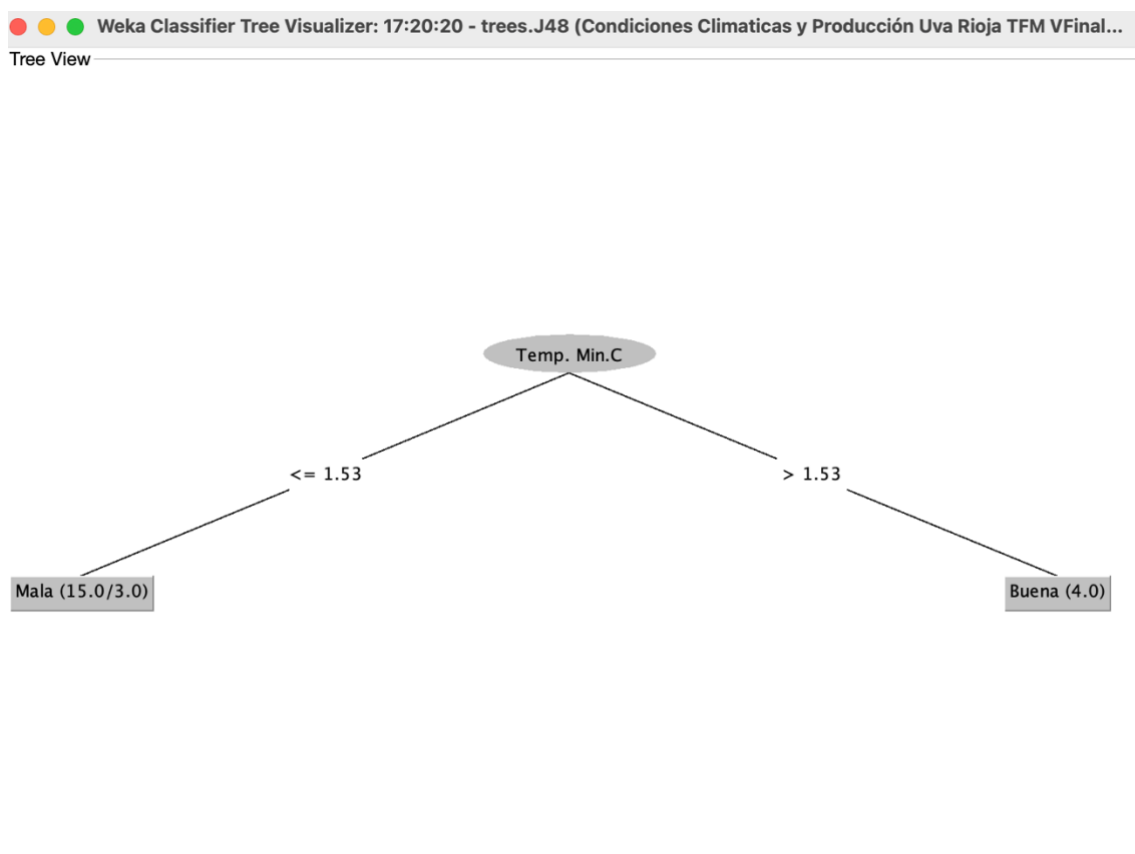
Fuente: Elaboración propia

5.5.2 J48

Se generaron dos modelos de árbol de decisión utilizando el algoritmo *J48*, con el objetivo de analizar cómo las variables climáticas influyen en la clasificación de la cosecha como buena o mala.

En el primer modelo se emplearon únicamente las seis variables más influyentes según la evaluación de atributos. Este árbol mostró una división basada en una sola variable: la temperatura mínima, identificada como la más influyente por *Weka*. El nodo de decisión indicaba que, si la temperatura mínima era mayor a 1.53 °C, se registraron 4 cosechas buenas, mientras que si era igual o inferior a 1.53 °C, se clasificaron 15 cosechas malas y 3 buenas. Aunque este modelo simplificado permitía identificar un patrón, su capacidad predictiva resultó limitada, ya que no reflejaba correctamente la proporción de cosechas buenas y malas en el dataset completo (12 malas y 7 buenas). Cabe destacar que para este modelo se aplicó una poda más estricta, considerando el tamaño reducido del conjunto de datos.

Figura 20. Árbol de decisión con las 6 variables más influyentes.

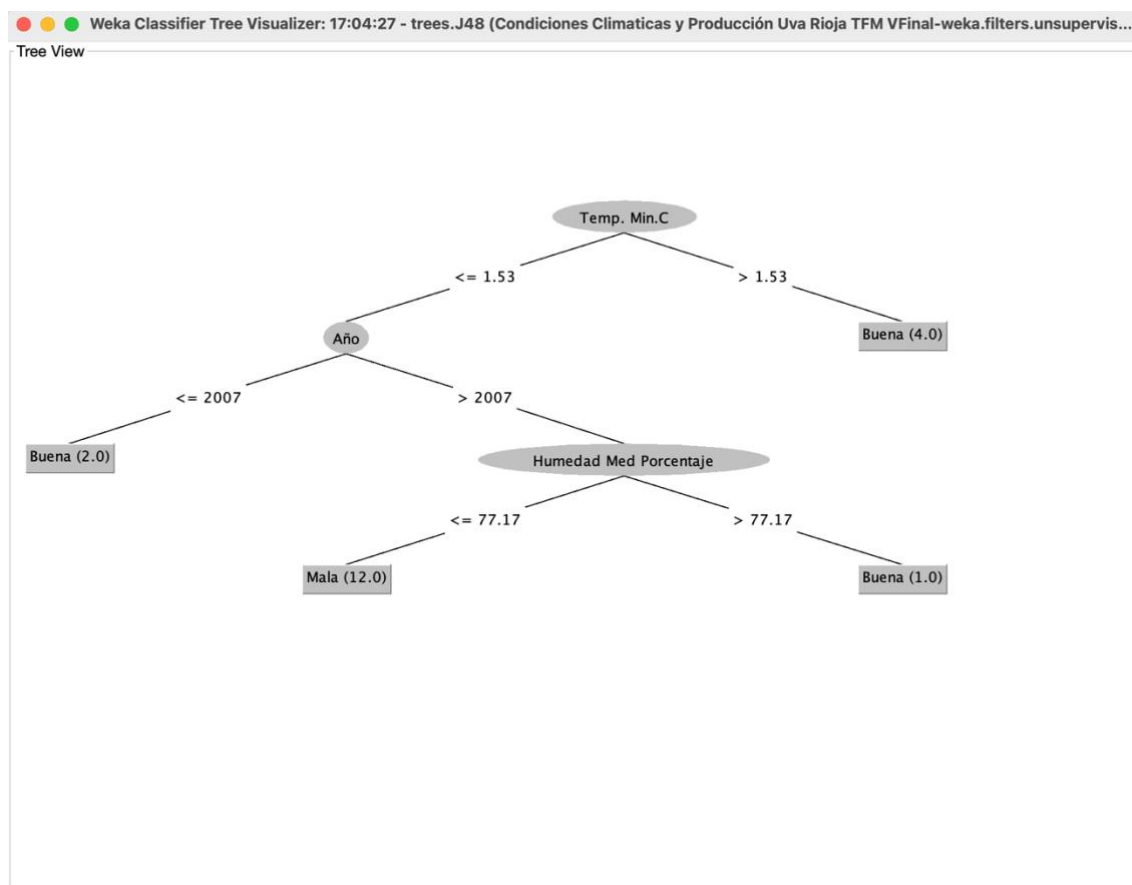


Fuente: Elaboración propia

En el segundo modelo se utilizaron todas las variables disponibles, excluyendo las de rendimiento, y se ajustó la poda de manera más moderada. El árbol generado mostró una estructura más compleja e interpretativa, comenzando con la temperatura mínima, nuevamente identificada como la variable más influyente y determinante. Las decisiones

siguientes se basaron en el año de la cosecha y en la humedad, respectivamente. La estructura del árbol indicaba que, si la temperatura mínima era mayor a 1.53 °C se registraron 4 cosechas buenas; si era igual o inferior, el siguiente nodo evaluaba el año (≤ 2007 , 2 cosechas buenas), y posteriormente la humedad (mayor a 77.17 % una buena cosecha, menor o igual a 77.17 % doce cosechas malas). Este segundo árbol permitió clasificar de manera más precisa la mayoría de las cosechas y facilitó una interpretación más completa de cómo interactúan las variables climáticas en la producción de uva.

Figura 21. Árbol de decisión con la totalidad de las variables.



Fuente: Elaboración propia

En conjunto, los resultados evidencian que la temperatura mínima es el factor climático más influyente, seguida por el año de cosecha y la humedad. Además, la inclusión de un mayor número de variables y un ajuste adecuado de la poda mejora la capacidad predictiva de *J48*, identificando patrones más realistas y relevantes para la toma de decisiones en el sector vitivinícola.

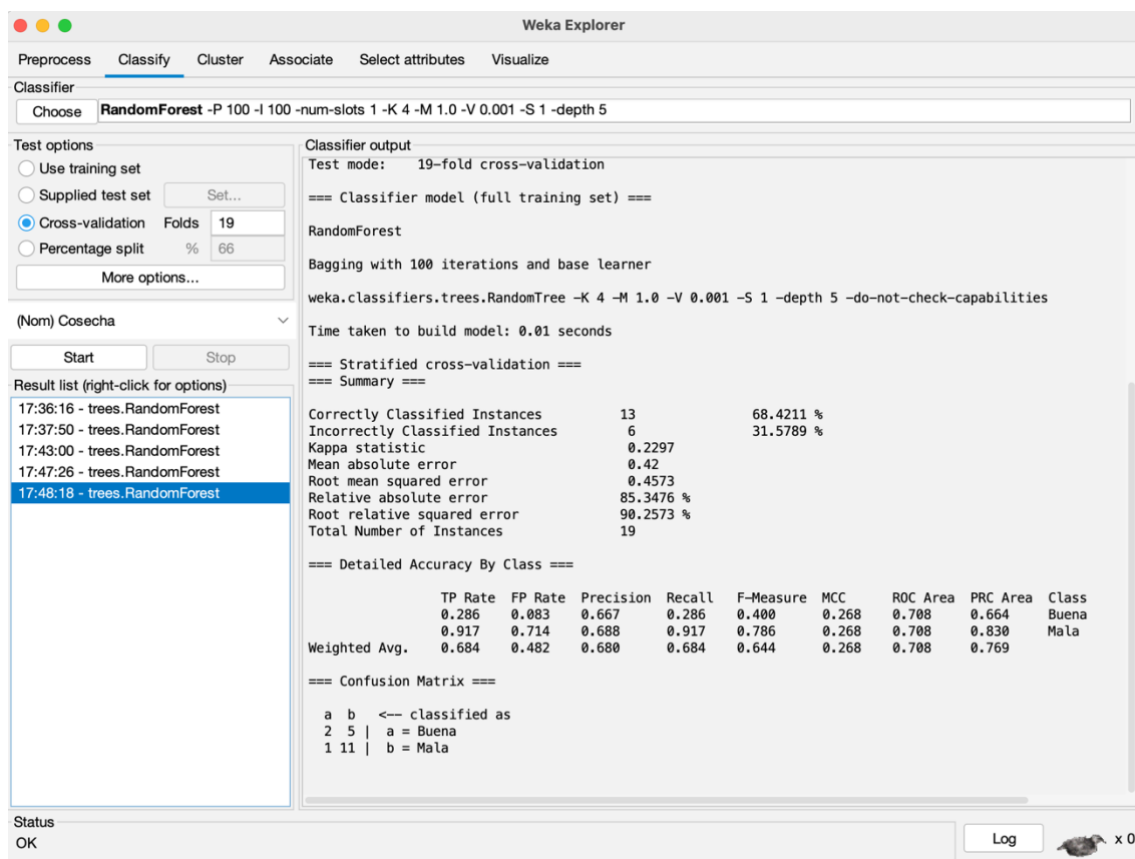
5.5.3 Random forest

Para el algoritmo *Random Forest* se realizaron dos pruebas principales con el objetivo de evaluar el comportamiento del modelo en la clasificación de cosechas buenas y malas, así como de identificar las variables más influyentes.

Prueba 1: Random Forest con la totalidad de las variables (excluyendo rendimiento) En este caso, el modelo obtuvo un 68.42% de instancias correctamente clasificadas (13 de 19), lo que equivale aproximadamente a dos de cada tres casos. El 31.57% restante (6 de 19) fueron clasificados de manera incorrecta. El valor del área bajo la curva ROC fue 0.708, lo cual se considera aceptable, ya que indica una capacidad razonable de discriminación entre cosechas buenas y malas.

En la matriz de confusión, se observa que para las cosechas buenas (clase a) únicamente 2 fueron clasificadas correctamente, mientras que 5 fueron clasificadas de forma errónea. Por el contrario, para las cosechas malas (clase b) el modelo tuvo un desempeño más sólido, con 11 clasificadas correctamente y solo 1 mal clasificada. Esto evidencia que el modelo tiene mayor facilidad para identificar cosechas malas, aunque presenta dificultades al clasificar las buenas.

Figura 22. Random Forest con la totalidad de las variables.



Fuente: Elaboración propia

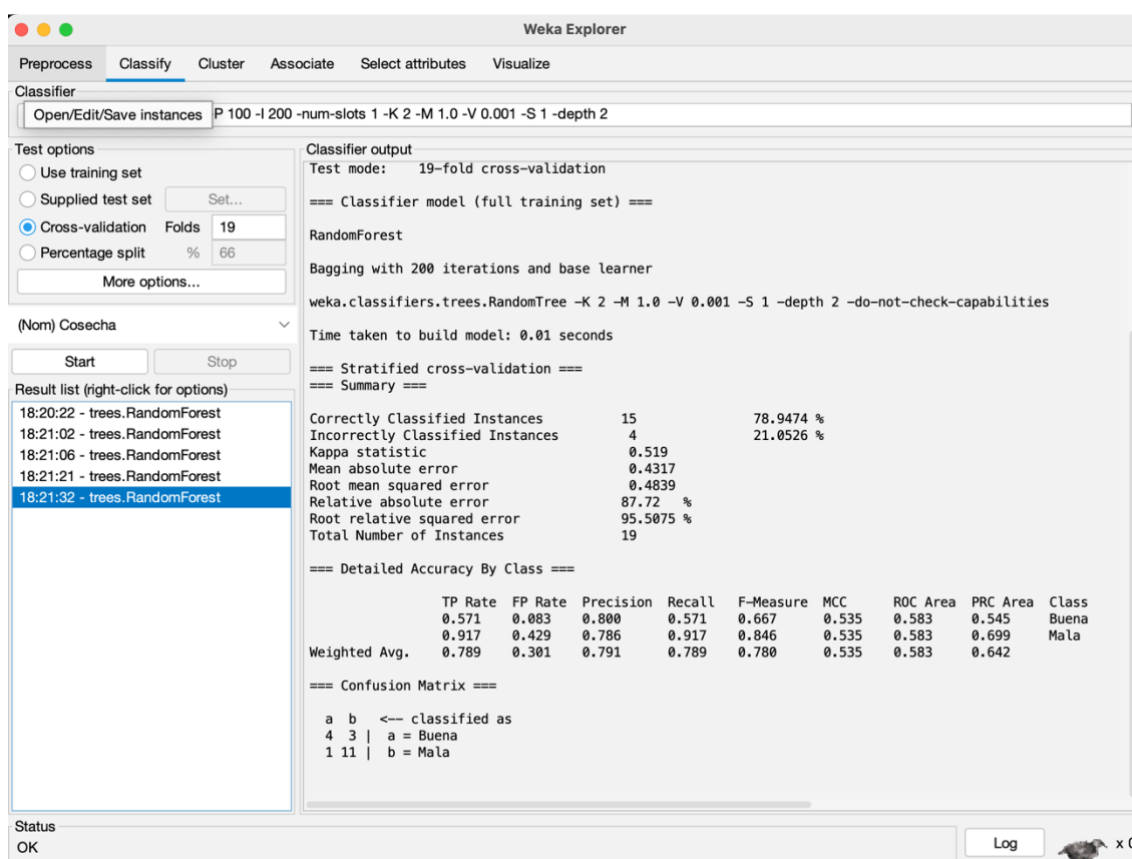
Al aplicar el método *WrapperSubsetEval* en combinación con *Random Forest*, se identificaron como variables más robustas la temperatura máxima, las heladas y la temperatura máxima del suelo, además de la variable objetivo cosecha. Esta selección refuerza la hipótesis de que las temperaturas extremas y los eventos de heladas son determinantes en el rendimiento agrícola.

Prueba 2: Random Forest con las seis variables más influyentes

En esta segunda prueba, el modelo alcanzó un 78.94% de instancias correctamente clasificadas (15 de 19), lo que representa una mejora respecto a la primera prueba, con un error del 21.05% (4 de 19). Sin embargo, el valor del área bajo la curva ROC fue 0.583, inferior al de la prueba anterior. Esto indica que, aunque el modelo aumenta su porcentaje de aciertos, presenta una menor capacidad de generalización y podría estar sobre ajustado al conjunto de datos.

La matriz de confusión refleja que en las cosechas buenas (clase a), el modelo logró 4 aciertos y 3 errores, mejorando respecto a la prueba anterior. Para las cosechas malas (clase b), el rendimiento se mantuvo estable con 11 aciertos y 1 error, consolidando la tendencia del modelo a clasificar mejor las malas cosechas.

Figura 23. Random Forest con las 6 variables más influyentes.



Fuente: Elaboración propia

El análisis mediante *WrapperSubsetEval* en esta prueba identificó como variables más robustas la temperatura máxima y la temperatura mínima, lo que confirma nuevamente la relevancia de las condiciones térmicas extremas como los factores climáticos más influyentes en el resultado de las cosechas.

En conjunto, los resultados sugieren que el algoritmo *Random Forest* logra un desempeño aceptable en la clasificación, con una mayor precisión en la identificación de cosechas malas. Asimismo, los distintos análisis de selección de atributos coinciden

en señalar a las temperaturas extremas (mínimas y máximas), junto con las heladas, como las variables más determinantes para predecir el éxito o fracaso de una cosecha.

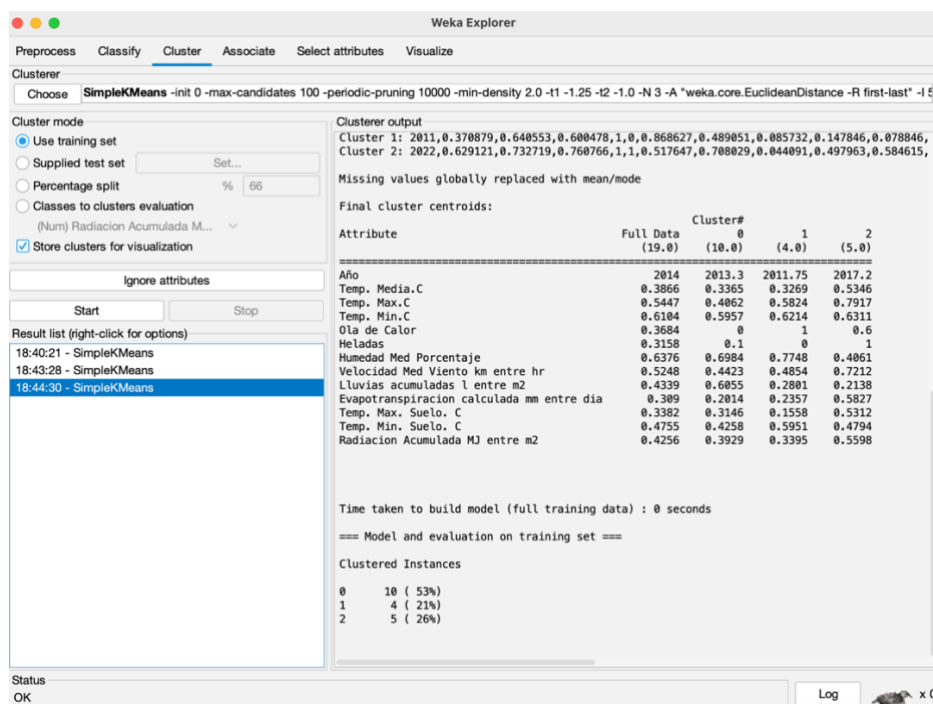
5.5.4 Simple K-Means

El algoritmo *Simple K-Means* es un método de aprendizaje no supervisado que agrupa las observaciones en función de su similitud, sin utilizar de manera explícita la clase objetivo (cosecha buena o mala). Su utilidad en este estudio radica en explorar si los datos presentan patrones naturales que puedan diferenciar entre años de cosechas favorables y desfavorables, lo cual aporta una validación adicional a los modelos supervisados.

Prueba 1: tres clusters con todas las variables, en esta primera prueba se incluyeron todas las variables normalizadas, y el algoritmo generó tres clústeres principales:

- **Cluster 0 (53% de los casos)**: agrupa la mayoría de los años, con valores intermedios en la mayoría de las variables. Representa un comportamiento climático medio, sin condiciones extremas.
- **Cluster 1 (21%)**: se caracteriza por presentar más eventos de ola de calor y heladas, con mayor variabilidad climática. Se podría asociar a años más extremos.
- **Cluster 2 (26%)**: concentra valores más altos en temperatura máxima, radiación acumulada y velocidad del viento, lo que sugiere años con condiciones más cálidas y exigentes.

Figura 24. Prueba 1 Simple K-Means con la totalidad de las variables.



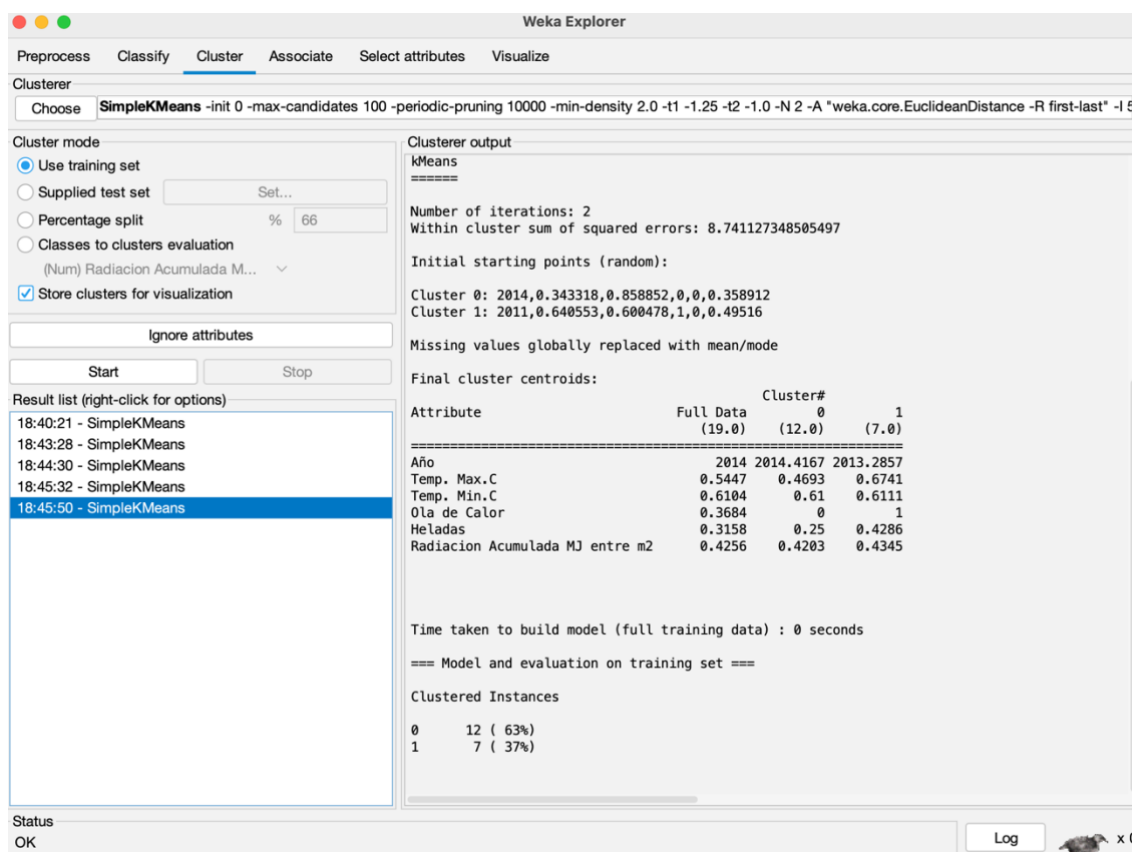
Fuente: Elaboración propia

El modelo detectó tres perfiles climáticos distintos en las cosechas, diferenciando entre años “promedio”, años “extremos” y años “más cálidos y radiantes”. Esto indica que el conjunto completo de variables permite capturar cierta complejidad y diversidad de escenarios agrícolas.

Prueba 2: dos clusters con las seis variables más influyentes, en la segunda prueba se redujo el análisis únicamente a las seis variables más relevantes (Temp. Máx., Temp. Mín., Ola de calor, Heladas, Radiación acumulada y Año), lo que llevó al algoritmo a agrupar los datos en dos clusters:

- **Cluster 0 (63%)**: caracterizado por valores más bajos en temperatura máxima y menor presencia de olas de calor y heladas. Se puede interpretar como años con condiciones más estables y favorables para la cosecha.
- **Cluster 1 (37%)**: concentra valores más altos de temperatura máxima y mayor frecuencia de olas de calor y heladas. Representa años con condiciones climáticas más extremas y, probablemente, menos favorables para la producción.

Figura 25. Prueba 2 Simple K-Means con las 6 variables más influyentes.



Fuente: Elaboración propia

Con menos variables el algoritmo generó agrupaciones más claras y compactas, separando de forma bastante evidente los años estables de los extremos.

La comparación entre ambas pruebas muestra que:

- **Con todas las variables (Prueba 1)**, los clusters capturan más diversidad de escenarios, pero la interpretación es menos clara.
- **Con las seis variables top (Prueba 2)**, los grupos son más nítidos y se alinean mejor con la hipótesis de que las temperaturas extremas y los eventos climáticos (heladas y olas de calor) son los principales determinantes de la cosecha.

En conjunto, *Simple K-Means* confirma que existen patrones diferenciables en los datos y que las temperaturas extremas y la variabilidad climática son las variables más influyentes en la segmentación natural de las cosechas.

Por último, aunque los modelos aplicados no otorgaron relevancia estadística significativa a la precipitación acumulada, el análisis exploratorio del dataset muestra que los años con mayores precipitaciones tienden a coincidir con cosechas catalogadas como buenas. Esta discrepancia sugiere que el efecto de la precipitación podría estar mediado por otras variables (como la temperatura mínima o la humedad) o que el tamaño reducido del dataset limita la capacidad de los modelos para reflejar esta relación.

6. CONCLUSIÓN

El presente estudio ha demostrado que la ciencia de datos puede aplicarse de manera efectiva al sector vitivinícola para analizar la influencia de los factores climáticos en la producción de uva. Los resultados obtenidos a través de los algoritmos *J48*, *Random Forest* y *Simple K-Means*, junto con el análisis exploratorio del dataset, muestran que la temperatura mínima es la variable más determinante, seguida por la humedad, las temperaturas extremas y las precipitaciones acumuladas, confirmando su relevancia en la diferenciación entre cosechas buenas y malas.

Desde un punto de vista técnico, el árbol de decisión *J48* ofreció interpretaciones claras de las relaciones entre las variables, lo que lo hace especialmente útil para comprender patrones de forma intuitiva. *Random Forest* aportó robustez, clasificando con buen desempeño las cosechas malas y mejorando el entendimiento global al identificar las variables con mayor impacto, aunque con cierta tendencia al sobreajuste. Por otra parte, *Simple K-Means* permitió agrupar los años en patrones climáticos coherentes, funcionando como una herramienta exploratoria que reforzó hallazgos previos y facilitó la predicción.

En relación con los objetivos planteados, este estudio ha permitido aplicar la ciencia de datos en la agricultura vitivinícola, demostrando el potencial de estas técnicas en el sector. El objetivo general de aplicar ciencia de datos al sector se ha cumplido, evidenciando cómo estas herramientas pueden aportar valor en la gestión agrícola. Esta investigación ha permitido visibilizar los retos climáticos que afectan a las cosechas de uva de la región y validar la utilidad práctica de diferentes técnicas de análisis y

predicción, mostrando que su integración en el sector vitivinícola puede favorecer decisiones más informadas y mejorar la eficacia y eficiencia de las cosechas.

Desde una perspectiva práctica, estos hallazgos ofrecen valor inmediato para agricultores y técnicos, ya que permiten priorizar la vigilancia de factores críticos como las temperaturas mínimas y las heladas, orientando la planificación de labores agrícolas, las decisiones de protección del viñedo y la gestión del riesgo. Aunque el control de las condiciones climáticas es limitado, la capacidad de anticiparse a sus efectos mediante modelos predictivos representa una ventaja competitiva en la gestión del viñedo.

Si bien el trabajo presenta ciertas limitaciones, estas deben entenderse como oportunidades para seguir avanzando en la investigación. El reducido tamaño del dataset (19 instancias) condicionó la robustez de los modelos, pero aun así permitió obtener resultados que abren nuevas preguntas y líneas de análisis. El hecho de que algunas variables, como la precipitación acumulada o la radiación solar, no mostraran la relevancia esperada constituye un hallazgo interesante que invita a explorar con mayor profundidad las razones de su limitada influencia. Además, la falta de datos de producción más detallados —al estar disponibles únicamente registros anuales y no mensuales o semanales— limitó la posibilidad de establecer correlaciones más precisas, lo que señala un camino claro para futuras investigaciones.

A futuro, sería recomendable ampliar la base de datos histórica, incorporar registros con mayor granularidad temporal y añadir variables relacionadas con el manejo agrícola, los tipos de uva cultivados y factores económicos. De esta manera, sería posible analizar no solo cómo el clima afecta a la producción, sino también si existe una relación directa entre los años de buena cosecha y el comportamiento de los precios o el valor del mercado. Igualmente, se debería experimentar con distintos algoritmos y enfoques de aprendizaje automático para contrastar resultados y extraer conclusiones más sólidas. Esto permitiría generar modelos predictivos más precisos y generalizables, contribuyendo a una planificación más eficiente y resiliente del sector vitivinícola.

En definitiva, el estudio confirma que la aplicación de la ciencia de datos a la viticultura ofrece un camino prometedor para anticipar el impacto del clima y optimizar la gestión agrícola. Si bien los resultados son preliminares, sientan una base sólida para futuras investigaciones y proporcionan un marco práctico de utilidad para la toma de decisiones en un contexto de creciente variabilidad climática.

7. REFERENCIAS

- Alfred, R., Obit, J. H., Chin, C. P.-Y., y Haviluddin, H. (29, Marzo 2021). *Towards Paddy Rice Smart Farming: A Review on Big Data, Machine Learning, and Rice Production Tasks*. Recuperado de IEEE Xplore: <https://ieeexplore.ieee.org/document/9389541>
- Antonis Tzounis, N. K. (2017). Internet of Things in agriculture, recent advances and future challenges, Biosystems Engineering. *ScienceDirect, Volumen 164*, 31-48.
- Alavanja, M. (2009). Pesticides Use and Exposure Extensive Worldwide. *National Institutes of Health*, 303-309.
- Bairoch, P. (1979). La agricultura de la revolución industrial 1700-1914. En C. M. Cipolla, *Historia Económica En Europa. La Revolución Industrial* (Vol. 3). Barcelona, España: Editorial Ariel.
- Carey, J. (2023). Unearthing the origins of agriculture. *Proceedings of the National Academy of Science of the United States of America*, 1-2.
- Cravero, A., y Sepúlveda, S. (2021). Use and Adaptations of Machine Learning in Big Data—Applications in Real Cases in Agriculture. *Electronics*, 4-5.
- Demográfico, M. (2025, Junio 11). *Impacto de los nitratos y pesticidas en el uso y calidad de las aguas*. Recuperado de Ministerio para la transición ecológica y el reto demográfico: <https://goo.su/GaVd0x>
- Donaldson, D., Kiely, T., y Grube, A. (1999). *Pesticide's industry sales and usage 1998-1999 market estimates*. Washington DC: US Environmental Protection Agency.
- Empresas, V. (Director). (2022). *Caso de éxito Vodafone | SAT Marigómez: cosechar todos los datos para mejorar los cultivos* [Motion Picture].
- Estadística, I. N. (2022). *Estadística estructural de empresas: sector industrial Año 2022*. Recuperado de INE: <https://www.ine.es/dyngs/Prensa/EEESI2022.htm>
- Europea, C. (2025, Junio 11). *Nitrates*. Recuperado de Comisión Europea: https://environment.ec.europa.eu/topics/water/nitrates_en
- Europea, C. d. (2025, Enero 1). *Cómo la invasión rusa de Ucrania ha agravado la crisis alimentaria mundial*. Recuperado de Consejo Europeo: <https://goo.su/YYuea>
- Group, L. C. (2022, Septiembre 22). *Agricultura de precisión: Qué es, beneficios y cómo aplicarla*. Recuperado de London Consulting Group: <https://londoncg.com/blog/agricultura-de-precision>
- Harlander, S. K. (2002). The Evolution of Modern Agriculture and Its Future with Biotechnology. *Journal of the American College of Nutrition*, 1-5.
- ICANE. (2024). *Perfil del Turista en Cantabria*. Recuperado de ICANE: <https://goo.su/M8tWh5C>
- Jeyarantnam, J. (1990). Acute pesticide poisoning: a major global health problem. *World Health Stat Q*, 139-144.
- Kamilaris, A. (2018). Deep learning in agriculture: A survey. *Elsevier*, 70-90.

- Khanna, A., y Kaur, S. (2019). Evolution of Internet of Things (IoT) and its significant impact in the field of Precision Agriculture. *Elsevier*, 218-231.
- Larson, C. S. (2023, Enero 17). The past 12,000 years of behavior, adaptation, population, and evolution shaped who we are today. *Proceedings of the National Academy of Science of the United States of America*, 120 (4), 2-3.
- Marr, B. (2025, Agosto 13). *The Incredible Ways The 4th Industrial Revolution And AI Are Changing Winemaking*. Recuperado de Bernard Marr: <https://goo.su/ilvLjS8>
- Microsoft. (2023). *¿Qué es la ciencia de datos?* Obtenido de Microsoft Azure: <https://goo.su/6sNxnc>
- Ministerio de Agricultura, P. y. (2023). *Informe anual de la industria alimentaria española periodo 2022 – 2023*. Recuperado de Ministerio de Agricultura, Pesca y Alimentación: <https://goo.su/aPkWzC>
- Morota, G., Ventura, R. V., Silva, F. F., Koyama, M., y Fernando, S. C. (2018). *Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture*. Oxford University Press, 1540-1550.
- Mueller, J. P., y Massaron, L. (2019). *Python for Data Science for dummies*. Hoboken: John Wiley & Sons, Inc.
- Nations, U. (2015, Julio 29). *UN projects world population to reach 8.5 billion by 2030, driven by growth in developing countries*. Recuperado de United Nations: <https://news.un.org/en/story/2015/07/505352>
- Networks, I. (2021, Septiembre 13). *Estos son los lenguajes de programación top en 2021 según el IEEE*. Recuperado de IMDEA Networks: <https://goo.su/Jeitl>
- Paul, P., Das, N., Hossain, R., y Saavedra, R. (2024). Agricultural 4.0: The Root for Agricultural Transformation— A Scientific Review. *International Journal of Information Science and Computing*, 1-12.
- Pulido, J., y Bocco, G. (2003). The traditional farming system of a Mexican indigenous community: the case of Nuevo San Juan Parangaricutiro, Michoacán, Mexico. *Geoderma. Elsevier*, 249-265.
- Rehman, A., Jingdong, L., Khatoon, R., Hussain, I., y Iqbal, M. S. (2017). Modern Agricultural Technology Adoption its Importance, Role and Usage for the Improvement of Agriculture. *Life Science Journal*, 70-74.
- Santandreu, A., y Rea, O. (2014). La gestión del conocimiento orientada al aprendizaje como motor de cambios en Agricultura Urbana: reconectando personas, sistemas sociales y sistemas ecológicos. . *Universidad de Vigo*, 10-11.
- Singh, R., y Singh, G. S. (2017). Traditional agriculture: a climate-smart approach for sustainable food production. *Springer Nature Link*, 2, 296–316.
- Siregar, R. R., Wahjuni, S., Seminar, K. B., y Santosa, E. (2022). Vertical Farming Perspectives in Support of Precision Agriculture Using Artificial Intelligence: A Review. *MDPI journals*, 1-2.
- Statista. (2024). *Ranking de los principales exportadores de vino del mundo en función del valor de las exportaciones en 2024*. Recuperado de Statista: <https://goo.su/MS9oiVj>

- Synox. (2025, Agosto 4). *Smart farming: what are the economic challenges for the agriculture of the future?* Recuperado de Synox Innovate Together: <https://www.synox.io/en/cat-smart-agriculture/smart-farming-agriculture-future/>
- Taneja, M., Byabazaire, J., Jalodia, N., Davy, A., Olariu, C., y Malone, P. (2020). Machine learning based fog computing assisted data-driven approach for early lameness detection in dairy cattle. *Elsevier*, 105-286.
- Thrall, P. H., Oakeshott, J. G., Fitt, G., Southerton, S., Burdon, J. J., Sheppard, A., . . . Denison, R. F. (2010). Evolution in agriculture: the application of evolutionary approaches to the management of biotic interactions in agro-ecosystems. *Evolutionary Applications*, 1-16.
- Vidal, J. (2014, Abril 7). *Consideraciones procesos ETL en entornos Big Data: Caso Hadoop*. Recuperado de Data Prix: <https://goo.su/t3Yzj71>
- Zambrano, J. C., Rodríguez, E., y Pire, R. (2002). Crecimiento, producción y extracción de NPK en plantas de pepino (*Cucumis sativus* L.) ante diferentes dosis de fertilizante. *Research Gate*, 85-88.