



Universidad de Valladolid

MODELIZACIÓN PREDICTIVA DEL RENDIMIENTO EN MAÍZ MEDIANTE ÍNDICES DE VEGETACIÓN

TRABAJO FIN DE MÁSTER DE INGENIERÍA AGRONÓMICA



Escuela Técnica Superior
de Ingenierías Agrarias **Palencia**

CURSO 2022/2023

ALUMNO: JOSÉ RODRÍGUEZ FERNÁNDEZ

Tutor: José Luis Marcos Robles

5.1.3.	Datos de rendimientos históricos	32
5.2.	Modelado predictivo del rendimiento	33
5.2.1.	Análisis de componentes principales (PCA en inglés)	33
5.2.2.	Árboles de decisiones.....	35
5.2.3.	Bosques de decisiones o modelos ensamblados	35
5.2.4.	Predicción con redes neuronales	36
5.2.5.	Evaluación de los modelos	37
6.	Conclusiones	38
7.	Bibliografía	39

1. Resumen

La provincia de León cuenta con una superficie de 135000 hectáreas de regadío y, en 2022 se sembró más de la mitad de maíz, unas 75000 ha, lo que da una idea de la gran importancia que tiene este cultivo en la zona.

Los modelos de simulación son una herramienta con la que se pueden estimar diferentes variables como la cantidad de insumos a utilizar, la producción de cada área cultivada o, simplemente prever tendencias a medio o largo plazo.

En este caso, puesto que trabajo en una de las empresas más importantes de semillas de maíz para la zona (*Corteva Agriscience* y su plataforma de semillas *Pioneer*), me ha parecido interesante utilizar algunas de las herramientas con las que asesoramos a nuestros agricultores para hacer una modelización predictiva del rendimiento de maíz en una parcela conocida. Para ello, también se van a utilizar otros instrumentos que forman parte de la llamada Agricultura Digital, tales como índices de vegetación, mapas de rendimiento, aplicaciones agrícolas... y otra serie de herramientas utilizadas en inteligencia artificial y Machine Learning, tan de moda en la actualidad.

2. Introducción

En las últimas décadas, se ha observado cómo las herramientas de localización y posicionamiento espacial, junto a las tecnologías de la información, han facilitado el estudio de la variabilidad espaciotemporal de los cultivos, lo que ha supuesto una revolución tecnológica para la agricultura que conocíamos. Para ello, la Agricultura de Precisión utiliza tecnologías y técnicas para el análisis de variabilidad, permitiendo una optimización en la utilización de inputs, y convirtiéndola así, en un tipo de agricultura mucho más sostenible y respetuosa con el medio ambiente (Benos et al., 2021).

No obstante, la utilización de todas estas herramientas tiene un coste para el agricultor, lo cual conlleva a preguntarse si todo este ahorro de insumos será capaz de amortizar dicha inversión o, al menos, si su explotación es lo suficientemente grande como para que la utilización de dichas herramientas le compense. Por este motivo, la adopción de todas estas técnicas está siendo más lenta de lo que debería.

Uno de los factores que más se ven influenciados por dicha variabilidad es el rendimiento final del cultivo, lo que lo convierte en un factor clave tanto para el productor en el sentido más estricto, como para el resto de la cadena alimentaria en un sentido más amplio. En los últimos años la utilización de modelos matemáticos estadísticos basados en los datos e informaciones pasadas capaces de modelizar el comportamiento están siendo de gran utilidad para predecir rendimientos futuros.

En este trabajo, se ha colaborado con la división de semillas *Pioneer* de la compañía *Corteva Agriscience* y su departamento Digital para generar un modelo de predicción del rendimiento.

2.1. Factores que causan la variabilidad

Principalmente, existen dos factores fundamentales que causan la variabilidad: espacio y tiempo.

En cuanto a la variabilidad espacial, entendida como la diferenciación productiva de diferentes zonas o ambientes presentes en el entorno. En el caso de la agricultura, serán aquellos que van ligados a la fertilidad del suelo de cada zona de estudio, compuestos por estructura, textura y capacidad de intercambio catiónico, básicamente. Como bien sabemos, ninguna parcela es homogénea, por lo que la elaboración de un mapeo del suelo junto con una interpretación agronómica adecuada puede dar a conocer dicha variabilidad. Evidentemente, el suelo es el punto de partida para el estudio de la variabilidad espacial, pero la agricultura, es un conjunto de interacciones entre suelo, clima y manejo, por lo que durante todo el crecimiento del cultivo intervienen toda esta serie de factores. Por este motivo, también parece útil la utilización de herramientas digitales durante todo el proceso, lo que permitirá la detección de plagas, enfermedades, requerimientos de cultivo y estado fenológico del mismo entre otros.

El estudio de la variabilidad, como factor diferenciador del rendimiento final, debe ser riguroso y suponer una ventaja frente a un manejo uniforme, ya sea en el aumento del rendimiento o en la reducción de los insumos a lo largo de las campañas.

El rendimiento predictivo de los cultivos varía espacialmente a lo largo de su desarrollo, pero es muy necesario tener presente la variación temporal, es decir, la variabilidad de ese mismo cultivo en campañas pasadas. Por ello, de cuanto mayor cantidad de datos pasados dispongamos, con mayor exactitud se podrá ejercer la predicción.

Esta variabilidad temporal pretende estudiar la interacción del mayor número de factores que la causan: plantas, animales, insectos, suelo, clima, agroquímicos y agua, con el fin de realizar una modelización lo más realista y precisa posible que sea capaz de anticipar eventos futuros.

Por tanto, se puede decir que conocer la variabilidad espaciotemporal de una parcela permite anticipar posibles problemas futuros y, por consiguiente, se podrá actuar preventivamente ante los mismos.

2.2. Transformación Digital en la agricultura

La transformación Digital en la agricultura ha sido capaz de llegar más allá de la experiencia e intuición de los agricultores, ya que pone en valor procesos y técnicas basadas en datos, permitiendo un aumento de la eficiencia, sostenibilidad y producción de los cultivos. Este proceso de incorporación de la tecnología es paulatino y está en constante desarrollo, por lo que podemos hablar de diferentes etapas desde el inicio de la agricultura hasta nuestros días o, incluso a un futuro más o menos cercano. En este proceso evolutivo se pueden diferenciar las siguientes fases (Valero, 2019):

- Agricultura 0.0: Fue la que se dio en los inicios de la civilización y se prolongó en un periodo largo, desde el Neolítico hasta la Edad Moderna. Estaba caracterizada principalmente por el uso de potencia humana o animal para la ejecución de las labores agrícolas y era una agricultura de subsistencia.
- Agricultura 1.0: Es la agricultura industrial, mecanizada, en la que se ha extendido el uso del tractor, las cosechadoras, los aperos y máquinas en general. Ya no se trata de una agricultura de subsistencia y necesita una externalización de inputs. Este tipo de agricultura permite un aumento de la producción respecto de la anterior, pero utiliza unos medios de producción homogéneos en toda la superficie de grandes extensiones.
- Agricultura 2.0: En ella las máquinas incorporan elementos electrónicos, que permiten un control preciso del trabajo. Es la conocida Agricultura de Precisión, con el uso generalizado del GPS, autoguiado de tractores, sistemas ISOBUS, etc. También aparece la utilización de satélites y/o drones con el fin de recuperar una información más detallada de las parcelas, expresándola en mapas, los cuales permiten dar tratamientos diferenciales a cada variabilidad de la finca.
- Agricultura 3.0: Es la que se conoce como Agricultura Digital y es tal y como se conoce la agricultura actualmente. En ella, existen cuadernos digitales, se realizan gran cantidad de estudios y aparece la importancia del dato o data Set (Cuántos más datos se dispongan, mayor será la calidad del análisis en cuestión).
- Agricultura 4.0: Se trata de una red permanente de intercambio de datos. La aparición de ecosistemas digitales inteligentes podría permitir que todos los sistemas estuvieran conectados con el fin de agilizar procesos y anticipar cualquier posible amenaza o beneficio.

2.3. Teledetección y nuevos sistemas de motorización

La creciente población mundial y la mayor concepción medioambiental de dicha población generan la controversia de la necesidad de aumentar la producción de alimentos con una disminución del impacto medioambiental. Evidentemente, para aumentar la producción solo existen dos maneras posibles, o bien se aumenta la superficie cultivada- lo que entraría en contraposición con el aspecto medioambiental- o se aumenta el rendimiento por unidad de superficie cultivada. En este aspecto, el seguimiento de la evolución temporal y la distribución

espacial de los cultivos proporcionan información relevante para un mejor manejo agronómico, pudiendo determinar la homogeneidad de este.

Una de las herramientas que nos permite analizar la homogeneidad-variabilidad de las parcelas agrícolas es la teledetección y, por consiguiente, el poder realizar una adecuada gestión sobre las mismas.

La teledetección combinada con herramientas de análisis automatizado de la información permite mejorar los procesos agrícolas y ayudar en la toma de decisiones, lo que jugará un papel crucial en la agricultura de los próximos años. Los satélites, aviones y drones donde se instalan los sensores multiespectrales son capaces de medir longitudes de onda más amplias que las del visible, lo que permite observar aspectos no visibles por el ojo humano.

El funcionamiento viene determinado por una interacción entre la superficie de la tierra y el sensor que registra la energía solar (radiación visible o infrarroja) reflejada por la superficie (conocida como reflectancia), o bien la reflejada por un haz de energía artificial (ej. radar), o la radiación emitida por la superficie (radiación térmica y microondas) (Teledetección, s. f.). La señal es transmitida a través de la atmósfera y capturada por los sensores, y la información es puesta finalmente a disposición para su posterior procesamiento en formato digital.

A nivel agrícola, la teledetección permite una observación no destructiva de los cultivos a gran escala y de manera remota. Gracias a la combinación de las diferentes bandas del espectro electromagnético se pueden elaborar una serie de índices que nos permitan hacer un seguimiento del estado fenológico de los cultivos, detectar patrones de crecimiento, anomalías como plagas y enfermedades, requerimientos de agua y demás factores que determinarán su producción.

Las plataformas de observación utilizadas como métodos de detección remotos son aviones, drones o satélites que orbitan alrededor de la tierra. Existen sensores pasivos, como los que se encuentran a bordo del satélite Sentinel 2 que reciben de forma natural las ondas electromagnéticas transmitidas por el objeto observado. En cambio, los sensores activos generan su propia radiación electromagnética, transmitiéndose desde el emisor hasta el objeto observado y de vuelta al sensor (Radar, LiDAR y microondas). Los sensores recogen los datos haciendo pasar la energía reflejada a través de filtros que la separan en bandas espectrales discretas (Bezares Sanfelip, F., 2019)

Las diferentes regiones en las que se divide el espectro electromagnético tienen diferentes aplicaciones, determinadas por sus capacidades. Usando combinaciones numéricas simples de la información espectral medida a diferentes longitudes de onda, principalmente en las regiones visibles e infrarrojas cercanas del espectro, es posible extraer información sobre el estado y la estructura de la vegetación, minimizando la perturbación causada por el suelo y las condiciones atmosféricas. Estas combinaciones se denominan índices de vegetación normalizados, que caracterizan el cultivo en sus diferentes estados fenológicos y permiten una actuación preventiva ante un determinado estrés o carencia, ya que cuantifican el verdor, densidad y salud de la vegetación en cada píxel de una imagen de satélite. A partir del reflejo de ciertos rangos del espectro electromagnético, se determina también el vigor del cultivo en cada píxel, por lo que permite conocer su estado actual, que luego podrá compararse con otra imagen temporal para observar su evolución en el tiempo.

Para su medición existen gran cantidad de índices y están en constante evolución, pero en la actualidad, los más utilizados para el seguimiento de los cultivos son el famoso NDVI y el GNDVI. Existen también otra serie de índices más adecuados para otros usos, como pueden ser (Alonso, 2023):

- EVI: Enhanced vegetation index (EVI) o Índice de Vegetación Mejorado, calculado por la diferencia de radiancia entre las bandas del Azul y Rojo y nos permite monitorizar el estado de la vegetación en caso de altas densidades de biomasa, como bosques tropicales.
- SAVI: Soil Adjusted Vegetation Index o Índice de Vegetación Ajustado al Suelo fue diseñado para minimizar las influencias del brillo del suelo. Se trata de un índice más preciso en la medición de cultivos jóvenes o áreas más áridas.
- NBR: Normalized Burn Ratio o Índice de Área Quemada, diseñado para resaltar áreas quemadas en grandes zonas de incendio. La fórmula es similar a NDVI, excepto que combina el uso de longitudes de onda de infrarrojo cercano (NIR) e infrarrojo de onda corta (SWIR).

Evidentemente, existen muchos más índices en la actualidad y su desarrollo está en constante crecimiento, ya que gran cantidad de empresas, universidades y científicos del mundo están continuamente estudiando nuevos índices de vegetación y/o modificando los ya existentes para hacerlos más precisos.

2.4. Justificación del trabajo

La división de semillas *Pioneer* de la empresa *Corteva Agriscience* dispone de un servicio de seguimiento del cultivo del maíz mediante imágenes por satélite periódicas que dan soporte al servicio agronómico de dicha plataforma. Normalmente, este servicio utiliza estas imágenes como asesoramiento a los agricultores en recomendaciones de riego, predicción de estados fenológicos determinados y detección de plagas. Sin embargo, con objeto de ampliar dicho servicio, en este trabajo se propone realizar una predicción precisa del rendimiento de las parcelas de los agricultores.

Este trabajo, ha tratado de complementar el seguimiento basado en índices espectrales de cultivo con otras herramientas como el radar a bordo de satélites, así como de obtener mediante técnicas de aprendizaje automático la predicción del rendimiento de una parcela característica de una zona productora de maíz. Obtener esta información permite a *Corteva* mantener un equipo mucho más informado, capaz de avanzar movimientos en el mercado, ayudando al cliente en la toma de decisiones a lo largo de la campaña y posicionarse de forma estratégica en el mercado. El empleo de plataformas como *Granular link* y *One Soil* para la obtención de los índices vegetativos y, herramientas de Machine Learning como *BigML* para la generación de modelos de predicción se justifica por la potencia de éstas y la novedad en su uso.

Además, se disponen de los rendimientos reales de la parcela, que han sido proporcionados por el agricultor, así como el mapeo de rendimiento de la última campaña. De esta manera, se va a comparar la eficacia del modelo predictivo con la producción real obtenida y ver así, cuan de fiable y preciso es.

3. Objetivos

3.1. Objetivo general

El objetivo general de este trabajo es monitorizar y modelizar el rendimiento de maíz en una determinada parcela sita en el municipio de Jabares de los Oteros, de la provincia de León. Dicho objetivo se va a llevar a cabo con los datos históricos más recientes, es decir, se tomarán como referencia los años correspondientes a las campañas de 2019, 2020, 2021 y 2022, con el fin de dar apoyo a su vez al servicio agronómico de *Corteva Agriscience*.

3.2. Objetivos específicos

Para la consecución del objetivo general se han de cumplir una serie de objetivos más específicos que culminarán en el logro del objetivo general:

- Explorar diferentes índices vegetativos con el fin de realizar una adecuada estimación del rendimiento final, así como el análisis de datos agroclimáticos.
- Desarrollar un modelo predictivo y multivariable basado en técnicas de aprendizaje automático supervisado.
- Implementar correctamente el modelo predictivo y la evaluación de su precisión.

4. Materiales y métodos

4.1. Introducción

Para alcanzar los objetivos propuestos en este trabajo se deberán desarrollar diferentes elementos relativos a la recopilación o captación de datos de diferentes fuentes para su posterior análisis avanzado. Se podría decir que la realización de una adecuada y fiable obtención de datos representaría más del 80% del éxito o fracaso de la modelización en sí.

Además, se han obtenido los parámetros climáticos que permiten generar el modelo multivariable que haga posible realizar una predicción sobre el rendimiento final en la parcela de estudio. Estos parámetros climáticos se han obtenido de forma manual a partir de la red de estaciones agroclimáticas de la SIAR (Sistema de Información Agroclimática para el Regadío, s.f.) y resultan una herramienta clave para el desarrollo del proyecto. Concretamente se han recogido de la estación más cercana localizada en el municipio de Cubillas de los Oteros (León) que dista a menos de 6 km de la parcela de estudio, lo que nos da una idea de la alta fiabilidad del dato obtenido.

Por último, se ha recopilado y proporcionado por parte de la compañía *Corteva Agriscience* un conjunto de datos históricos relativos al rendimiento en campañas anteriores de las parcelas en estudio, con tal de generar un modelo etiquetado de aprendizaje supervisado que se detallará en esta sección.

Además, en el año 2022 el agricultor ha realizado un mapa de rendimiento de la parcela, es decir, la cosecha del maíz se ha realizado con una cosechadora georreferenciada que integra un medidor de caudal instantáneo, por lo que es capaz de determinar la producción del cultivo por área cosechada.

En los siguientes apartados se profundizará en la descripción de los desarrollos, sistemas, materiales y metodología empleada en este trabajo.

4.2. Área de estudio

4.2.1. Localización y caracterización de la parcela

Para la realización de este trabajo, se ha seleccionado la parcela llamada Camino los Barrios, compuesta por las parcelas N.º 94, 95, 96, 97 y 98 del polígono 405, pertenecientes a la pedanía de Jabares de los Oteros, que a su vez pertenece al municipio de Cabreros del Río. Este conjunto de parcelas conforma un total de 14,30 ha cultivables en un único recinto y está situada a la margen izquierda de la vega del río Esla, perteneciendo a su vez a la comarca Esla-Campos de la provincia de León. Se trata de una zona llana y de vega, es decir, caracterizada por suelos profundos constituidos por los aportes aluviales del río Esla.



Ilustración 1. Detalle del área de estudio sin escalas. Región de la Vega del Esla (Visor SigPac, s/f)

Sobre esta parcela, se ha recopilado información relativa a su localización, superficie, híbrido sembrado en la última campaña, densidad de siembra y, quizá lo más relevante en la modelización, un histórico de rendimientos medios totales de campañas anteriores que se detallará más adelante en el presente documento. Cabe destacar que variables como densidad de siembra, elección del híbrido y abonado han sido similares en los años que se va a realizar el estudio.

En la siguiente ilustración se incluye la ficha técnica de la parcela, así como el híbrido seleccionado, la localización de la parcela y la densidad de siembra elegida.



Ilustración 2. Detalles de la parcela (Escala 1:6250) (Visor SigPac, s/f)

4.2.2. Clima y Suelo de la zona de estudio

El clima de la provincia de León pertenece al tipo mediterráneo de influencia continental, matizado en algunos sectores por la influencia atlántica. Las principales características son la elevada amplitud térmica, la presencia de inviernos muy fríos y largos, el efímero significado de la primavera y otoño y las temperaturas moderadas de los cortos veranos que padece la provincia. La temperatura media anual es de 11,3°C, con una precipitación media anual de unos 504 mm. En cuanto a la humedad relativa, se puede definir como húmedo, ya que tiene un valor próximo al 70%. El período más lluvioso se produce en el otoño, concretamente en los meses de octubre y noviembre, seguido de la primavera.

Los suelos de la zona de estudio son de tipo franco-arenoso, lo que otorga una textura ideal para el cultivo del maíz. Son por lo general profundos, constituidos por los aportes aluviales del río como consecuencia de la erosión del suelo sufrido aguas arriba. Pertenecen al tipo de suelo de los entisoles, conformando los suelos más jóvenes de la superficie terrestre. Poseen una buena aptitud para el cultivo, debido a su adecuado equilibrio estructural, baja compactación y buena capacidad de retención de agua y nutrientes. Sin embargo, el contenido de materia orgánica presenta un valor bajo. En cuanto al resto de valores, como nutrientes y valores de pH, se pueden considerar normales, ya que se encuentran dentro de unos valores medios y un pH neutro.

De manera complementaria, en la Ilustración 3. Resultado del análisis edáfico de la parcela y recomendación de abonado, se muestra el resultado del análisis edáfico de la parcela en cuestión, así como las recomendaciones de abonado para una producción esperada de 15000kg/ha, según las necesidades del cultivo.



ANÁLISIS N. SEV26371		CARTAAGRONOMICA 50267	
DATOS ANALISIS 09/02/2022		FINCA CAMINO LOS BARRIOS	
DEVOLVERA C00852246		SUPERFICIE (ha) 14,30 RIEGO SI	
LEON		CULTIVO PRÓXIMO MAIZ	
24224 JABARES DE LOS OTEROS ES24		ABONO APLICADO (kg/ha) N P ₂ O ₅ K ₂ O	
TECNICO PIONEER		FERTILIZANTE ORGANICO APLICADO	
		ESTIERCOL OVEJA (q/ha) 50000	
		FECHA MUESTREO 25/01/2022	

GRANULOMETRIA		Ø	
Arena	(2,0 - 0,05 mm)	67,2 %	
Limo	(0,05 - 0,002 mm)	16,7 %	
Arcilla	(< 0,002 mm)	16,2 %	
Textura	(clasificación USDA)	FRANCO ARENOSO	

PARAMETROS ANALITICOS			
pH (H ₂ O)	7,1	neutro	
Carbonato cálcico equi	2,0 %	insapreciable	
Carbono Organico	0,6 %		
Dotación			
BAJA MEDIA ALTA MUY ALTA			
Nitrógeno Nitrico(N-NO ₃) mg/kg	6,70 mg/kg	[Barra de dotación]	
Fósforo asimilable (Método Olsen)	22,9 ppm P ₂ O ₅	[Barra de dotación]	
Potasio asimilable (ext. Ac NH ₄)	279,0 ppm K ₂ O	[Barra de dotación]	
Materia orgánica	1,10 %	[Barra de dotación]	
Na intercambiable (ext. Ac NH ₄)	8,0 ppm Na ⁺	[Barra de dotación]	
Ca intercambiable (ext. Ac NH ₄)	838,0 ppm Ca ²⁺	[Barra de dotación]	
Mg intercambiable (ext. Ac NH ₄)	58,0 ppm Mg ²⁺	[Barra de dotación]	

RECOMENDACIÓN DE ABONADO		Atención:
N	350 kg/ha	
P ₂ O ₅	80 kg/ha	
K ₂ O	100 kg/ha	

- La recomendación de abonado es válida para la campaña actual
- La dosis máxima de Nitrógeno pueden estar sujetas a las normativas de las Zonas Vulnerables en las distintas Comunidades Autónomas.

Ilustración 3. Resultado del análisis edáfico de la parcela y recomendación de abonado (Fuente: Servicio agronómico de Pioneer)

Como se puede apreciar en la ilustración anterior, las recomendaciones de abonado para una producción de 15 tm/ha son de 350 unidades de nitrógeno, 80 unidades de fósforo y 100 unidades de potasa. El nitrógeno es un elemento muy móvil y fácilmente lixiviable, por lo que en la mayoría de los casos se suele fertilizar según las extracciones. Sin embargo, el fósforo y el potasio son elementos poco móviles en el perfil del suelo, por lo que se suelen aportar según las necesidades de cada suelo. Por regla general, los suelos de la provincia de León presentan unos valores altos de fósforo y normales o bajos de potasio, por lo que a la hora de realizar una planificación de abonado se ha de tener en cuenta. También hay que tener muy en cuenta los valores de pH y contenidos de calcio y magnesio en el suelo, ya que, por lo general,

tienden a ser bajos (puede llegar a haber suelos con un $\text{pH} < 5$). Por ello, suele ser muy común la aplicación de enmiendas calizas, ya sea en forma de espumas de azucarera, dolomitas o carbonato cálcico, bien sea como corrección o mantenimiento.

4.2.3. Fenología y cultivo del maíz

4.2.3.1. La planta de maíz

El maíz es una planta que proviene por selección de los agricultores de la planta ancestral del teocinte, que con sólo cinco mutaciones genéticas dio lugar a la planta de maíz. Se trata de una planta hermafrodita, es decir, hay flores masculinas (panoja o espiga) y femeninas (mazorca) en la misma planta (Endicott et al., 2015). Como se puede apreciar en la Ilustración 4. Situación de la panoja y mazorca o estigmas en la planta de maíz.



Ilustración 4. Situación de la panoja y mazorca o estigmas en la planta de maíz. (Imagen de la izquierda: detalle de la panoja e imagen central: detalle de los estigmas) (Endicott et al., 2015)

Como se muestra en la imagen de la derecha, hay una separación vertical de alrededor de 1 metro entre las flores, lo que puede contribuir al desafío de una polinización exitosa. El desprendimiento de polen se produce de forma discontinua durante un período de aproximadamente cinco a ocho días, y solo sucede cuando las condiciones de temperatura y humedad son favorables. Según el ambiente, un estigma continúa creciendo durante aproximadamente siete días o hasta que intercepta los granos de polen, dando lugar cada seda a un único grano en la mazorca. Es importante que el desprendimiento del polen y la emergencia de los estigmas sucedan al mismo tiempo para garantizar una correcta polinización

4.2.3.2. Crecimiento y desarrollo

Se entiende por crecimiento como el aumento de tamaño y como desarrollo, la progresión de las diferentes etapas vegetativas y reproductivas. Las etapas vegetativas (V) se caracterizan por la presencia del cuello de una hoja en hojas emergidas, comienzan con la emergencia (VE) y continúan de forma numérica con cada hoja sucesiva hasta que emerge la panoja (VT). Las etapas reproductivas se caracterizan por la emergencia de granos en desarrollo en la mazorca, excepto por la primera etapa reproductiva (R1), que se identifica únicamente por la emergencia de estigmas de las chalas. En la Ilustración 5 se muestran las diferentes etapas vegetativas y reproductivas.

ESTADOS VEGETATIVOS Y REPRODUCTIVOS DE UNA PLANTA DE MAÍZ

ESCALA RITCHIE Y HANWAY, 1982.

ESTADOS VEGETATIVOS

- VE** Emergencia.
- V1** 1^{era} hoja desarrollada.
- V2** 2^{da} hoja desarrollada.
- V3** 3^{era} hoja desarrollada.
- V(n)** "n" ésima hoja.
- VT** Panojamiento.

ESTADOS REPRODUCTIVOS

- R1** Emergencia de estigmas.
- R2** Ampolla (blister).
- R3** Grano lechoso.
- R4** Grano pastoso.
- R5** Grano dentado.
- R6** Madurez fisiológica.

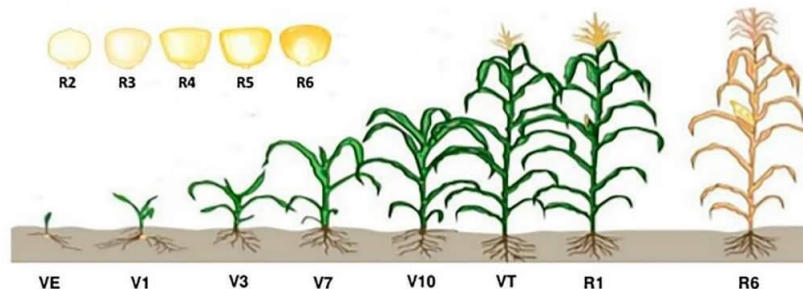


Ilustración 5. Etapas vegetativas y reproductivas del maíz (Fuente: Servicio agronómico de Pioneer)

4.2.3.3. Grados-Día de crecimiento

La radiación solar es un aporte fundamental para el crecimiento y desarrollo de la planta, ya que es utilizada por hojas y tallo como fuente de energía para la realización de la fotosíntesis. Se ha demostrado que el tiempo que el maíz requiere para pasar de una etapa de desarrollo a otra depende de la cantidad de calor acumulado (Gilmore & Rogers, 1958). Existen diferentes medios para el cálculo del calor acumulado o integral térmica, tomando como base la temperatura mínima de 10°C y la máxima de 30°C, es decir, fuera de estos umbrales se detiene el crecimiento de la planta. El híbrido promedio de la provincia de León desarrollará estigmas (R1) aproximadamente de 60 a 65 días calendario (1050 GDD) después de la VE y alcanzará la madurez (R6) aproximadamente de 100 a 105 días calendario (1950 GDD) después de la VE. Los híbridos de maíz varían en el tiempo requerido para alcanzar la madurez fisiológica cuando se calculan mediante el uso de los días calendario o GDD.

Tabla 1. Clasificación de los híbridos de maíz según el índice de precocidad de la FAO (Fuente: Servicio agronómico de Pioneer)

Índice FAO	Integral térmica (GDD)	Días
200	<1800	85-93
300	1801-1900	94-99
400	1901-2000	100-105
500	2001-2075	106-111
600	2076-2125	112-117
700	2126-2176	118-123
800	2177-2228	124-129
900	2229-2279	130-135
1000	>2280	136-141

Aproximadamente, se necesitan 90-120 GDD (grados día de desarrollo) para la emergencia de la planta. Tras una siembra normal en la zona, de primeros de abril, el crecimiento de las hojas y el tallo coincide en los meses de abril hasta julio, y es en este mes cuando se produce la fase de emisión del penacho y floración. Tras ella, la fase de llenado del grano, seguida de la madurez fisiológica (o punto negro) se alcanzan a mediados de septiembre hasta que finalmente, se produce la senescencia del cultivo y la posterior cosecha a partir de la segunda quincena de octubre. Evidentemente, cualquier estrés durante el desarrollo del cultivo, ya sea hídrico, nutricional o de otro tipo, causará una merma en la producción final, variando también la cuantía de dicha merma en función de la sensibilidad de la etapa en la cual se haya producido.

En la Ilustración 6 se muestra un simulador gráfico para el cálculo de integral térmica en el maíz, lo que da una orientación de los GDD que necesita acumular hasta la aparición de las sedas (floración) y la llegada a punto negro o madurez fisiológica. En este simulador se elige la fecha de siembra (1 de abril en este caso), el híbrido escogido (La variedad *Pioneer P0023* es un ciclo con CRM100 o FAO 400) y la estación meteorológica más cercana. De esta manera, la aplicación realiza una estimación de la fecha de floración (Hacia el 30 de junio con una cantidad de 935 GDD) y la de madurez (Sobre el 7 de septiembre con 1915 GDD). También se muestra el riesgo de heladas que existe en la zona para que se complete el ciclo vegetativo del híbrido, siendo muy bajo en este caso, ya que hasta la llegada a madurez fisiológica nos movemos con valores de poco más del 20%.

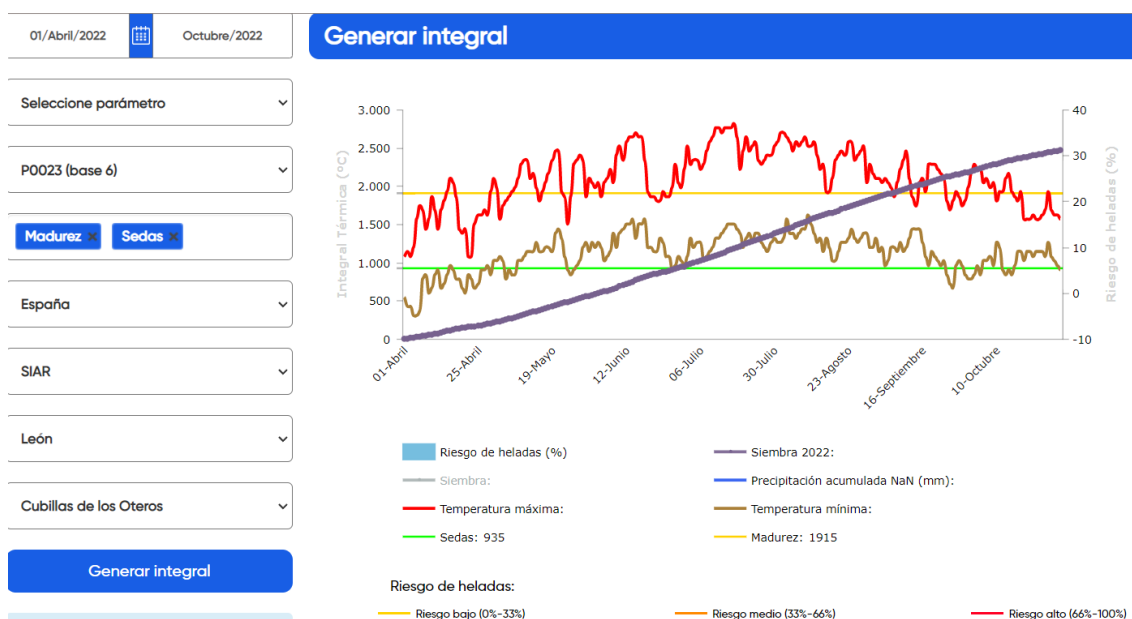


Ilustración 6. Simulación de la integral térmica del maíz en la zona en función de la fecha de siembra y el ciclo del híbrido seleccionado (Fuente: Servicio agronómico de Pioneer)

Con este tipo de simuladores se pueden establecer fechas límites de siembra para los distintos ciclos. Evidentemente, cuanto mayor sea el ciclo escogido, el potencial esperado será más alto, pero siempre y cuando no exista riesgo de heladas antes de la llegada a madurez fisiológica. De esta manera, con este tipo de simuladores y la experiencia del agricultor-asesor es posible adecuar la utilización de distintos ciclos en función de las diferentes fechas de siembra, optimizando así el potencial productivo de cada parcela.

4.2.3.4. Cultivo del maíz

El maíz se siembra cuando la temperatura del suelo es superior a 8°C a partir del mes de abril y hasta finales de mayo. La profundidad de siembra debe ser entre 3-5 cm en función del tipo y estado del terreno y a una densidad de 95000-105000 semillas por hectárea con una separación de líneas de 50-75 cm. Para conseguir una adecuada nascencia es recomendable garantizar un buen lecho de siembra y tener cierta humedad en el suelo.

En cuanto a la fertilización, es necesario realizar una aplicación en fondo con todo o gran parte del fósforo y potasio necesario y entre el 10-30 % del nitrógeno requerido. El fósforo ayudará a la nascencia y llenado del grano y el potasio, en gran medida al desarrollo vegetativo de la planta. El nitrógeno es utilizado para la producción de almidones, azúcares, proteínas y aminoácidos y otros procesos celulares de la planta.

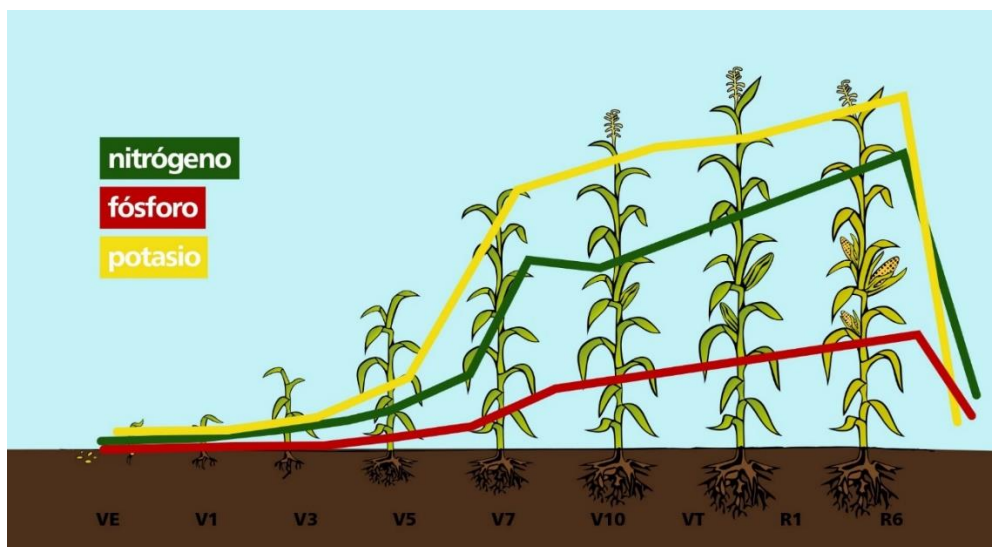


Ilustración 7. Curva de absorción de nutrientes NPK durante el ciclo del maíz (Fertilización en maíz: La necesidad de magnesio, s. f.)

Las necesidades nutricionales del cultivo se calculan en función de la cosecha esperada y en relación con el resultado del análisis de suelo si es que disponemos de él. Para ello, se muestra la siguiente ilustración, con las necesidades NPK por tm esperada.

Extracciones por ha y tm grano

Cultivo	N	P ₂ O ₅	K ₂ O
Maíz	22-27 (17-20)	9-11 (6-7)	20-25 (5-7)
Sorgo	35	14	33
Girasol	50	18	100
Colza	44	25	24

(Fuentes, 1999)



Ilustración 8. Extracciones del maíz por tm/ha esperada (Fuentes, 1999)

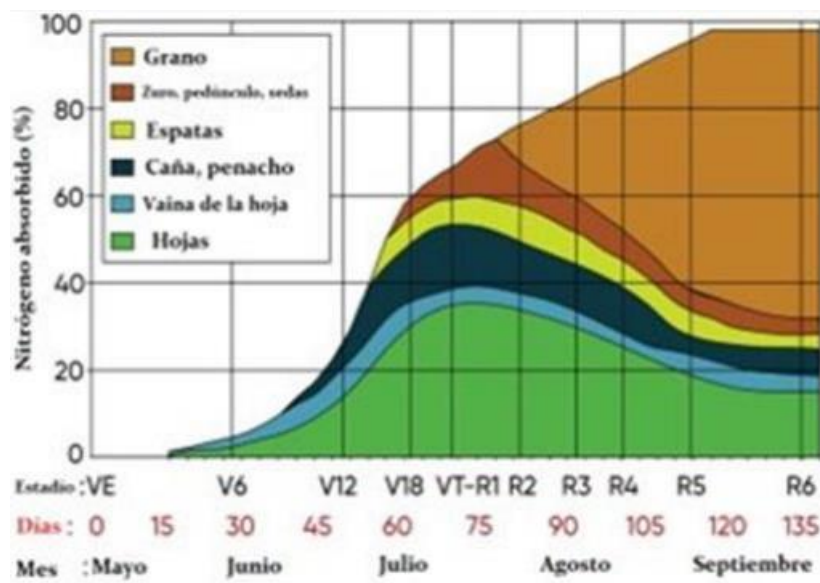


Ilustración 9. Evolución de las extracciones de nitrógeno absorbido por cada parte de la planta (Fuente: Servicio agronómico de Pioneer)

Las aplicaciones de abonado nitrogenado se suelen recomendar a partir de la 6ª hoja desplegada, momento en el que la curva de absorción de nutrientes permite que se aproveche la mayor cantidad de nitrógeno aportado. Es necesario asegurar la presencia de nitrógeno hasta el momento final, ya que después la floración aún es necesario un 30 % de las necesidades totales, de las que aproximadamente la mitad será traslocada de los tejidos de la propia planta y la otra mitad será absorbida del suelo.

Las necesidades hídricas del cultivo son elevadas, sobre todo durante el período reproductivo, coincidiendo además con la época de mayor evapotranspiración. Aproximadamente, la dotación hídrica para el cultivo del maíz está cifrada en 5000-6000 m³/ha en función de la zona y tipo de terreno. Obviamente, las necesidades hídricas varían en función del estadio en el que se encuentre la planta y la evapotranspiración, siendo máximas en la floración (Hasta 55 litros/semana en nuestra latitud). Como regla general, existen ciertos momentos críticos en los que no debe faltarle agua al cultivo:

- V0, VE-V5: Desde la siembra a emergencia y hasta las 5 hojas, las necesidades de agua serán mínimas pero suficientes para conseguir una adecuada nascencia y enraizamiento de la planta (1^{er} componente del rendimiento = N^o plantas nacidas)
- V6 a Vn: Coincide con la época de mayor crecimiento vegetativo. A partir de este estadio se está definiendo la mazorca, es decir, longitud y número de filas de óvulos fértiles en la mazorca (2^o componente del rendimiento = N^o óvulos fértiles /planta)
- VT a R2: Época de floración, lo que coincide con la fertilización de dichos óvulos, tratando de evitar que haya granos abortados (2^o componente del rendimiento = N^o de granos u óvulos fecundados /planta)
- R2 a R6: Llenado del grano hasta la llegada a punto negro. La línea de llenado de almidones o paso del estado lechoso-pastoso a duro empieza desde la parte exterior del grano hasta la interior, finalizando con la llegada a la necrosis o punto negro (Véase la ilustración 10). En esta fase tenemos que conseguir un elevado peso específico del grano. (3^{er} componente del rendimiento = Peso específico del grano)



Ilustración 10. Línea de llenado de almidones y punto negro (Fuente: Elaboración propia)

De esta manera hemos calculado la fórmula del rendimiento final esperado:

$$\text{Rendimiento esperado} = \frac{N^{\circ}\text{plantas emergidas}}{\text{ha}} \times \frac{N^{\circ}\text{granos}}{\text{planta}} \times \text{peso granos}$$

Se puede decir que el maíz ha completado su ciclo cuando aparece la línea de punto negro en los granos, momento en el cual se corta el suministro de nutrientes de la planta al grano y libramos ya el riesgo de heladas. A partir de este momento, se considera que el grano está hecho (Humedad grano \approx 35%) y únicamente se necesita perder la humedad para poder ser cosechado (Pastore \approx 29-32% y grano seco $<$ 23%). Por lo general, el maíz se cosecha con una humedad superior al 14%, por lo que es necesario pasarlo por secadero para poder garantizar una adecuada conservación.

4.3. Generación del dataset

Para la generación de datos, tan importante es generar una gran cantidad de datos como que todos esos datos sean útiles para la consecución del objetivo final. En este momento, es cuando entra en vigor el papel del buen ingeniero agrónomo estableciendo qué datos van a ser determinantes para la producción del maíz. Por ese motivo, se requiere de un conocimiento pragmático adecuado para la selección de datos determinantes en el rendimiento final del maíz. Con este objetivo, se han determinado una serie de variables descriptivas que pueden tener más o menos influencia en el resultado final. Estas variables se derivan de tres fuentes de datos relevantes:

- Datos agroclimáticos o meteorológicos que afectan al cultivo.
- Rendimientos históricos medios obtenidos, que actuarán como etiquetas para el aprendizaje.
- Evolución vegetativa del cultivo en las diferentes campañas mediante índices de vegetación derivados de la teledetección.

Granularidad de los datos

La granularidad de los datos hace referencia al nivel de detalle que posee un dato o una estructura de datos (Zeus, s. f.). En esta se evalúa la complejidad del tipo de dato y el tamaño en el que se dividen los campos de datos.

Los datos climáticos se encuentran disponibles con una frecuencia diaria, los datos relativos a los índices de vegetación mantienen una frecuencia semanal y los datos relativos al rendimiento histórico tienen una frecuencia anual, ya que únicamente se realiza una cosecha al año. En los apartados a continuación se detalla cómo se trabaja con esta variabilidad en la granularidad, para ajustar la resolución temporal de los mismos a una frecuencia quincenal que permita generar un set de datos y un modelado detallado y riguroso.

A continuación, se detallan las tres fuentes de datos empleadas a partir de las cuales se han obtenido el set de datos.

4.3.1. Datos meteorológicos

Como bien se ha mencionado anteriormente, simplemente por ser la planta de maíz un ser vivo, cualquier cambio anómalo que se produzca en el ambiente, puede ser causa de una modificación en el rendimiento final del cultivo. Obviamente, un estrés en plena floración causará mayores pérdidas que en una etapa de senescencia, por ejemplo, así que se puede afirmar que en función del período en el que se produzca variará la cuantía del daño sobre el rendimiento final. Por este motivo, los datos climáticos constituyen una parte muy importante en la construcción del modelo predictivo del rendimiento, debido a su interacción con el mismo durante todas las fases vegetativas.

Los datos han sido obtenidos de la estación agroclimática más cercana a la parcela de estudio y con similares características climáticas. Concretamente, la estación se encuentra en el municipio de Cubillas de los Oteros (Ilustración 11), situado a una distancia de escasos 6 km de la parcela y pertenece a la red de estaciones agroclimáticas de la SIAR (Sistema de Información Agroclimática para el Regadío). Estas estaciones agroclimáticas juegan un papel determinante en los trabajos de asesoramiento de las zonas de riego modernizado con el fin de optimizar la utilización eficiente del agua por parte de los agricultores. Además, otorgan información sobre las diferentes variables meteorológicas de interés agrícola, tales como temperaturas, humedades relativas, precipitación, radiación, dirección y velocidad del viento, evapotranspiración (Eto). Por tratarse de una zona de regadío cobra gran importancia el conocimiento de la Eto, a través de la cuál y multiplicado por el factor de cultivo (kc) se calculan las necesidades hídricas (NH) del cultivo o la cantidad de agua que es necesario aportar mediante la irrigación ($NH = Eto * kc$). El factor de cultivo es una variable constante que depende del cultivo que se trate y del estadio vegetativo en el que se encuentre. Normalmente presenta valores bajos al inicio del cultivo y va creciendo hasta el momento de plena floración, reduciéndose a medida que se va alcanzando la senescencia de este (InfoRiego. s/f).



Ilustración 11. Red de estaciones agroclimáticas del SIAR y descarga de ficheros históricos (Sistema de Información Agroclimática para el Regadío, s.f.)

A partir de la estación agroclimática de Cubillas de los Oteros se han incorporado cinco variables relevantes al dataset, como son las temperaturas máxima y mínima, la humedad relativa media, la precipitación acumulada y la Eto. Para este set de datos agroclimáticos no se han determinado relevantes la utilización de temperaturas medias, velocidades de viento y radiación, ya que se considera que prácticamente no influyen sobre el rendimiento final del maíz.

Como se ha explicado anteriormente en la granularidad de la información, para la construcción del set de datos únicamente se han cogido como referencia los meses en los que hay crecimiento del maíz (desde el 1 de abril hasta el 1 de octubre) y se han agrupado dichos datos por quincenas, haciendo las medias de temperaturas y humedades y el acumulado para las precipitaciones.

4.3.2. Índices de Vegetación

La obtención de los índices de vegetación relativos al cultivo del maíz se ha llevado a cabo empleando como fuente las imágenes satelitales correspondientes a los satélites de la misión *Copernicus Sentinel 2* de la Agencia Espacial Europea. El programa Copérnico, anteriormente llamado *Global Monitoring for Environment and Security*, también conocido como *Copernicus*, es un proyecto coordinado y gestionado por la Comisión Europea, que pretende lograr una completa, continua y autónoma capacidad de observación terrestre de alta calidad cuyos resultados sean accesibles libremente por la comunidad científica o cualquier otra persona interesada. El objetivo general es la obtención de información exacta, fiable y continua, para, entre otras cosas, mejorar la gestión y conservación del medio ambiente, comprender y mitigar los efectos del cambio climático y asegurar la seguridad civil. Pretende agrupar diferentes fuentes de información de satélites medioambientales y bases terrestres para proporcionar una visión global del «estado de salud» de la Tierra. La misión *Copernicus Sentinel 2* comprende una constelación de dos satélites en órbita polar colocados en la misma órbita en fase de 180° entre sí. Su objetivo es monitorear la variabilidad en las condiciones de la superficie terrestre y su amplio ancho de franja (290 km) y tiempo de revisita alto (10 días en el ecuador con un satélite y 5 días con 2 satélites en condiciones libres de nubes, lo que resulta en 2-3 días en latitudes medias) apoyará el seguimiento de los cambios en la superficie de la Tierra.

Sin embargo, la utilización de imágenes por satélite también presenta una serie de limitaciones:

- La resolución temporal, de una frecuencia de hasta 10 días en el ecuador con el Sentinel-2, puede limitar la ventana de observación ante problemas del cultivo. Sin embargo, en las latitudes que nos encontramos no ocasiona ningún problema.
- La resolución espacial de 10 metros puede limitar la visibilidad en cultivos que no cubran la superficie del suelo. En el caso del maíz, por tratarse de un cultivo extensivo y con una cobertura total del suelo, tampoco presenta un problema.
- La presencia de nubes puede alterar o inutilizar la imagen tomada. Para ello, se utilizan una serie de filtros que suprimen las imágenes con este problema. En caso de que hubiera muchos días seguidos con este problema, perderíamos gran cantidad de información. Sin embargo, por tratarse el maíz de un cultivo de verano, se reduce el problema de esta incidencia.
- Las imágenes tomadas pueden ser alteradas por la presencia de otros agentes atmosféricos o de otro tipo como polvo y humedad, lo que se soluciona utilizando los índices vegetativos adecuados, capaces de paliar dichas alteraciones.

Para extraer información sobre la vegetación a partir de imágenes de satélites es necesario transformar el volumen de datos, permitiendo así mejorar su disposición y hacerla útil para el consumidor. Estas transformaciones son las conocidas como índices de vegetación. Por tanto, un índice de vegetación normalmente es el resultado de una fórmula que utiliza una o varias bandas del espectro electromagnético. La relación entre estas bandas tiene detrás un estudio empírico que demuestra la relación directa entre el valor numérico captado por el sensor y la variable de la planta a medir (normalmente biomasa o vigor vegetal).

Existen gran cantidad de índices y están en constante evolución. Las empresas, universidades y científicos del mundo están continuamente estudiando nuevos índices de vegetación, modificando los ya existentes para hacerlos más precisos. Hay todo un mundo por hacer, desarrollar y descubrir. En este caso se van a emplear dos índices muy conocidos, el NDVI y el GNDVI, por el hecho de ser los índices que mejor representan el crecimiento del maíz (rápido y con gran cantidad de biomasa). Se explican a continuación:

- NDVI: Son las siglas de Normalized Difference Vegetation Index. Fue creado por Tucker en 1979 para el monitoreo de cubiertas vegetales y pese a su antigüedad, aún es el índice más empleado y conocido. Utiliza las bandas del rojo y el infrarrojo cercano y nos permite calcular el vigor de una planta en un estado fenológico concreto. Este vigor nos indica el estado del cultivo. Si está por debajo de lo normal, el índice nos dirá que la planta está sufriendo (o no está en su punto óptimo) y tendremos que averiguar por qué. Puede ser por estrés hídrico, por una plaga o enfermedad, por falta de nutrientes. Los valores se encuentran entre -1,0 y +1,0 dependiendo de la captación de luz por el sensor, siendo mínima cuando la captura de luminosidad es nula (zonas sin vegetación) y máxima cuando se captura toda la energía luminosa disponible con un rango de valores (Muñoz Herranz, 2021). Las causas pueden ser muchas pero lo que el índice nos dice de forma prematura es que a la planta le está ocurriendo algo para que podamos llegar a tiempo y corregirlo. El NDVI brinda a los agricultores la oportunidad de evaluar la biomasa y el rendimiento de los cultivos. Estas mediciones de reflectancia indirecta se han utilizado para estimar la biomasa y el rendimiento de la planta (Tamás et al., 2023). Se calcula de la siguiente manera:

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

- GNDVI: Son las siglas de Green Normalized Difference Vegetation Index. Se trata de una variación del NDVI que utiliza la banda del verde en vez del rojo para sus cálculos y es hasta cinco veces más sensible que el NDVI en la detección de la variación de clorofila. Hay cultivos que responden mejor a unas bandas que a otras y en cultivos de arroz o en maíz, en ocasiones este índice da mejores resultados, por ejemplo, en la predicción del rendimiento de cosecha en función del vigor calculado por este método. Su fórmula es la siguiente:

$$GNDVI = \frac{NIR - Green}{NIR + Green}$$

Con estos dos índices se va a construir el set de datos relativo a la monitorización de las parcelas en estudio, y se buscará asociar su influencia y su evolución temporal como posibles indicadores del rendimiento final obtenido en el cultivo del maíz.

Como se puede observar, la utilización de estas herramientas e índices vegetativos son capaces de georreferenciar el problema, pero no de averiguar la causa que lo ha originado, por lo que son una herramienta más para la figura del agrónomo, pero no de sustituir su trabajo. Se puede decir que estas herramientas complementan y optimizan el trabajo diario del agrónomo, siendo hoy en día casi imprescindibles, ya que le permiten ir directamente al epicentro del problema y poder analizar la causa que lo originó. Actualmente es posible estimar la ET para diferentes cultivos, brindando información espacial y temporal sobre un área amplia, utilizando los diferentes índices de vegetación (Gavilán et al., 2019).

4.3.3. Serie histórica de rendimientos

En el modelado del rendimiento que se pretende realizar, se incorporan variables correspondientes al desarrollo vegetativo del propio cultivo (índices de vegetación), así como de parámetros climáticos, actuando todas ellas como descriptoras del rendimiento final. De esta forma, es clave evaluar la evolución de cada uno de estos parámetros el dato final de rendimiento.

La utilización de herramientas de Inteligencia Artificial o Machine Learning, no son más que un modelo al que debemos enseñar para poder lograr el objetivo que pretendamos. En este caso, deberemos entrenar el modelo con una serie de variables climáticas y vegetativas para que nos prediga la producción final. Para “entrenar” el modelo debemos introducir una serie de entradas que darán lugar a un rendimiento final que será nuestro objetivo. Así, el modelo determinará que la combinación ponderada de cada una de estas variables da como resultado una producción final determinada. En el caso que nos ocupa, vamos a utilizar los tres primeros años como entrenamiento del sistema para poder predecir el rendimiento del cuarto año con unas variables climáticas y de vegetación ya conocidas. Finalmente, como el rendimiento del cuarto año es también conocido, podremos calcular la precisión/fiabilidad del modelo predictivo realizado.

La información de la serie histórica de rendimientos ha sido proporcionada directamente por el agricultor, ya que está en permanente contacto con los asesores agronómicos de Pioneer.

4.4. Análisis y modelado predictivo

4.4.1. Plataforma análisis cloud: *BigML*

Las tecnologías de predicción basadas en datos se han simplificado tanto que se han puesto al alcance no sólo de las grandes empresas, sino también de particulares. Herramientas como las desarrolladas por *BigML* (Bigml.com, s/f) están aproximando el Machine Learning a estos pequeños consumidores, pero no hay que olvidar que la materia prima para cualquier sistema predictivo son los datos. No en vano, el esfuerzo que estamos dedicando a la fase de recogida y tratamiento de datos previa a la algoritmia es entre un 80% y un 90% del tiempo total del proyecto.

BigML es un sistema de aprendizaje automático en la nube que sirve para crear modelos predictivos. Utilizando una sencilla interfaz web, *BigML* permite la creación de árboles de decisión, análisis de clúster y detectores de anomalías, dotados de una visualización interactiva que permite su fácil exploración y puede ayudar en la toma de decisiones basadas en grandes cantidades de datos (Big Data es la denominación actual para definir modelos de negocio y tecnologías especializadas en la captación, almacenamiento, tratamiento, análisis y presentación de enormes cantidades de datos obtenidos a partir de diversas fuentes).

Las técnicas estadísticas y de aprendizaje automático supervisado que incorpora la plataforma, aplicadas al ámbito de la agricultura, permiten contribuir a la toma de decisiones por parte de los técnicos de forma rápida y flexible.

Una vez se han recogido y estructurado los datos necesarios para modelizar la producción de maíz en las cinco parcelas de estudio, estos se han utilizado como input, modelando el dataset empleando la herramienta de *BigML* para la definición de variables, tal y como se muestra en la Ilustración 12.

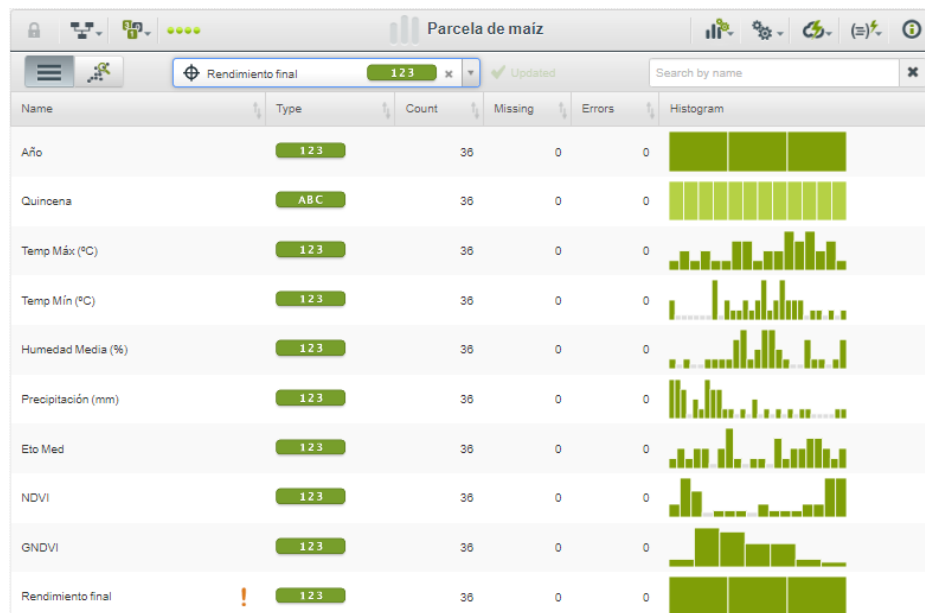


Ilustración 12. Ejemplo del set de datos obtenido en *BigML*, mediante índices vegetativos y datos climáticos. En la parte superior se observa cómo el rendimiento final ha sido marcado como objetivo (Bigml.com, s/f)

Con los datos estructurados se pueden realizar análisis de las diferentes variables que se han estudiado. Por ejemplo, el programa da la opción de visualizar un *scatter plot* o gráfico de dispersión para poder ver la correlación que existe entre cada una de las variables seleccionadas. Se muestran a continuación un par de ejemplos.

Como se puede apreciar en la Ilustración 13, se observa que la correlación existente entre las temperaturas máximas y la Eto es muy alta, ya que los puntos tienden a acercarse a la recta bisectriz del gráfico. Los colores de cada punto representan los diferentes años. Además, en la parte inferior izquierda del gráfico, se muestran los coeficientes de correlación de Pearson y Spearman, que permiten cuantificar el grado de ajuste y de relación lineal entre dos variables. El primero es un coeficiente paramétrico, es decir, infiere sus resultados a la población real, lo que hace necesario que la distribución de nuestra muestra se asemeje a la distribución real, es decir, que haya normalidad. Sin embargo, el coeficiente de Spearman es no paramétrico, pues la distribución muestral no se ajusta a una distribución conocida, por lo que los estimadores muestrales no son representativos de los parámetros poblacionales.

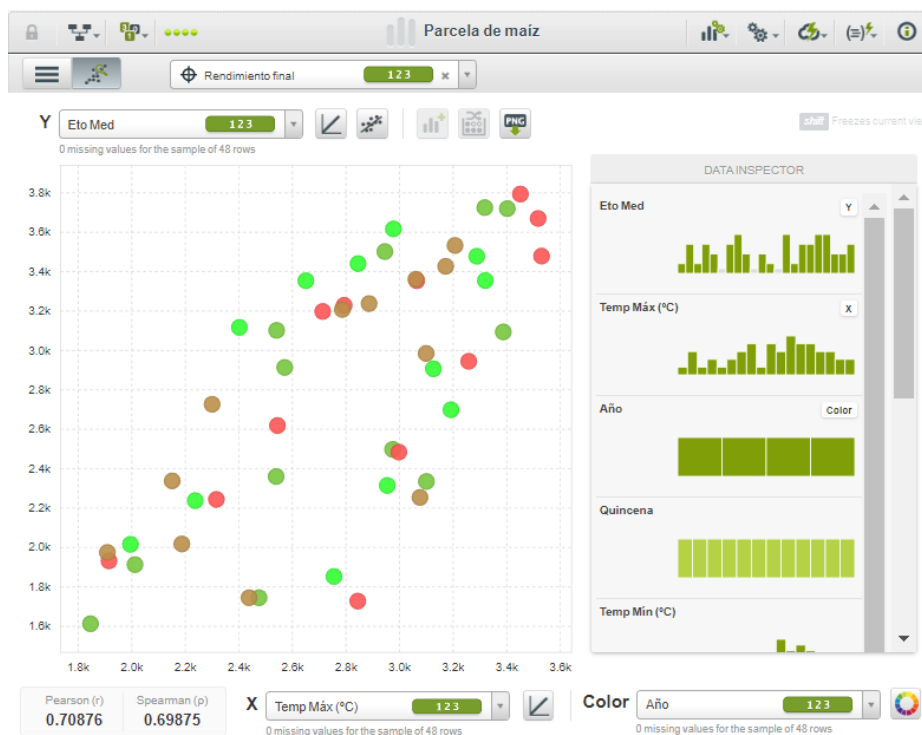


Ilustración 13. Scatter plot con T^a Máx en el eje X y Eto en el eje Y (Bigml.com, s/f)

Como se puede observar en la Ilustración 14, la correlación que hay entre los índices vegetativos y las temperaturas máximas también es muy alta. Esto se debe a que como ya se ha explicado con anterioridad, la planta de maíz basa su crecimiento en la acumulación de grados-día, por lo que cuanta más temperatura diaria alcancemos, mayor será su crecimiento vegetativo.

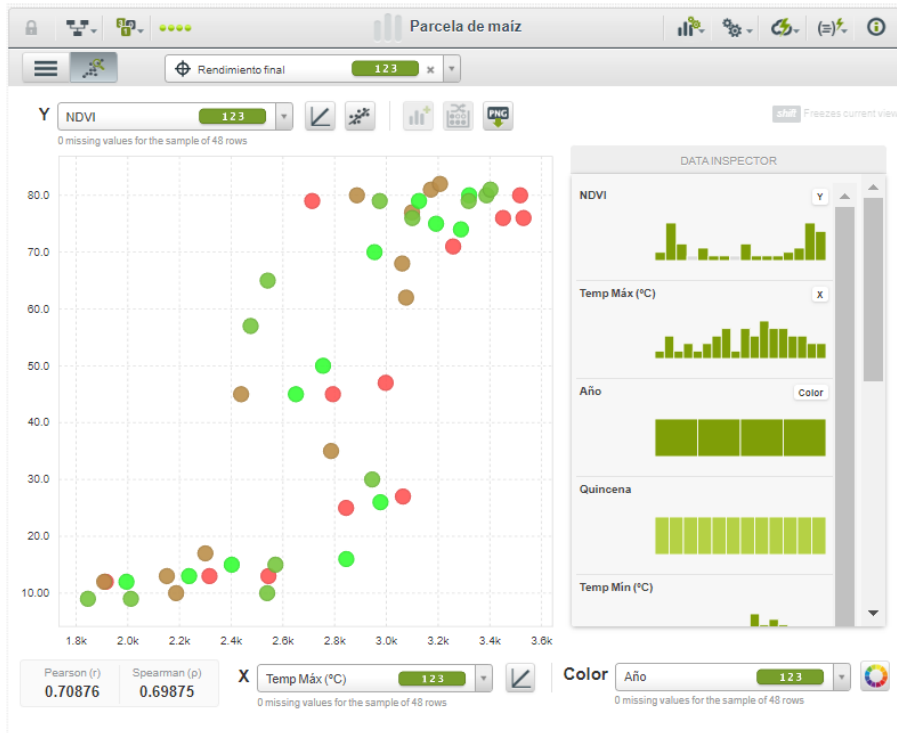


Ilustración 14. Scatter plot con T^a Máx en el eje X y NDVI en el eje Y (Bigml.com, s/f)

Para realizar un *scatter plot* con el rendimiento final hay que tener en cuenta la granularidad de los datos y realizar las medias anuales de cada dato, tal y como se muestra en la Ilustración 15.

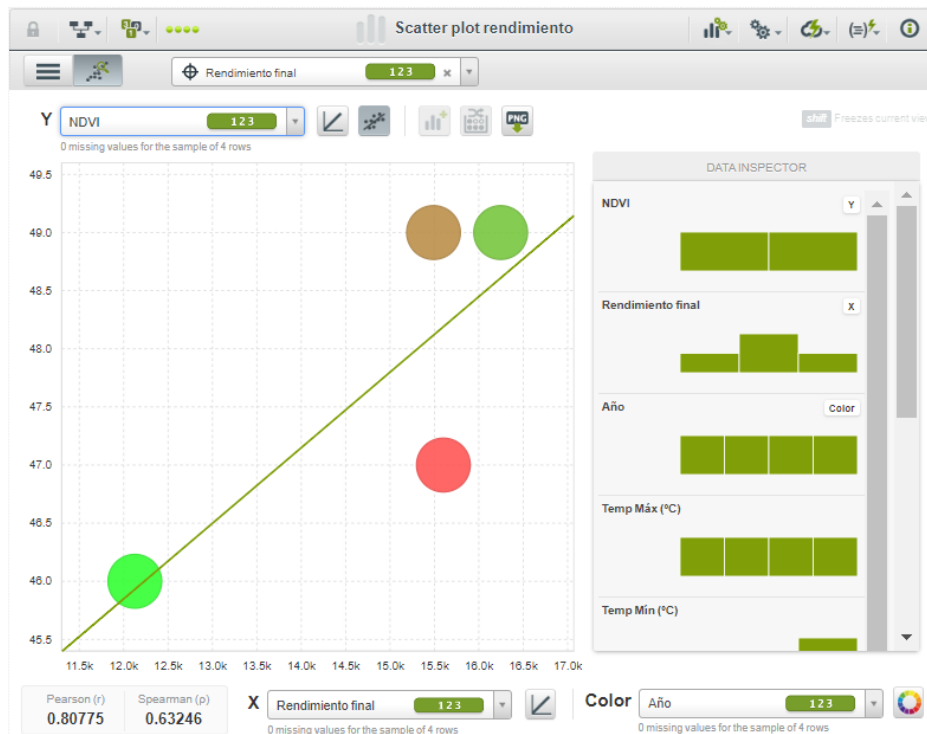


Ilustración 15. Scatter plot con Rendimiento final en el eje X y NDVI en el eje Y (Bigml.com, s/f)

Como se puede observar en la ilustración 15 la correlación entre NDVI y rendimiento final es muy alta, ya que el desarrollo de biomasa de cualquier cultivo tiene una estrecha analogía con su capacidad productiva. Con este ejemplo, se puede visualizar la importancia de estos índices vegetativos para la cuantificación de cosechas o la elaboración de mapeos predictivos de rendimiento variable.

4.4.2. Modelos de aprendizaje supervisado

Una vez integrados los datos en la plataforma de *BigML*, se emplean los diferentes algoritmos y técnicas de Machine Learning integrados para obtener un modelo de predicción ajustado.

Para la modelización predictiva de la producción de maíz es necesario la construcción de un dataset o conjunto de datos a partir del cual, entrenar un algoritmo que permita minimizar el error en la predicción.

Las técnicas de Machine Learning de aprendizaje supervisado buscan patrones en los datos de entrenamiento, de forma que utiliza un modelo pregunta-respuesta conocidas (etiquetadas) para deducir una función matemática que relacione dichos parámetros (Acuña et al., 2022).

Tras su fase de entrenamiento, se obtiene un modelo (en nuestro caso predictivo), capaz de ofrecer una respuesta a un nuevo input o parámetro de entrada, se realiza una validación con una muestra representativa para evaluar su eficacia (Véase Ilustración 16). En nuestro caso, estos parámetros de entrada una vez entrenado el modelo, corresponden con la información de la campaña 2022 y su respuesta deberá ser el dato predicho de producción de maíz.

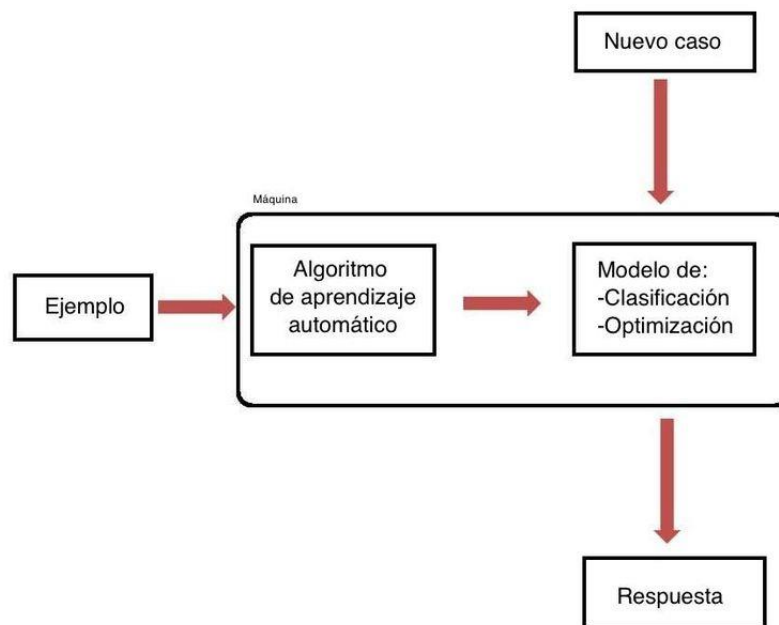


Ilustración 16. Diagrama modelo de aprendizaje supervisado (Colaboradores de Wikipedia, 2023)

Los modelos evaluados con el conjunto de datos para la predicción del rendimiento final van a ser mediante el análisis de componentes principales, el árbol de decisiones y el uso de redes neuronales.

4.4.2.1. *Análisis de componentes principales*

En estadística, el análisis de componentes principales (en español ACP, en inglés, PCA) fue creado por Karl Pearson en 1901 y es una técnica utilizada para describir un conjunto de datos en términos de nuevas variables («componentes») no correlacionadas (Zubcoff, J. J., & de Estadística, s/f). Los componentes se ordenan por la cantidad de varianza original que describen, por lo que la técnica es útil para reducir la dimensionalidad de un conjunto de datos. En el modelo que se crea como resultado del análisis PCA se otorgan diferentes pesos a los datos de entrada de las quincenas. El análisis valora en qué quincenas de los meses estudiados el valor de determinada variable es el que tiene el peso más importante para el rendimiento.

Los pasos para realizar el análisis de componentes principales son (Marcos Robles, J. L., 2016):

- Análisis de la matriz de correlaciones: un PCA tiene sentido si existen altas correlaciones entre las variables, ya que esto es indicativo de que existe información redundante y, por tanto, pocos factores explicarán gran parte de la variabilidad total.
- Selección de los factores: la elección de los factores se realiza de tal forma que el primero recoja la mayor proporción posible de la variabilidad original; el segundo factor debe recoger la máxima variabilidad posible no recogida por el primero, y así sucesivamente. Del total de factores se elegirán aquéllos que recojan el porcentaje de variabilidad que se considere suficiente. A éstos se les denominará componentes principales.
- Análisis de la matriz factorial: una vez seleccionados los componentes principales, se representan en forma de matriz. Cada elemento de ésta representa los coeficientes factoriales de las variables (las correlaciones entre las variables y los componentes principales). La matriz tendrá tantas columnas como componentes principales y tantas filas como variables.

4.4.2.2. *Técnica predictiva mediante árbol de decisiones y bosques de decisión o modelos ensamblados*

Se trata de un modelo predictivo que mapea observaciones sobre un artículo a conclusiones sobre el valor objetivo del artículo. Es uno de los enfoques de modelado predictivo utilizado en estadísticas, minería de datos y aprendizaje automático.

Una vez determinadas cuáles son las variables más significativas no correlacionadas entre sí como resultado del análisis PCA, se han generado dos predicciones diferentes empleando los árboles de decisión y los bosques de decisión. Los árboles de decisión presentan una estructura ramificada que muestra las diferentes opciones y sus consecuencias. Los puntos en los que hay que tomar decisiones se muestran como nodos, las ramas unen estos nodos y las decisiones últimas son las hojas, donde termina el camino. Se trata de la técnica predictiva más simple.

Con respecto al uso de los modelos ensamblados, se tratan de técnicas que buscan obtener una gran calidad en las predicciones mediante la combinación de varios árboles de decisiones. Este método analiza los datos en forma de regresión, de forma que busca mejorar la predicción empleando un conjunto de muchos árboles de decisión y los generaliza minimizando una función de pérdida. La Ilustración 17 explica el proceso, en el que partiendo de un set de datos, se genera el modelo de árbol de decisión, se analiza el error cometido en forma de regresión, y se vuelve a aplicar un árbol de decisión.

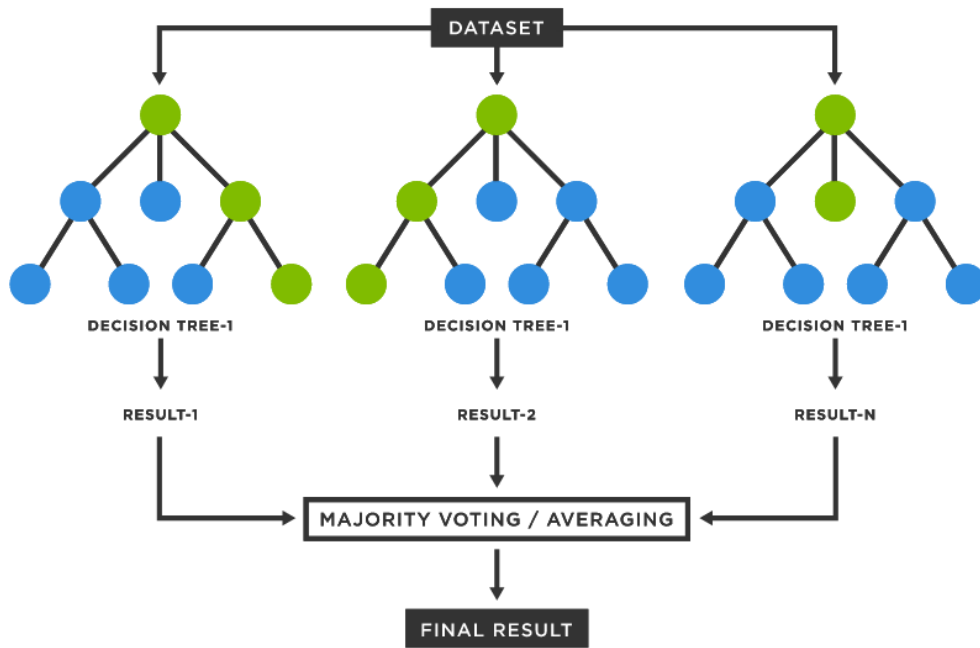


Ilustración 17. Bosque de decisión (¿Qué es un bosque aleatorio?, s. f.)

4.4.2.3. Predicción mediante redes neuronales

Es un método de la inteligencia artificial que enseña a los ordenadores a procesar datos de una manera que está inspirada en la forma en que lo hace el cerebro humano. Se trata de un tipo de proceso de Machine Learning llamado aprendizaje profundo, que utiliza los nodos o las neuronas interconectados en una estructura de capas que se parece al cerebro humano (Ovando et al., 2005). Crea un sistema adaptable que los ordenadores utilizan para aprender de sus errores y mejorar continuamente. De esta forma, las redes neuronales artificiales intentan resolver problemas complicados, como la realización de resúmenes de documentos o el reconocimiento de rostros, con mayor precisión. El conjunto de datos de entrenamiento permite que la red neuronal aprenda de forma automática, aquellos rasgos o variables que tienen interés en el modelo predictivo. El conjunto de validación entra en juego una vez el modelo ha realizado el aprendizaje en base a los inputs de entrada. Este subconjunto de datos permite evaluar el modelo, midiendo el porcentaje de acierto o fallo y realizando ajustes sobre el mismo.

Finalmente, el conjunto de datos de testeo o prueba es el que permite emplear el modelo para el fin que se ha construido, siendo en este caso el de predecir el dato de rendimiento de la campaña 2022 de maíz.

5. Resultados y discusión

Dentro de este apartado se muestra todo el proceso de trabajo, desde la obtención del conjunto de datos hasta la consecución de los objetivos.

- Construcción de un dataset que integra las variables climáticas, índices de vegetación y rendimientos históricas.
- Modelización predictiva del rendimiento. Con estos datos crearemos las entradas para que el modelo predictivo nos calcule nuestro objetivo, el rendimiento esperado en este caso.

5.1. Dataset desarrollado para el modelo predictivo

Todo el conjunto de datos obtenidos se va a utilizar como entrenamiento y validación del modelo predictivo. Estos datos o entradas se han obtenido de medios públicos y a partir de los históricos por contacto directo con el agricultor.

Como bien se ha comentado antes, el conjunto de datos presentaba una granularidad variable, ya que variaba la resolución temporal o frecuencia con la que se obtenía cada dato. Mientras había datos diarios para las variables climáticas, otras como los índices vegetativos, presentaban datos semanales y hasta anuales en el caso del rendimiento. Por este motivo, se ha decidido reunir los datos en quincenas y únicamente durante el crecimiento vegetativo del maíz, desde el 1 de abril hasta el 1 de octubre.

5.1.1. Datos climáticos

Para la construcción del conjunto de datos climáticos se han escogido las variables climáticas que tienen mayor relación con el rendimiento final del maíz. Pese a ser una planta C_4 (posee una mayor eficiencia en el uso del agua), el maíz es una planta subtropical de rápido crecimiento, por lo que requiere grandes necesidades de temperatura y agua para su adecuado desarrollo. No obstante, las temperaturas extremas, por debajo de 7°C y por encima de 38°C , tampoco le son favorable, ya que paralizan su crecimiento. Evidentemente, la sensibilidad a estas temperaturas extremas varía según el estadio de desarrollo del maíz, siendo crítico durante el período reproductivo o floración (Endicott et al., 2015).

Por todo ello, las variables elegidas han sido: temperaturas máximas, temperaturas mínimas, humedad relativa, precipitación y evapotranspiración ellas han sido calculadas como valores medios para la quincena, excepto la precipitación que es el acumulado de lluvia (Véase tabla 2). Obviamente la selección de cada una de estas no ha sido aleatoria, si no que se ha analizado cuidadosamente cuáles tienen mayor relación con la producción del maíz.

Tabla 2. Dataset climático agrupado por quincenas (Fuente: Elaboración propia)

Año	Quincena	Tª Máx (°C)	Tª Mín (°C)	Hum Med (%)	Precipit (mm)	Eto Med
2019	1-15 abril	19,95	-2,00	68,58	16,00	20,16
	15-30 abril	22,37	1,62	70,46	25,62	22,38
	1-15 mayo	24,02	2,61	59,85	2,46	31,17
	15-31 mayo	28,45	2,94	54,59	3,90	34,41
	1-15 junio	29,77	4,44	56,14	3,08	36,17
	15-31 junio	26,50	5,69	60,61	5,54	33,54
	1-15 julio	32,88	12,11	69,64	33,49	34,78
	15-31 julio	33,20	8,75	66,52	17,52	33,55
	1-15 agosto	31,26	7,35	67,97	5,74	29,07
	15-31 agosto	31,92	8,50	67,62	8,81	26,99
	1-15 sept	29,54	4,30	66,47	10,20	23,15
	15-30 sept	27,55	6,35	75,21	28,80	18,53
2020	1-15 abril	18,46	1,65	78,96	41,20	16,13
	15-30 abril	20,12	5,57	78,24	29,80	19,13
	1-15 mayo	25,39	5,53	74,16	36,80	23,60
	15-31 mayo	25,71	5,17	67,48	14,60	29,14
	1-15 junio	29,45	7,98	61,27	15,80	35,02
	15-31 jun	25,41	6,62	63,87	3,40	31,02
	1-15 jul	33,18	9,97	64,57	8,20	37,25
	15-31 jul	34,02	11,12	64,60	14,60	37,20
	1-15 agosto	33,87	10,13	68,44	13,20	30,94
	15-31 agosto	29,74	7,29	73,87	47,40	24,98
	1-15 sept	31,00	5,27	69,65	0,20	23,35
	15-30 sept	24,75	7,02	77,15	16,80	17,45
2021	1-15 abril	21,87	-1,93	63,77	13,20	20,18
	15-30 abril	19,09	1,55	68,62	18,20	19,75
	1-15 mayo	21,51	3,14	73,26	44,20	23,38
	15-31 mayo	23,00	1,66	65,12	13,60	27,27
	1-15 junio	27,86	4,93	64,19	21,00	32,08
	15-31 junio	30,61	8,18	68,44	58,20	33,62
	1-15 julio	28,86	5,69	67,08	4,60	32,38
	15-31 julio	32,07	7,01	62,31	0,60	35,33
	1-15 agosto	31,72	6,69	63,29	0,00	34,27
	15-31 agosto	30,99	7,78	64,35	0,00	29,85
	1-15 sept	30,76	8,41	73,55	12,40	22,54
	15-30 sept	24,38	3,90	78,98	57,00	17,45
2022	1-15 abril	19,15	-4,88	65,51	12,22	19,32
	15-30 abril	23,16	0,57	67,96	21,87	22,44
	1-15 mayo	25,44	2,58	65,90	3,55	26,19
	15-31 mayo	30,64	8,58	58,98	3,94	33,54
	1-15 junio	27,93	5,86	61,52	9,65	32,29
	15-31 junio	34,51	9,73	59,07	25,21	37,94
	1-15 julio	27,13	6,93	66,57	15,17	31,98
	15-31 julio	35,17	10,01	58,75	1,53	36,70
	1-15 agosto	35,30	11,95	59,89	10,04	34,79
	15-31 agosto	32,58	9,56	63,92	0,00	29,45
	1-15 sept	29,97	6,95	68,61	12,21	24,84
	15-30 sept	28,44	4,31	73,01	30,34	17,28

5.1.2. Datos vegetativos

A partir de las imágenes satelitales se han calculado los índices vegetativos a través de los cuales obtendremos los datos para el dataset. Como se ha explicado anteriormente, únicamente se han utilizado los índices NDVI y GNDVI por representar de manera más precisa el desarrollo vegetativo del maíz.

Los valores de NDVI están comprendidos entre -1 y 1. Sin embargo, para cualquier cultivo se puede afirmar que están entre 0 y casi 1. Valores negativos corresponderían a cuerpos con gran capacidad de refracción, tales como agua y metales, mientras que valores positivos en torno a 0 representan zonas de suelo desnudas. A medida que el cultivo crece, desarrolla más biomasa y es capaz de cubrir mayor área de suelo, por lo que los valores de NDVI también aumentan. En el caso del maíz, empieza con unos valores de 0,10 en la siembra, 0,15 en la emergencia y a partir de mediados de junio es cuando produce una explosión en el crecimiento (crecimiento exponencial y NDVI > 0,50). A mediados de julio es cuando alcanza la plena floración, con un valor NDVI > 0,70. Conforme el cultivo va llegando a la senescencia y perdiendo actividad vegetativa, los valores de NDVI también decrecen, hasta llegar a un valor de 0,15 aproximadamente, en el momento de la cosecha. Todos estos valores son recogidos en la tabla 3.

Tabla 3. Dataset de evolución de NDVI y GNDVI a lo largo de las campañas (Fuente: Elaboración propia)

Año	Mes	Abril		Mayo		Junio		Julio		Agosto		Septiembre	
		1-15	15-30	1-15	15-31	1-15	15-30	1-15	15-31	1-15	15-31	1-15	15-30
2019	NDVI	0,12	0,13	0,15	0,16	0,26	0,45	0,74	0,80	0,79	0,75	0,70	0,50
	GNDVI	0,20	0,20	0,20	0,20	0,30	0,30	0,50	0,50	0,50	0,40	0,40	0,30
2020	NDVI	0,09	0,09	0,10	0,15	0,30	0,65	0,79	0,81	0,80	0,79	0,76	0,57
	GNDVI	0,10	0,10	0,20	0,20	0,30	0,30	0,50	0,50	0,60	0,40	0,40	0,30
2021	NDVI	0,10	0,12	0,13	0,17	0,35	0,68	0,80	0,82	0,81	0,77	0,62	0,45
	GNDVI	0,20	0,20	0,20	0,20	0,30	0,30	0,50	0,70	0,60	0,40	0,40	0,30
2022	NDVI	0,12	0,13	0,13	0,27	0,45	0,76	0,79	0,80	0,76	0,71	0,47	0,25
	GNDVI	0,20	0,20	0,20	0,20	0,30	0,30	0,60	0,50	0,50	0,40	0,30	0,20

Para verlo de manera más visual se presenta la ilustración 18 en la que aparece la imagen satelital de cada campaña de mediados de julio, que es cuando el maíz suele alcanzar su NDVI máximo, período inmediatamente posterior a la plena floración (NDVI ≈ 0,80). Es en este periodo cuando la planta de maíz alcanza su máximo crecimiento vegetativo y, por tanto, la cobertura del suelo es máxima.



Ilustración 18. Evolución del NDVI a lo largo de las diferentes campañas (Fuente: Elaboración propia)

Como se puede apreciar en la comparación de las distintas gráficas y por el hecho de ser el maíz un cultivo que basa su desarrollo en la acumulación de grados día, en el año 2022 se alcanzaron unas temperaturas estivales muy altas y durante bastantes días, lo que provocó que el maíz acumulara más grados día en menos tiempo. Esto se traduce en una curva NDVI con una pendiente más pronunciada que explicaría ese crecimiento tan rápido del cultivo.

5.1.3. Datos de rendimientos históricos

Para poder modelizar con un sistema de aprendizaje supervisado es necesario tener un dataset etiquetado que permita este modelo de pregunta-respuesta conocidas y se pueda extraer un conocimiento en base a estos datos.

El rendimiento medio de las últimas campañas ha sido facilitado por el agricultor. Por consiguiente, se van a utilizar los datos de los años 2019, 2020 y 2021 como entrenamiento para validar el modelo y el rendimiento real del 2022 para comprobar la predicción, pudiendo así calcular el error de esta, tal y como se muestra en la tabla 4.

Tabla 4. Datos de los rendimientos históricos de cada campaña expresados en kg secos/ha (Fuente: Elaboración propia)

Entrenamiento	2019	12126
	2020	16246
	2021	15490
Validación	2022	15601

En un futuro inmediato/presente se podría optimizar la predicción de este modelo con la información que facilitan los mapas de rendimiento, es decir, la disponibilidad del dato de rendimiento georreferenciado. Los mapas de rendimiento permiten determinar la variabilidad espacial de la parcela y poder así realizar aplicaciones de inputs variables en función del potencial de la zona. Cuanta más cantidad de mapas de rendimiento vayamos teniendo, mayor conocimiento tendremos de la variabilidad de la parcela y podremos así optimizar cada operación. En este caso, y solo como curiosidad se dispone del mapa de rendimiento de la última campaña

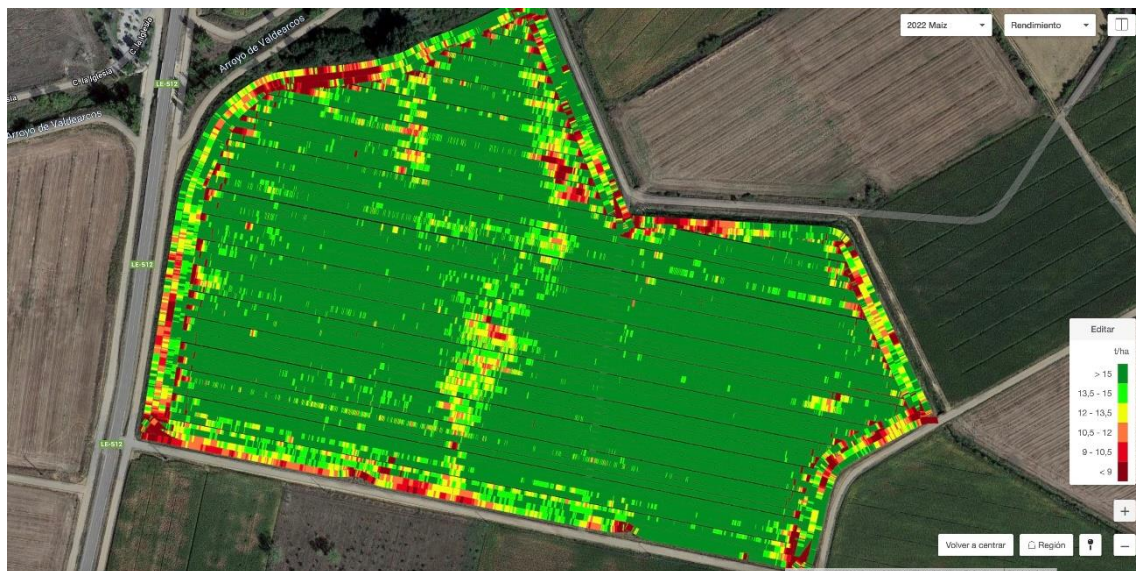


Ilustración 19. Mapa de rendimiento variable de maíz de la cosecha 2022 (Fuente: Mapa rendimiento John Deere Harvest Doc™)

En la ilustración 19 se aprecia la variabilidad de rendimiento en función de una escala de colores que representa una serie de rangos de rendimiento. Concretamente, según la paleta de colores elegida, colores más verdes oscuro representan rendimientos más elevados, mientras que tonos más rojizos, rendimientos más bajos. La obtención de estos datos se realiza a través de una cosechadora georreferenciada con un medidor de caudal instantáneo del maíz cosechado. Los sensores están situados en la garganta de la máquina y miden el caudal de grano que pasa por la misma. De esta manera, si conocemos la posición de la máquina, podremos conocer el rendimiento de cada punto de la parcela. Evidentemente, los sensores deberán estar adecuadamente calibrados con los datos de humedad y peso específico para no falsear el resultado del mapeo.

5.2. Modelado predictivo del rendimiento

5.2.1. Análisis de componentes principales (PCA en inglés)

Con este tipo de análisis se trata de encontrar qué tipo de variables tienen mayor influencia sobre el rendimiento final. El modelo permite generar varias componentes principales en función de la varianza acumulada que establezcamos con un selector manual. Las componentes principales son las combinaciones lineales de las variables originales que explican la varianza en los datos. El número máximo de componentes extraídos siempre es igual al número de variables. Los vectores propios, compuestos por los coeficientes que corresponden a cada variable, se utilizan para calcular las puntuaciones de los componentes principales. Los coeficientes indican la ponderación relativa de cada variable en el

componente. Cuanto mayor sea el valor absoluto del coeficiente, más importante será la variable correspondiente en el cálculo del componente (Interpretar todos los estadísticos y gráficas para Análisis de los componentes principales - Minitab, s. f.).

En este caso hemos cogido un total de 5 componentes principales, siendo la componente principal 1 la combinación de variables que más explica la varianza existente con un 25,49%. La ilustración 20 recoge los resultados del PCA.

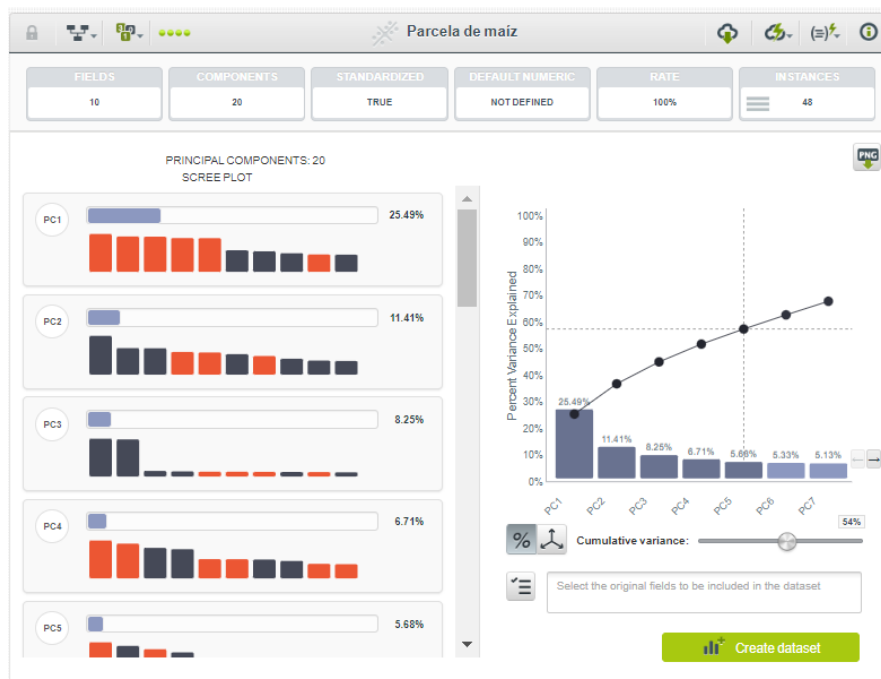


Ilustración 20. Resultado gráfico del análisis de componentes principales (Bigml.com, s/f)

Las cinco variables que cuentan con un mayor peso, ordenadas de más a menos importancia, han sido:

- Temperaturas máximas
- NDVI
- GNDVI
- Eto media
- Temperaturas mínimas

El resultado obtenido en el análisis de componentes principales da una idea de qué factores están más relacionados con el rendimiento final. Como se puede apreciar, la temperatura máxima es el factor más importante según el PCA 1, lo que es perfectamente posible, ya que como se ha explicado antes, el maíz es una planta que necesita acumular una serie de GDD para completar su ciclo de vida. La segunda y tercera variable se corresponde a los índices de vegetación, que dan una idea la cantidad de biomasa o crecimiento vegetativo existente, por lo que también están estrechamente relacionados con el rendimiento final. En cuanto a la Eto es una variable que está directamente con la temperatura, por lo que era de suponer que tuviera una relación similar con el rendimiento. De la misma manera, las temperaturas mínimas también determinan de manera indirecta la integral térmica del maíz, por lo que era de suponer su importancia.

5.2.2. Árboles de decisiones

Un árbol de decisión en Machine Learning es una estructura similar a un diagrama de flujo donde un nodo interno representa una característica, la rama representa una regla de decisión y cada nodo hoja representa el resultado. El nodo superior en un árbol de decisión en Machine Learning se conoce como el nodo raíz. Es una visualización como un diagrama de flujo que imita fácilmente el pensamiento a nivel humano, por lo que son fáciles de entender e interpretar. Los árboles de decisión clasifican los ejemplos clasificándolos por el árbol desde la raíz hasta algún nodo hoja, con el nodo hoja proporcionando la clasificación al ejemplo, este enfoque se llama Enfoque de arriba hacia abajo (Data, 2022).

En este caso, se ha utilizado un árbol de decisión para la predicción del rendimiento final. Las hojas verdes del árbol son las producciones de los distintos años, que están unidas con la predicción del rendimiento final esperado para la campaña 2022, en este caso 14621kg/ha. La utilización de árboles de decisión es la forma más simple y, por tanto, se supone que menos precisa para la estimación de la producción final tal y como se muestra en la ilustración 21.

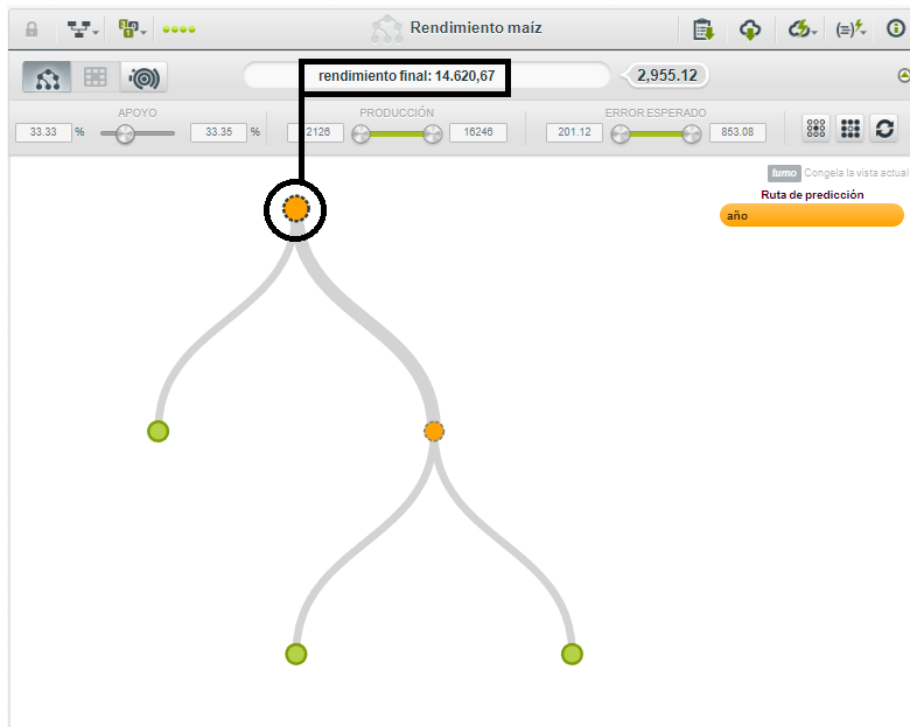


Ilustración 21. Predicción simple del rendimiento final mediante un árbol de decisión (Bigml.com, s/f)

5.2.3. Bosques de decisiones o modelos ensamblados

El ensamblado de modelos (árboles de decisión) distintos genera predicciones más robustas. Los grupos de árboles de clasificación se combinan y se deduce una única predicción votada en democracia por la población de árboles. Evidentemente, la predicción del bosque de decisiones ha de ser más precisa que el caso anterior. En este caso, el programa hace una predicción para el rendimiento final del año 2022 de 15260 kg/ha. En la parte de la derecha de la ilustración 22 se pueden seleccionar los distintos años y verse el peso de cada dato sobre la producción final.

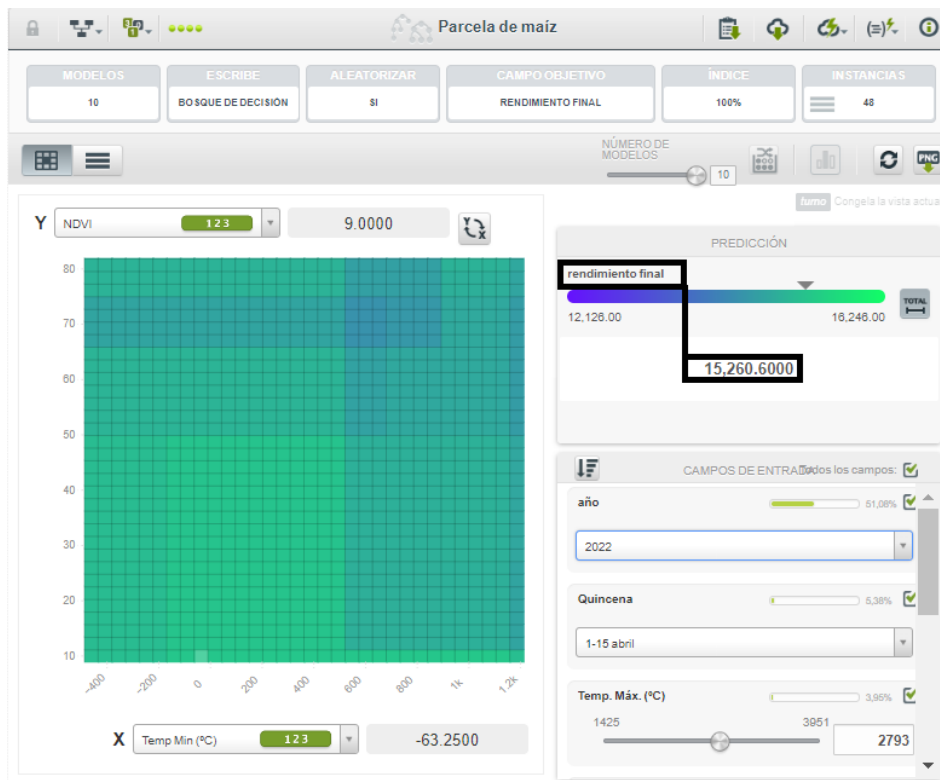


Ilustración 22. Predicción del rendimiento final mediante bosques de decisiones (Bigml.com, s/f)

5.2.4. Predicción con redes neuronales

Mediante la utilización de las redes neuronales se ha generado un nuevo modelo para la predicción del rendimiento final esperado. En la descripción de la red neuronal se ha empleado el algoritmo de ADAM (ADaptive Moment estimation) de carácter estocástico creando la red neuronal con 2 capas ocultas y una tasa de aprendizaje del 0,01 (Kingma & Lei Ba, 2015). En este caso los datos ordenados que toman mayor peso sobre el resultado final son:

- Evapotranspiración media
- GNDVI
- Temperaturas mínimas
- NDVI
- Temperaturas máximas

De nuevo, vuelven a destacar las temperaturas e índices vegetativos en el modelo, lo que garantiza el papel que juegan en la producción final.

La predicción obtenida con el empleo de redes neuronales ha sido de 15527 kg/ha tal y como se muestra en la Ilustración 23.

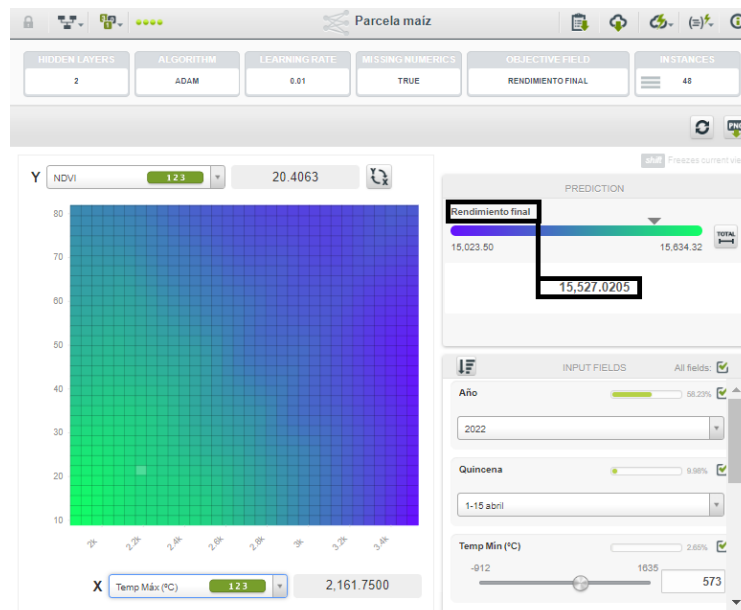


Ilustración 23. Predicción del rendimiento con la utilización de una red neuronal (Bigml.com, s/f)

5.2.5. Evaluación de los modelos

Una vez realizadas todas las predicciones anteriores hay que comparar cada una de las mismas con el dato real. De esta manera, se va a determinar el error de cada una de ellas respecto a la producción real y, por tanto, la precisión de cada método para la consecución de dicho objetivo.

Para el cálculo del error se ha restado la producción real de la predicha y después se ha calculado el porcentaje de error, en valor absoluto, sobre la producción real.

Como se puede observar en la tabla 5, el método más preciso ha sido el de redes neuronales, mientras que el de árbol de decisiones ha sido el menos certero. También puede apreciarse que todas las predicciones han sido inferiores al valor real, por lo que podríamos vaticinar que el tiempo pasado desde el último dato cogido de cada año hasta el momento de cosecha ha sido muy favorable.

Tabla 5. Resultados comparativos de cada predicción con la producción real (Fuente: Elaboración propia)

Predicciones	Árbol de decisiones	Modelos ensamblados	Redes neuronales
	14621 kg/ha	15260 kg/ha	15527 kg/ha
Producción Real	15601 kg/ha		
Diferencia	-980 kg/ha	-341 kg/ha	-74 kg/ha
/ Error (%) /	6,28%	2,19%	0,47%

Como puede observarse, cuanto más complejo ha sido el método utilizado, mayor ha sido su grado de precisión. No obstante, se podría decir que el porcentaje de error es muy bajo en los modelos ensamblados y redes neuronales y medio-bajo en el árbol de decisiones.

De cara a predicciones futuras, sería interesante disponer de mapas de rendimiento para aumentar la fiabilidad del sistema. Además, cuanto más amplio sea el histórico que dispongamos, más fácil será afinar la precisión de cada método.

6. Conclusiones

Una vez realizado el estudio y alcanzados los objetivos que se pretendían, se han extraído las siguientes conclusiones:

- El cultivo del maíz tiene una relación directa con las variables climáticas que contribuyen directamente al rendimiento final, sobre todo aquellas que influyen sobre su integral térmica.
- También se ha plasmado la utilidad de las imágenes por satélite y los índices de vegetación en el cultivo del maíz que son capaces de anticipar un mapa de rendimiento a la cosecha. Por lo general, una imagen por satélite de cualquier cultivo en el momento de pleno apogeo (normalmente a partir de la floración) tiene una correlación directa con un mapa de rendimiento, siempre y cuando la senescencia del cultivo entre dentro de la normalidad.
- El uso de herramientas de Machine Learning otorga la capacidad de predecir las producciones con gran fiabilidad momentos antes de la cosecha. Esto puede ayudar a realizar predicciones, manejos logísticos y demás variables que influyan con el volumen de cosecha.
- Quizá en un futuro más bien cercano y con la llegada de la Agricultura 4.0 se produzca un intercambio permanente de datos a tiempo real que simplifique aún más todo este tipo de predicciones. La utilización de mapas de rendimiento, aplicaciones variables de insumos y de manejo... permite optimizar todos los recursos existentes a la vez que se aumenta la competitividad y el cuidado del medio ambiente.
- Se podría afirmar que la predicción con cualquiera de los modelos utilizados tiene una alta fiabilidad en relación con los resultados reales. No obstante, el método que más se ha aproximado al rendimiento final, con un error muy por debajo del 1% (0,47%) ha sido la predicción mediante redes neuronales, lo que da una idea de la importancia y fiabilidad que tienen estas herramientas de inteligencia artificial.

7. Bibliografía

- Acuña-Rello, L., Spavento, E., Casado-Sanz, M., Basterra, L.-A., López-Rodríguez, G., Ramón-Cueto, G., Relea-Gangas, E., Morillas-Romero, L., Escolano-Margarit, D., Martínez, R. D., & Balmori, J. A. (2022). Assessment of machine learning algorithm-based grading of *Populus x euramericana* I-214 structural sawn timber. *Engineering Structures*, 254(113826), 113826. <https://doi.org/10.1016/j.engstruct.2021.113826>
- Alonso, D. (2023). Los 6 Índices de Vegetación para completar el NDVI. MappingGIS. <https://mappinggis.com/2020/07/los-6-indices-de-vegetacion-para-completar-el-ndvi/>
- Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, 21(11), 3758. <https://doi.org/10.3390/s21113758>
- Bezares Sanfelip, F. (2019). Herramienta de segmentación automática de masas forestales a partir de sensores remotos: LiDAR y Sentinel 2. Universidad de Valladolid.
- Bigml.com. (s/f). Bigml.com - Machine Learning made easy. Recuperado el 25 de marzo de 2023, de <https://bigml.com/>
- Colaboradores de Wikipedia. (2023). Aprendizaje automático. *Wikipedia, la enciclopedia libre*. https://es.wikipedia.org/wiki/Aprendizaje_autom%C3%A1tico
- Data, S. B. (2022, 26 septiembre). Árbol de decisión en Machine Learning (Parte 1) - sitiobigdata.com. [sitiobigdata.com. https://sitiobigdata.com/2019/12/14/arbol-de-decision-en-machine-learning-parte-1/#](https://sitiobigdata.com/2019/12/14/arbol-de-decision-en-machine-learning-parte-1/#)
- Dietterich, T. G. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Machine Learning*, 40(2), 139–157. <https://doi.org/10.1023/A:1007607513941>
- Endicott, S., Brueland, B., Keith, R., Schon, R., Bremer, C., Farnham, D., DeBruin, J., Clausen, C., Strachan, S., & Carter, P. (2015). Corn growth and development. Dupont Pioneer, https://www.pioneer.com/CMRoot/International/Latin_America_Central/Chile/Servicios/Informacion_tecnica/Corn_Growth_and_Development_Spanish_Version.pdf.
- Fertilización en maíz: La necesidad de magnesio. (s. f.). <https://www.fertiberia.com/es/blog/2017/marzo/fertilizacion-en-maiz-la-necesidad-de-magnesio/>
- Fuentes, J.L., 1999. El Suelo y los Fertilizantes. MAPA
- Gavilán, V., Lillo-Saavedra, M., Holzapfel, E., Rivera, D., & García-Pedrero, A. (2019). Seasonal crop water balance using harmonized Landsat-8 and Sentinel-2 time series data. *Water*, 11(11), 2236. <https://doi.org/10.3390/w11112236>
- Gilmore, E. C., J. S. Rogers. 1958. Heat Units as a Method of Measuring Maturity in Corn. *Agronomy Journal*, Vol. 50 No. 10, p. 611-615. College Station, TX.
- InfoRiego. (s/f). Inforiego. [Inforiego.org. Recuperado el 25 de marzo de 2023, de https://www.inforiego.org/opencms/opencms/info_meteo/construir/index.html](https://www.inforiego.org/opencms/opencms/info_meteo/construir/index.html)

Interpretar todos los estadísticos y gráficas para Análisis de los componentes principales - Minitab. (s. f.). (C) Minitab, LLC. All rights Reserved. 2023. <https://support.minitab.com/es-mx/minitab/21/help-and-how-to/statistical-modeling/multivariate/how-to/principal-components/interpret-the-results/all-statistics-and-graphs/#:~:text=La%20gr%C3%A1fica%20de%20sedimentaci%C3%B3n%20muestra%20el%20%C3%BAmero%20del%20componente%20principal,varianzas%20de%20los%20componentes%20principales.>

Kingma, D. P., & Lei Ba, J. (2015). ADAM: A Method for stochastic optimization. ICLR 2015, <https://arxiv.org/pdf/1412.6980.pdf>.

Lobell, D. B. (2013). The use of satellite data for crop yield gap analysis. Field Crops Research. <https://doi.org/10.1016/j.fcr.2012.08.008>

Mapa rendimiento John Deere Harvest Doc™

Marcos Robles, J. L. (2016). El control de la supervivencia en plantaciones agroforestales con waterboxx mediante vehículos aéreos no tripulados e información multiespectral. <http://uvadoc.uva.es/handle/10324/16511>

Máster Ingeniería Agronómica, ETS Ingenierías Agrarias (Palencia). Apuntes.

Muñoz Herranz, G. (2021). Caracterización agronómica y espectral de un viñedo situado en la denominación de origen Ribera del Duero. Universidad de Valladolid.

Ovando, G., Bocco, M., & Sayago, S. (2005). REDES NEURONALES PARA MODELAR PREDICCIÓN DE HELADAS. Agricultura Técnica. <https://doi.org/10.4067/s0365-28072005000100007>

Pioneer- Corteva. Apuntes técnicos internos cultivo maíz.

¿Qué es un bosque aleatorio? (s. f.). TIBCO Software. <https://www.tibco.com/es/reference-center/what-is-a-random-forest>

Servicio agronómico de Pioneer

Sistema de Información Agroclimática para el Regadío. (s. f.). <https://portal.mapa.gob.es/websiar/ResultadoConsultaDatos.aspx>.

Teledetección. (s. f.). <https://www.um.es/geograf/sig/teledet/fundamento.html>

Valero, C. (2019). La evolución de la agricultura de precisión. Fruticultura, 68, 30-34. https://oa.upm.es/56889/1/Fruticultura_N_68_pags_30_39.pdf

Visor SigPac. (s/f). Gob.Es. Recuperado el 25 de marzo de 2023, de <https://sigpac.mapama.gob.es/fega/visor/>

Zeus. (s. f.). La relevancia de la granularidad del dato en los negocios – Upicus. <https://www.upicus.com/granularidad-del-dato/>

Zubcoff, J. J., & de Estadística, A. (s/f). Principal Component Analysis. Rua.ua.es. Recuperado el 25 de marzo de 2023, de <https://rua.ua.es/dspace/bitstream/10045/72567/1/Analisis-multivariante-PCA.pdf>