

Trabajo Fin de Máster

Máster en Ingeniería Electrónica, Robótica y Automática

Diseño de sistema de inteligencia artificial para la implantación de un modelo de agricultura eco sostenible empleando técnicas de aprendizaje automático.

Autor: Lucía Baeza Moreno

Tutor: Ramón González Carvajal

Cotutor: Rubén Martín Clemente

Dpto. Ingeniería Electrónica
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla



Sevilla, 2021



Trabajo Fin de Máster
Máster en Ingeniería Electrónica, Robótica y Automática

Diseño de sistema de inteligencia artificial para la implantación de un modelo de agricultura eco sostenible empleando técnicas de aprendizaje automático

Autor:

Lucía Baeza Moreno

Tutor: Ramón González Carvajal

Cotutor: Rubén Martín Clemente

Dep. Ingeniería Electrónica

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2021

Trabajo Fin de Máster: Diseño de sistema de inteligencia artificial para la implantación de un modelo de agricultura eco sostenible empleando técnicas de aprendizaje automático.

Autor: Lucía Baeza Moreno

Tutor: Ramón González Carvajal

Cotutor: Rubén Martín Clemente

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2021

El Secretario del Tribunal

A mi familia

Agradecimientos

A mis padres, por darme su cariño y apoyo incondicional.

A Ramón González, mi tutor, por creer y apostar por mi para el desarrollo de este proyecto.

A Rubén Martín, por brindarme su ayuda y compartir sus conocimientos.

A Pedro Blanco, por ser el mejor compañero que se puede tener.

Lucía Baeza Moreno

Sevilla, 2021

Resumen

En este proyecto se ha desarrollado un sistema de inteligencia artificial para la implantación de un nuevo modelo productivo de agricultura ecosostenible.

Entre las tareas a realizadas, se encuentra la elaboración de un algoritmo de aprendizaje automático que analiza datos de sensores instalados en una finca experimental y que simula el conocimiento de los expertos para informar al agricultor de las distintas situaciones que se estén dando en las plantaciones. Se analizan datos de cultivos como el almendro o tomate, y se miden magnitudes como la temperatura ambiental, cantidad de precipitaciones, humedad, presión o salinidad.

Con estas medidas y su análisis, se han encontrado relaciones que han permitido predecir riesgos bióticos como plagas o enfermedades y riesgos abióticos como estrés hídrico, estrés por radiación solar, estrés por salinidad, escaso desarrollo vegetativo o fallo en el cuajado de frutos. El algoritmo alerta al agricultor de los posibles riesgos que puedan darse en el futuro anticipándose y de esta forma permite evitar posibles daños en los cultivos. También, dependiendo del riesgo, se hace la recomendación correspondiente que señalen los especialistas.

Asimismo, se han analizado datos relacionados con el manejo de los cultivos, tales como el volumen, la dosis y el número de aplicaciones de los insumos agrícolas. Se ha buscado modelar su relación con la producción, el peso, el calibre de los frutos o el volumen entre otras magnitudes.

Otro de los objetivos que se han contemplado en el desarrollo de este trabajo es la determinación la sostenibilidad del cultivo analizando las medidas de los sensores. Se han prestado principal atención a los colocados a distintos niveles de profundidad junto a las raíces de los cultivos, pues proporcionan la información si sobre los fertilizantes aplicados se están adhiriendo únicamente a la planta o si por el contrario se está produciendo algún filtrado al suelo y como consecuencia el cultivo es poco sostenible. La función del sistema de inteligencia artificial será evaluar la situación y avisar a el agricultor para que tome las medidas oportunas que indiquen los expertos.

Abstract

In this project, an artificial intelligence system has been developed for the implementation of a new eco-sustainable agriculture production model.

Among the tasks to be carried out, is the development of a machine learning algorithm that analyzes data from sensors installed in an experimental farm and that simulates the knowledge of the experts to inform the farmer of the different situations that are occurring in the plantations. Data from crops such as almond or tomato are analyzed, and magnitudes such as environmental temperature, amount of rainfall, humidity, pressure or salinity are measured.

With these measures and their analysis, relationships have been found that have made it possible to predict biotic risks such as pests or diseases and abiotic risks such as water stress, solar radiation stress, salinity stress, poor vegetative development or fruit set failure. The algorithm alerts the farmer of the possible risks that may arise in the future by anticipating and in this way allows to avoid possible damage to the crops. Also, depending on the risk, the corresponding recommendation indicated by the specialists is made.

Likewise, data related to crop management have been analyzed, such as volume, dose and number of applications of agricultural inputs. It has been sought to model its relationship with production, weight, fruit size or volume, among other magnitudes.

Another objective that has been considered in the development of this work is to determine the sustainability of the crop by analyzing the measurements of the sensors. Main attention has been paid to those placed at different depth levels next to the roots of the crops, since they provide information if the applied fertilizers are adhering only to the plant or if, on the contrary, some filtration is taking place to the soil and as a consequence the cultivation is not very sustainable. The function of the artificial intelligence system will be to evaluate the situation and notify the farmer to take the appropriate measures indicated by the experts.

Índice

Agradecimientos	x
Resumen	xii
Abstract	xiv
Índice.....	xvi
ÍNDICE DE TABLAS	xix
ÍNDICE DE FIGURAS	xxi
Notación	xxiv
1 INTRODUCCIÓN Y OBJETIVOS.....	1
1.1 Introducción	1
1.2 Objetivos.....	1
2 ESTADO DE LA TÉCNICA	3
2.1 Modelos estrés hídrico.....	3
2.2 Algoritmos de Machine Learning	5
3 METODOLOGÍA.....	9
3.1 Instalación de sensores.....	10
3.2 Fuentes de datos.....	14
3.2.1 Odin.....	14
3.2.2 Estaciones climáticas.....	16
3.2.3 Ephem.....	18
3.3 Entorno de desarrollo.....	19
3.4 Algoritmos.....	20
3.4.1 Árboles de decisión.....	20
3.4.2 Redes Neuronales Artificiales.....	22
4 IMPLEMENTACIÓN	27
4.1 Entradas.....	27
4.2 Salidas.....	28
4.3 Scripts.....	28
4.3.1 Entrenamiento riesgos bióticos.....	30
4.3.2 Entrenamiento riesgos abióticos.....	32
4.3.3 Ejecución riesgos bióticos.....	33
4.3.4 Ejecución riesgos abióticos.....	34
4.3.5 Salidas scripts.....	35
5 RESULTADOS	37
5.1 Predicción de riesgos bióticos y abióticos	37

5.2	Análisis de insumos	46
5.3	Estudio variables de manejo de tomate Rubenza	48
6	CONCLUSIONES	52
6.1	Conclusiones	52
6.2	Trabajos futuros.....	52
	REFERENCIAS.....	53

ÍNDICE DE TABLAS

Tabla 1 Índices espectrales de vegetación.	5
Tabla 2 Algoritmos de ML aplicados a agricultura.	6
Tabla 3 Precisiones de los distintos algoritmos puestos a prueba.	39

ÍNDICE DE FIGURAS

Figura 1 Algoritmo SVM.....	7
Figura 2 Características pluviómetro.....	10
Figura 3 Características barómetro.....	11
Figura 4 Características termómetro.....	11
Figura 5 Características higrómetro.....	12
Figura 6 Características de contador de pulsos.....	12
Figura 7 Características de sonda.....	13
Figura 8 Banco de ensayos.....	14
Figura 9 Interfaz de los sensores.....	15
Figura 10 Fichero json con dispositivos de la finca.....	15
Figura 11 Datos del dispositivo “Presión de vapor”.....	16
Figura 12 Interfaz web IFAPA.....	17
Figura 13 Modelo IFAPA para datos diarios.....	18
Figura 14 Modelo IFAPA para datos mensuales.....	18
Figura 15 Cálculo de las horas diarias de sol con la librería Ephem.....	19
Figura 16 Entorno de programación utilizado.....	19
Figura 17 Jupyter notebook.....	20
Figura 18 Diagrama de flujo del diseño de la inteligencia artificial.....	21
Figura 19 Representación esquemática de una neurona artificial.....	23
Figura 20 Red neuronal multicapa.....	23
Figura 21 Diagrama de flujo para la determinación de riesgos bióticos.....	25
Figura 22 Diagrama de flujo para la determinación de riesgos abióticos.....	25
Figura 23 Entradas json.....	27
Figura 24. Organigrama generación/uso de los modelos de la plataforma.....	29
Figura 25. Esquema entradas/salidas de los modelos.....	30
Figura 26. Organigrama script entrenamiento riesgos bióticos.....	31
Figura 27. Organigrama script entrenamiento riesgos abióticos.....	32
Figura 28. Organigrama script ejecución riesgos bióticos.....	33
Figura 29. Organigrama script ejecución riesgos abióticos.....	34
Figura 30 Ficheros json de salida.....	36
Figura 31 Árbol de decisión para riesgos bióticos.....	37
Figura 32. Ejemplo de evaluación de riesgos. Comparación entre el estimado por los técnicos y el determinado por la inteligencia artificial. Se observa una coincidencia completa.....	38

Figura 33 ANN Keras - Multi Class Entropy Loss Precisión ANN (Precisión 97.57%).....	39
Figura 34. ANN Keras - Sparse multiclass cross entropy loss (Precisión ANN (97.95%).....	39
Figura 35 Ejemplo de evaluación de riesgos abióticos.	40
Figura 36. Árbol de decisión para la evaluación de riesgos abióticos.	40
Figura 37. Ejemplo de correo al agricultor informando de la evaluación de riesgos bióticos.	41
Figura 38. Análisis automatizado de las medidas de los sensores instalados en La Añoreta.....	42
Figura 39 Ejemplo de correo día 20/03/2021.....	43
Figura 40 Óptimo desarrollo vegetativo sin afectaciones bióticas.....	43
Figura 41 Correo con alertas consistentes de riesgo medio 21/03/2021.	44
Figura 42 Primeros síntomas de planta debilitada por acción de pulgón.	44
Figura 43 Correo con alertas sostenidas de riesgo alto 24-30/03/2021.....	45
Figura 44 Planta severamente afectada por la población de pulgón.	45
Figura 45. Identificación de insumo en las medidas de humedad a distintas profundidades.	46
Figura 46. Ejemplo de correo al agricultor informando sobre los insumos del entorno A.	47
Figura 47. Ejemplo de correo al agricultor informando sobre los insumos del entorno B.....	48
Figura 48. Gráfica correlación entre peso y calibre del fruto para la Tesis 1.....	49
Figura 49. Gráfica correlación entre peso del fruto y días del experimento para la Tesis 1.	49
Figura 50. Gráfica correlación entre calibre del fruto y aplicaciones de los insumos para la Tesis 2.	50
Figura 51. Resultados primer algoritmo.....	50
Figura 52. Resultados segundo algoritmo.....	51
Figura 53. Predicción del calibre de fruto empleando el segundo algoritmo.....	51

Notación

ML	Machine Learning
ET	Evo transpiración
PM	Penman-Monteith
CWSI	Crop Water Stress Index
TDR	Time Domain Reflectometry
FDR	Frequency Domain Reflectometry
RF	Random Forest
DANS	Degrees Above Non-Stressed
DACT	Degrees Above Canopy Threshold
VPD	Vapour-pressure déficit
SVM	Support Vector Machine

1 INTRODUCCIÓN Y OBJETIVOS

1.1 Introducción

En el futuro, la humanidad deberá enfrentarse a grandes desafíos, en los cuales la agricultura ocupará un papel central. El sistema alimentario mundial tendrá que proporcionar sustento a una población que pasará de siete a casi diez mil millones de personas en 2050. Al mismo tiempo, aumentar la producción de las cosechas será más costoso, puesto que el calentamiento global, la escasez de agua y la urbanización masiva dificultarán la consecución de este objetivo [1].

El calentamiento global trae consigo olas de calor y nuevas plagas y enfermedades, por lo que los cultivos, que están en constante exposición a los cambios medioambientales, se verán notablemente perjudicados. Su productividad y supervivencia peligran, ya que podrán darse situaciones a las que los cultivos no puedan adaptarse, estresándolos. Existe una amplia gama de estreses tanto bióticos como abióticos. El estrés biótico es causado por organismos vivos (hongos, virus y bacterias ...), mientras que el estrés abiótico se atribuye a condiciones físicas y/o químicas del ambiente.

Para conseguir una determinada productividad de los cultivos se necesita recurrir al empleo de fertilizantes. Estos se aplican con el fin de que los suelos tengan un contenido adecuado de minerales y, de esta forma, se proporcione a los vegetales los nutrientes necesarios para estimular su crecimiento. Por tanto, los fertilizantes aseguran que la producción de alimentos sea suficiente para abastecer a una población mundial en crecimiento constante. Asimismo, ayudan a crear una cubierta vegetal que mantiene la humedad y previene la erosión del suelo, reduciendo la deforestación y contribuyendo a reducir el cambio climático. Sin embargo, se ha de evitar el uso en exceso de fertilizantes no sostenibles, ya que el nutriente sobrante puede contaminar las aguas superficiales o subterráneas.

La digitalización del mundo rural y el aprovechamiento de la tecnología en el mundo agrícola es uno de los máximos objetivos de la sociedad actual ante un escenario de crecimiento poblacional, cambio climático y limitaciones de insumos aplicables en agricultura por su impacto en el medio y una de las tecnologías más prometedoras para hacer frente a los problemas agrícolas mencionados anteriormente es el aprendizaje automático o *Machine Learning* (ML). Se espera de éste que impulse la productividad de los cultivos, al tiempo que minimiza su impacto ambiental **¡Error! No se encuentra el origen de la referencia..** El ML proporciona una variedad amplia de herramientas para el análisis de datos [3] [4]. Estos datos, que serán proporcionados por muchos tipos de sensores embebidos en los cultivos, van a permitir una mejor comprensión del entorno y de las operaciones agrícolas. La gestión de las explotaciones pasará a estar basada, por lo tanto, en recomendaciones en tiempo real elaboradas por una inteligencia artificial [5][6]. Así, por ejemplo, mediante técnicas de ML será posible predecir las situaciones de estrés que afectan a la planta o ciertos factores de calidad de los cultivos, tales como el peso y calibre de los frutos, proporcionando de esta forma a los agricultores herramientas con las que mejorar el rendimiento de sus cosechas [7] .

Este documento presenta el análisis realizado para seleccionar algoritmos que, analizando datos obtenidos a través de sensores, generen información útil para las explotaciones agrícolas. Además de estudiar técnicas de aprendizaje automático, se tendrán en cuenta herramientas de análisis de series temporales, las características estadísticas de las señales y su evolución, distintos algoritmos de clasificación, etc. Se optará por aquellas soluciones que, proporcionando un nivel suficiente de precisión, sean capaces de satisfacer las necesidades particulares de cada caso de estudio considerado.

1.2 Objetivos

El objetivo del proyecto consiste en la elaboración de un modelo de inteligencia artificial capaz de realizar una predicción sobre los posibles riesgos de estreses bióticos y abióticos que pueden darse a partir de las medidas climáticas que aportan sensores instalados. De esta forma se tiene una herramienta con la que asesorar al agricultor en la toma de decisiones de sus cultivos.

Además, también se pretende analizar datos relacionados con el manejo de los cultivos, tales como el volumen, la dosis y el número de aplicaciones de los insumos agrícolas y de esta forma modelar su relación con la producción, el peso, el calibre de los frutos o el volumen entre otras magnitudes.

El último de los objetivos consiste es la determinación la sostenibilidad del cultivo analizando las medidas de los sensores. Se han prestado principal atención a los colocados a distintos niveles de profundidad junto a las raíces de los cultivos, pues proporcionan la información si sobre los fertilizantes aplicados se están adhiriendo únicamente a la planta o si por el contrario se está produciendo algún filtrado al suelo y como consecuencia el cultivo es poco sostenible. La función del sistema de inteligencia artificial será evaluar la situación y avisar a el agricultor para que tome las medidas oportunas que indiquen los expertos.

2 ESTADO DE LA TÉCNICA

En este capítulo se contempla algunas de las herramientas y métodos del ámbito de la inteligencia artificial aplicados el sector de la agricultura.

2.1 Modelos estrés hídrico

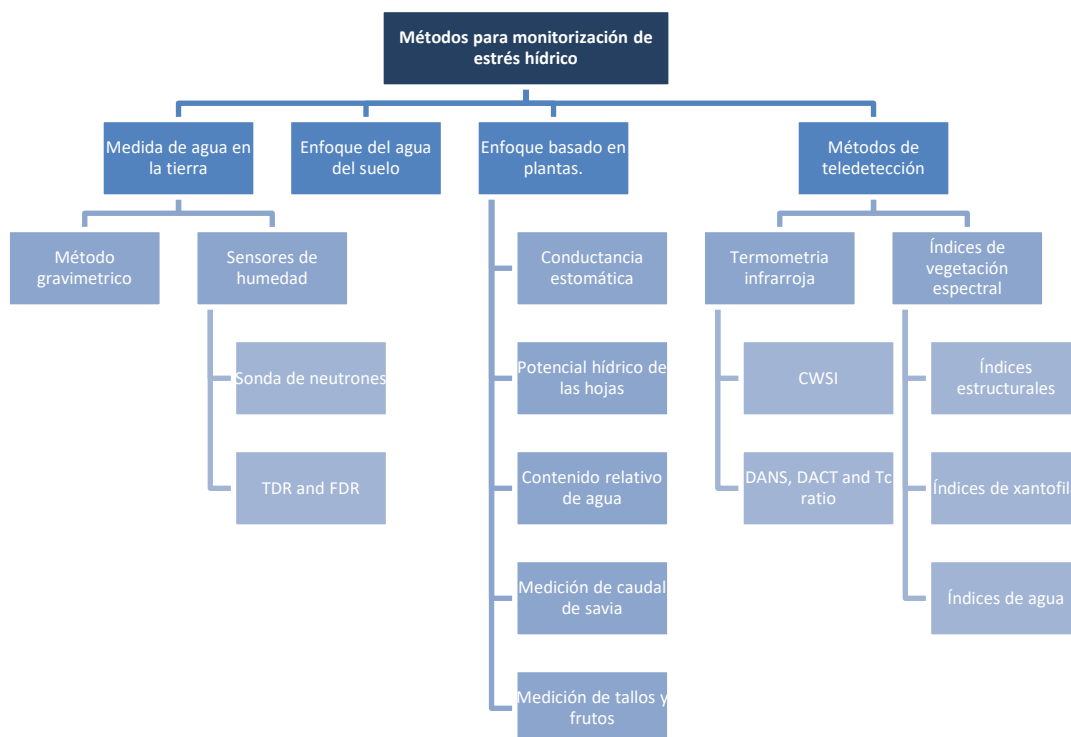
Existen numerosos métodos para detectar el tipo de estrés en los cultivos. La mayoría de las técnicas existentes se basan en la monitorización del estrés hídrico. Estas se basan en mediciones in situ de la humedad de la tierra y en las variables meteorológicas.

Uno de los métodos para predecir el estrés hídrico de la planta es emplear los modelos de evapotranspiración (ET). Con estos es posible realizar una predicción sobre los parámetros del tiempo que pueden afectar al estado del agua de la planta. Los modelos de ET más usados son el de modelo Penman-Monteith y el modelo Hangreaves.

Por otro lado, el enfoque más utilizado para estimar las necesidades del agua y la monitorización del balance hídrico suelo-planta es el método FAO-56, en el cual la ET se estima a partir de la combinación de ETo (calculado con alguno de los modelos anteriores) y coeficientes de cultivo. Hay dos tipos de enfoque dentro de este método:

- Único: Expresa la transpiración de la planta y la evaporación del suelo en un único coeficiente de cultivo, K_c .
- Dual: Utiliza un coeficiente para la transpiración, K_{cb} , y otro para la evaporación del suelo, K_e . K_{cb} se multiplica un coeficiente del estrés del agua, K_s ($K_s=1-CWSI$) y de esta forma tener en cuenta la reducción de ET debido al agotamiento de la humedad del suelo.

A continuación, se muestra un esquema donde aparece de forma resumida los principales métodos para monitorizar el estrés hídrico en las plantas:



Entre los métodos más interesantes se encuentran:

- Sensores TDR y FDR: Basado en la diferencia entre la constante dieléctrica del agua y del suelo. Método preciso y fácil en la práctica, lecturas pueden automatizarse. Se requieren muchos sensores lo que puede suponer un alto coste de instalación de sensores.
- Tensiómetros: Mide el potencial del agua del suelo. Son precisos y fáciles en la práctica. Útiles en suelos de textura gruesa o en riegos de alta frecuencia.
- Cálculo del Crop Water Stress Index (CWSI): Los métodos basados en las medidas de temperatura del dosel (canopy temperature, T_c) son los mayormente usados para programación de riegos de muchos cultivos. A medida que las plantas absorben radiación solar, la T_c aumenta, pero esta se reduce cuando se produce la evapotranspiración. Las plantas con estrés hídrico tienen una transpiración reducida y una reducción en la conductancia estomática, esto provoca una temperatura de las hojas más alta comparando con cultivos no estresados.

Existe una relación lineal entre las temperaturas del dosel (medidas mediante termometría infrarroja) y la temperatura del aire y el déficit de presión de vapor, que permite cuantificar el estrés hídrico de los cultivos. El CWSI empírico:

$$CWSI = \frac{[(T_c - T_a) - (T_{nws} - T_a)]}{[(T_{dry} - T_a) - (T_{nws} - T_a)]}$$

Donde:

- T_c : temperatura del dosel (°C)
- T_a : temperatura del aire (°C)
- T_{nws} (non-water stressed canopy Temperature): temperatura del dosel sin estrés hídrico (°C)
- T_{dry} (waterstressed canopy Temperature): temperatura del dosel con estrés hídrico (°C).

El CWSI proporciona un valor de 0 a 1 dependiendo del nivel de estrés hídrico del cultivo. Las zonas más estresadas corresponden a valores próximos de CWSI próximo a 1, y las zonas bien regadas tienen valores próximos a 0.

- Cálculo de T_c ratio: Relación sobre la temperatura del dosel, que se demostró que puede sustituir al coeficiente K_s usado para el cálculo del ET (método FAO-56 dual).
- Cálculo de Degrees Above Non-Stressed (DANS): Se basa únicamente en la temperatura del dosel (T_c).

$$DANS(h) = T_c(h) - T_{cNS}(h)$$

- Cálculo de Degrees Above Canopy Threshold (DACT). Solo requiere una única medida de temperatura del dosel.

$$DACT(h) = \max\{0, T_c(h) - T_{critical}\}$$

Si se compara los índices CWSI, DANS, DACT, y T_c , se llega a la conclusión de que los índices de DANS y DACT responden al estrés hídrico y no requieren tantos datos como CWSI. Por otro lado, las desventajas de calcular CWSI y los otros parámetros son: se asume que las condiciones del tiempo son constantes (los datos tomados varían mucho con estas) y que las medidas se realizan cerca del mediodía y en cielo despejado.

En la bibliografía se encuentra que existe y se usan con frecuencia una gran cantidad de indicadores de las plantas utilizando información de reflectancia basada en las regiones del espectro. En la siguiente tabla extraída de [8] se observan una serie de índices espectrales que han sido correlacionados con el estrés hídrico de las plantas:

Tabla 1 Índices espectrales de vegetación.

Spectral vegetation indices that has been correlated to plant water stress.

Reflectance indices		Formula	Plant water stress indicators
Names	Abbreviations		
Xanthophyll indices			
Photochemical Reflectance Index	PRI	$\frac{(R_{570} - R_{531})}{(R_{570} + R_{531})}$	Chlorophyll fluorescence and Stomatal conductance.
Normalized Photochemical Reflectance Index	PRI _{norm}	$\frac{PRI}{[RDVI * (R_{700}/R_{670})]}$	Chlorophyll fluorescence and Stomatal conductance.
Structural indices			
Normalized Difference Vegetation Index	NDVI	$\frac{R_{800} - R_{670}}{R_{800} + R_{670}}$	Stomatal Conductance, Leaf water potential
Renormalized Difference Vegetation Index	RDVI	$\frac{R_{800} - R_{670}}{\sqrt{R_{800} + R_{670}}}$	Stomatal Conductance, Leaf water potential
Transformed Chlorophyll Absorption in Reflectance Index	TCARI	$3[(R_{700} - R_{670}) - 0.2(R_{700} - R_{550}) * (R_{700}/R_{670})]$	Stomatal Conductance, Leaf water potential
Optimized Soil Adjusted Vegetation Index	OSAVI	$\frac{(1 + 0.16)(R_{800} - R_{670})}{(R_{800} + R_{670}) + 0.16}$	Stomatal Conductance, Leaf water potential
	TCARI/OSAVI	$\frac{3[(R_{700} - R_{670}) - 0.2(R_{700} - R_{550}) * (R_{700}/R_{670})]}{(1 + 0.16)(R_{800} - R_{670}) / (R_{800} + R_{670} + 0.16)}$	Stomatal Conductance, Leaf water potential
Water indices			
Normalized Difference Water Index	NDWI	$\frac{(R_{860} - R_{1240})}{R_{860} + R_{1240}}$	Leaf water potential
Simple Ratio Water Index	SRWI	$\frac{R_{860}}{R_{1240}}$	Leaf water potential
Water Index	WI	$\frac{R_{900}}{R_{970}}$	Leaf water potential

Where R represents the reflectances at the respective wavelengths, nm.

El estrés biótico puede cuantificarse igualmente a través de medidas de temperatura de la planta como la temperatura del dosel, puesto que al darse una interrupción en la corriente de transpiración (provocada por acciones de organismos) se produce un aumento de la temperatura como consecuencia. Por lo tanto, índices como el CWSI que cuantifican el aumento de temperatura también pueden indicar este tipo de estrés.

Por otro lado, para determinar el estrés por salinidad, el cual se debe a una alta frecuencia de riego, también se han empleado medidas de temperatura del dosel del suelo y del aire. Concretamente se ha demostrado que existe una diferencia entre la temperatura del dosel y el aire como consecuencia de un aumento de salinidad correspondiente. El índice de déficit de presión (VPD) puede utilizarse de igual forma como indicador de estrés en relación con la conductividad eléctrica del suelo como demostró [9] en plantas de algodón, si el VPD era inferior a 3.5 kPa no había estrés, sin embargo, para valores superiores de VDP las plantas mostraron estrés (con condiciones de agua adecuadas). Por su parte el estrés por deficiencia de nutrientes se identifica a través de información espectral. Por ejemplo, para el nitrógeno se identificó que la reflectancia roja visible aumentaba y la reflectancia cercana al infrarrojo disminuía a medida que disminuía la cantidad de nitrógeno.

2.2 Algoritmos de Machine Learning.

Debido a la necesidad de crear el aprendizaje automático es un campo de gran potencial, que puede beneficiar a la agricultura. De hecho, hasta la fecha son numerosas las aportaciones del ML a este sector. A continuación, se puede observar una tabla con algunos ejemplos de aplicación de las técnicas de ML a este sector.

Tabla 2 Algoritmos de ML aplicados a agricultura.

Tipo de algoritmo	Área de aplicación
Nonlinear Regression	Predicción de cosecha de maíz
Markov Chain Approach	Predicción de cosecha de algodón
Linear Regression	Estimación de cosecha de grano de arroz maduro
Belief Networks	Producción de cosecha futura
Neuro-Fuzzy Modeling	Predicción de cosecha
Second Order Markov Chains	Predicción de cosecha
Polynomial Regression	Factores influyentes en la cosecha de invierno de cereales
Deterministic and Probabilistic Prediction Approaches	Predicción de clima
Artificial Neural Network	Predicción de cosechas de soja y maíz
Neural Network	Monitorización cosecha de arroz
Fuzzy logic network	Cultivo de olivo en Andalucía
Regression	Predicción de azúcar de caña usando variables climáticas
Decision Tree	Productividad de soja
Time Series	Predicción de cosecha
Fuzzy Cognitive Map learning	Predicción de cosechas de manzanas
Regression and Neural Network	Predicción de producción de cosecha
Modified K-Means Clustering	Predicción de cosecha
Artificial Neural Network	Predicción de cosecha
Markov logic model	Rotaciones de cultivos para el mapeo temprano de cultivos

Según [6] algunos de los algoritmos más destacados son:

- Support Vector Machine (SVM):
Enfoque de aprendizaje estadístico para clasificar datos heterogéneos (con mayor precisión que los

clasificadores estadísticos tradicionales. Este es un método supervisado, no paramétrico, que también se puede usar para la regresión.

El clasificador SVM separa los datos etiquetados en clases en un espacio multidimensional. El SVM tiene como objetivo lograr los hiperplanos de separación óptimos, que son los límites de decisión entre clases que minimiza el error de clasificación en el entrenamiento al tener el margen máximo. Al margen se le conoce como la distancia entre las muestras de datos de las clases, este se maximiza gracias a la ayuda de los vectores de soporte (datos que se encuentran más cerca del margen).

SVM está diseñado para clasificación binaria, pero puede extenderse a clasificación múltiple usando distintos métodos.

La capacidad de SVM con trabajar con pequeñas muestras de entrenamiento sin comprometer la precisión del clasificador hace que sea de los métodos más utilizados.

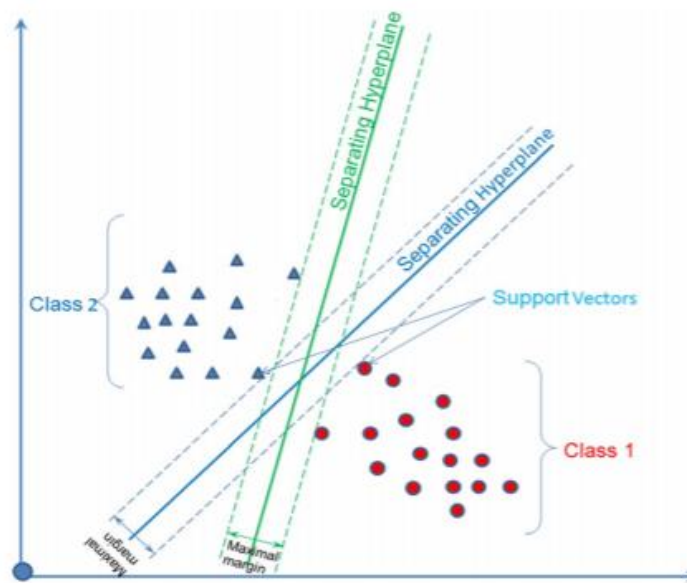


Figura 1 Algoritmo SVM.

- Random forest classifier (RF):
Clasificador que se construye a partir de la construcción de árboles de decisión débiles para clasificación y para regresión (una de las mayores ventajas). Es un algoritmo no paramétrico.

Las muestras para el entrenamiento son aleatoriamente seleccionadas del conjunto de datos original para construir el árbol. El conjunto de datos original se divide en: 2/3 de los datos para entrenamiento y 1/3 para validación cruzada (estimación del error).

Cada árbol se construye de manera independiente en función de dos atributos definidos por el usuario (hiper-parámetros): número de árboles (Ntree) y número de características para dividir cada nodo (Mtry) (puede tomar cualquier valor, pero normalmente se asigna como la raíz cuadrada del número de características). Con RF se produce bajo error de generalización. Es una herramienta eficaz para la predicción, ya que no desborda.

Cada árbol vota por una clase, la clase que reciba el mayor número de votos será la clase seleccionada.

- eXtreme gradient boosting (XGBoost):

Consiste en reforzar el aprendizaje a través del gradiente. El aumento del gradiente se implementa usando el clasificador XGBoost, diseñado para una mejora en el rendimiento y la velocidad. XGBoost utiliza información proporcionada como la de retroalimentación de los árboles e intenta reducir el error en las próximas iteraciones.

- **Rotation forest:**
Otro enfoque de conjunto basado en árboles es el bosque de rotación. Este difiere de RF en que se consideran diferentes subconjuntos de características en los árboles de entrenamiento. La extracción de características se realiza en un espacio rotado usando análisis de componentes principales (PCA). La decisión que se toma es similar a la decisión tomada en RF.
- **Oblique random forest:**
Crea árboles mediante bagging y selecciona variables aleatorias para la división de nodos. El análisis discriminante lineal (LDA), PCA, regresión ridge, los mínimos cuadrados parciales (PLS) y SVM se utilizan para dividir el nodo. A diferencia de RF, la RF oblicua aprende la dirección de división óptima utilizando todas las variables seleccionadas.
- **Artificial Neural network (ANN):**
La teledetección genera una gran cantidad de datos procedentes de múltiples sensores. Este tipo de problema no lineal se puede analizar aplicando ANN, debido a su capacidad para modelar una relación lineal y altamente no lineal entre conjuntos de datos de entrada y salida.

Las ANN se modelan así mismas a través de la selección de valores apropiados para los pesos. El modelo de ANN es prometedor en la agricultura para numerosas aplicaciones como: el modelado de información térmica para evaluar el estrés hídrico; mapeo de vegetación; predicción de cosecha y predicciones de estrés por nitrógeno.

3 METODOLOGÍA

El modelo de inteligencia artificial construido se encarga de analizar datos de distintos sensores en tiempo real, para proporcionar una información de ayuda al agricultor en la toma de decisiones y de esta forma poder mejorar la sostenibilidad y el rendimiento en el manejo de los cultivos. El cultivo seleccionado para este proyecto fue el cultivo del almendro. Este es un cultivo homogéneo que ha iniciado su época productiva (4º año desde la implantación). Estos detalles hacen que este cultivo sea el más indicado para tomar datos de interés y representativos para la elaboración de los modelos.

Como se ha mencionado previamente, el objetivo general es el diseño un sistema de aprendizaje automático que se adapte a las características de la finca de producción y del propio agricultor, y de que de esta forma permita implantar un nuevo modelo productivo de agricultura ecosostenible. Como resultado, se ha generado un sistema automático de alarmas que alerta en tiempo real de los riesgos bióticos/abióticos que corren las plantas; se supervisa la absorción de los fertilizantes y se pueden hacer predicciones sobre el rendimiento de cierto cultivo de ensayo (tomate).

Para conseguir dicho objetivo general, la presente tarea tiene los siguientes objetivos específicos:

- Con las medidas tomadas y su posterior análisis, los algoritmos extraerán la información relevante de los datos, encontrando relaciones con el fin de generar información de confianza para predecir riesgos bióticos como plagas o enfermedades y riesgos abióticos como estrés hídrico, estrés por radiación solar, estrés por salinidad, escaso desarrollo vegetativo o fallo en el cuajado de frutos.
- La inteligencia artificial deberá alertar al agricultor de los posibles riesgos que puedan darse en el futuro, anticipándose, y de esta forma permitiendo evitar posibles daños en los cultivos. También, dependiendo del riesgo, se llevará a cabo una notificación de forma personalizada de eventos dependiendo de las características de los cultivos y del manejo que esté empleando el productor, realizando la recomendación correspondiente que señalen los especialistas en cada caso.
- Asimismo, se analizarán datos relacionados con el manejo de los cultivos, tales como el volumen, la dosis y el número de aplicaciones de los insumos agrícolas. Se buscará modelar su relación con la producción, el peso, el calibre de los frutos o el volumen entre otras magnitudes. Las correlaciones que encuentre el algoritmo se utilizarán para la predicción de las variables, con el fin de alertar al agricultor en caso de que sea necesario llevar a cabo acciones correctoras.
- Finalmente, se estudiará si el cultivo es sostenible o no analizando las medidas de los sensores. Se prestará principal atención a los sensores de humedad y salinidad colocados junto a las raíces de los cultivos, a distintos niveles de profundidad, pues permiten inferir si los fertilizantes aplicados se están adhiriendo a la planta o si, por el contrario, se están filtrando al suelo y, como consecuencia, el cultivo es poco sostenible. La función del sistema de inteligencia artificial será evaluar la situación y avisar a el agricultor para que tome las medidas oportunas que prescriban los expertos. De esta forma se podrá reducir el impacto ambiental de la producción de los alimentos.

3.1 Instalación de sensores

La agricultura de precisión es una asignatura pendiente en la mayoría de las explotaciones agrícolas y actualmente no existe demanda de sensores de control de un cultivo. Los servicios que existen actualmente sólo trabajan centrados en ciertos factores que afectan al cultivo, en concreto dos de ellos: factores climáticos o manejo del agua de riego.

- Los factores climáticos se monitorizan con estaciones climatológicas, las cuales suelen ser modulares, según los parámetros que interese medir.
- Las opciones que permiten monitorizar el manejo del agua se basan en la “capacidad de campo” del suelo en el que está instalado el cultivo. Se basan en sensores de humedad que regulan el riego según la humedad que tenga el suelo, permitiendo mantener una humedad constante.

Para conseguir los objetivos previstos en este trabajo, se ha instalado un sistema experimental formado por una red de sensores en la finca de análisis los cuales miden magnitudes como la temperatura ambiental, la cantidad de precipitaciones o la humedad, presión y salinidad del suelo. Los sensores instalados permiten integrar los diferentes parámetros que afectan a un cultivo: factores ambientales (climatología), factores de manejo (salinidad y riego) condiciones del suelo (humedad y temperatura del suelo). Para el montaje de los equipos de medida se realizó previamente un estudio previo de las particularidades de la zona de instalación, se procedió a seleccionar que sensores eran los más ideales para obtener los datos de los indicadores fijados. Los sensores que monitorizan las diferentes variables mencionadas anteriormente permiten no sólo ver los valores que arroje cada una de las variables, sino las interacciones que están ocurriendo entre las estas. Los sensores propuestos están integrados en la misma plataforma, constituyendo una estación climática completa. Esta está compuesta de los siguientes sensores:

- **Pluviómetro:** Permite recoger los datos de las lluvias caídas en la zona con una sensibilidad de 0.2 mm/m² El pluviómetro presenta un error de unos $\pm 0,2$ mm/m² que equivale a un volumen de lluvia de 200 ml/m².

Typical switching times for PCB 9602

Spoon	Terminal 1-2 NC	Terminal 3-4 NO
2ml	443ms, $\sigma = 14mS$	352ms, $\sigma = 14mS$
4ml	307ms $\sigma = 12mS$	278ms, $\sigma = 12mS$
5ml	326ms $\sigma = 11mS$	301ms, $\sigma = 12mS$
10ml	322ms, $\sigma = 7mS$	305ms, $\sigma = 8mS$

σ = Standard deviation

Filter design: Max 200msec recommended.

Accuracy:

+/- 2%

Resolutions

0.10 mm, 0.20 mm, 0.25 mm, 0.50 mm

Figura 2 Características pluviómetro.

- **Barómetro:** Permite medir la presión atmosférica de la zona en KPa. El sensor de presión atmosférica tiene un error que se encuentra entre $\pm 0,01$ kPa y $\pm 0,28$ kPa.

Vapor Pressure	
Range	0–47 kPa
Resolution	0.01 kPa
Accuracy	Sensor measurement accuracy is variable across a range of temperatures and RH. Refer to the chart in Figure 5 .

HUMIDITY (%RH)	TEMPERATURE (°C)								
	0	10	20	30	40	50	60	70	80
100	± 0.03	± 0.05	± 0.09	± 0.16	± 0.27	± 0.44	± 0.69	± 1.33	± 2.38
90	± 0.03	± 0.05	± 0.09	± 0.15	± 0.26	± 0.42	± 0.66	± 1.26	± 2.24
80	± 0.03	± 0.04	± 0.07	± 0.12	± 0.21	± 0.34	± 0.63	± 1.20	± 2.10
70	± 0.02	± 0.04	± 0.07	± 0.12	± 0.20	± 0.32	± 0.50	± 1.13	± 1.96
60	± 0.02	± 0.03	± 0.06	± 0.11	± 0.18	± 0.30	± 0.47	± 1.06	± 1.82
50	± 0.02	± 0.03	± 0.06	± 0.10	± 0.17	± 0.28	± 0.45	± 0.99	± 1.68
40	± 0.02	± 0.03	± 0.05	± 0.09	± 0.16	± 0.26	± 0.42	± 0.76	± 1.54
30	± 0.01	± 0.03	± 0.05	± 0.09	± 0.15	± 0.24	± 0.39	± 0.69	± 1.40
20	± 0.01	± 0.02	± 0.04	± 0.08	± 0.14	± 0.23	± 0.36	± 0.62	± 1.26
10	± 0.01	± 0.02	± 0.04	± 0.07	± 0.12	± 0.21	± 0.33	± 0.55	± 1.13
0	± 0.01	± 0.02	± 0.04	± 0.06	± 0.11	± 0.19	± 0.30	± 0.48	± 0.99

Figure 5 Vapor pressure sensor accuracy

Barometric Pressure	
Range	1–120 kPa
Resolution	0.01 kPa
Accuracy	±0.05 kPa at 25 °C
Equilibration Time (τ , 63%)	<10 ms
Long-Term Drift	<0.1 kPa/year, typical

Figura 3 Características barómetro.

- Termómetro: Para medir la temperatura ambiental del cultivo. El sensor de temperatura ambiental tiene un error de $\pm 0,2$ °C.

Temperature	
Range	–40 to 80 °C
Resolution	0.1 °C
Accuracy	±0.2 °C
Equilibration Time (τ , 63%)	<165 s (response time in 1 m/s air stream)
Long-Term Drift	<0.03 °C/year, typical

Figura 4 Características termómetro.

- Higrómetro: Para medir la humedad ambiental relativa que rodea al cultivo. El sensor instalado para arrojar el dato de HR ambiental, tiene un error de $\pm 1,5\%$ en los valores habituales de un cultivo (de 0 - 50°C y entre 30-70% de humedad).

Relative Humidity (RH)	
Range	0–100% RH (0.00–1.00)
Resolution	0.10% RH
Accuracy	Sensor measurement accuracy is variable across a range of RH. Refer to the chart in Figure 4.

HUMIDITY (%RH)	TEMPERATURE (°C)								
	0	10	20	30	40	50	60	70	80
100	±2.0%	±2.0%	±2.0%	±2.0%	±2.0%	±2.0%	±2.0%	±2.0%	±2.0%
90	±2.0%	±2.0%	±2.0%	±2.0%	±2.0%	±2.0%	±2.0%	±2.0%	±2.0%
80	±2.0%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±2.0%	±2.0%	±2.0%
70	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±2.0%	±2.0%
60	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±2.0%	±2.0%
50	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±2.0%	±2.0%
40	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±2.0%
30	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±2.0%
20	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±2.0%
10	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±2.0%
0	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±1.5%	±2.0%

Figure 4 RH sensor accuracy

Equilibration Time (τ , 63%)	<25 s (response time in 1 m/s air stream)
Hysteresis	±0.80% RH, typical
Long-Term Drift	±0.25% RH/year, typical

Figura 5 Características higrómetro.

La misma estación climática permitirá conectar los sensores de suelo y de control del riego. Estos sensores eran los siguientes:

- Contador de pulsos: permite calcular la cantidad de agua que se aporta al cultivo. El error de medición del contador de pulsos instalado en la goma de goteo es de un ±3%, con una exactitud de indicación de 0.00005 m³.

Parámetro			JS 1,6-02 JS90 1,6-02	JS 2,5-02 JS90 2,5-02	JS 2,5-G1-02 JS90 2,5-G1-02	JS 4-02 JS90 4-02
Diámetro nominal	DN	mm	15		20	
Caudal Nominal	Q ₃	m ³ /h	1,6		2,5	4
Caudal Máximo	Q ₄	m ³ /h	2		3,125	5
Caudal de Transición	para el agua fría	H R100	Q ₂	L/h	25,6	64
		V R50			51,2	128
	para el agua caliente	H R80			32	80
		V R40			64	160
Caudal Mínimo	para el agua fría	H R100	Q ₁	L/h	16	40
		V R50			32	80
	para el agua caliente	H R80			20	50
		V R40			40	100
Caudal de Arranque	-	L/h	4-6		6-8	10-15
Relación Q ₂ /Q ₁	-	-			1,6	
Clase de temperatura (Temperatura nominal en funcionamiento)	-	-			t30 / t90	
Clase de resistencia al perfil de flujo	-	-			U0 / D0	
Escala de indicaciones	-	m ³			10 ⁵	
Exactitud de indicaciones	-	m ³			0,00005	
Presión máxima	P _{max}	MPa			1,6	
Pérdida máxima de presión	A ^P	kPa			100	
Error de límite admisible en la escala: Q ₂ ≤ Q ≤ Q ₄	E	%			±2 para el agua fría ±3 para el agua tibia	
Error de límite admisible en la escala: Q ₁ ≤ Q ≤ Q ₂	E	%			±5	
Conexión rosca	G	cal	G ³ / ₄		G1	
Altura	H	mm	68,5			
Longitud	L	mm	110	115	115	130
Diámetro	D	mm	72			
Peso (sin los elementos de instalación)	-	kg	0,5	0,5	0,6	0,6

Modelo 02 (totalizador de ocho tambores, contador adaptado para el montaje de emisor de radio o de pulsos).

Figura 6 Características de contador de pulsos.

- Sonda de 60 cm: Esta permite tomar 3 datos diferentes. Humedad del suelo, temperatura del suelo y conductividad (salinidad del suelo). Esta sonda tiene la particularidad de que mide estos parámetros en seis puntos diferentes del perfil del suelo, cada 10 cm. Esta característica de la sonda es la que mayor valor tiene, no sólo por aportar 6 datos con una medición, sino por permitir analizar cómo se comporta la raíz del cultivo en todo el perfil. La sonda edafológica, con mediciones cada 10 cm hasta un perfil de 60 cm, tiene un error de $\pm 0,03\%$ en el dato de humedad relativa del suelo y un error de ± 2 °C en la temperatura del suelo.

Probe Length	10.5cm/4"	30cm/12"	60cm/24"	90cm/36"	120cm/48"
Number of Sensors	1	3	6	9	12
Outer Probe Diameter Top: Bottom:	24.5mm 24mm	30mm 28.75mm	30mm 27.5mm	30mm 26.25mm	30mm 25mm
Connection Protocols Available	SDI-12 only	SDI-12 RS232 RS485 Sentek Bluetooth	SDI-12 RS232 RS485 Sentek Bluetooth	SDI-12 RS232 RS485 Sentek Bluetooth	SDI-12 RS232 RS485 Sentek Bluetooth
Resolution	Moisture = 1:10000 Salinity = 1:3000 Temperature = 0.3°C				
Moisture Precision	$\pm 0.03\%$ vol.				
Temperature Accuracy	± 2 °C @ 25 °C				
Operating Temperature	-20°C to +60°C				

Figura 7 Características de sonda.

El banco de ensayos formado por los sensores mencionados anteriormente está alimentado con un panel solar y cuenta con: dos entradas digitales para el contador y el pluviómetro; cuatro entradas analógicas y una entrada SDI-12 para sensores multiparamétricos.



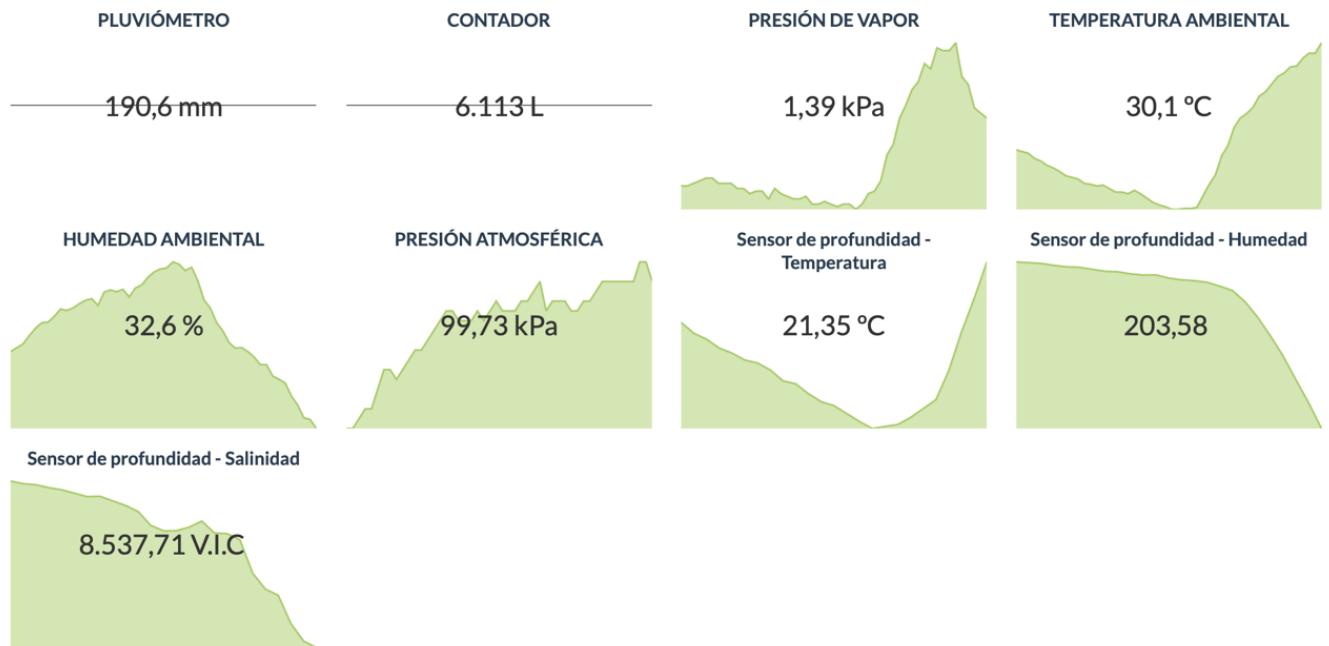
Figura 8 Banco de ensayos.

3.2 Fuentes de datos.

3.2.1 Odin.

La solución de sensórica viene acompañada de una plataforma de gestión de datos que se encarga de recopilar y almacenar la información generada por los sensores configurados en la misma la cual puede verse en la Figura 1. Los datos son enviados hasta esta plataforma mediante un Gateway instalado en campo. Los datos, generados con un periodo de 15 minutos, se exponen hacia plataformas externas mediante una API Rest. Dicha API, implementa el estándar NGSI de FIWARE. Utilizando este acceso remoto, ha sido posible descargar a diario, y de manera automática, todos los datos generados durante la jornada. Esta frecuencia de recopilación permitirá recoger una cantidad suficiente de datos para que los algoritmos puedan funcionar, al mismo tiempo que refleja de forma precisa todos los eventos que estén ocurriendo en el cultivo monitorizado.

Para la solución tecnológica escogida las 25 variables que se puede obtener son las siguientes: pluviómetro, contador, presión de vapor, temperatura ambiental, humedad ambiental, presión atmosférica, nivel de batería, estratos de temperatura (T0, T1, T2, T3, T4, T5), estratos de humedad (H0, H1, H2, H3, H4, H5), estratos de salinidad (S0, S1, S2, S3, S4, S5).



Como resultado de la petición de la API Rest se generan un json como respuesta donde se recopilan los datos recopilados por los dispositivos de la finca.

```
{
  "id": "Mex06:00086",
  "type": "Device",
  "category": [ ],
  "refDeviceModel": { },
  "name": "Econatur",
  "serialNumber": "MA0101F16000200086",
  "location": { },
  "dateModified": "2021-06-07T11:30:25.540Z",
  "batteryLevel": { },
  "solarPanelLevel": { },
  "rssi": { },
  "pulseCounter_5fd7477feadee895b125bbf2": { },
  "pulseCounter_5fd7477feadee895b125bbf9": { },
  "analogInput_5fd7477feadee895b125bbfc": { },
  "analogInput_5fd7477feadee895b125bbff": { },
  "analogInput_5fd7477feadee895b125bc02": { },
  "analogInput_5fd7477feadee895b125bc05": { },
  "sdi12_5fd7477feadee895b125bc19_Temperature": { },
  "sdi12_5fd7477feadee895b125bc19_Humidity": { },
  "sdi12_5fd7477feadee895b125bc19_Salinity": { }
}
```

Dentro de la estructura, se encuentra a su vez una subestructura para cada dispositivo configurado. En este caso se dispone de 9 dispositivos: pluviómetro, contador, presión de vapor, temperatura ambiental, humedad ambiental, presión atmosférica, sensor de profundidad – temperatura, sensor de profundidad – humedad y sensor de profundidad – salinidad. Para que se puedan mapear correctamente los datos de cada dispositivo, el nombre del campo de cada *device* en la plataforma debe coincidir con el nombre que se le ha dado en la plataforma de Odin, ya que esta será la clave de búsqueda que se utilizará para recuperar un dato concreto del json recibido. Este texto se buscará en el campo “metadata | description”. El valor recuperado se tomará del campo “value”.

```
"analogInput_5fd7477feadee895b125bbfc":{  
  "type":"Float",  
  "value":1.518,  
  "configuration":{+},  
  "metadata":{  
    "timestamp":{+},  
    "description":{  
      "value":"PRESIÓN DE VAPOR"  
    }  
  },  
  "unitCode":{+}  
}
```

3.2.2 Estaciones climáticas.

Además de los sensores instalados, se cuenta con la información aportada por estaciones climáticas. Conocer el estado climatológico es fundamental para la determinación y predicción de factores que afecten al estado de los cultivos, la presencia de plagas y la sostenibilidad de este. La agencia estatal de meteorología (AEMET) tiene desplegadas una red de estaciones climatológicas de distinta tipología por todo el territorio nacional:

- Estaciones completas
- Estaciones termométricas
- Estaciones pluviométricas
- Estaciones automáticas

En paralelo a la red nacional de estaciones climáticas, cada comunidad autónoma puede disponer de redes de estaciones de recolección de datos. Concretamente, Andalucía dispone de la Red de Información Agroclimática de Andalucía (RIA), cuya infraestructura se desplegó con el objetivo facilitar a agricultores y técnicos información meteorológica para calcular las necesidades de agua de los cultivos y programar sus riegos. Esta Red suministra información de las principales variables meteorológicas, como son la temperatura y humedad relativa del aire, la velocidad y dirección del viento, la radiación solar y la precipitación en forma de lluvia. Sus medidas permiten calcular la evapotranspiración de referencia (ET_o), variable imprescindible para estimar las necesidades de riego.

La infraestructura, tecnología e información generada por dicha red de estaciones es desarrollada y mantenida por el Instituto Andaluz de Investigación y Formación Agraria, Pesquera, Alimentaria y de la Producción Ecológica (IFAPA). Uno de los factores diferenciales de esta red de estaciones agroclimáticas automáticas (EAA) es que dispone de una interfaz tipo API Rest abierta que permite el acceso tanto a datos actuales como históricos, convirtiéndose en una fuente de datos de contexto especialmente valiosa. Las EAA generan una muestra de datos diariamente. El amplio número de EAA desplegadas facilita que se puedan conseguir datos representativos de casi cualquier ubicación andaluza. Para la elaboración de los modelos se utilizará la estación agroclimática de Santaella, la cual está ubicada a unos 7 Km de la finca experimental.

Las 20 variables que se puede obtener de las EAA de la red RIA son las siguientes: temperatura ambiental

media; temperatura ambiental máxima; temperatura ambiental mínima; hora temperatura ambiental máxima; hora temperatura ambiental mínima; radiación; velocidad del viento; humedad relativa media; humedad relativa máxima; humedad relativa mínima; hora humedad relativa máxima; hora humedad relativa mínima; velocidad viento media; dirección viento; velocidad viento máxima; hora velocidad viento máxima; dirección viento velocidad máxima; precipitaciones y evapotranspiración.

Los datos recopilados en las estaciones agroclimáticas automáticas (EAA) de la red RIA están expuestos públicamente a través de una interfaz de tipo API Rest. Los datos pueden ser también consultados en la propia web de IFAPA.

← → ↻ juntadeandalucia.es/agriculturaypesca/ifapa/riaweb/web/estacion/14/7

Estación Meteorológica de Santaella

Provincia: Córdoba
Código de Estación: 7
Primer día registrado: 17-10-2000
Último día registrado: 06-06-2021
Datos Localización
Latitud: 37° 31' 20" N
Longitud: 04° 53' 07" W
Altitud: 196
Coordenadas UTM
X: 333382
Y: 4154500

Último Dato Registrado	Descarga de Ficheros	Datos Históricos
Fecha de Medida	06-06-2021	Día Juliano 157
Temp. Máxima (°C)	33.0	Velocidad Viento Media (m/s) 1.0
Hora Temp. Máxima	17:28	Dirección Viento Media (°) 83.8
Temp. Mínima (°C)	13.4	Radiación (MJ/m ² d) 29.2
Hora Temp. Mínima	4:54	Precipitación (mm) 0.0
Temp. Media (°C)	23.4	ETo (mm/día) 6.0
Humedad Relativa Máxima (%)	100.0	
Humedad Relativa Mínima (%)	21.7	Precipitación Acumulada (mm) 357.0
Humedad Relativa Media (%)	56.7	ETo Acumulada (mm) 775.0

Figura 12 Interfaz web IFAPA.

Todos los modelos de datos y métodos disponibles están detallados en la web de IFAPA:

- provincia-controller: métodos para consulta de ID de provincias.
- estacion-controller: métodos para consulta de ID de estaciones.
- dato-diario-controller: consulta de datos diarios de una estación identificada por su ID de provincia y estación. Los datos pueden ser recuperados para una fecha concreta o para un periodo (históricos).
- dato-mensual-controller: consulta de datos mensuales de una estación identificada por su ID de provincia y estación.

```

DatoDiario v {
  bateria          number($double)
  dia              integer($int32)
  dirViento        number($double)
  dirVientoVelMax  number($double)
  et0             number($double)
  fecha           string($date-time)
  fechaUtlMod     string($date-time)
  horMinHumMax    string
  horMinHumMin    string
  horMinTempMax   string
  horMinTempMin   string
  horMinVelMax    string
  humedadMax      number($double)
  humedadMedia    number($double)
  humedadMin      number($double)
  precipitacion   number($double)
  radiacion       number($double)
  tempMax         number($double)
  tempMedia       number($double)
  tempMin         number($double)
  velViento       number($double)
  velVientoMax   number($double)
}

```

Figura 13 Modelo IFAPA para datos diarios.

```

DatoMensual v {
  anyo            integer($int32)
  bateria         number($double)
  diaHorMinTempMax string($date-time)
  diaHorMinTempMin string($date-time)
  diahorMinHumMax string($date-time)
  diahorMinHumMin string($date-time)
  diahorMinVelMax string($date-time)
  dirViento       number($double)
  dirVientoVelMax number($double)
  fechaUtlMod     string($date-time)
  humedadMax      number($double)
  humedadMedia    number($double)
  humedadMin      number($double)
  mes             integer($int32)
  numDias         integer($int32)
  precipitacion   number($double)
  radiacion       number($double)
  tempMax         number($double)
  tempMedia       number($double)
  tempMin         number($double)
  velViento       number($double)
  velVientoMax   number($double)
}

```

Figura 14 Modelo IFAPA para datos mensuales.

3.2.3 Ephem.

En relación con el cálculo de las horas diarias de luz, las horas de salida y puesta del sol están determinadas por 3 factores: la posición en la tierra (latitud, longitud y altitud), la fecha y los elementos posiblemente ocultos como las montañas. Este puede resolverse de forma local mediante el uso de alguna librería específica de software. En este caso y dado que la algoritmia de aprendizaje se desarrollará en Python, se optó por el uso de PyEphem. Las horas de luz diarias podrán servir para cuantificar la incidencia de otros factores como la de radiación recibida o la cantidad de energía que ha procesado las plantas del cultivo, entre otros.

La librería de Python Ephem podrá ser utilizada para calcular las horas diarias de sol. Los datos históricos de otras variables de entrada (para los algoritmos de aprendizaje) podrán ser enriquecidos con las horas diarias de sol, cuyo cálculo estará basado en la hora de salida y puesta del sol para una ubicación concreta.

Para que dicha librería esté disponible en los scripts tendrá que ser instalada previamente en el servidor de la plataforma web. Esto se consigue ejecutando el comando “pip install ephem”. Para realizar la consulta, inicializamos un objeto tipo “Sun” y otro tipo “Observer” con las coordenadas de la ubicación que se desea consultar. Utilizamos los métodos “previous_rising” y “next_setting” y calculamos la diferencia en horas.

```

#Observer initialization
city=ephem.Observer()
city.lat='37.5788'
city.lon='-4.8291'
city.elevation = 196 # meters
city.date = datetime.datetime.now()
#Star reference
sun = ephem.Sun()
p_sunrise=ephem.localtime(city.previous_rising(sun))
n_sunset=ephem.localtime(city.next_setting(sun))
#Daily timing
sunrise,sunset=p_sunrise,n_sunset
print("Previous sunrise was: ",sunrise)
print("Next sunset will be: ",sunset)
#Light hours
time_delta = (sunset - sunrise)
total_seconds = time_delta.total_seconds()
sun_hours = total_seconds/3600

print("Today there will be ",sun_hours," sun hours")

Previous sunrise was: 2021-05-28 07:01:00.225606
Next sunset will be: 2021-05-28 21:32:36.662476
Today there will be 14.526788019444444 sun hours

```

Figura 15 Cálculo de las horas diarias de sol con la librería Ephem.

Una vez recopilados los datos de las medidas captadas por los sensores o bien de la estación climática de IFAPA. Los técnicos proceden a clasificarlos y etiquetarlos, aplicando su conocimiento experto. De esta forma, se ha podido asociar a cada conjunto de medidas un nivel de riesgo biótico/abiótico o cuantificar la sostenibilidad de los fertilizantes utilizados **¡Error! No se encuentra el origen de la referencia..** Finalmente, se ha transformado dichas estimaciones en reglas de aprendizaje y entrenamiento para la inteligencia artificial.

3.3 Entorno de desarrollo.

Para la implementación de los algoritmos de inteligencia artificial se ha usado el lenguaje de programación Python [10][11]. Se ha escogido este lenguaje por ser el que está más extendido y soportado por la comunidad de investigadores y desarrolladores en el ámbito del aprendizaje automático. A su vez se ha establecido joblib como la extensión permitida para el almacenamiento de los modelos generados.

Entre las librerías usadas destacan las específicas de aprendizaje automático, *Scikit-learn* [12] y *Tensorflow* [13]. La librería *Scikit-learn*, es probablemente una de las librerías más útiles para ML en Python debido a que facilita una amplia gama de algoritmos de aprendizaje supervisados y no supervisados. Además, esta librería se integra fácilmente con otras como *Numpy*, para el cálculo numérico y *Pandas*, especializada en el manejo y análisis de estructuras de datos, las cuales se usan en la etapa de pre-procesado de los datos. Por otro lado, *Tensorflow* es una librería gratuita para el aprendizaje automático profundo (*Deep learning*) desarrollada por Google que cuenta también con numerosos algoritmos de inteligencia artificial. Para ejecutar los algoritmos de aprendizaje de esta biblioteca se empleará la interfaz de alto nivel *Keras* [14], capaz de ejecutarse sobre *Tensorflow*.



En cuanto al entorno de desarrollo se ha empleado *Jupyter Notebook*. Este es una aplicación web de código abierto que permite crear y compartir documentos que contienen código en vivo, ecuaciones, visualizaciones y texto narrativo. Los usos incluyen limpieza y transformación de datos, simulación numérica, modelado estadístico, visualización de datos, aprendizaje automático y mucho más.



Figura 17 Jupyter notebook.

3.4 Algoritmos.

En base a la experiencia previa, y tras ensayos preliminares, se ha optado por desarrollar la inteligencia artificial a partir de las siguientes tecnologías de ML.

3.4.1 Árboles de decisión.

Se trata de uno de los algoritmos de aprendizaje supervisado de referencia en ML para la predicción, debido a los buenos resultados que obtiene ante problemas muy diversos[15][16].

Los árboles de decisión construyen modelos elaborados a partir de reglas binarias (sí/no), con las cuales se produce un reparto de las observaciones en función de los datos de entrada del árbol, es decir de los atributos considerados[17][18]. A partir de estos datos y aplicando el modelo del árbol se consigue obtener una predicción del valor de la variable/s de salida [19].

Este tipo de algoritmo consta de una raíz, la cual es el nodo del que cuelgan múltiples ramas, tantas ramas como atributos de entrada. A su vez estas ramas vuelven a subdividirse en hojas que son los nodos finales, y que indican la respuesta del algoritmo (sí/no).

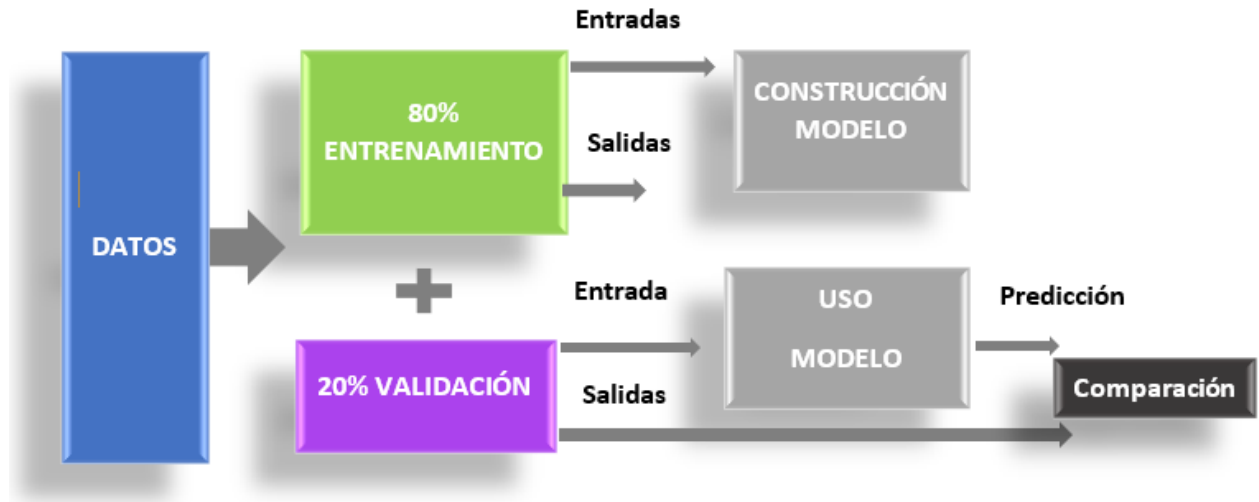
Existen dos tipos de árboles:

- Regresión: en este tipo de árbol la/s variable/s a predecir es de tipo continua.
- Clasificación: en este tipo de árbol la/s variable/s a predecir es de tipo categórica.

Los ventajas de los modelos de árboles de decisión son : la facilidad de interpretación de estos incluso cuando las correlaciones entre predictores son complejas; admiten representación de forma gráfica; la capacidad de tratar predictores tanto categóricos como numéricos; no necesitan ninguna distribución específica al tratarse de métodos no paramétricos; necesitan un menor preprocesado y una menor depuración en contraposición con otras técnicas de aprendizaje automático; no se ven influenciados por valores atípicos (outliers), etc.

En el empleo de los árboles de decisión se distinguen dos etapas principales: el entrenamiento y la predicción. En primer lugar, los datos de entrada deben dividirse en datos de entrenamiento y datos de

validación (normalmente se suele seguir el criterio de 80% del conjunto para el entrenamiento y 20% para la validación). Posteriormente, se crea el modelo y se entrena usando los datos de entrenamiento. Para la evaluación del modelo, se realiza una predicción con el algoritmo entrenado usando los datos de validación y se comparan con las salidas que deberían tener estos.



La formulación matemática de este algoritmo es la siguiente:

Dados los vectores de entrenamiento $x_i \in \mathbf{R}^n$, $i=1, \dots$ y el vector de etiquetas y \mathbf{R}^1 , un árbol de decisión divide de forma recursiva el espacio de características de modo que las muestras con las mismas etiquetas o valores similares se agrupan.

Los datos en el nodo m son representados por Q_m con N_m muestras. Para cada dato θ existe una característica j y un umbral t_m $\theta = (j, t_m)$, partiendo los datos en dos subconjuntos, $Q_m^{izq}(\theta)$ y $Q_m^{drcha}(\theta)$.

$$Q_m^{izq} = \{(x, y) | x_j \leq t_m\}$$

$$Q_m^{drcha}(\theta) = Q_m \setminus Q_m^{izq}(\theta)$$

La calidad de una división de un nodo candidato se calcula utilizando la función de pérdida, cuya elección depende de la tarea que se desee resolver (clasificación o regresión)

$$G(Q_m, \theta) = \frac{N_m^{izq}}{N_m} H(Q_m^{izq}(\theta)) + \frac{N_m^{drcha}}{N_m} H(Q_m^{drcha}(\theta))$$

Se selecciona los parámetros que minimizan el error

$$\theta^* = \operatorname{argmin}_{\theta} G(Q_m, \theta)$$

de los subconjuntos $Q_m^{izq}(\theta)$ y $Q_m^{drcha}(\theta)$ hasta que se alcanza la máxima profundidad permitida $N_m <$

$\min_{\text{simples}} \text{ or } N_m=1.$

Si el objetivo es una salida de clasificación que toma valores $0,1,\dots, K-1$ para el nodo m , la proporción de observaciones de la clase k en el nodo m . Si m es un nodo final, predict_proba para esta región se establece en:

$$p_{mk} = \frac{1}{N_m} \sum_{y \in Q_m} I(y = k)$$

Las mediciones de error son:

- Gini:

$$H(Q_m) = \sum_k p_{mk}(1 - p_{mk})$$

- Entropía:

$$H(Q_m) = - \sum_k p_{mk} \log(p_{mk})$$

- Clasificación

$$H(Q_m) = 1 - \max(p_{mk})$$

Si el objetivo es una variable continua, entonces para el nodo m , el criterio común para minimizar, así como para determinar futuras localizaciones es el error cuadrático medio (MSE),

$$\bar{y} = \frac{1}{N_m} \sum_{y \in Q_m} y$$

$$H(Q_m) = \frac{1}{N_m} \sum_{y \in Q_m} (y - \bar{y})^2$$

o el error medio absoluto (MAE)

$$\begin{aligned} \text{median}(y)_m &= \text{median}(y) \\ H(Q_m) &= \frac{1}{N_m} \sum_{y \in Q_m} |y - \text{median}(y)_m| \end{aligned}$$

3.4.2 Redes Neuronales Artificiales.

Por otro lado, las redes neuronales artificiales son un sistema de aprendizaje automático formado por un conjunto de unidades denominadas neuronas artificiales, las cuales están conectadas unas con otras a través de unos enlaces mediante los que comparten información [20]. En estos enlaces, a la salida de la neurona se obtiene como resultado la suma de las entradas multiplicadas por sus pesos, filtrada mediante de la función de activación. Al construir la red, se intenta optimizar una función objetivo, para lo cual los pesos se van actualizando de manera constante. Este método se conoce como *back-propagation*.

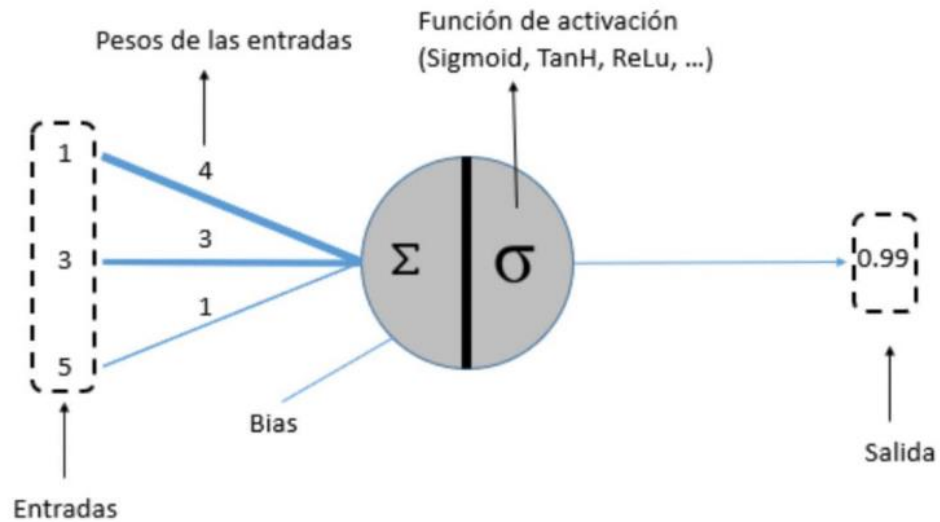


Figura 19 Representación esquemática de una neurona artificial.

Las distintas neuronas artificiales que forman la red se agrupan en distintos niveles, denominados capas. Las capas son por tanto conjuntos de neuronas, cuyas entradas provienen de una capa anterior y cuyas salidas son las entradas de una capa posterior. La primera capa, debido a que es la encargada de recibir los datos de entrada se denomina capa de entrada. Por el contrario, la última capa es la encargada de recibir la salida de la red, de ahí que se designe como capa de salida. Las capas que se encuentran entre la capa de entrada y la capa de salida se denominan capas ocultas. Una red puede contener varias capas ocultas.

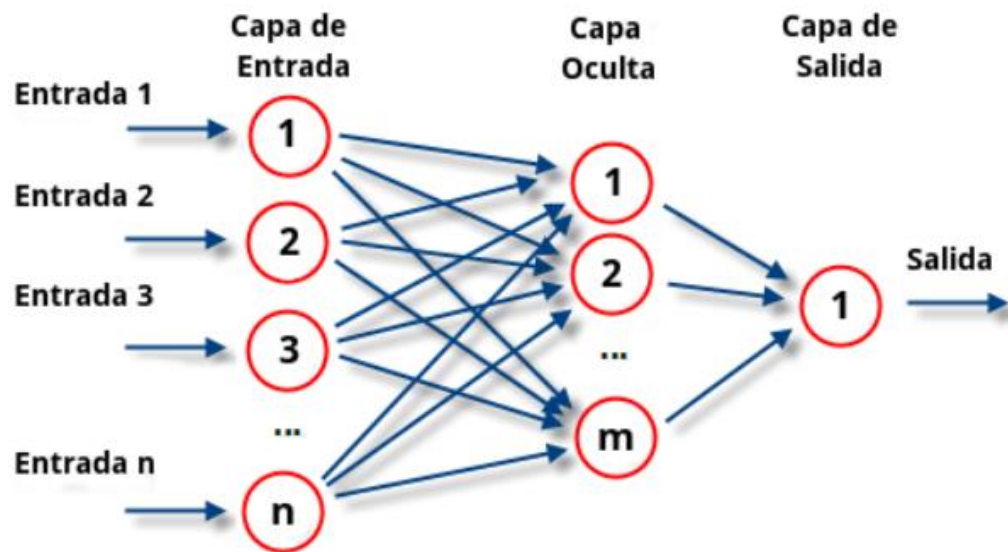


Figura 20 Red neuronal multicapa.

Respecto a la construcción de la red, sigue una metodología parecida a la de los árboles de decisión ya que hay dos etapas: el entrenamiento y la predicción. Los datos de entrada se dividen en datos de entrenamiento y datos de validación. En el entrenamiento se realiza un ajuste de cada uno de los pesos de entradas de todas las neuronas que conforman la red para que el valor que se obtiene a la salida se ajuste lo más posible a las salidas que usamos en el entrenamiento.

La formulación matemática de este algoritmo es la siguiente:

Dado un conjunto de ejemplos de entrenamiento $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ donde $x_i \in \mathbf{R}^n$ y $y_i \in \{0, 1\}$, una única capa oculta y una única neurona MLP aprende de la función $f(x) = W_2 g(W_1^T x + b_1) + b_2$ donde $W_1 \in \mathbf{R}^m$ y W_2, b_1, b_2 representa el bias añadido a la capa oculta y a la capa de salida respectivamente $g(\cdot): \mathbf{R} \rightarrow \mathbf{R}$ es la función de activación, conjunto por defecto como la hiperbólica. Esto resulta en:

$$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

Para la clasificación binaria, $f(x)$ pasa a través de la función logística $g(z) = 1/(1+e^{-z})$ para obtener valores de salida entre cero y uno. Un umbral, se establece a 0.5, asignará muestras de salidas más grandes o iguales a 0.5 a la clase positiva y el resto a la clase negativa.

Si hay más de dos clases, $f(x)$ por sí misma será un vector de tamaño $(n_classes)$ en lugar de pasar a través de la función logística, pasará a través de la función softmax, que se describe como,

$$softmax(z)_i = \frac{\exp(z_i)}{\sum_{l=1}^k \exp(z_l)}$$

Donde z_i representa el elemento i de la entrada de softmax, el cual corresponde a la clase i , y K es el número de clases. El resultado es un vector que contiene las probabilidades de que la muestra x pertenezca a cada clase. La salida es la clase con mayor probabilidad.

En regresión, la salida remanece como $f(x)$, así que, la función de activación es la función de identidad. MLP usa diferentes funciones de pérdida dependiendo del tipo de problema. La función de pérdida para la clasificación es la cross-entropy, la cual en el caso binario es dada como:

$$Loss(\hat{y}, y, W) = -y \ln \hat{y} - (1 - y) \ln (1 - \hat{y}) + \alpha \|W\|_2^2$$

Empezando desde unos pesos iniciales $\nabla Loss_W$ aleatorios, la MLP (multi layer perceptrón) minimiza la función de pérdida actualizando repetidamente estos pesos. Después de calcular la pérdida, un paso backward se propaga desde la capa de salida hasta las capas previas proporcionando a cada parámetro de peso un valor de actualización destinado a disminuir la pérdida. En el descenso de gradiente, el gradiente de la pérdida con respecto a los pesos es calculado y deducido desde W . Más formalmente, esto es expresado como,

$$W^{i+1} = W^i - \epsilon \nabla Loss_W^i$$

Donde i es el paso de iteración, y ϵ es la velocidad de aprendizaje con un valor mayor que 0. El algoritmo para cuando alcanza un número máximo de iteraciones, o cuando la mejora en la pérdida está por debajo a un cierto número pequeño.

De estas dos técnicas, se ha elegido aquella que mejor ha caracterizado el sistema en base a las métricas de precisión habituales. Los algoritmos han de ser capaces de analizar e interpretar los indicadores definidos en tareas anteriores. Estos indicadores se han empleado con el fin de predecir la incidencia de factores relevantes para la producción. Dependiendo de lo que se desee predecir, se ha tenido en cuenta parámetros distintos. En este caso de estudio en particular, se desea predecir la aparición de estreses de origen biológico como enfermedades y plagas (riesgos bióticos) o con origen ambiental, como el estrés hídrico y el estrés de salinidad. Para los riesgos bióticos se ha determinado que los indicadores óptimos son: la

temperatura ambiental; la temperatura ambiental máxima; humedad relativa máxima; humedad relativa media y las precipitaciones. Los riesgos abióticos por su parte tienen en cuenta: la temperatura ambiental; la temperatura ambiental máxima; humedad relativa mínima; humedad relativa media; salinidad; radiación y número de horas de radiación.

El algoritmo escogido ha de realizar una predicción de forma automatizada sobre los niveles de los riesgos bióticos y abióticos del cultivo en base a las entradas mencionadas anteriormente. Estos niveles pueden tomar tres valores distintos que reflejan la probabilidad de incidencia de los riesgos. Los niveles son los siguientes: riesgo alto; riesgo medio y riesgo bajo

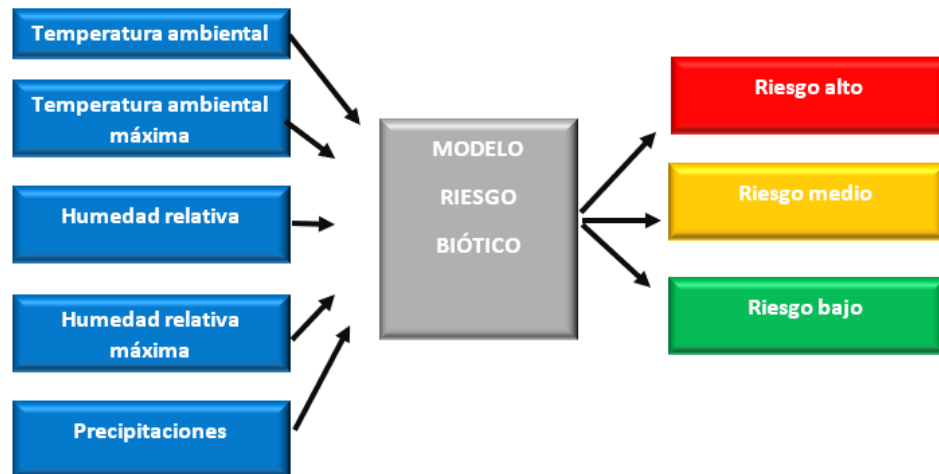


Figura 21 Diagrama de flujo para la determinación de riesgos bióticos.

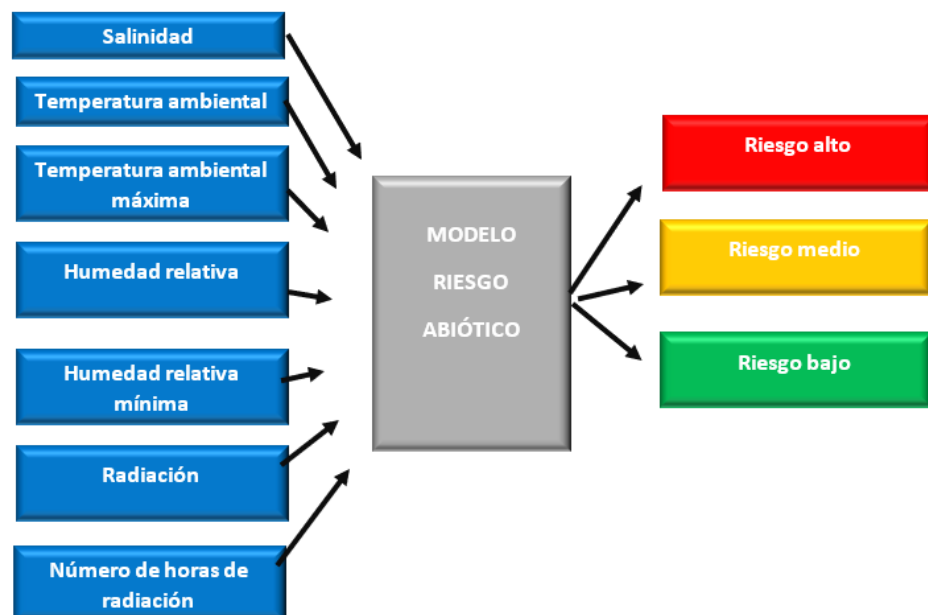


Figura 22 Diagrama de flujo para la determinación de riesgos abióticos

Además de la elaboración de los algoritmos de inteligencia artificial para predecir los riesgos bióticos y abióticos, se han realizado dos tareas más:

- La primera consiste en la preparación de algoritmo que analice la sostenibilidad del cultivo

analizando las medidas de humedad de los sensores instalados a distintas profundidades. El programa debe capturar los insumos aplicados y analizar el alcance de estos (deben quedarse en niveles superiores, junto a las raíces de las plantas).

- Finalmente, con el objetivo de estudiar de manera más exhaustiva la influencia de los fertilizantes en los cultivos, se han examinado datos referentes al manejo de cultivos, como la dosis, el volumen y el número de aplicaciones de fertilizantes. La finalidad de este estudio es encontrar correlaciones de los insumos con parámetros como el peso, el calibre de los frutos, la producción, etc. Se ha escogido el cultivo de tomate de la variedad Rubenza debido a que es del que se posee una mayor cantidad de información. Se va a estudiar concretamente la relación de este con las siguientes variables: tipo de fertilizante; número de aplicaciones de los fertilizantes; volumen de agua usada con fertilizante para el riego conocido como volumen de caldo; cantidad de fertilizante empleada; peso del fruto; calibre del fruto; producción; número de días que dura el experimento.

4 IMPLEMENTACIÓN

En este capítulo se estudiará de forma más exhaustiva las características y factores para tener en cuenta en la construcción del sistema de inteligencia artificial. Estos son los indicadores que se han usado como entradas, las salidas y la estructura jerárquica de los distintos modelos.

4.1 Entradas.

Se ha establecido que los datos se intercambien mediante serialización a través de objetos json. Se implementa una API específica para cada fuente de datos: IFAPA (Estaciones Agroclimáticas Automáticas) y Odin (sensórica instalada en la finca experimental). Estas APIs se encargan de interpretar la estructura del documento intercambiado (típicamente json). A continuación, se muestra un ejemplo de dicho json:

<pre>[{ "code": "Disp_14_7_1", "name": "IFAPA_14_7_1", "descriptions": "IFAPA 14_7 section 01", "fields": [{ "name": "tempMedia", "values": [13.3, 10.8,] }, ... { "name": "radiacion", "values": [19.8, 18.1] }, { "name": "TimeStamp", "values": ["2021-03-18T08:01:49.193", "2021-03-19T08:02:40.182"] }] }, ...]</pre>	<pre>{ "code": "Disp_14_7_2", "name": "IFAPA_14_7_2", "descriptions": "IFAPA 14_7 section 02", "fields": [{ "name": "humedadMedia ", "values": [45.1, 48.7,] }, ... { "name": "precipitacion", "values": [0.0, 0.4.] }, { "name": "TimeStamp", "values": ["2021-03-18T08:01:49.193", "2021-03-19T08:02:40.182"] }] }</pre>
--	--

Figura 23 Entradas json.

Volviendo al proceso de integración para particularizar sobre los algoritmos finalmente integrados, cuando los datos son servidos a los scripts, estos se presentan en formato json con tantas estructuras como *variables* de entrada se hayan seleccionado. Se recopilarán los datos de estas junto con un array adicional para las marcas de tiempo. La longitud del array dependerá del periodo de tiempo seleccionado y de la frecuencia de recopilación de los datos. Al recibir esta información, se hace la lectura de las variables que intervienen en los algoritmos de inteligencia artificial. Estos indicadores han sido seleccionados en base a la experiencia de los expertos y tienen el fin de proveer información sobre las condiciones climáticas de la zona donde se encuentra el cultivo. Estas variables interactúan entre ellas, afectando al cultivo en mayor o menor medida. Tener estas variables controladas es clave para poder hacer frente a posibles estreses en el cultivo. Estas son las siguientes:

- **Temperatura:**

Se puede englobar diferentes datos de temperatura, temperatura ambiental media, temperatura ambiental máxima y temperatura ambiental mínima; además de las horas con los valores máximos de cada una de ellas. Cuando las temperaturas alcanzan niveles extremos (tanto temperaturas altas como temperaturas bajas), la planta puede sufrir fenómenos de estrés. Las temperaturas altas generan dos tipos de estreses:

 - Estrés hídrico: Las plantas cierran estomas (órgano de las hojas encargado del intercambio de gases) para evitar perder agua por calor y en consecuencia a esto, las raíces no pueden ejercer una actividad normal. Al no poder la planta hacer un uso normal del agua disponibles, está entra en estrés hídrico.
 - Estrés lumínico: Cuando la temperatura es demasiado alta, las plantas no pueden realizar bien la fotosíntesis, la cadena de transporte de electrones no funciona correctamente y estos se desvían. Para evitar que estos electrones generen ROS (Especies Reactivas de Oxígeno, en sus siglas en inglés) la planta redirige estos electrones para formar carotenos (pigmento protector lumínico). Si las condiciones que generan este fenómeno persisten, la planta empieza a acumular ROS y acelera el envejecimiento.
- **Radiación:** Esta es una variable íntimamente relacionada con la anterior, ya que esta representa la intensidad con la que incide la luz en el cultivo, provocando los mismos efectos estresante en la planta que la temperatura.
- **Humedad relativa:** Esta cuando es ambiental, es un factor clave para el desarrollo de plagas y enfermedades. El rango óptimo de la humedad esta entre 60 -70%.
 - Humedad relativa alta: Cuando la humedad relativa es alta y viene acompañada con días de temperaturas altas (generalmente en primavera), se crean las condiciones ideales para generar enfermedades fúngicas o bacterianas.
 - Humedad relativa baja: Cuando la humedad relativa es baja y viene acompañada de días con temperatura alta (generalmente se da a finales de primavera/inicio de verano) se dan las condiciones ideales para la aparición de plagas en el cultivo.
- **Precipitaciones:** Las precipitaciones, pueden generar diferentes problemas si son muy abundantes. Estrés hídrico por afixia de la raíz al anegar el suelo o pérdida de producción por caída de frutos.
- **Salinidad:** La salinidad presente en el suelo es la que limita la cantidad de insumos que es capaz de admitir el suelo y que sea asimilable para la planta. Una salinidad alta, aunque no genere un estrés hídrico, si puede significar tener que aportar insumos de manera más espaciada y aumentando el número de tratamientos y de agua aportados (siendo al final el cultivo menos rentable). Las consecuencias del exceso de salinidad son la toxicidad debida al exceso de iones Na⁺, lo que lleva consigo un déficit hídrico como consecuencia de la pérdida de agua por procesos de ósmosis

4.2 Salidas.

Para los scripts en los que se usan los modelos, se generará como salida un fichero json con los resultados de las predicciones realizadas por los algoritmos entrenados. Para cada riesgo evaluado se mostrarán los siguientes parámetros:

- Nivel de riesgo.
- Porcentaje de nivel de riesgo.

4.3 Scripts.

Los modelos de aprendizaje automático se generan para la evaluación de distintos tipos de riesgos. Para la generación y uso de estos, se usarán scripts de Python con extensiones de tipo py. Habrá cuatro scripts independientes, enumerados a continuación:

- `model_train_biotico.py`: Encargado de la generación y entrenamiento del modelo de aprendizaje automático que predice los riesgos bióticos.
- `model_run_biotico.py`: Destinado a ejecutar el modelo previamente entrenado de riesgos bióticos.

- `model_train_abiotico.py`: Cuyo objetivo se trata de la generación y entrenamiento del modelo de aprendizaje automático que predice los riesgos abióticos.
- `model_run_abiotico.py`: Encargado de ejecutar el modelo entrenado de riesgos abióticos.

Estos scripts siguen el siguiente esquema general:

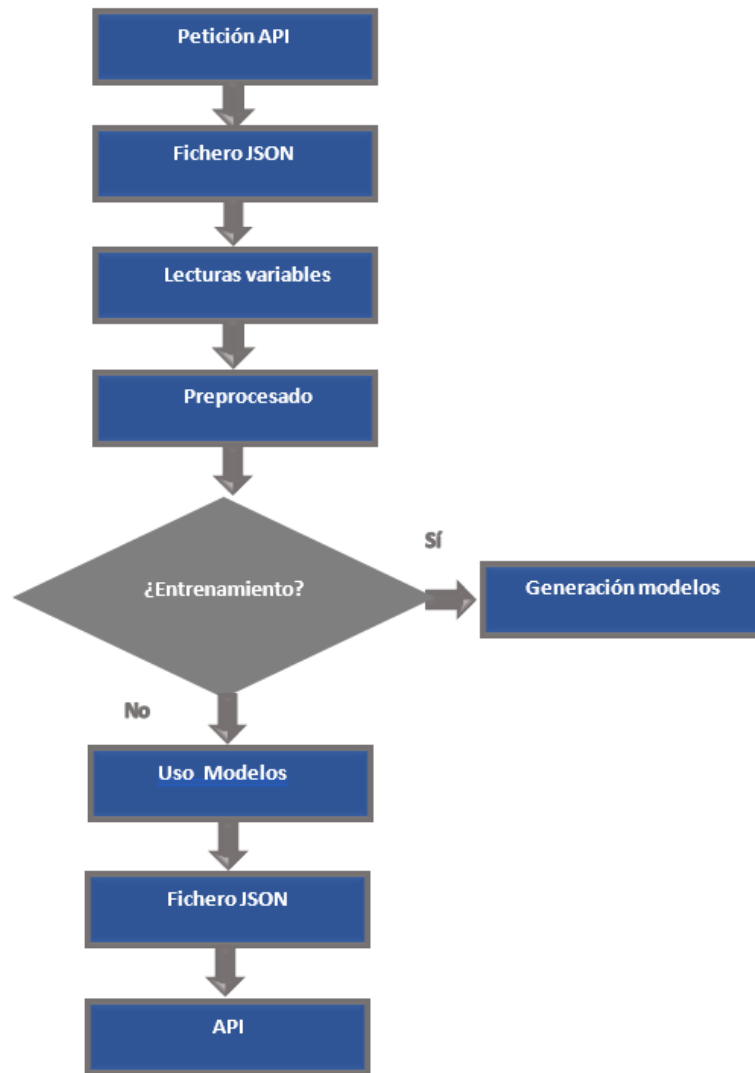


Figura 24. Organigrama generación/uso de los modelos de la plataforma.

Para elaborar los modelos de inteligencia artificial se probaron con distintas técnicas de redes neuronales (DNNClassifier, KerasClassifier o el Sequential) y árboles de decisión (DecisionTreeClassifier y DecisionTreeRegressor). Finalmente, se optó por la modalidad de árboles de decisión debido a su mejor precisión (100%) y ajuste al problema que se plantea. Esto se estudiará de forma más detallada en el siguiente capítulo de Resultados.

Como se ha mencionado anteriormente, hay dos tipos de scripts: de entrenamiento o de ejecución. Además, se contemplan dos modelos distintos: el modelo para los riesgos bióticos, en el que se consideran factores biológicos de los cultivos como plagas y enfermedades, y el modelo para los riesgos abióticos, donde se predice los riesgos de estreses ocasionados por factores no biológicos como la radiación y la sequía.

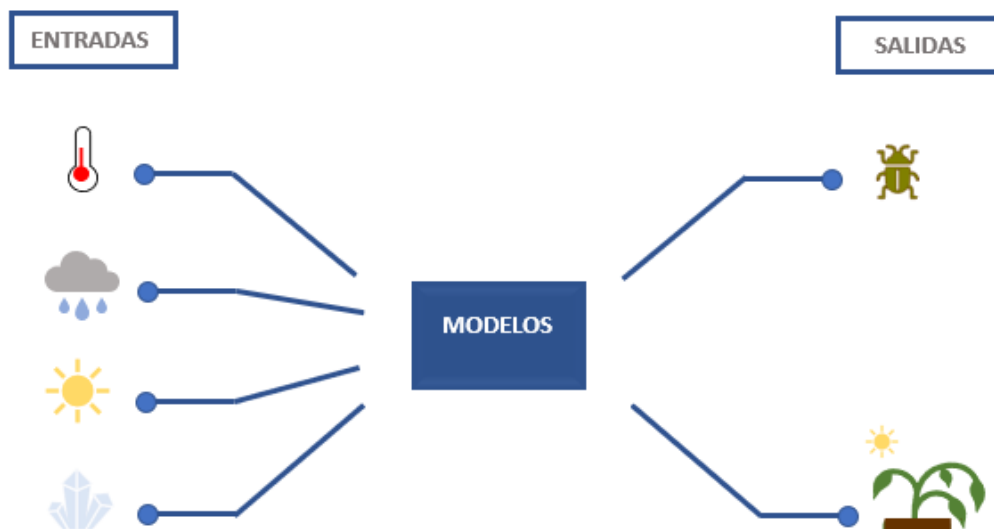


Figura 25. Esquema entradas/salidas de los modelos.

A continuación, se van a comentar los cuatro programas de extensión py de manera individual. Para el análisis de estos se ha optado por organigramas, ya que de forma visual se explica con mayor claridad la metodología y el razonamiento empleado en cada script.

4.3.1 Entrenamiento riesgos bióticos.

En primer lugar, se procede a analizar el programa “`model_train_biotico.py`”. El objetivo de este es la generación y el entrenamiento de un algoritmo capaz de predecir los riesgos bióticos. Concretamente, dentro de los riesgos bióticos, los expertos señalaron la distinción de riesgos de plagas y riesgos de enfermedades. Primeramente, se incluyen librerías entre las que destacan: Scikit-learn, para los algoritmos de Machine Learning; Pandas, para la manipulación y análisis de datos y Numpy para el cálculo numérico. El siguiente paso es la recopilación de la información para el entrenamiento del modelo. En este caso, para la creación de los modelos se emplearon los datos tomados en diciembre. Por una parte, mediante peticiones a la API, se toman las entradas de los modelos, es decir, las distintas medidas de las variables climatológicas que proporcionan los sensores. La temperatura ambiental, las precipitaciones y la humedad son los tres parámetros que afectan a los riesgos bióticos. Por otro lado, los expertos dan información de las salidas con las que se va a entrenar el modelo, dicho de otra forma, de las predicciones de riesgos de plagas y enfermedades que ellos estiman en base a su experiencia. Las medidas de los sensores junto con la información de los expertos conforman los datos necesarios para el entrenamiento del modelo. Estos datos se subdividirán en dos conjuntos de distinto tamaño. El 80 % de la información se destina al entrenamiento y el 20 % restante para la validación. Para hacer esta partición se ejecuta la siguiente orden (X , son las entradas, y las salidas):

```
X_train, X_test, y_train, y_test= train_test_split(X, y, test_size=0.2)
```

Como se aprecia anteriormente dentro de los subconjuntos de entrenamiento (train) y de validación(test), los datos se dividen a su vez en entradas (X_{train} , X_{test}) y salidas (y_{train} , y_{test}). Una vez preparados los datos, se construye el modelo del árbol.

```
model=DecisionTreeClassifier()
```

Para el entrenamiento del árbol se utiliza el siguiente comando:

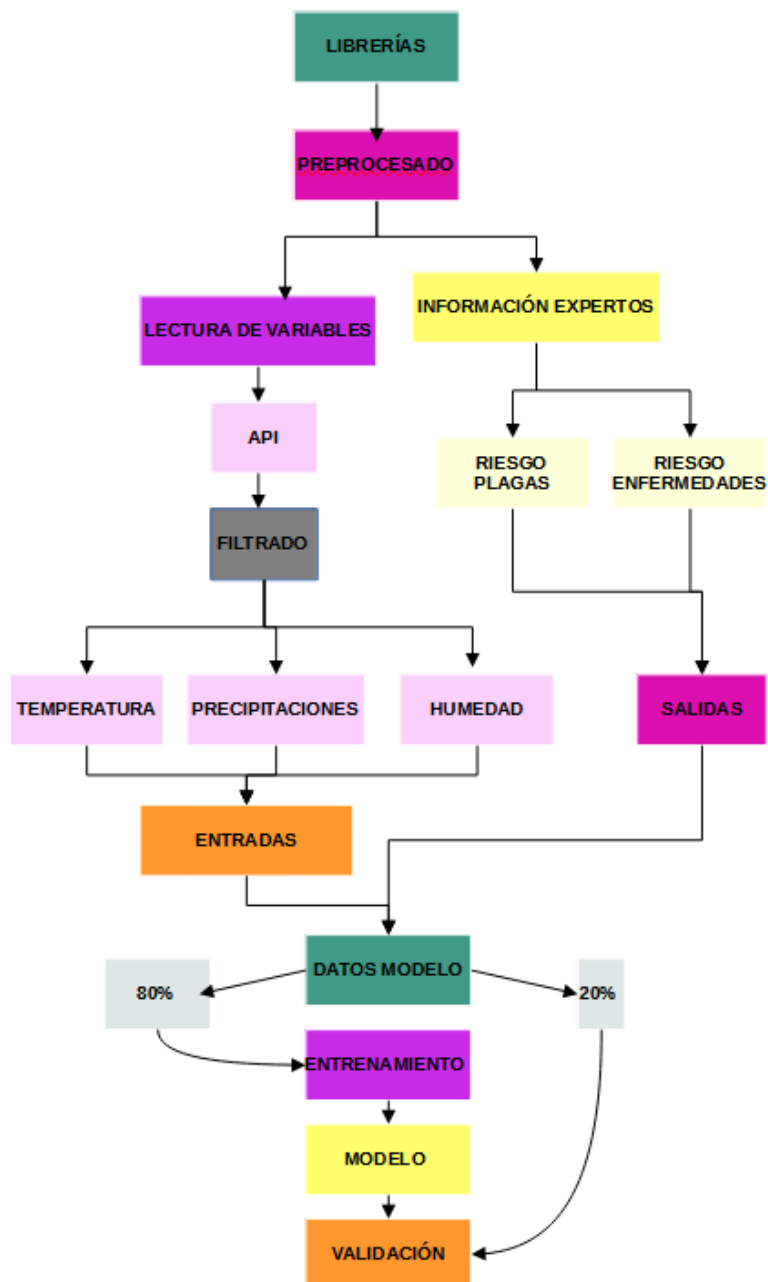

```
model.fit(X_train,y_train)
```

Por otro lado, para validar el modelo se emplea el error cuadrático medio.

```
predictions=model.predict(X_test)  
mean_squared_error(y_test,predictions)
```

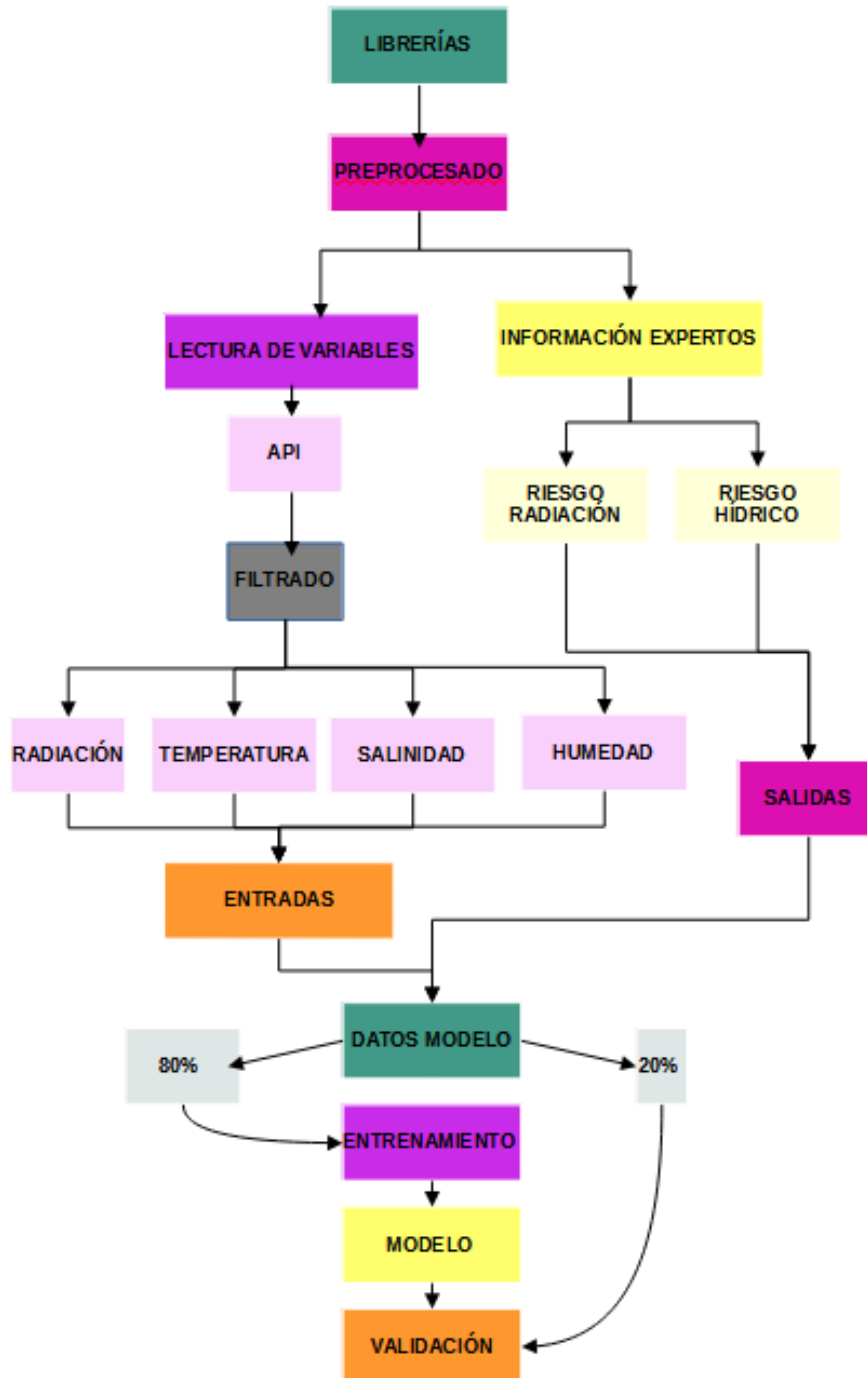
Si el resultado de la validación es favorable, se procederá a guardar el modelo en formato joblib para su posterior uso.

```
joblib.dump( model,'biotico.joblib')
```



4.3.2 Entrenamiento riesgos abióticos.

Para el script “model_train_abiotico.py” se sigue un procedimiento similar al anterior programa. En este caso, la tarea a realizar es la predicción de riesgos abióticos, concretamente se escogieron dos tipos, el riesgo de estrés por radiación y de estrés hídrico. Las variables utilizadas como entradas en el árbol de decisión son distintas al modelo previo. Para los riesgos abióticos, las variables ambientales son: radiación, temperatura, salinidad y humedad.



4.3.3 Ejecución riesgos bióticos.

En el script “model_run_biotico.py” también se sigue un razonamiento análogo a los anteriores scripts. Para este ejemplo, la finalidad no es generar un algoritmo de aprendizaje, es ejecutarlo. El árbol de decisión que se ejecuta es el que se ha entrenado en el script “model_train_biotico.py”. Al igual que en los programas de generación de modelos, también se incluyen las librerías empleadas: Sckit-learn, Pandas, Numpy, etc. El modelo solo necesita como entradas las siguientes variables medidas por los sensores: temperatura, precipitaciones y humedad. No se introduce la información de los expertos puesto que el modelo empleado ya ha aprendido de esta en la etapa del entrenamiento. El árbol es capaz de imitar al conocimiento de los expertos y hacer una predicción de los riesgos de plagas y enfermedades en base a las variables de entrada. Para cargar el modelo entrenado se emplea la siguiente orden:

```
model=joblib.load( 'biotico.joblib')
```

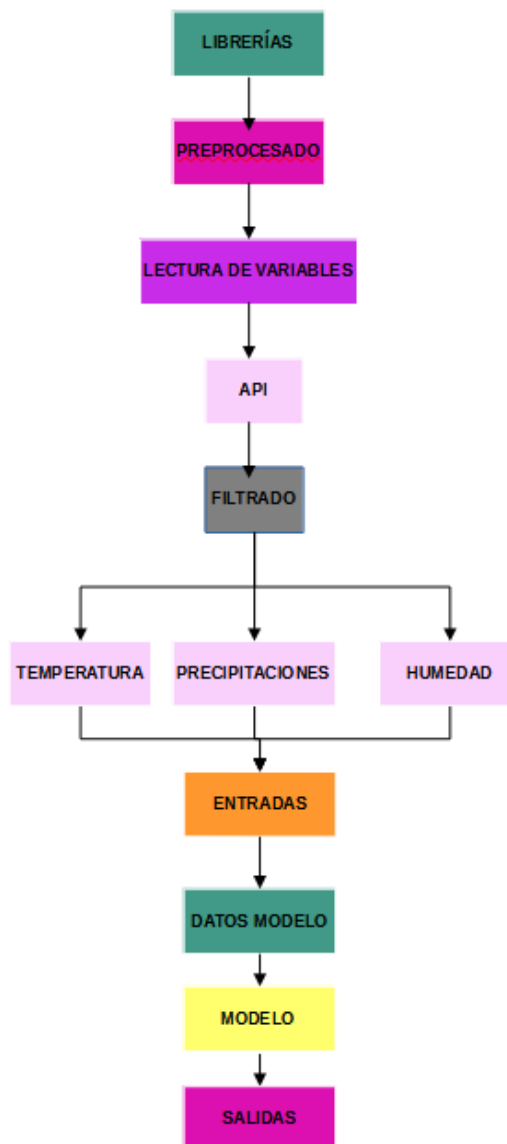


Figura 28. Organigrama script ejecución riesgos bióticos.

4.3.4 Ejecución riesgos abióticos.

Al igual que ocurría con los scripts de entrenamiento, el script “model_run_abiotico.py” sigue un patrón similar al script “model_run_biotico.py”. La diferencia es que en este caso se ejecuta el árbol de decisión de los riesgos abióticos generado en el programa “model_train_abiotico.py”. Las variables que se toman como entradas son: radiación; temperatura; salinidad y humedad. El modelo hace las predicciones del riesgo de radiación y riesgo hídrico que la plataforma se encargará de notificar.



4.3.5 Salidas scripts.

Una vez, generados y ejecutados los algoritmos de aprendizaje automático, se obtendrán las siguientes variables de salida:

- Riesgos bióticos:
 - Nivel riesgo plagas. (ALTO, MEDIO, BAJO)
 - % riesgo plagas.
 - Nivel riesgo enfermedades. (ALTO, MEDIO, BAJO)
 - % riesgo enfermedades.
- Riesgos abióticos:
 - Nivel riesgo radiación. (ALTO, MEDIO, BAJO)
 - % riesgo radiación.
 - Nivel riesgo hídrico. (ALTO, MEDIO, BAJO)
 - % riesgo hídrico.

Formateando estas salidas con formato tipo json, para ser devuelta a través de la API. La estructura usada para esta salida se muestra a continuación:

Json salida run_biotic.py	Json salida run_abiotic.py
<pre>{ "name": "AI_Out_01", "fields": [{ "name": "diseasesRiskLevel", "values": [1, 2] }, { "name": "diseasesProbabilityLevel", "values": [55, 20] }, { "name": "plagueRiskLevel", "values": [2, 0,] }, { "name": "plagueProbabilityLevel", "values": [60, 80] }, { "name": "TimeStamp", "values": ["2021-03-18T08:02:50.374", "2021-03-19T08:03:41.512"] }] }</pre>	<pre>{ "name": "AI_Out_02", "fields": [{ "name": "radiationRiskLevel", "values": [0, 0,] }, { "name": "radiationProbabilityLevel", "values": [30, 40] }, { "name": "hydricStressRiskLevel", "values": [1, 1,] }, { "name": "hydricStressProbabilityLevel", "values": [20, 60] }, { "name": "TimeStamp", "values": ["2021-03-18T08:02:50.374", "2021-03-19T08:03:41.512"] }] }</pre>

<pre>] }] }</pre>	<pre>] }] }</pre>
--------------------	--------------------

Figura 30 Ficheros json de salida.

5 RESULTADOS

En este capítulo se explican los resultados obtenidos usando distintos tipos de algoritmos de aprendizaje automático para realizar una predicción sobre los riesgos de estrés bióticos y abióticos a partir de las medidas climatológicas tomadas por los sensores. Además, se contempla un algoritmo que estudia la sostenibilidad del cultivo a través de las medidas de humedad de distintas profundidades. Por último, se muestran los resultados del estudio realizado sobre la correlación existente entre las variables de manejo del cultivo de tomate de la variedad Rubenza y variables relacionadas con la producción de este.

5.1 Predicción de riesgos bióticos y abióticos

Con la ayuda de las librerías de Sckit-learn y de Tensorflow se pusieron a prueba los algoritmos mencionados en el punto anterior. Se necesitan dos modelos: uno para predecir los riesgos bióticos y otro para predecir los riesgos abióticos.

Para los árboles de decisión se ha experimentado con el módulo *DecisionTreeClassifier()* incluido en la librería de Sckit-learn. Por otro lado, se usaron distintos modelos de redes neuronales artificiales incluidos en la librería de aprendizaje automático de Tensorflow como el *DNNClassifier*, el *KerasClassifier* o el *Sequential* empleando distintas funciones de pérdida.

Para la evaluación de riesgos de origen biótico, la inteligencia artificial ha determinado el árbol de decisión óptimo que se muestra en la 31. En ella se aprecian las reglas extraídas tras el análisis de los datos. Por ejemplo, si se estudia la bifurcación que está más a la izquierda señalada en la imagen se extrae la siguiente regla: “Cuando la temperatura registrada es menor o igual a diez grados y la humedad relativa máxima es menor del noventa y cinco por ciento, el nivel de riesgo biótico es bajo”.

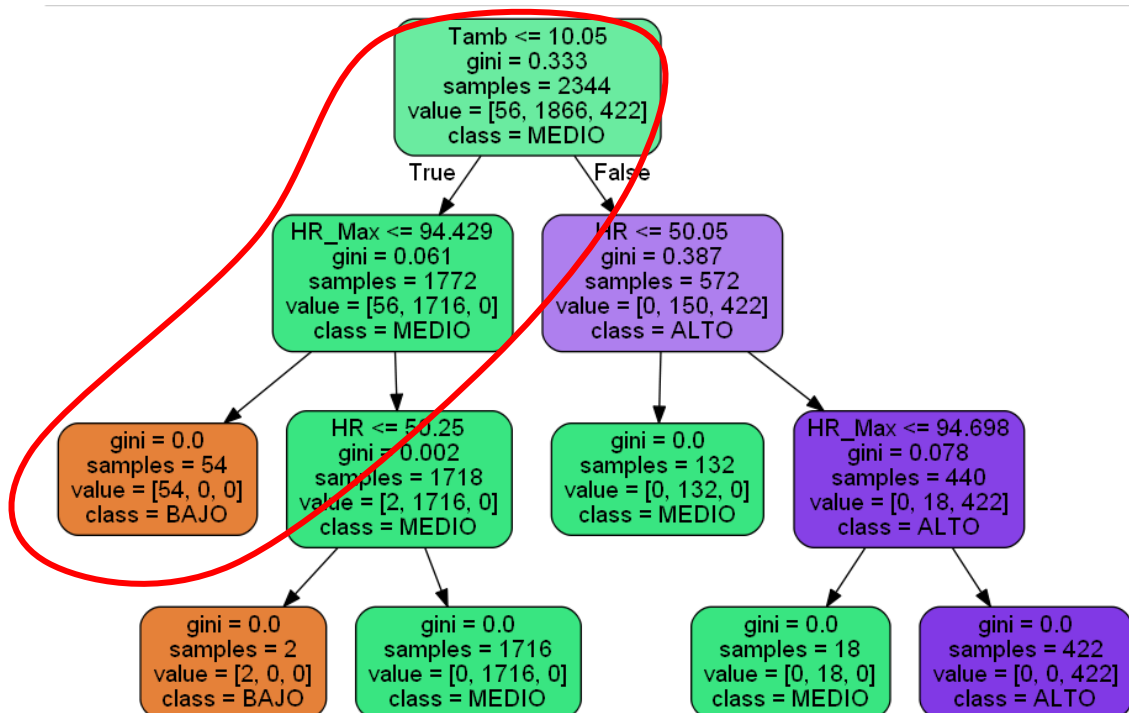


Figura 31 Árbol de decisión para riesgos bióticos

Para evaluar el modelo se muestra a continuación una gráfica en la que se hace la comparación entre el valor de la predicción y el valor real del riesgo. En el eje de abscisas se representa una de las variables de entrada

introducidas en el modelo, en este caso se ha tomado la humedad relativa. Por otro lado, en el eje de ordenadas se muestra el nivel del riesgo, correspondiendo el nivel de riesgo dos a riesgo alto, el nivel uno a riesgo medio y el nivel cero a riesgo bajo. Como se puede contemplar en la siguiente gráfica, el modelo de árbol es capaz de predecir todos los valores y posee una precisión del cien por cien. En esta gráfica, se puede observar cómo van conmutando continuamente los niveles de riesgo.

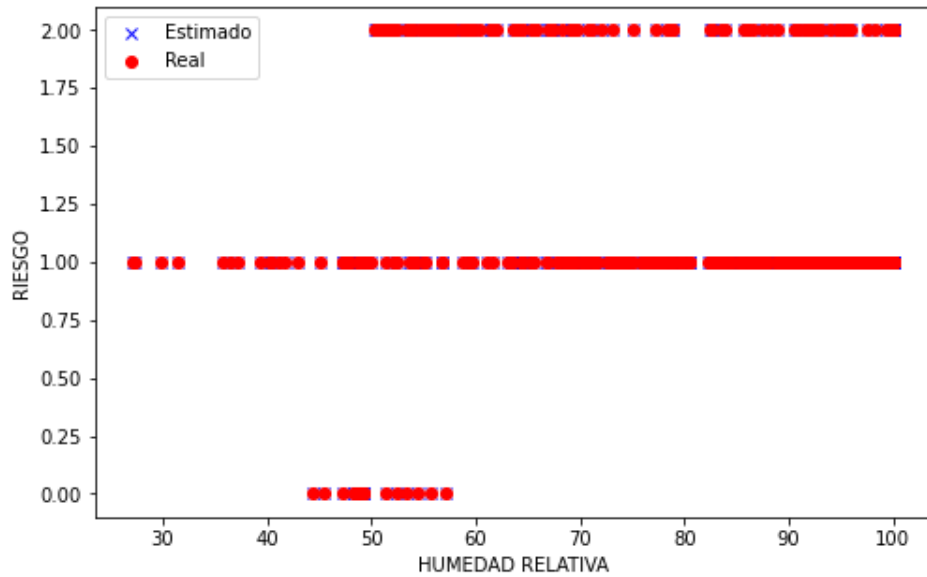
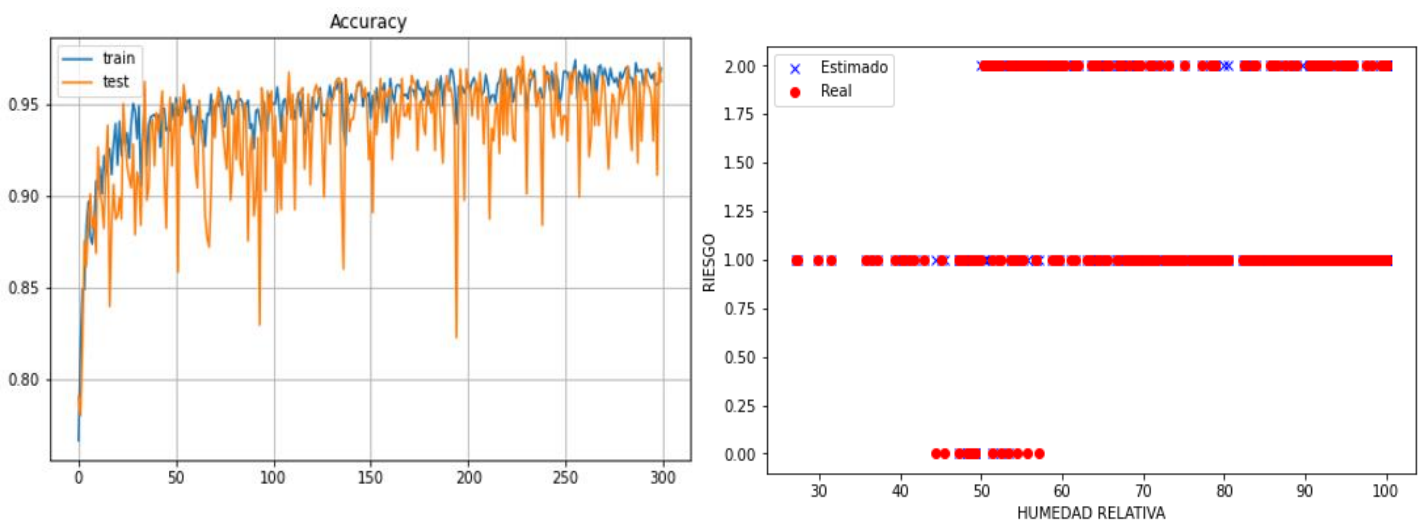
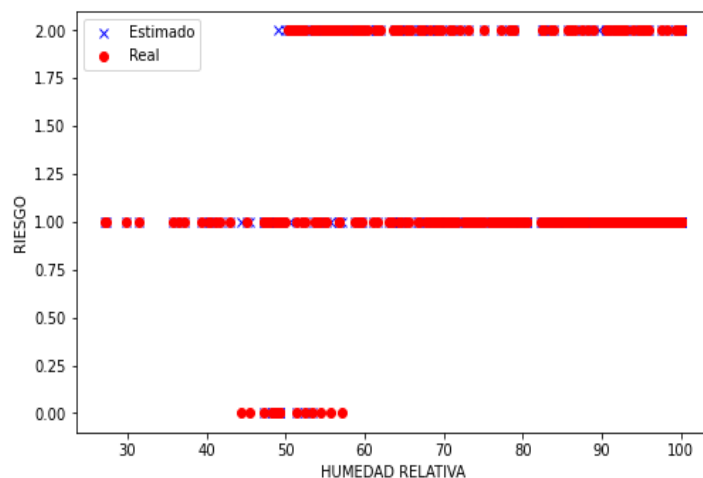
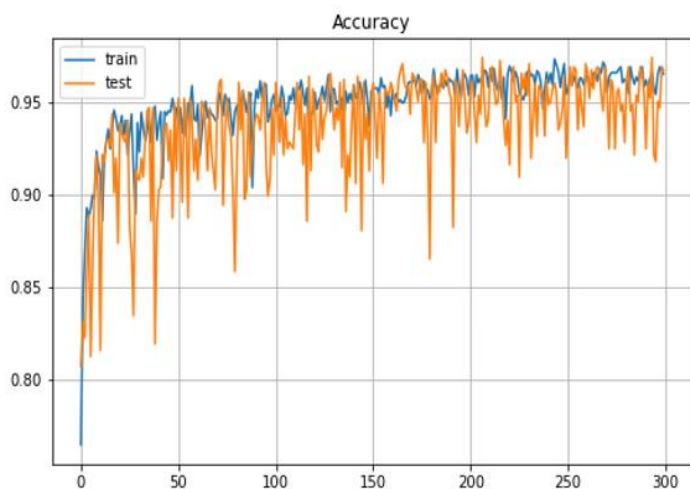


Figura 32. Ejemplo de evaluación de riesgos. Comparación entre el estimado por los técnicos y el determinado por la inteligencia artificial. Se observa una coincidencia completa

Las distintas técnicas de redes neuronales artificiales mencionadas con anterioridad aplicadas para el problema planteado no lograron la eficacia de los árboles de decisión a pesar de que se utilizaron métricas, en principio, más eficientes. Se consiguieron las precisiones más altas con el modelo de KerasClassifier aplicando las funciones de pérdida de 'categorical_crossentropy' (Multi Class Entropy Loss) y de 'sparse_categorical_crossentropy' (Sparse multiclass cross entropy loss) – ver Figura 33 e Figura 34.





En siguiente tabla se reflejan los resultados de la evaluación de los distintos métodos estudiados:

Tabla 3 Precisiones de los distintos algoritmos puestos a prueba.

MODELOS	Decission TreeClassifier	DNN Classifier	Keras- Kullback	Keras Multi Class	Keras Sparse
PRECISIÓN	100 %	87,4 %	79,35 %	97,57 %	97,95 %

Cómo se ha dicho anteriormente, según los resultados experimentales, el algoritmo de árbol de decisión es el más adecuado para el resolver el problema que se plantea. Por este motivo se ha elegido trabajar con este tipo de modelo de aprendizaje automático. Seguidamente se muestra el resultado obtenido para el árbol de decisión de los riesgos abióticos. Al igual que ocurría con el árbol para los riesgos bióticos, el árbol de los riesgos abióticos tiene una precisión del cien por cien

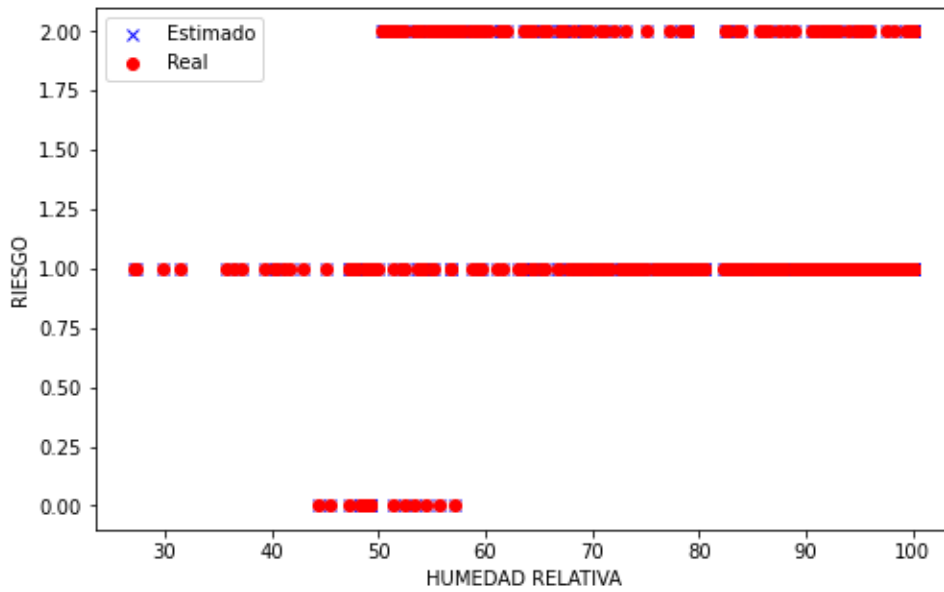
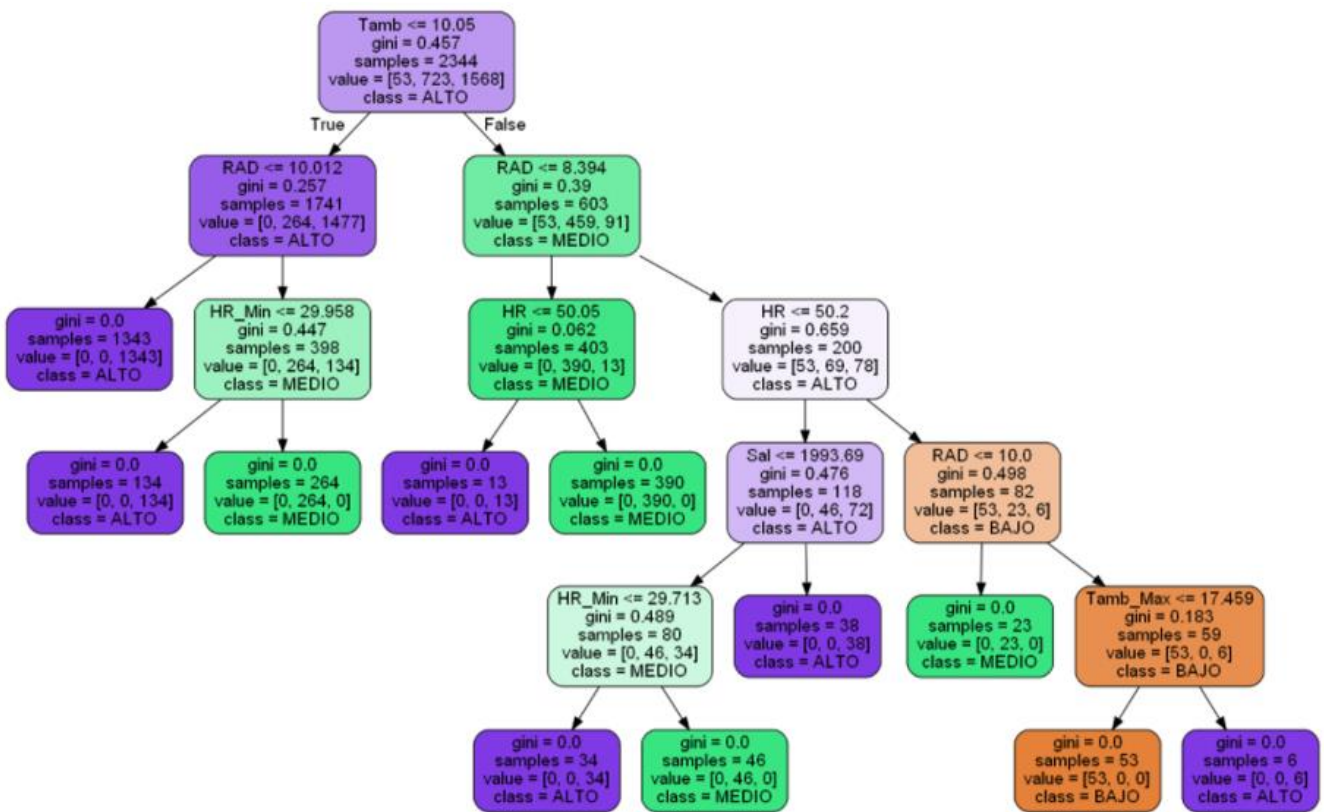


Figura 35 Ejemplo de evaluación de riesgos abióticos.



Complementariamente, se ha implementado el envío automático de correos diarios informando a los expertos de los resultados de los modelos. De esta forma, los especialistas han podido evaluar su eficacia de estos y participar en el ajuste de la inteligencia artificial. En el correo (ver

Figura 37) aparecen dos gráficas para cada tipo de riesgo, la de la izquierda se muestra la evolución del riesgo a lo largo del día y la de la derecha el comportamiento del riesgo en los últimos siete días. En estas gráficas se muestra en el eje x la escala de tiempo y en el eje y el porcentaje de sufrir un determinado nivel de riesgo. Para hacer una distinción entre los posibles niveles de riesgo (alto, medio y bajo), se ha empleado un color para cada uno: rojo para el nivel de riesgo alto; amarillo para el nivel medio y verde para el nivel bajo.

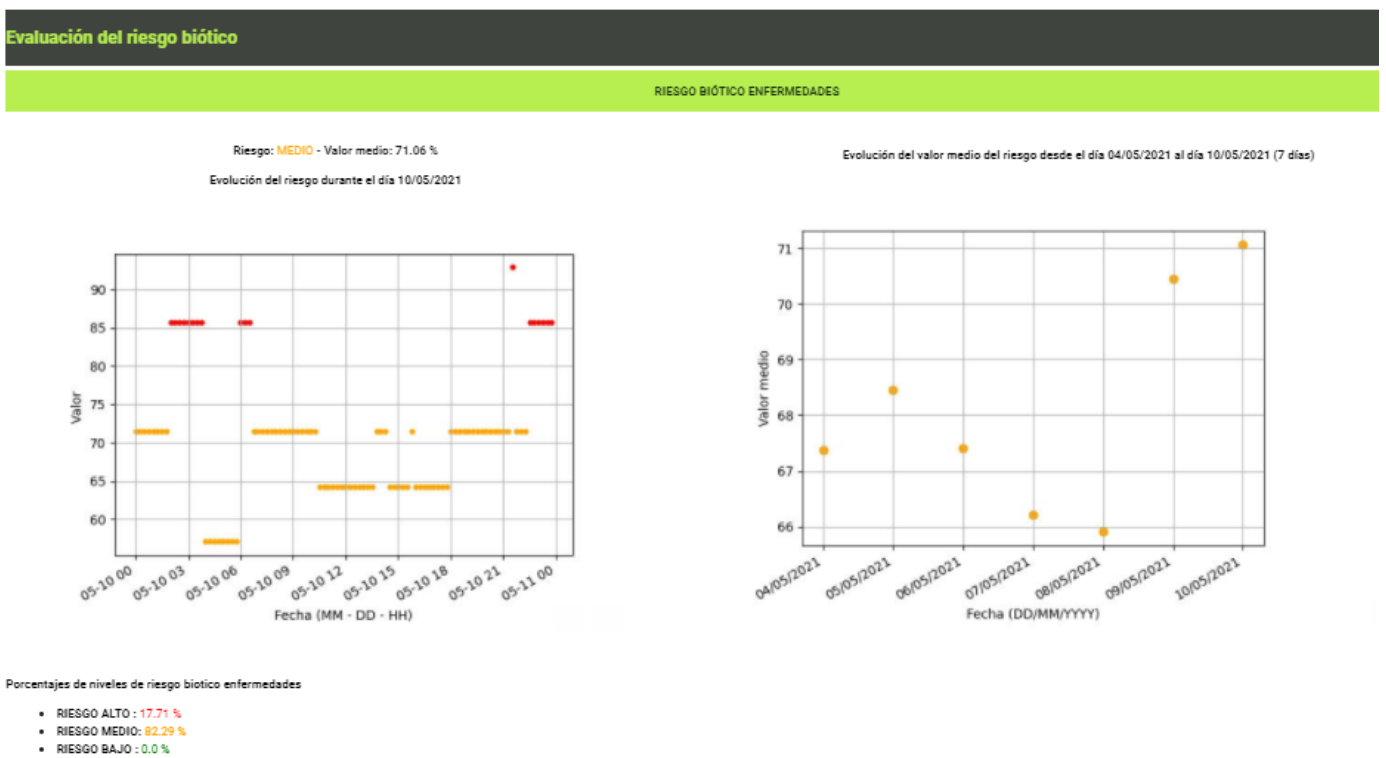


Figura 37. Ejemplo de correo al agricultor informando de la evaluación de riesgos bióticos.

En el correo también se muestran representadas en gráficas las variables empleadas como entradas en los modelos (temperatura, humedad, precipitaciones y salinidad). Se empleará una tonalidad más fuerte cuando el nivel de la variable sea alto, una media para un valor intermedio y una clara para un valor bajo.

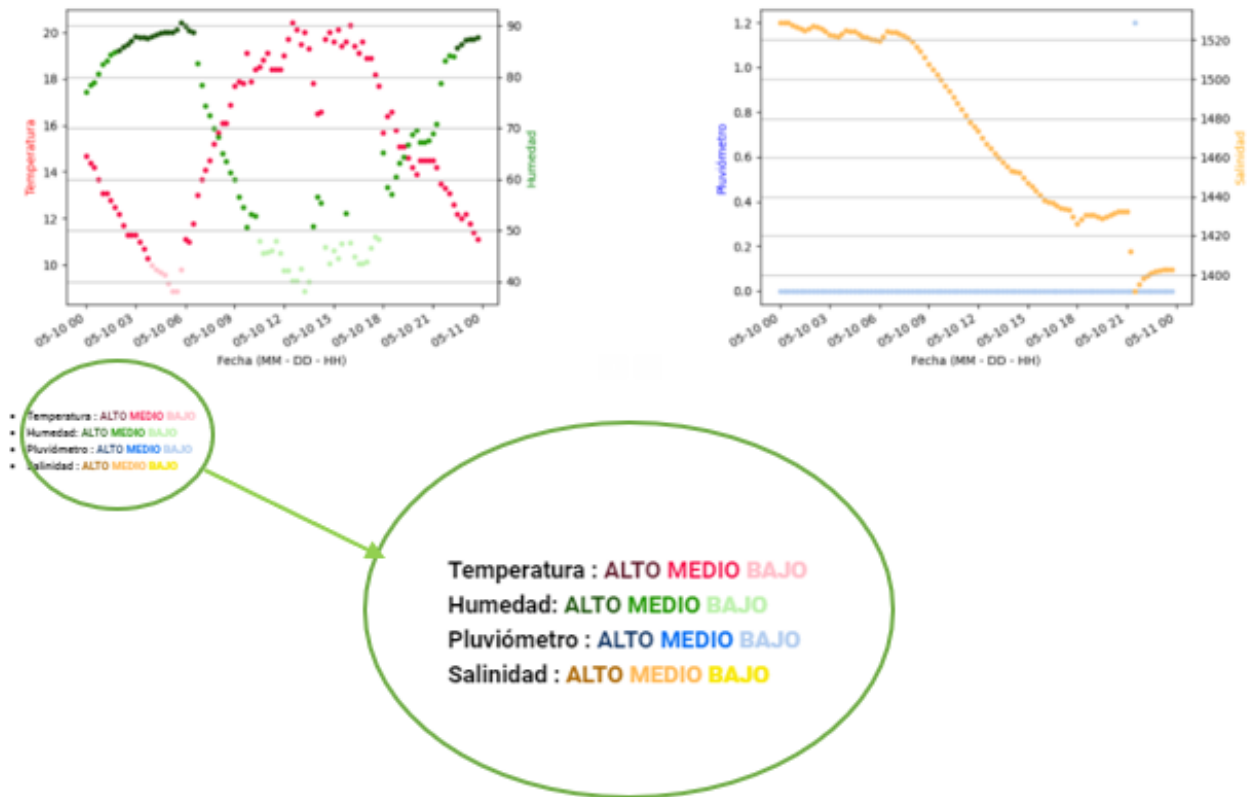


Figura 38. Análisis automatizado de las medidas de los sensores instalados en La Añoreta.

Los expertos validaron la eficacia de los modelos. El cultivo del almendro se mantuvo sano durante la etapa fenológica de floración y cuajado del fruto, sin incidencia de pulgón (0 pulgones/hoja). Esta tendencia se mantuvo durante el desarrollo del cultivo hasta el día 20/03/2021 (Figura 39 y 40).

INFORME 20/03/2021

Evaluación del riesgo biótico

RIESGO BIÓTICO ENFERMEDADES

Riesgo: **BAJO** - Valor medio: 63.76 %

Evolución del riesgo durante el día 20/03/2021

Porcentajes de niveles de riesgo biótico enfermedades

- RIESGO ALTO : 4.17 %
- RIESGO MEDIO: 40.62 %
- RIESGO BAJO : 55.21 %

Evolución del valor medio del riesgo desde el día 14/03/2021 al día 20/03/2021 (7 días)

RIESGO BIÓTICO PLAGAS

Riesgos: **BAJO** - Valor medio: 68.01 %

Evolución del riesgo durante el día 20/03/2021

Porcentajes de niveles de riesgo biótico plagas

- RIESGO ALTO : 4.17 %
- RIESGO MEDIO: 40.62 %
- RIESGO BAJO : 55.21 %

Evolución del valor medio del riesgo desde el día 14/03/2021 al día 20/03/2021 (7 días)

Figura 39 Ejemplo de correo día 20/03/2021

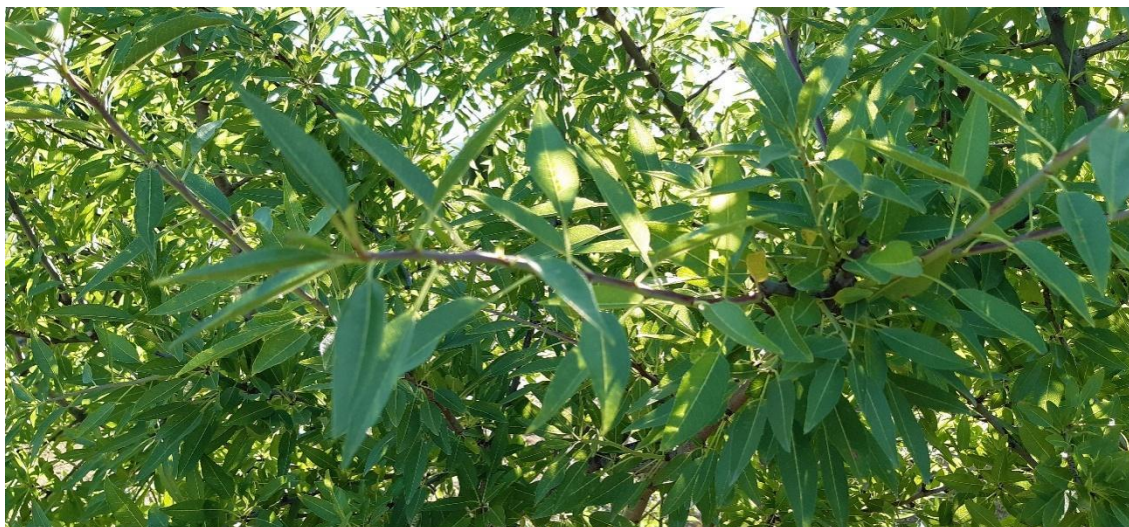


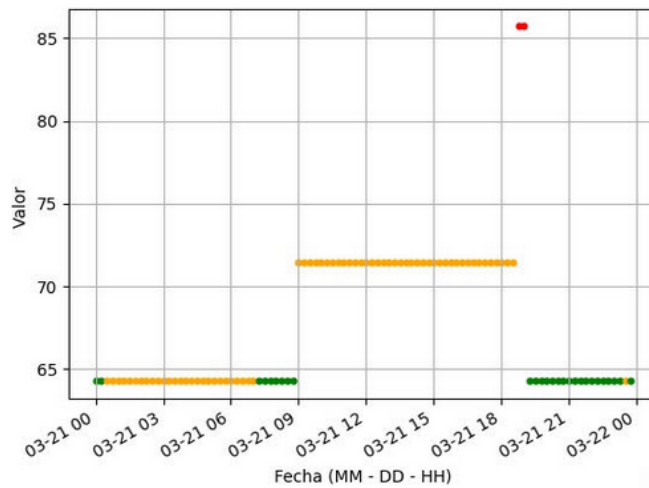
Figura 40 Óptimo desarrollo vegetativo sin afectaciones bióticas.

El día después, los modelos hicieron la predicción de riesgo biótico de plagas con nivel medio (Figura 41 y 42) que coincidió con el inicio de los conteos positivos de adultos de pulgón, aunque aún a niveles bajos (1-3 pulgones/hoja).

RIESGO BIÓTICO PLAGAS

Riesgos: **MEDIO** - Valor medio: 67.63 %

Evolución del riesgo durante el día 21/03/2021



Porcentajes de niveles de riesgo biótico plagas

- RIESGO ALTO : 2.08 %
- RIESGO MEDIO : 69.79 %
- RIESGO BAJO : 28.12 %

Evolución del valor medio del riesgo desde el día 15/03/2021 al día 21/03/2021 (7 días)

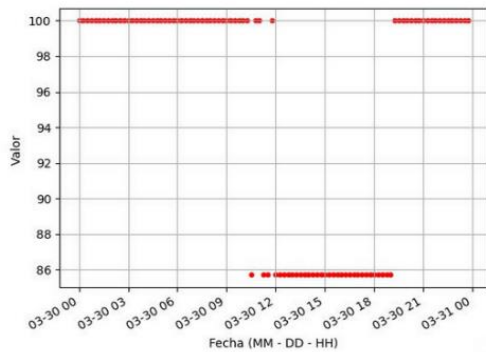
Figura 41 Correo con alertas consistentes de riesgo medio 21/03/2021.



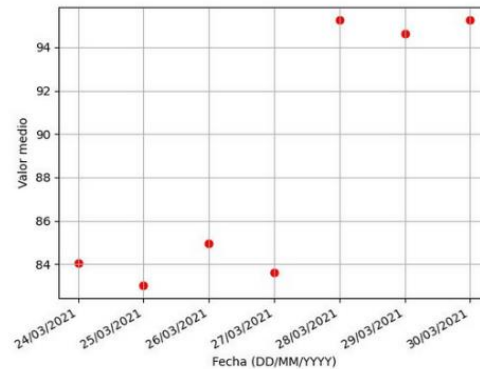
Figura 42 Primeros síntomas de planta debilitada por acción de pulgón.

Los días posteriores a esta alarma, la aplicación arrojaba datos de que esta alerta se mantenía. Esta tendencia se mantuvo hasta el día 30/03/2021, cuando la alerta que proporcionaba la herramienta ya era de riesgo alto de aparición de alguna plaga en el cultivo del Almendro (Figura 7). Los días posteriores a esta alarma se evaluó de nuevo el cultivo de almendro y se confirmó la presencia de la plaga de pulgón con un mayor nivel de severidad (3-9 pulgones/hoja) (Fig. 6)

Riesgos: **ALTO** - Valor medio: 95.24 %
Evolución del riesgo durante el día 30/03/2021



Evolución del valor medio del riesgo desde el día 24/03/2021 al día 30/03/2021 (7 días)



Porcentajes de niveles de riesgo biótico plagas

- RIESGO ALTO : 100.0 %
- RIESGO MEDIO: 0.0 %
- RIESGO BAJO : 0.0 %

Figura 43 Correo con alertas sostenidas de riesgo alto 24-30/03/2021



Figura 44 Planta severamente afectada por la población de pulgón.

Frente a riesgos abióticos no hubo ninguna alarma reseñable, ya que las alarmas de posible estrés hídrico por exceso de humedad se debían a la presencia de riegos largos en el cultivo del Almendro, hasta que este llegaba a capacidad de campo.

La experiencia de uso de un modelo predictivo basado en reglas de toma de decisiones y el empleo del análisis avanzado de datos y la inteligencia artificial ha sido satisfactoria, apoyando en la toma de decisiones en base a pronósticos ajustados a las observaciones realizadas en campo. El empleo de este tipo de tecnologías podrá incrementar la eficacia de las aplicaciones de insumos y facilitar el manejo de los cultivos a técnicos y agricultores. A su vez, fueron contrastadas observaciones previas sobre la relevancia del uso de fertilizantes avanzados y

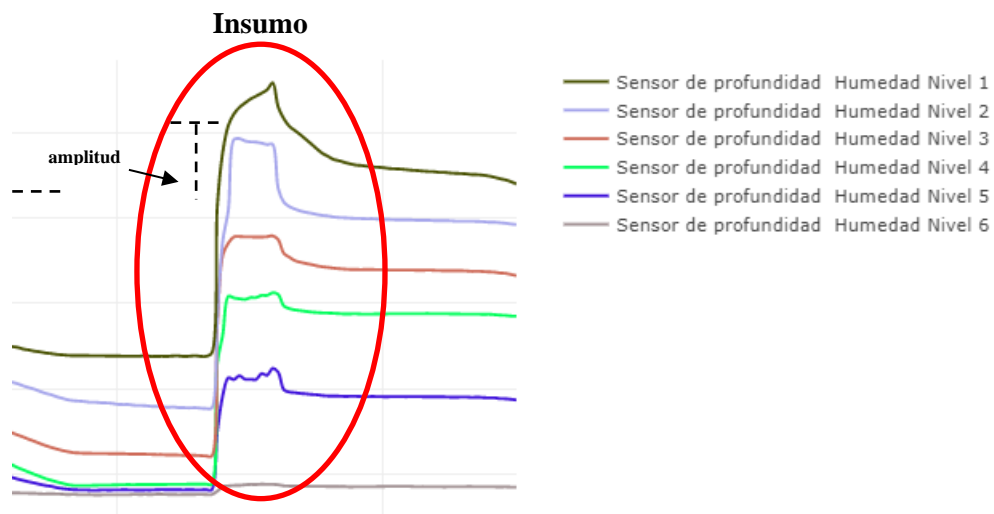
sostenibles para optimizar las aplicaciones realizadas y minimizar los riesgos de pérdida de producto y contaminación de acuífero.

5.2 Análisis de insumos.

Entre los objetivos del proyecto se contempla igualmente llevar a cabo un estudio de la sostenibilidad de los cultivos a partir de las medidas proporcionadas por los sensores colocados a distintos niveles de profundidad junto a las raíces de los cultivos.

Los expertos han considerado necesario estudiar el comportamiento del suelo cuando se aplica un fertilizante sostenible y cuando se usa uno no sostenible. Para realizar esta tarea se tomaron datos de dos entornos experimentales, uno en el que se estaba siguiendo un manejo sostenible y otro en el que no. Partiendo de su experiencia, los expertos determinaron que los indicadores más significativos para el análisis son la humedad y salinidad del suelo. Cuando se aplican los insumos, en las medidas de los sensores de humedad de las distintas profundidades se aprecian picos de incremento. Estos aumentos proporcionan evidencia indirecta sobre la absorción de los productos aplicados. Si el fertilizante llega a niveles profundos (nivel 4 o superior) indica que aquél no se está adhiriendo únicamente a la raíz de la planta, la cual se encuentra en los niveles superiores de profundidad (nivel 1 y nivel 2) sino que, por el contrario, se está filtrando al suelo provocando un riesgo de contaminación de las aguas subterráneas y como consecuencia perjudicando gravemente al ecosistema, ya que estas aguas podrían desembocar en ríos o mares y acabar con la fauna y flora del lugar. La salud de las personas también puede verse afectada, puesto que las aguas contaminadas pueden ser manantiales destinados al consumo humano.

Planteado el problema, se ha diseñado un algoritmo que evalúa la sostenibilidad del insumo suministrado. Para lograr identificar los instantes en los que se ha añadido fertilizante se ha empleado la derivada de las señales de humedad. Para cuantificar el insumo, se mide la amplitud de los picos previamente detectados (ver Figura 16).



Para notificar sobre el análisis del insumo, cada vez que el algoritmo detecte una nueva aplicación en alguna de las dos fincas de estudio se envía un correo con la información de los últimos cien días. En este correo aparece

tanto el comportamiento de los insumos realizados en el entorno de manejo sostenible (entorno A) como en aquél donde se han empleado técnicas no sostenibles de fertilización (entorno B). Para cada uno aparece una tabla donde se detalla la hora de aplicación del insumo y las distintas amplitudes de los picos de humedad correspondientes a los cinco niveles de profundidad presentes. Junto a la tabla aparece de forma resumida la información del porcentaje de días en los que los insumos han llegado a los distintos niveles de penetración además de información estadística referente al valor medio de los picos máximos y de la desviación estándar. En el entorno A, en el cual han usado fertilizantes sostenibles, se ha comprobado que durante ninguna de las aplicaciones se ha originado filtración de producto, puesto que en ningún día las amplitudes de los picos de los niveles inferiores han alcanzado niveles significativos.

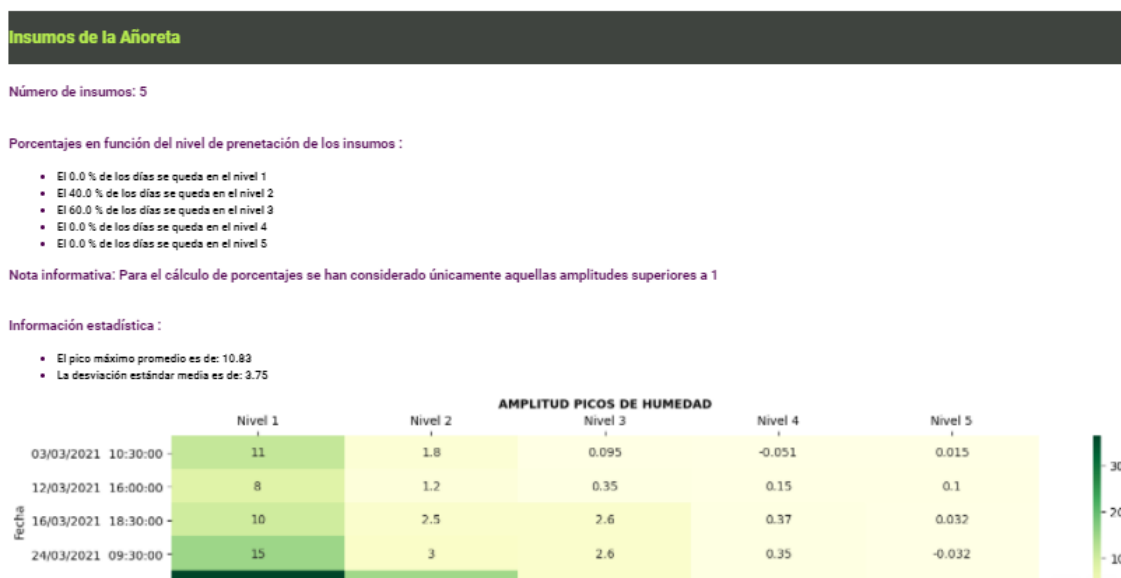


Figura 46. Ejemplo de correo al agricultor informando sobre los insumos del entorno A.

Por otro lado, en el entorno B, según los resultados obtenidos con el algoritmo, el manejo realizado no ha sido sostenible, puesto que el veinticinco por ciento de los días en los que se ha aplicado fertilizantes se ha filtrado el producto y ha quedado en el nivel cuatro. Incluso el dos por ciento de las aplicaciones han llegado hasta el nivel cinco, que se trata del nivel más delicado puesto que es el más próximo a las aguas subterráneas que se encuentran en el suelo

Insumos de la Halconera

Número de insumos: 56

Porcentajes en función del nivel de prenetación de los insumos :

- El 0.0 % de los días se queda en el nivel 1
- El 0.0 % de los días se queda en el nivel nivel 2
- El 73.0 % de los días se queda en el nivel 3
- El 25.0 % de los días se queda en el nivel 4
- El 2.0 % de los días se queda en el nivel 5

Nota informativa: Para el cálculo de porcentajes se han considerado únicamente aquellas amplitudes superiores a 1

Información estadística :

- El pico máximo promedio es de: 8.68
- La desviación estándar media es de: 1.83

	AMPLITUD PICOS DE HUMEDAD				
	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5
04/02/2021 17:00:00	3	4.5	5	7.4	2.2
08/02/2021 13:30:00	2.8	4.7	5	6.1	0.46
12/02/2021 13:30:00	3	5	5	4.8	0.047
15/02/2021 13:30:00	3.1	5.3	5	5.2	0.1
19/02/2021 13:30:00	3.6	6.1	4.8	0.45	-0.047
22/02/2021 13:30:00	3.5	6	4.8	0.54	-0.04
24/02/2021 13:30:00	4.5	6.8	5.8	5	0
26/02/2021 13:30:00	4.5	7.1	5.9	2.4	-0.053
01/03/2021 13:30:00	3.3	7.9	4.6	-0.006	-0.06
03/03/2021 13:30:00	4.3	7.9	5.3	-0.024	-0.046
05/03/2021 13:30:00	4.3	8	5.1	-0.036	-0.053
07/03/2021 10:30:00	-0.13	3.2	5.5	0.4	-0.046
15/03/2021 14:00:00	1.3	7.4	3.7	-0.053	-0.094
17/03/2021 13:30:00	1.9	6.7	2.3	-0.088	-0.073
19/03/2021 13:30:00	1.5	6.8	4.7	-0.047	-0.06
22/03/2021 10:30:00	2.2	8	3.7	-0.088	-0.059
23/03/2021 11:30:00	3.4	6.9	4.9	0.018	-0.046
24/03/2021 09:30:00	2.9	6.7	5.2	1.2	-0.02
25/03/2021 09:30:00	4.1	7.6	5.6	1.8	0

Figura 47. Ejemplo de correo al agricultor informando sobre los insumos del entorno B.

En este correo de insumos, se puede observar también otro factor que interviene en la sostenibilidad de los cultivos, el riego. Mientras que, en los últimos cien días, en el entorno A se han realizado cinco riegos, en el B se han hecho cincuenta y seis. Esto indica que el entorno B, además de usar unos fertilizantes menos respetuosos con el medio ambiente, tiene una huella hídrica mayor que el A.

5.3 Estudio variables de manejo de tomate Rubenza.

Por último, se va a analizar los resultados del estudio realizado sobre la relación existente entre las variables de manejo del cultivo de tomate de la variedad Rubenza y variables relacionadas con la producción de este. En la investigación, se han hecho tres ensayos distintos los cuales son:

- Tesis 1: No se usa ningún tipo de insumo.
- Tesis 2: Se aplica fertilizante PLUS con una dosis de 0.5 L/ha y fertilizante CARBOFIT con una dosis de 2L/ha.
- Tesis 3: Se aplican los mismos tipos de fertilizantes que los aplicados en la tesis 2. La diferencia está en la cantidad aplicada del producto CARBOFIT, en este caso se usa una dosis de 3 L/ha.

Tras un análisis preliminar de las variables de aplicación de fertilizante y el calibre obtenido del fruto, se determinó que existía correspondencia entre ambos. Para la tesis 1, en la cual no se utilizó ningún tipo de

fertilizante, se obtuvo un calibre medio del fruto de 7,4 cm. Por otro lado, para la tesis 2 y para la tesis 3, en las que sí se aplicaron insumos, se consiguieron tras al menos 8 aplicaciones, calibres medios de 7.73 cm y 7.83 cm respectivamente. Los resultados demuestran que los ensayos en los que se emplearon fertilizantes se lograron frutos de mayor calibre. Incluso se contempla que a mayor cantidad de insumo mayor es el calibre obtenido como se refleja en el ensayo 3.

A continuación de forma más detallada se procederá a estudiar algunas de las relaciones más significativas entre los parámetros tenidos en cuenta en el análisis. En primer lugar, se observa que entre el peso y el calibre del fruto existe una relación lineal, a pesar de que se cabría esperar una relación cúbica, ya que el calibre depende del radio y el peso del radio al cubo.

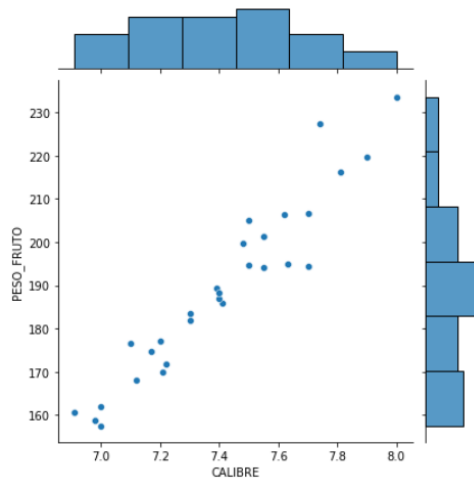


Figura 48. Gráfica correlación entre peso y calibre del fruto para la Tesis 1.

También se prevé que exista una fuerte dependencia entre el número de días y el peso del tomate. A más días que transcurren, la planta tiene más tiempo para crecer y mayor es el peso del fruto.

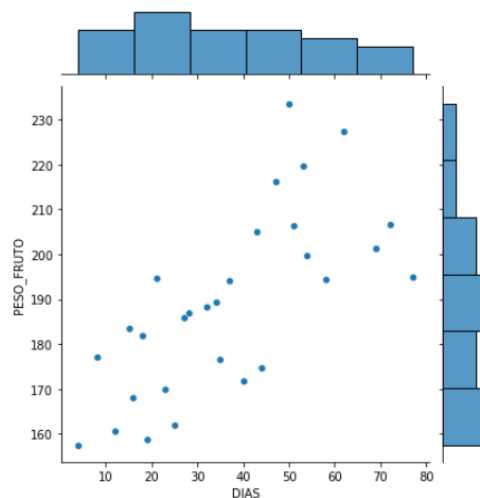


Figura 49. Gráfica correlación entre peso del fruto y días del experimento para la Tesis 1.

Como se ha mencionado con anterioridad, el uso de insumos está ligado al calibre del fruto. En la siguiente gráfica, se observa la relación lineal que existe entre el número de aplicaciones de insumos y el calibre. A más aplicaciones, mayor es el calibre del tomate.

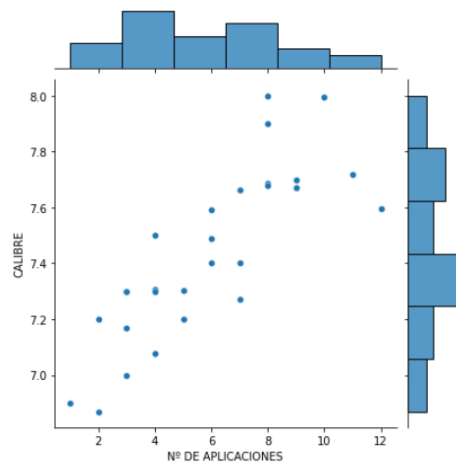


Figura 50. Gráfica correlación entre calibre del fruto y aplicaciones de los insumos para la Tesis 2.

Por último, se pasó a elaborar un algoritmo de aprendizaje automático empleando el método de regresión lineal. El objetivo del modelo de inteligencia artificial es predecir el calibre del tomate a partir de variables como el tipo de ensayo, el fertilizante, el número de días o el número de aplicaciones. A continuación, se muestra un resumen de las características de modelo construido. Según el valor de R-squared del modelo de 0.646, la precisión del algoritmo es bastante buena. Observando la columna de $P > |t|$, la cual indica la importancia de las variables (cuanto más pequeño más importante), las variables con mayor peso son: la TESIS, que indica que ensayo de los tres es, y la variable DÍAS, lo cual guarda coherencia, ya que el calibre aumenta con el número de días de los ensayos como se ha visto anteriormente.

Dep. Variable:	CALIBRE	R-squared:	0.621			
Model:	GLS	Adj. R-squared:	0.607			
Method:	Least Squares	F-statistic:	45.56			
Date:	ju., 04 feb. 2021	Prob (F-statistic):	1.14e-27			
Time:	09:25:01	Log-Likelihood:	44.592			
No. Observations:	145	AIC:	-77.18			
Df Residuals:	139	BIC:	-59.32			
Df Model:	5					
Covariance Type:	nonrobust					
	coef	std err	t	P> t 	[0.025	0.975]
Intercept	6.7162	0.095	70.861	0.000	6.529	6.904
C(PRODUCTO_CONTROL)[T.1]	0.2772	0.092	3.017	0.003	0.096	0.459
C(PRODUCTO_PLUS)[T.1]	-0.0005	0.034	-0.014	0.989	-0.068	0.067
TESIS	0.1249	0.034	3.698	0.000	0.058	0.192
DIAS	0.0078	0.001	5.660	0.000	0.005	0.011
APLICACIONES	0.0297	0.011	2.688	0.008	0.008	0.052
Omnibus:	2.943	Durbin-Watson:	1.572			
Prob(Omnibus):	0.230	Jarque-Bera (JB):	2.481			
Skew:	0.209	Prob(JB):	0.289			
Kurtosis:	2.514	Cond. No.	356.			

Figura 51. Resultados primer algoritmo.

Si se construye el modelo a partir de los ensayos en los que se aplica fertilizante únicamente se obtienen los resultados que se muestran a continuación. En este modelo se consigue una predicción de 0.66, mejor que la del anterior modelo. Aquí la variable de mayor peso es el número de aplicaciones.

Dep. Variable:	CALIBRE	R-squared:	0.660
Model:	GLS	Adj. R-squared:	0.644
Method:	Least Squares	F-statistic:	42.68
Date:	ju., 04 feb. 2021	Prob (F-statistic):	3.02e-24
Time:	09:25:01	Log-Likelihood:	42.542
No. Observations:	116	AIC:	-73.08
Df Residuals:	110	BIC:	-56.56
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	6.9759	0.151	46.213	0.000	6.677	7.275
C(TESIS)[T.3.0]	0.1336	0.045	2.955	0.004	0.044	0.223
C(PRODUCTO_PLUS)[T.1]	-0.0290	0.132	-0.220	0.826	-0.290	0.232
DOSIS	-0.0065	0.064	-0.101	0.919	-0.134	0.121
DIAS	0.0029	0.002	1.349	0.180	-0.001	0.007
APLICACIONES	0.0622	0.015	4.024	0.000	0.032	0.093

Omnibus:	3.422	Durbin-Watson:	1.531
Prob(Omnibus):	0.181	Jarque-Bera (JB):	2.391
Skew:	0.177	Prob(JB):	0.302
Kurtosis:	2.392	Cond. No.	564.

Figura 52. Resultados segundo algoritmo

En la siguiente gráfica se muestra el resultado de la predicción del algoritmo de inteligencia artificial anteriormente visto.

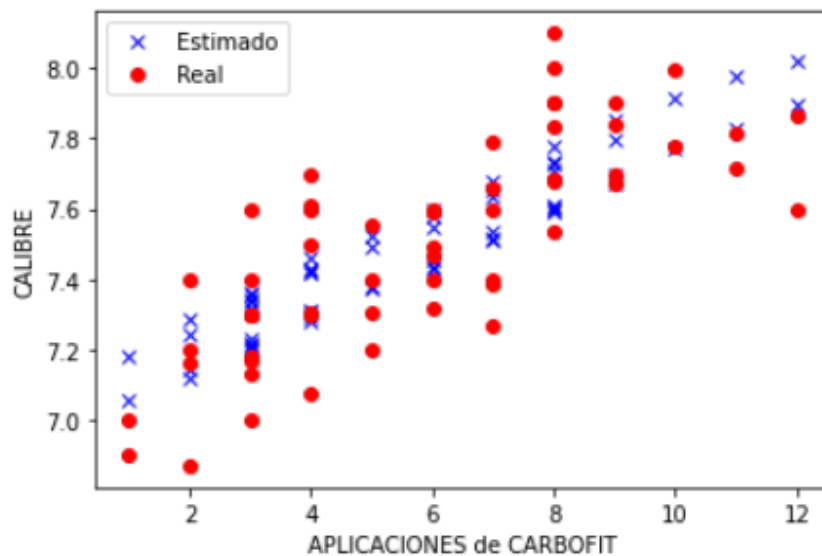


Figura 53. Predicción del calibre de fruto empleando el segundo algoritmo

6 CONCLUSIONES

6.1 Conclusiones

En conclusión, se ha conseguido elaborar un algoritmo de inteligencia artificial capaz de aprender de la experiencia de los técnicos. Este algoritmo de clasificación automática es capaz de hacer una predicción sobre los porcentajes y niveles de riesgos bióticos y abióticos que tiene el cultivo en cuestión en base a las condiciones climatológicas medidas con los sensores instalados. Los resultados obtenidos con el modelo suponen un soporte para el agricultor en la toma de decisiones del manejo de sus cultivos. Al tener una predicción de las plagas y enfermedades que pueden darse, el productor puede aplicar los insumos solo cuando el algoritmo indique que exista riesgo. De esta forma, se reduce la cantidad de fertilizantes aplicados a los cultivos y por consecuencia sus posibles impactos al medioambiente.

De igual modo, se ha construido un algoritmo capaz de analizar la sostenibilidad de los cultivos a través de las distintas medidas de los sensores de humedad colocados junto a las raíces de los cultivos. Este algoritmo detecta los niveles de penetración en los que se queda los insumos. El agricultor puede saber de esta forma de la sostenibilidad de su cultivo a partir del algoritmo. Si se detecta que los insumos han alcanzado niveles muy profundos, el cultivo no es sostenible puesto que provocar la contaminación de las aguas subterráneas.

6.2 Trabajos futuros

En un futuro se contempla la elaboración de otros modelos de inteligencia artificial que empleen técnicas de análisis de imágenes. El objetivo de aplicar un sistema de visión de aprendizaje automático sería que este aprendiera desde las imágenes a reconocer factores influyentes como son la existencia de plagas como los pulgones o de enfermedades.

REFERENCIAS

- [1] La agricultura mundial en la perspectiva del año 2050. (2009,12 de octubre). fao.org.
- [2] Andrade, M.A (2018). Machine learning algorithms applied to the forecasting of crop water stress indicators. In Proceeding in technical irrigation show. California, USA: Irrigation Association.
- [3] Chlingaryan, A. (2018, 1 agosto). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *ScienceDirect*.
- [4] Dimitriadis, S. (2008, 1 agosto). Applying Machine Learning to Extract New Knowledge in Precision Agriculture Applications. *IEEE Conference Publication / IEEE Xplore*.
- [5] Pratyush,K (2020).IoT based Smart Agriculture using Machine Learning, Second International Conference on Inventive Research in Computing Applications (ICIRCA).
- [6] Virnodkar, S. S. (2020, 17 febrero). Remote sensing and machine learning for crop water stress determination in various crops: a critical review. *Precision Agriculture*.
- [7] Behmann, J. (2014, 31 agosto). A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precision Agriculture*.
- [8] Ihuoma, S.O. (2017, 1 septiembre). Recent advances in crop water stress detection. *ScienceDirect*.
- [9] Howell, T. A. (1984, 1 abril). *Response of cotton water stress indicators to soil salinity*. Irrigation Science.
- [10] Rossum, G (2009). Python 3 reference manual Python documentation manual part 1. SoHoBooks.
- [11] Rossum, G (2009). Python 3 reference manual Python documentation manual part 2. SoHoBooks.

- [12] Paper, D. (2020). Hands-on Scikit-Learn for Machine Learning Applications Data Science Fundamentals with Python
- [13] Silaparasetty, N. (2020). Machine learning concepts with Python and the Jupyter Notebook environment : using Tensorflow 2.0 .
- [14] Moolayil, J. (2019). Learn Keras for Deep Neural Networks A Fast-Track Approach to Modern Deep Learning with Python.
- [15] Breiman, L. (2001, 1 octubre). Random Forests. *Machine Learning*.
- [16] Chopda, J (2018) Cotton Crop Disease Detection using Decision Tree Classifier. International Conference on Smart City and Emerging Technology .
- [17] Papageorgiou, E (2011, 1 junio). Fuzzy cognitive map based approach for predicting yield in cotton crop production as a basis for decision support system in precision agriculture application. *ScienceDirect*.
- [18] Wu, J (2009). The Development and Application of Decision Tree for Agriculture Data .2009 Second International Symposium on Intelligent Information Technology and Security Informatics.
- [19] Yi-yang, G (2009). Data mining and analysis of our agriculture based on the decision tree. ISECS International Colloquium on Computing, Communication, Control, and Management.
- [20] Jiang, D. (2004). An artificial neural network model for estimating crop yields using remotely sensed information. *International Journal of Remote Sensing*.