

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/282076145>

A support system for predicting vine mildew fungus

Research · September 2015

DOI: 10.13140/RG.2.1.2263.0888

CITATIONS

0

READS

839

1 author:



[Roi Levoso](#)

Universidad del Pais Vasco / Euskal Herriko Unibertsitatea

1 PUBLICATION 0 CITATIONS

SEE PROFILE



UNIVERSIDAD DE A CORUÑA

FACULTAD DE INFORMÁTICA

Mención en Computación

GRADO EN INGENIERÍA INFORMÁTICA

Sistema de apoyo a la predicción del hongo mildiu en la vid

Autor: Roi Levoso Fernández

Directoras: Amparo Alonso Betanzos
Verónica Bolón Canedo

A Coruña, Junio de 2015

14 de junio de 2015
UNIVERSIDAD DE A CORUÑA

FACULTAD DE INFORMÁTICA
Campus de Elviña s/n
15071 - A Coruña (España)

Aviso legal:

No está permitida la reproducción total o parcial de este documento, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier medio, ya sea electrónico, mecánico, por fotocopia, por registro u otros medios, sin el permiso previo y por escrito del autor del mismo.

A mis abuelos.

Agradecimientos

Gracias, en primer lugar, a mis directoras, a Amparo por aceptar este proyecto nada más proponérselo y a Verónica que a pesar de estar lejos durante el tramo final siguió ofreciéndome apoyo. Agradecer a ambas la paciencia que han tenido conmigo.

Gracias a mis padres, sobre todo mi padre que fue el culpable de que terminara haciendo este proyecto, y a mi familia en general por apoyarme.

Gracias a Susana por darme toda la información y a prestarme su tiempo para ayudarme a entender como se combate el mildiu, a Julián por toda la información que me ha proporcionado y que me ha ayudado a entender más esta enfermedad y a Jose

Antonio por ayudarme con los datos.

Gracias a mis compañeros de piso Iván y Juan por aguantarme cada vez que les hablaba del proyecto.

Gracias a todos mis amigos.

Roi Levoso Fernández

Junio de 2015

Resumen

El Mildiu de la Vid es un hongo que ha provocado muchos daños en las plantaciones de vid durante el último siglo, sobre todo en zonas húmedas. Es por ello que en Galicia la zona del Ribeiro, situada en los valles del Miño y el Avia, y la zona de Rías Baixas hayan sido muy vulnerables a esta enfermedad. Por esta razón es importante trabajar en una mejora en las diferentes estrategias de protección existentes. Desde que se detectó por primera vez esta enfermedad, a finales del siglo XIX, hasta finales del XX, las estrategias de protección consistían en el uso indiscriminado y sistemático de productos químicos que, si bien protegían la planta contra la enfermedad, disminuían la calidad del producto y aumentaban el coste de producción. En los últimos años, y gracias a que se le está dando una mayor importancia por parte de las autoridades a la calidad sanitaria de los alimentos, se está trabajando en nuevos métodos que permitan minimizar el uso de productos químicos en las plantas, dañinos para el medio ambiente y, por ende, para el ser humano. Estos métodos se valen de datos meteorológicos para crear modelos que permiten describir el comportamiento del hongo y así dar lugar a una toma de decisiones, en cuanto a los tratamientos, más eficiente.

Este trabajo tiene por objetivo ser una ayuda a la hora de tomar las decisiones con respecto a la fecha en que se darán los tratamientos. Para cumplir este objetivo, en primer lugar, se estudiarán una serie de modelos existentes que ya han sido probados y validados en otras áreas vitivinícolas. En segundo lugar, se propondrán y analizarán nuevos métodos basados en aprendizaje máquina. Finalmente se evaluará cada uno de estos métodos y modelos para determinar cuál de ellos proporciona mejores resultados en cuanto a la predicción del riesgo de infección del hongo, lo que permitirá determinar de manera óptima la fecha de tratamiento. Una vez determinado el método que proporcione mejores resultados, se procederá a la implementación e integración de este en un sistema que ayudará al experto a la hora de tomar decisiones en cuanto a tratamientos.

El sistema que se desarrollará en este trabajo contará con una interfaz amigable que permita la visualización de datos meteorológicos, proporcionados por los diferentes sensores que están colocados estratégicamente en las viñas de la *Cooperativa Vitivinícola do Ribeiro*. También proporcionará datos relativos a los tratamientos y, como ya se ha

adelantado, incorporará los resultados obtenidos por el método de predicción más eficaz.

Palabras clave: *mildiu, modelo, clasificación, datos meteorológicos*

Índice general

1. Introducción	1
1.1. Objetivos	2
1.2. Planificación	3
1.3. Estructura de la memoria	4
2. Materiales y métodos	7
2.1. Herramientas software	7
2.1.1. Matlab	7
2.1.2. Weka	8
2.1.3. Python	8
2.1.4. Perl	9
2.2. Técnicas de evaluación	9
2.3. Herramientas de campo	10
3. Mildiu. Modelos actuales	17
3.1. Historia del patógeno	17
3.2. Biología del Mildiu	17
3.3. Estrategias de protección	18
3.3.1. EPI	20
3.3.2. Modelo de Goidanich	22
3.3.3. PMI	25
3.3.4. Otros modelos	26
4. Conjuntos de datos, métodos de clasificación y resultados	29
4.1. Datos meteorológicos	29
4.1.1. Datos históricos	31
4.2. Métodos de clasificación	32
4.2.1. C4.5	33
4.2.2. Naïve Bayes	34
4.2.3. IB1	36
4.2.4. SVM	37
4.2.5. Regresión Logística	39

4.3. Resultados del modelo de Goidanich	39
4.4. Comparación de resultados	40
5. Desarrollo del sistema	43
5.1. Interfaz	46
6. Conclusiones y trabajo futuro	51
6.1. Trabajo futuro	52
I. Código	53
I.1. Implementar petición al servidor	53
II. Instalación	55
II.1. Requisitos software	55
II.2. Instalación	55

Índice de figuras

1.1. Diagrama de Gantt del proyecto.	6
2.1. Sensor de temperatura y humedad ambiente.	11
2.2. Sensor de humedad de suelo.	12
2.3. Sensor de temperatura de suelo.	12
2.4. Sensor de humedad foliar.	13
2.5. Sensor de velocidad y dirección del viento.	14
2.6. Sensor de radiación solar.	14
2.7. Pluviómetro.	15
3.1. Ciclo de vida del mildiu.	18
3.2. Relación entre la humedad, la temperatura y el porcentaje de desarrollo del hongo.	24
5.1. Diagrama de clases de cliente Resful.	45
5.2. Diagrama de casos de uso.	45
5.3. Interfaz de usuario	46
5.4. Código de colores	47
5.5. Cuaderno de tratamientos.	48
5.6. Caso de uso tratamiento.	49
5.7. Caso de uso parámetro.	49

Índice de tablas

1.1. Tabla de costes del proyecto.	4
4.1. Matriz de confusión de C4.5	34
4.2. Matriz de confusión de Naïve Bayes	36
4.3. Matriz de confusión de IB1	37
4.4. Matriz de confusión de SVM	38
4.5. Matriz de confusión de Regresión Logística	39
4.6. Matriz de confusión del modelo de Goidanich	40
4.7. Comparación de los principales métodos	40

CAPÍTULO 1

Introducción

Durante los últimos años la aplicación de las tecnologías de la información se ha ido incrementando, abriendo un amplio campo de investigación y haciendo posible la modernización de esta actividad. Esta nueva tendencia, que se conoce como *e-agriculture*, engloba la concepción, diseño, desarrollo, evaluación y aplicación de formas innovadoras de utilización de las tecnologías de la información en la agricultura. La aplicación de esta técnica está relacionada con numerosos procesos agrícolas, como son la gestión del agua, control de la cadena de producción o control de enfermedades, entre otros. El control de enfermedades haciendo uso de las tecnologías de la información se incluye en el Manejo Integrado de Plagas y Enfermedades (MIP), tomando especial relevancia en la etapa de prevención.

El Mildiu de la Vid es una enfermedad fúngica producida por el hongo *Plasmopara viticola*, originaria de América del Norte y que afecta hoy en día a multitud de cultivos de vid. El tratamiento indiscriminado con productos químicos ha sido durante años la solución preferida por agricultores que veían como, de esta manera, se erradicaba la incidencia de este patógeno. A medida que se ha ido poniendo en práctica el MIP, el tratamiento indiscriminado con químicos ha ido siendo rechazado y se ha optado por un firme seguimiento del hongo para determinar los momentos críticos en los cuales aplicar un tratamiento adecuado. En este contexto han aparecido diversos modelos que han intentado describir lo más fielmente posible el desarrollo de este hongo, algunos de los cuales se describen en la sección 3.3. El control de este patógeno está directamente relacionado con las condiciones meteorológicas por lo que la aplicación de técnicas de *e-agriculture* parece razonable.

La aplicación de técnicas de inteligencia artificial también tiene su espacio en este campo, aprovechando la gran cantidad de datos proporcionados por los diferentes sensores disponibles. El objetivo de estas técnicas es utilizar estos datos e intentar que el sistema aprenda a hacer una predicción precisa de los momentos más críticos del desarrollo del hongo, de la misma manera que hacían los modelos que antes se han

citado. Para ello se ha recurrido a diversos esquemas supervisados de clasificación, que han sido entrenados y evaluados con conjuntos de datos históricos recogidos por las mismas estaciones meteorológicas que alimentan el sistema.

Por otra parte se ha implementado uno de los modelos propuestos, utilizando los datos meteorológicos disponibles y se ha desarrollado un sistema con una interfaz amigable que responda a las necesidades básicas necesarias para la aplicación del *MIP* en el caso concreto de este hongo.

Durante todo el proceso se ha trabajado con el consejo de expertos de la *Cooperativa Vitivinícola do Ribeiro* que han proporcionado información y bibliografía, además de participar en el diseño de la interfaz. Los datos meteorológicos han sido proporcionados por la empresa *Monet Viticultura*, que ha implementado un API para el acceso privado a estos datos, recogidos por los diferentes sensores colocados en las viñas de la *Cooperativa Vitivinícola do Ribeiro*.

1.1. Objetivos

El objetivo de este proyecto es desarrollar un sistema que ayude en la tarea de control del hongo mildiu, enmarcada en el Manejo Integrado de Plagas y Enfermedades. Para ello se trata de hacer una predicción de los niveles de riesgo de este hongo mediante diversos métodos que serán analizados en este proyecto. Los tres objetivos principales para alcanzar esta meta son los siguientes:

- **Modelos disponibles en la actualidad:** Se estudiarán algunos de los modelos más usados en la actualidad y que ya han demostrado su validez en otras áreas vitivinícolas.
- **Estudio y aplicación de técnicas de Inteligencia Artificial:** Se propondrán y analizarán diferentes técnicas de clasificación con el objetivo de crear una alternativa a los modelos conocidos.
- **Desarrollo de un sistema de apoyo al experto:** Una vez cumplidos los dos objetivos anteriores y tras evaluar cual de los métodos proporciona un resultado óptimo, se desarrollará un sistema que permitirá visualizar, por medio de una interfaz, los datos meteorológicos proporcionados por los sensores, los datos re-

ferentes a los tratamientos aplicados y el resultado del método más eficaz. Estos elementos serán importantes para el experto a la hora de tomar decisiones en cuanto a las fechas de los tratamientos.

1.2. Planificación

Para el correcto desarrollo de un proyecto es necesario un estudio de viabilidad en términos de coste y tiempo. Con el objetivo de conseguir este propósito es necesario definir las diferentes fases del proyecto, así como definir la duración de cada una de ellas, para así poder calcular el coste asociado. Las diferentes fases de las que ha constado el proyecto son las siguientes:

- **Fase 1:** Obtención de la documentación adecuada para realizar un estudio del estado del arte en el campo de trabajo.
- **Fase 2:** Recogida de casuística histórica para la elaboración de la base de datos de predicción.
- **Fase 3:** Entrevista con expertos para la obtención de información sobre modelos.
- **Fase 4:** Implementación del modelo que ha sido seleccionado en base a diferentes criterios.
- **Fase 5:** Evaluación de las técnicas de Inteligencia Artificial propuestas.
- **Fase 6:** Implementación de la interfaz.
- **Fase 7:** Análisis de resultados y conclusiones.
- **Fase 8:** Documentación y elaboración de la memoria del trabajo fin de grado.

La duración de estas fases se describe en el diagrama de Gantt, representado en la figura 1.2. Los costes asociados, descritos en la tabla 1.1, se corresponden tanto a los materiales utilizados como a los recursos humanos. Dentro de los costes de los materiales no se mencionan los relacionados con el acceso a los datos, ya que estos dependen directamente de la *Cooperativa Vitivinícola do Ribeiro* y además son utilizados para otros cometidos.

Tipo de recurso	Nombre	Coste
<i>Software</i>	8.4.0.150421(Matlab 2014b)	2000 €
	Perl	0€
	Python	0€
	Weka	0€
	TexMaker	0€
	MacTex	0€
<i>Hardware</i>	PC	1200€
<i>Recursos humanos</i>	Ingeniero Informático	14720€
<i>Total</i>		17920€

Tabla 1.1: Tabla de costes del proyecto.

1.3. Estructura de la memoria

- **Capítulo 1: Introducción.** Se explican la motivación y objetivos de este proyecto. Además se incluye la organización y planificación del mismo.
- **Capítulo 2: Materiales y métodos.** Se describen las herramientas software empleadas en el desarrollo de este proyecto, así como las técnicas de evaluación y medidas de rendimiento empleadas. Además se hace una breve descripción de los sensores utilizados.
- **Capítulo 3: Mildiu. Modelos actuales.** En este capítulo se describe el hongo del cual tratamos en este proyecto y se describen los principales modelos que ayudan en la lucha contra esta enfermedad.
- **Capítulo 4: Conjuntos de datos, métodos de clasificación y resultados.** Aquí se describen los datos que disponemos para el desarrollo de este proyecto, los diferentes esquemas de aprendizaje máquina utilizados y los resultados de los experimentos hechos con estos y con uno de los modelos propuestos.
- **Capítulo 5: Desarrollo del sistema.** Describe diferentes aspectos del sistema desarrollado.
- **Capítulo 6: Conclusiones y trabajo futuro.** Recoge las conclusiones obtenidas del desarrollo del trabajo y los resultados obtenidos de los experimentos. Finalmente, propone varias vías de continuación del proyecto.

- **Apéndice 1: Código.** Muestra algunos extractos de código interesantes.
- **Apéndice 2: Instalación.** Describe el proceso de instalación y los requisitos necesarios.

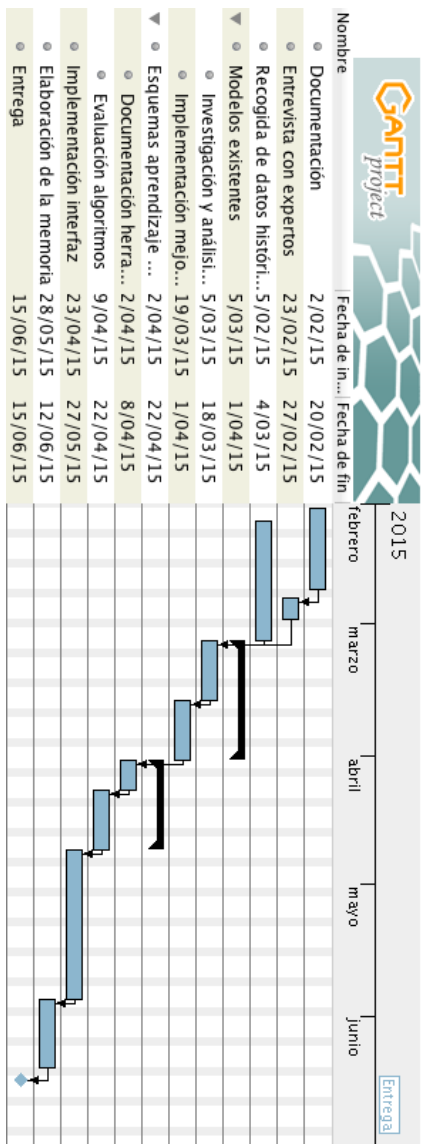


Figura 1.1: Diagrama de Gantt del proyecto.

Materiales y métodos

En este capítulo se describirán las herramientas software que se han utilizado para la experimentación e implementación del proyecto así como las métricas que se han utilizado para la evaluación. Además se describirá el material hardware que proporciona los datos meteorológicos.

2.1. Herramientas software

Para el desarrollo de las pruebas y la implementación de este proyecto se ha recurrido a las siguientes herramientas:

- Matlab, lenguaje que nos facilitará el manejo de datos.
- Weka, herramienta que nos permitirá utilizar una amplia variedad de métodos de clasificación.
- Perl, lenguaje de script que facilitará la tarea de pasar ficheros xls con varios tipos de datos a Matlab.
- Python, para la implementación del cliente rest y preprocesamiento de los datos.

2.1.1. Matlab

Matlab [3] [4] es un lenguaje de alto nivel y un entorno interactivo pensado para el cálculo numérico, visualización y programación, creado en 1984 por *TheMathWorks*. El dato elemental es la matriz por lo que resulta especialmente útil a la hora de manipular datos con estructura de array, proporcionando así una herramienta útil en el ámbito

científico y de la ingeniería. Cuenta además con un extenso número de librerías, o *toolboxes*, que permiten trabajar con facilidad en un amplio rango de aplicaciones, como son el procesamiento de señal, procesamiento de imágenes y video, sistemas de control o biología computacional, entre otras. Otra característica importante es la gran cantidad de usuarios con los que cuenta esta herramienta.

Para este proyecto se ha utilizado la versión 8.4.0.150421 (Matlab 2014b) con el fin de procesar una cantidad importante de datos de manera sencilla, además de brindarnos la posibilidad de manejar ficheros xls.

2.1.2. Weka

Weka (Waikato Environment for Knowledge Analysis) [1] [2] es una colección de algoritmos de aprendizaje máquina y herramientas de preprocesado de datos desarrollados en lenguaje Java y distribuido bajo la licencia GNU (*General Public License*). Resulta ideal para aplicaciones de minería de datos, pues proporciona un gran soporte para todo el proceso de experimentación, incluyendo la preparación de los datos de entrada, la evaluación de los diferentes esquemas de aprendizaje y la visualización de todos los procesos. Los algoritmos y herramientas de los que dispone abarcan una amplia gama de técnicas entre las que se incluyen la clasificación, el clustering o la regresión.

Para este proyecto se ha utilizado la versión de Weka 3.6.12, que es compatible con la de Matlab 2014b, con la intención de evaluar diversos algoritmos de clasificación sobre datos históricos.

2.1.3. Python

Python [5] es un lenguaje interpretado multiparadigma creado por Guido Van Rossum a finales de la década de los ochenta. Este lenguaje ha ido ganando popularidad por su sintaxis clara y legible, con lo que hoy en día cuenta con una amplia comunidad de desarrolladores que generan multitud de librerías.

Se ha optado por usar Python en este proyecto, en su versión 2.7, para la descarga y preprocesado de los datos, utilizando además diversas librerías especializadas en el

servicio RESTful y el tratamiento de datos en formato JSON. Dicho formato se usa con el objetivo de automatizar la descarga de datos, que en el caso del formato xls es necesaria la interacción con el usuario.

2.1.4. Perl

Perl [6] es un lenguaje interpretado creado por Larry Wall a finales de los ochenta. Las características de este lenguaje son similares a las de Python, aunque no cuenta con la misma popularidad. El propósito de usar este lenguaje en este proyecto es facilitar el paso de ficheros xls que contengan diversos tipos de datos a Matlab, beneficiándose del hecho de que esta herramienta cuenta con soporte para la ejecución de ficheros en lenguaje Perl.

2.2. Técnicas de evaluación

La evaluación es la clave a la hora de realizar un análisis en minería de datos. El objetivo principal de esta es determinar la calidad de los resultados y poder comparar entre diversos esquemas de aprendizaje. Para ello se hace hincapié en dos problemas fundamentales. Uno de ellos hace referencia a la manera de escoger los conjuntos de test y de entrenamiento, y el otro a la manera en que se mide la idoneidad de cada uno de los esquemas.

En cuanto al primer aspecto, al poseer una componente temporal y teniendo en cuenta que para el proyecto se disponía de tres años de datos históricos, se ha optado por utilizar *validación simple*, tomando dos años como conjunto de entrenamiento y un año para el conjunto de test.

A la hora de determinar la calidad de los resultados hay numerosas métricas y, puesto que los esquemas que se van a utilizar son supervisados, partiremos de los elementos que conforman una matriz de confusión:

- **VP** (número de verdaderos positivos): muestras que pertenecen a la clase positiva y fueron clasificadas correctamente dentro de esa clase.

- **FP** (número de falsos positivos): muestras que no pertenecen a la clase positiva y fueron clasificadas dentro de ella.
- **VN** (número de falsos negativos): muestras que pertenecen a la clase negativa y fueron clasificadas correctamente dentro de esa clase.
- **FN** (número de verdaderos negativos): muestras que no pertenecen la clase negativa y fueron clasificadas dentro de ella.

A partir de estos datos disponemos de tres métricas que los relacionan y las cuales nos dan una apreciación de lo adecuado que es el algoritmo evaluado. A continuación se explica cada una de estas métricas:

- **Sensibilidad**

Nos indica la proporción de casos con valor real positivo que han sido clasificados correctamente.

$$S = \frac{VP}{VP + FN}$$

- **Especificidad**

Nos indica la proporción de casos con valor real negativo que han sido correctamente clasificados.

$$E = \frac{VN}{VN + FP}$$

- **Precisión**

Nos indica el ratio de de muestras correctamente clasificadas.

$$P = \frac{VN + VP}{VN + FP + VP + FP}$$

2.3. Herramientas de campo

En este apartado describiremos brevemente los sensores que se han utilizado para la recogida de los datos y que se ha llevado a cabo en las fincas de la *Cooperativa Vitivinícola do Ribeiro*.

■ **Sensor de temperatura y humedad modelo Honeywell**

Este sensor digital, que se muestra en la figura 2.1, se encarga de medir la temperatura y humedad ambiente de forma totalmente autónoma. Los rangos de medida son los siguientes:

- Temperatura: desde -25°C hasta $+85^{\circ}\text{C}$ con una precisión de $0,5^{\circ}\text{C}$.
- Humedad relativa: desde 0 a 100 % con una precisión de 4 %



Figura 2.1: Sensor de temperatura y humedad ambiente.

■ **Sensor de humedad de suelo modelo *Watermark/Irrrometer 200ss***

Este sensor, que podemos ver en la figura 2.2, está formado por dos electrodos recubiertos de un material especial, todo ello envuelto por una cobertura de acero inoxidable. El material especial que recubre los electrodos es el encargado de reflejar los cambios en la humedad del suelo a los electrodos. De esta forma, la humedad del suelo se ve reflejada en los datos del sensor tanto cuando éste se seca como cuando se humedece. Así, cuando el suelo se seca, la humedad en el material del sensor disminuye, y por lo tanto la resistencia eléctrica entre los electrodos se incrementa.

Este sensor de Watermark proporciona medidas precisas entre 0 y 200 centibares (cbar).



Figura 2.2: Sensor de humedad de suelo.

- **Sensor de temperatura de suelo modelo *NTC Termistor 10k***

Este sensor, que podemos ver en la figura 2.3, está basado en una resistencia eléctrica variable en función de la temperatura que la rodea. Proporciona una lectura de 10Kohm a 25°C, y tiene una constante térmica inferior a los 15 segundos. Su rango de medida va desde los -55°C hasta los 125°C.



Figura 2.3: Sensor de temperatura de suelo.

- **Sensor de humedad foliar modelo *Monet***

Este sensor, que se muestra en la figura 2.4, está basado en una resistencia eléctrica

variable en función de la humedad que exista sobre cualquiera de sus caras.



Figura 2.4: Sensor de humedad foliar.

- **Sensor de velocidad y dirección de viento modelo *Davis 7911***

Este anemómetro y veleta de Davis, que podemos ver en la figura 2.5, permite medir velocidades de viento entre 1 y 322 km/h, a la vez que nos permite determinar de qué dirección proviene ese viento con valores entre 0 y 359 grados. La precisión de estas medidas es de 3 km/h para la velocidad y de 7° en la dirección del viento. La precisión de estas medidas es de 3 km/h para la velocidad y de 7° en la dirección del viento.



Figura 2.5: Sensor de velocidad y dirección del viento.

- **Sensor de radiación solar modelo *Davis 7911***

Este sensor de radiación solar (mostrado en la figura 2.6) permite medir radiación global; es decir, la suma, en el punto de medida de la radiación solar directa y difusa. La respuesta espectral del sensor abarca desde los 400 a los 1100 nanómetros (nm). Las lecturas que se proporcionan en el servidor van desde los 0 a los $1800W/m^2$.



Figura 2.6: Sensor de radiación solar.

- **Sensor de precipitaciones. Pluviómetro modelo *Davis 7857M***

Este pluviómetro de cazoletas de Davis, que se muestra en la figura 2.7, permite medir las precipitaciones con una precisión de 0.2 mm, lo cual quiere decir que precipitaciones inferiores a los 0.2 l/m^2 no quedarán registradas, o se registrarán cuando la precipitación acumulada en la cazoleta alcance esos 0.2 mm.



Figura 2.7: Pluviómetro.

Mildiu. Modelos actuales

3.1. Historia del patógeno

El mildiu de la vid [15] [16] es una enfermedad originaria de América que fue observada en Francia por Planchon en 1878. Está producida por un hongo, *Plasmopara viticola*, que se desarrolla sobre todos los órganos verdes: pámpanos, hojas, racimos y zarcillos. Al principio fue la causa de la pérdida catastrófica de cosechas reduciendo la producción de vino hasta una cuarta parte del volumen medio de la época, pero poco a poco se ha ido controlando gracias a los diversos productos anti-mildiu que han ido surgiendo.

3.2. Biología del Mildiu

El mildiu es un hongo microscópico que pasa el invierno en el suelo, en el interior de las hojas muertas, bajo forma de *huevos de invierno* muy resistentes.

En cuanto la temperatura sobrepasa los 11 °C y después de una lluvia, los huevos emiten un filamento al final del cual se forma una *macroconidia*¹ que revienta en presencia de agua y libera pequeños gérmenes llamados *zoosporas*. Estas *zoosporas* son proyectadas hacia los órganos herbáceos por las salpicaduras que producen con las lluvias. En presencia de agua y a una temperatura superior a 11 °C, cada *zoospora* emite un filamento o *micelio* que penetra por los estomas² de las hojas: es la *contaminación* que determina un foco *primario*. Después el micelio se desarrolla en el interior de la hoja: es la *incubación*, que dura como mínimo diez días para las contaminaciones a partir

¹Se refiere a la estructura microscópica del hongo que se encarga de la reproducción asexual, que en este caso tiene un tamaño relevante dando lugar a la expulsión de miles de esporas llamadas conidios

²Se denomina estomas a ciertos pequeños orificios o poros que atraviesan la epidermis de las plantas de forma que comunican el ambiente gaseoso del interior de la planta con el del exterior

de los huevos de invierno, o siete días para los que proceden de los gérmenes de verano (conidias). Después de este período aparecen las *manchas de aceite* en la cara superior (haz) y el polvo blanco en la cara inferior (envés) de las hojas: es la *invasión*. El polvo blanco está formado por filamentos emitidos por el micelio y llamados *conidióforos*; estos llevan los gérmenes de verano o *conidias*. Estas conidias se desprenden y, transportadas por el viento o la lluvia, contaminan otros órganos herbáceos. En presencia de agua, revientan y liberan zoosporas. Las zoosporas germinan y penetran en los tejidos, lo que se llama *contaminación*; posteriormente el micelio se desarrolla en los tejidos y dura siete días por lo menos, lo que da lugar al periodo de *incubación*. Finalmente los conidióforos salen por los extremos llevando conidias dando lugar a la *invasión*.

Durante el verano se producen contaminaciones sucesivas que determinan otros focos secundarios. En otoño ya no se produce emisión de polvo blanco; el huevo de invierno se forma en el interior de la hoja después de la fecundación del oogonio por el anterido, órgano reproductor diferenciado en el seno del micelio. El huevo es muy resistente a la intemperie y permanece en los residuos de las hojas después de la caída. Todo el ciclo descrito se puede observar en la figura 3.1

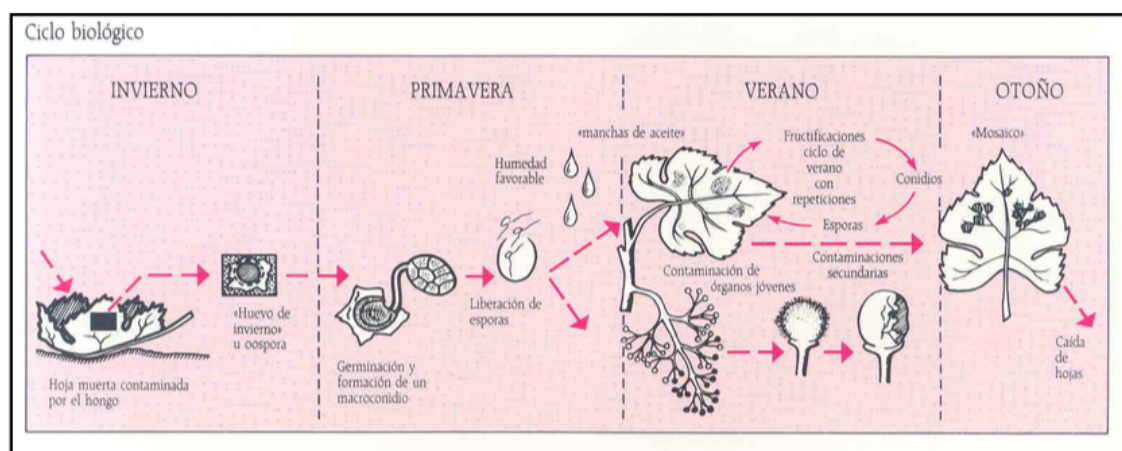


Figura 3.1: Ciclo de vida del mildiu.

3.3. Estrategias de protección

La utilización de variedades resistentes sería una solución pero dado que todas la variedades de *V. vinífera* son sensibles a este hongo, esta medida no es viable, aunque

se está trabajando en la modificación genética de las plantas de la vid para que sean resistentes a la enfermedad. Las estrategias viables se dividen en dos grupos:

■ **Medidas profilácticas de carácter ecológico**

En este caso no se utilizan productos químicos y se realizan en dos fases:

- Reducir las condiciones favorables para la formación de focos primarios mediante la supresión de encharcamientos, eliminación de la vegetación baja, destruir las plantas jóvenes nacidas en viñas de no cultivo o recoger las aportaciones de orujo sin fermentar, entre otras.
- Destruir los focos primarios por la recogida de hojas enfermas y despuntar después de la floración las hojas jóvenes receptivas.

■ **Medidas profilácticas de carácter químico**

A pesar de que la lucha puramente ecológica sería deseable, en la práctica resulta inviable, tanto en términos económicos, como por la eficiencia en la protección contra el hongo. Por lo tanto la lucha química es indispensable y resulta especialmente preventiva depositando el fungicida sobre los órganos sanos de la vid o, dependiendo del tipo de producto, conseguir una acción curativa. Los diferentes productos químicos se dividen en tres grandes categorías:

- *Fungicidas de contacto*

Este tipo se distribuye solamente por la superficie de la planta, lo que impide penetrar a los parásitos al interior de los órganos de la vid, inhibiendo el proceso de respiración del hongo al principio de su desarrollo. Este tipo de fungicidas son puramente preventivos y son muy poco resistentes al *lavado* a causa de precipitaciones. A pesar de esto son muy utilizados en la práctica, tanto solos como asociados con otros tipos. Los principales componentes de estos fungicidas son el cobre, los ditiocarbamatos o la dicarboximida, entre otros.

- *Fungicidas penetrantes*

Estos fungicidas llegan a penetrar en los órganos de la vid, con lo que la resistencia al lavado es mayor que en el caso de los anteriores. A pesar de poder penetrar en la planta, los órganos formados después del tratamiento no están protegidos. Los principales componentes de estos fungicidas son cimoxanilo, dimetomorf y azoxistrobin, entre otros.

- *Fungicidas sistémicos*

Este tipo penetran en los tejidos y son vehiculados por la savia, lo que

permite proteger los órganos jóvenes formados después de un tratamiento y por lo tanto aumentar la duración de la acción del producto. A pesar de esto, este tipo de funguicidas nunca se utilizan solos, sino que son mezclados con productos de contacto o penetrante mejorando así su acción. Los principales componentes de estos funguicidas son los basados en fosetil de aluminio y los basados en anilidas o fenilamidas.

Estas medidas, en la práctica, suelen ir de la mano, reduciendo así el riesgo de una infección que podría ser desastrosa en caso de no detectarse a tiempo. A la hora de llevar a cabo las medidas de carácter químico, el viticultor debe tomar la decisión de la fecha en la cual se va a dar el tratamiento y el producto a utilizar, para así conseguir la máxima seguridad y el menor número de tratamientos. Históricamente esta toma de decisiones se resumía en el tratamiento indiscriminado y sistemático con productos en periodos regulares de tiempo, o atendiendo a factores meteorológicos como la lluvia, que provocaba el *lavado* de estos productos y que en su mayoría pertenecían al grupo de funguicidas de contacto. Poco a poco, con la utilización de diversos aparatos de medición precisos, como higrómetros o pluviómetros entre otros, fueron apareciendo los primeros modelos que hacían una predicción aproximada del crecimiento hongo. Con la llegada de las nuevas tecnologías, que han permitido la instalación de pequeñas estaciones meteorológicas con numerosos sensores a bajo coste, se ha conseguido investigar más en la relación entre las condiciones climáticas y el hongo, desarrollando más modelos. A pesar de ello, cada nuevo modelo atiende a condiciones características de la zona donde ha sido creado y validado. A continuación se describirán algunos de los modelos que se utilizan hoy en día.

3.3.1. EPI

El modelo EPI mildiu (l'Etat Potentiel d'Infection) [21] [18] fue concebido por Serge Strizyk en Francia en 1983, aunque se han ido incorporando diversas modificaciones a lo largo de los años. Este sistema se basa en la idea de que, estudiando las condiciones que favorecen la aparición de la enfermedad a lo largo de varios años, es posible predecir cuando se va a dar un foco en los años siguientes, mejorando a medida que poseamos más datos históricos. Para ello se requiere una serie climática de entre veinte y treinta años que aporte la pluviometría media mensual, temperatura media mensual y el número de días de lluvia de cada mes.

El modelo se divide en dos fases, la invernal que representa la reproducción sexual del hongo y permite prever la agresividad potencial de este al inicio de la fase vegetativa de la vid, y la fase estival que representa la reproducción asexual del hongo que calcula el riesgo potencial de una infección secundaria. La primera fase se inicia en Octubre y termina al final de Marzo, es decir cuando la temperatura media ronda los 10 grados, mientras que la segunda fase comprende los otros seis meses. Para el cálculo del riesgo en cada una de las fases se utilizan las siguientes fórmulas.

■ *Fase invernal*

$$E_p = \left(2^{ct} \left(\sqrt{H} - \sqrt{\frac{HM \times 95}{100}} \right) \right) + \left(0.2 \left(\sqrt{H} \times \sqrt{t} - \left(\sqrt{\frac{HM \times 95}{100}} \times \sqrt{\theta c} \right) \right) \right) - \left(\left(\sqrt{\frac{NGP \times 1.5}{18}} \right) \times \log \frac{h}{ngp} \right)$$

ct : 1.2 en Octubre y Noviembre, 1 en Diciembre y 0.8 en Enero, Febrero y Marzo

HM : Pluviometría de la progresión histórica en cada mes [mm]

θc : Temperatura media de la progresión histórica en cada mes [°C]

NGP : Número de días de lluvia de la progresión histórica en cada mes

t : Temperatura media del mes [°C]

H : Pluviometría de cada mes [mm]

ngp : Número de días de lluvia en la década

■ *Fase estival*

$$E_c = 0.012 \times \left(\frac{\left(\frac{5Ucn+3Ud}{8} \right)^2 \times \sqrt{T}}{100} - \frac{Ud^2 \times \sqrt{\theta c}}{100} \right)$$

Ucn : Humedad relativa e la progresión histórica en cada mes [%]

θc : Temperatura media de la progresión histórica en cada mes [°C]

Ud : Humedad relativa diurna entre las 10:00 horas y las 18:00 horas [%]

T : Temperatura media diaria [°C]

Con el cálculo de estos dos riesgos podemos obtener el valor EPI que pronosticará el nivel de riesgo de infección con el valor del EPI resultante:

$$EPI = \sum_{\text{Octubre}}^{\text{Marzo}} Ec + \sum_{\text{Abril}}^{\text{Septiembre}} Ep$$

A partir de Abril el area de riesgo se sitúa entre $-10 < EPI < 0$.

Al poseer dos fases es posible determinar la agresividad del hongo al principio de la fase vegetativa de la planta. Para ello se debe utilizar solamente el valor Ep .

A pesar de tratar de modelar la adaptación ecológica de la planta, este esquema resulta débil frente a variaciones climáticas. Además, se enfrenta a la dificultad de tratar de modelar las infecciones primarias, las cuales implican problemas complejos, como la salpicadura de las gotas de lluvia de una planta a otra.

3.3.2. Modelo de Goidanich

Este modelo fue diseñado por Gabriele Goidanich en 1959 [14] y todavía se utiliza en muchas Estaciones de Avisos Agrícolas. Este modelo se basa en el cálculo del desarrollo del hongo una vez constatada la primera infección procedente de las oosporas invernales. Para la puesta en marcha de este modelo es necesario tener las siguientes consideraciones:

- Disponer de una red de estaciones meteorológicas extensa que permita recoger las diferentes subzonas climáticas dentro de cada comarca.
- Detección temprana de las primeras manchas de mildiu en las fincas, es decir, detectar las infecciones primarias, o bien mediante observación, o bien mediante un modelo como EPI o DMCAst. Un método tradicional de aproximar estas infecciones es la regla de los tres diez, que pronostica que las infecciones primarias se darán una vez la temperatura media haya superado los 10 °C ³, la brotación supere los 10 cm y en un día haya al menos 10 mm de lluvia. Esta regla, aunque simple, nos proporciona una idea aproximada de cuándo empezar a buscar esas infecciones primarias.

³En la práctica esta temperatura suele ser un poco superior a diez grados

- Los avisos que produzca este modelo deben de ser preventivos, indicando que hay riesgo con un margen suficiente antes de llegar al 100 % de desarrollo.

Este modelo tiene en cuenta principalmente tres parámetros: la humedad relativa media, la temperatura media y la lluvia acumulada diariamente. A partir de los dos primeros se puede calcular el porcentaje diario de desarrollo del hongo que se presenta en la tabla 3.2, mientras que el último nos da la posibilidad de saber si un funguicida ha sido *lavado*, en caso de haber utilizado un producto anti-mildiu. El modelo funciona de manera que si la suma acumulada del desarrollo diario supera el 70 %, el riesgo es alto, por lo que hay la necesidad de dar un tratamiento lo antes posible. En caso de aplicar un producto funguicida, el desarrollo del hongo se reinicia y queda en suspensión un determinado número de días, o hasta que el producto haya sido lavado por la lluvia. Este periodo depende del tipo de tratamiento e incluso depende del componente activo. En este proyecto únicamente se diferencian los periodos según el tipo de tratamiento, proporcionando el experto una aproximación basada en los productos más utilizados de cada tipo. Estos periodos se distribuyen de la siguiente manera:

- *Productos de contacto*

Su efectividad dura en torno a los seis días. Si las precipitaciones acumuladas durante un máximo de tres días son superiores a $10l/m^2$ el producto es lavado y pierde toda su efectividad.

- *Productos penetrantes*

Su efectividad dura un máximo de diez días. Si hay precipitaciones antes de pasar una hora después del tratamiento este pierde la efectividad. En otro caso, el producto mantiene toda su efectividad mientras las precipitaciones acumuladas en un máximo de tres días no superen los $40l/m^2$. En caso de que este umbral se supere, la efectividad se perderá de manera ponderada hasta llegar a los $100l/m^2$ pudiendo reducirse hasta un 20 %, lo que se traduce en que el periodo máximo de efectividad dure sólo ocho días. En caso de superar los $100l/m^2$ (situación improbable), se perderá sistemáticamente la efectividad del producto y volverá a empezar el ciclo de desarrollo, en caso de no haber perdido la cosecha.

- *Productos sistémicos*

El comportamiento de estos productos en cuanto al periodo de acción es similar al de los penetrantes, con la única diferencia de que el periodo máximo pasa a ser de catorce días y por lo tanto el mínimo será de diez.

T media (°C)		% diario desarrollo Hm<75%	% diario desarrollo Hm>75%
12		0,00	5,25
	25	4,40	5,75
	50	4,70	6,20
	75	5	6,70
13		5,30	7,10
	25	5,70	7,70
	50	6	8
	75	6,30	8,50
14		6,6	9
	25	6,8	9,40
	50	7,10	9,70
	75	7,30	10,20
15		7,60	10,60
	25	7,80	10,80
	50	8,10	11,10
	75	8,30	11,30
16		8,50	11,70
	25	9	12
	50	9,30	12,50
	75	9,60	12,90
17		10	12,25
	25	10,30	16,60
	50	10,50	14,30
	75	10,75	14,75
18		11,10	15,30
	25	11,48	15,20
	50	11,70	16
	75	12,10	16,30

T media (°C)		% diario desarrollo Hm<75%	% diario desarrollo Hm>75%
19		12,50	16,60
	25	12,90	17,50
	50	13,40	18,30
	75	13,70	19,30
20		14,20	20
	25	14,50	20,50
	50	14,80	21
	75	15	21,50
21		15,30	22,20
	25	15,70	22,20
	50	16	22,20
	75	16,30	22,20
22		16,60	22,20
	25	17	22,60
	50	17,30	23,50
	75	17,70	24,40
23		18,10	25
	25	18,10	25
	50	18,10	25
	75	18,10	25
24		18,10	25
	25	17,70	24,30
	50	17,30	23,50
	75	16,60	23,20
25			22,20

Nota.- A partir de 25°C el desarrollo es constante

Figura 3.2: Relación entre la humedad, la temperatura y el porcentaje de desarrollo del hongo.

Este modelo ha sido, y sigue siendo, el más utilizado por su flexibilidad a la hora de plantear riesgos. Al ser un esquema basado el crecimiento del hongo en condiciones controladas la problemática que hay para adaptar el modelo a cada zona desaparece parcialmente. En todo caso la flexibilidad reside en la particularidad de poder marcar el umbral de riesgo en el porcentaje de crecimiento que creamos que se adapta mejor. En el caso de este proyecto se ha decidido que el umbral de riesgo debe estar entorno al 70 % basándonos en observaciones sobre los datos históricos y en el conocimiento aportado por los expertos.

3.3.3. PMI

El modelo PMI (Power Model Index) [19] fue ideado en la Universidad de California a finales de los años noventa. El objetivo de este modelo es evaluar el riesgo de desarrollo de la enfermedad, relacionándolo con condiciones climáticas, prever la fortaleza del ataque y así ayudar a tomar la decisión que debe tomar el agricultor para determinar cuándo tratar. Al igual que el modelo anterior es necesario determinar previamente las infecciones primarias y, una vez determinado el momento en que estas comienzan, recoger datos durante los 7-10 días siguientes. Las consideraciones a tener en cuenta a la hora de plantear este modelo son las siguientes:

- Las temperaturas óptimas para el crecimiento del hongo son de 21°C a 29°C
- Si la temperatura es superior a 35°C durante 12 horas seguidas se paraliza el crecimiento

Este método se divide en dos fases, una fase de inicialización del índice con la que se determina el riesgo de desarrollo del hongo, y una segunda fase llamada calendario de tratamiento donde se mide la fortaleza de este hongo. La mecánica de ambas fases se describe a continuación:

- *Fase de inicialización del índice*
 - Añadir **20** puntos cada día que tenga más de **6 horas** con temperatura entre **21 y 29°C**
 - Si un día tiene menos de **6 horas** con temperatura entre **21 y 29°C** el PMI se reinicia a **0**.
 - Una vez que el **PMI=60** se procede con la fase dos, el calendario de tratamiento.
- *Calendario de tratamiento*

En este caso se evalúa la temperatura de cada día y se ajusta el índice del día anterior. El valor de PMI está acotado entre 0 y 100, y no se pueden añadir más de 20 puntos por día ni restar menos de 10. El método es el siguiente:

- Si hay menos de **6 horas** con temperatura entre **21 y 29°C** entonces decrementar el índice **10** puntos.

- Si hay más de **6 horas** con temperatura entre **21 y 29°C** entonces aumentar el índice **20** puntos.
- Si la temperatura es superior a **35°C** durante más de **15 min** entonces decrementar el índice **10** puntos.
- En el caso de que haya más de **6 horas** con temperatura entre **21 y 29°C** y la temperatura sea superior a **35°C** durante más de **15 min** entonces se procederá a incrementar el índice **10** puntos.

Teniendo en cuenta los resultados diarios se puede establecer la fortaleza del patógeno, el estado de este y las pautas de tratamiento⁴. La siguiente tabla resume las correspondencias entre el índice y estos parámetros.

PMI	FORTALEZA	ESTADO DEL PATÓGENO
0-30	BAJO	PRESENTE
30-50	MEDIO	SE REPRODUCE CADA 15 DÍAS
> 50	ALTO	SE REPRODUCE CADA 5 DÍAS

Este modelo ha sido validado en diversos estudios en California. A pesar de ello es sensible a las condiciones locales, ya que no tiene en cuenta las condiciones de humedad, que son claves a la hora de determinar la rapidez con la que se desarrolla el hongo.

3.3.4. Otros modelos

■ *Milvit*

Este modelo [18] fue desarrollado por el Servicio Francés de Protección de Plantas y predice el desarrollo de la enfermedad simulando diferentes estados de su ciclo de vida. Para hacer esta simulación se usan como datos de entrada la humedad relativa y la temperatura media, y el riesgo resultante se mide según el número de conidias en la simulación. Igual que el modelo de Goidanich es necesario establecer en primer lugar el momento de las infecciones primarias. El resultado que produce este modelo no indica el momento exacto del tratamiento sino que se debe de establecer según las condiciones del lugar.

⁴Estas pautas serán variables en función del funguicida utilizado

- *Plasmo*

Modelo desarrollado por Rosa y Orlandini en 1997 [18], que se caracteriza por combinar un modelo epidemiológico con uno de crecimiento de la planta. El modelo epidemiológico usa como parámetros de entrada la humedad relativa media, la temperatura media y la humedad foliar mientras que el de crecimiento de la planta⁵ sólo usa la temperatura media. Este hecho hace referencia a estados de la planta en los que es más sensible a un ataque, en concreto el periodo de floración es el más crítico. Para que funcione correctamente es necesario que diversos parámetros sean calibrados la primera vez que es puesto en marcha en una viña. Al igual que muchos de estos modelos, no propone una pauta a la hora de dar el tratamiento.

Al llegar a este punto tenemos una perspectiva suficiente a la hora de elegir un método que ayude al viticultor a la hora de decidir las fechas de los tratamientos. A pesar de haber más modelos que los seleccionados⁶ se han presentado los más relevantes en cuanto a uso y eficacia, exponiendo sus ventajas e inconvenientes. En este proyecto se ha optado por implementar el modelo de Goidanich. Esta decisión se ha tomado atendiendo a varias razones:

- Los datos históricos disponibles en el entorno real en el que se va a aplicar son sólo de tres años, por lo que modelos como EPI, que requieren una gran cantidad de años de datos recogidos, son utilizables.
- El modelo de Goidanich es el más flexible de los propuestos.
- Los tratamientos que reflejan los datos históricos están orientados por una ejecución manual del modelo Goidanich.

En el siguiente capítulo veremos la otra perspectiva desde la que se ha tratado el problema, aplicando esquemas de aprendizaje máquina para poder determinar mediante clasificadores las fechas de los tratamientos.

⁵En realidad es un modelo del ciclo de vida de la planta, es decir, predice desde el momento del brote hasta la caída de la hoja basándose en la temperatura media

⁶PRO, DMCast o UCSC entre otros

Conjuntos de datos, métodos de clasificación y resultados

La agricultura es una actividad estrechamente relacionada con el clima y este juega un papel muy importante en cuanto a la adaptación de las plantas al entorno. Conocer la información meteorológica actualizada es fundamental para la toma de decisiones en la agricultura moderna, ya que cada vez se requiere más productividad a la vez que más calidad en los productos. Para ello se está haciendo un gran esfuerzo durante los últimos años para poner en marcha redes de estaciones meteorológicas distribuidas estratégicamente para la obtención de datos fiables y de calidad. Este cometido se hace cada vez más sencillo gracias a la reducción de los costes de estas estaciones. Estas estaciones se componen de diversos sensores que responden a estímulos electrónicos y que tienen la capacidad de registrar y recolectar información meteorológica en forma automática y en tiempo real. Gracias a esto podemos monitorizar numerosos parámetros importantes a la hora de hacer un seguimiento de enfermedades y del estado de la planta.

4.1. Datos meteorológicos

Son numerosos los datos meteorológicos que se pueden extraer en una estación meteorológica, pero para este proyecto no son muchos los relevantes, por ello describiremos en profundidad los más relevantes y se discutirá su posible repercusión en el desarrollo del mildiu.

- **Temperatura ambiental**

La temperatura es la medida del contenido de calor de un cuerpo o del medio ambiente. El calor equivale a la energía calorífica que contiene. Se mide con el sensor

denominado *termómetro* y se expresa en grados centígrados ($^{\circ}\text{C}$). Este parámetro es fundamental tanto en el desarrollo de la planta, como en el desarrollo de enfermedades. Un aumento de temperaturas a finales del invierno puede provocar un despertar prematuro de la planta, que influye en todo su ciclo de desarrollo y en la incidencia del hongo. Como se ha podido apreciar en el capítulo anterior, todos los modelos tienen en cuenta este parámetro a la hora de hacer sus predicciones y simulaciones, por lo que resultará muy importante.

- **Humedad ambiental**

Este parámetro hace referencia a la cantidad de vapor de agua presente en el aire. Se puede expresar de forma absoluta o de forma relativa denominándose humedad relativa o grado de humedad. Normalmente se emplea la segunda y se expresa como porcentaje (%) respecto a un ambiente saturado. Se mide normalmente con un sensor llamado *psicrómetro*. Este es otro de los parámetros fundamentales a la hora de hacer un seguimiento del mildiu. Entornos húmedos implican un crecimiento más acelerado del hongo.

- **Precipitación pluvial**

La precipitación es la caída directa de gotas de agua o de cristales de hielo sobre la superficie terrestre. La cantidad de precipitación se mide en milímetros (mm) con un sensor llamado *pluviómetro*. Este parámetro es muy importante para el desarrollo de la planta (y de todas las plantas), por lo que incide directamente en su ciclo de desarrollo. Además es especialmente importante tenerlo controlado a la hora de monitorizar la aplicación de tratamientos, ya que este fenómeno meteorológico puede provocar el *lavado* del producto con el que se ha tratado.

- **Velocidad y dirección del viento**

El viento se describe por dos características: la velocidad y la dirección. El instrumento que se usa para medir la velocidad del viento se llama *anemómetro* y se mide en m/s. La dirección del viento hace referencia al punto del horizonte de donde viene o sopla. Para distinguir la dirección se le aplica los grados de acimut con respecto al Norte Magnético. El instrumento que se usa para medir la dirección es la *veleta* y se expresa en grados de acimut. Este parámetro es importante a la hora de monitorizar infecciones secundarias ya que nos puede dar información de la propagación de la enfermedad. El viento arrastra las conidias, que se depositan en otras plantas y pueden llegar a liberar zoosporas que infecten a la planta.

El resto de parámetros no son tan importantes para la monitorización de este hongo, pero pueden serlo para la monitorización del estado de la planta, lo que incide en el desarrollo del mildiu, así como en el desarrollo de otras enfermedades. Por ello en este proyecto se han utilizado tanto para el ámbito de la clasificación como para la monitorización. Los enumeramos a continuación:

- Temperatura suelo [$^{\circ}\text{C}$]
- Presión barométrica [mbar]
- Humedad foliar [counts]
- Radiación solar [W/m^2]
- Punto de rocío [$^{\circ}\text{C}$]
- Humedad en suelo a 30 cm [cbar]
- Contenido de agua [%]

4.1.1. Datos históricos

Para el entrenamiento y validación se han utilizados datos meteorológicos históricos correspondientes a tres años, en los periodos de Primavera-Verano. Estos datos han sido extraídos de diversos ficheros xls, donde se encontraban en formato minutal, es decir, aproximadamente una medición por minuto. Posteriormente se han preprocesado para obtener la media y desviación típica de cada día. La media nos proporciona parámetros que son clave para los modelos anteriormente expuestos, como son la temperatura media diaria o la humedad relativa diaria, en cambio la desviación típica nos proporciona una medida de dispersión que nos da una idea de la variabilidad de estos datos, proporcionándonos una medida de la regularidad climática a lo largo de un día, la cual en ciertos casos puede llegar a ser relevante en el desarrollo de la enfermedad.

Con el fin de añadir una componente temporal se procedió a calcular las diferencias con los siete días previos a cada medición. La decisión de utilizar siete días se fundamentó en resultados experimentales, con pruebas que abarcaron desde un día previo hasta catorce, coincidiendo así con el plazo más largo de efectividad de un funguicida. Esta medida coincide aproximadamente con la media de desarrollo del hongo después de darse las condiciones favorables, que suele durar entre 5 y 10 días.

Por último se ha añadido el porcentaje de desarrollo que nos proporciona el modelo teórico de Goidanich y además se ha añadido un parámetro binario que indica los días que se ha dado un tratamiento con el objetivo de tener dos clases diferenciadas y poder aplicar un esquema de aprendizaje supervisado. Los datos de tratamiento se han sacado del cuaderno de campo proporcionado por un experto.

Antes de proceder con la evaluación de los algoritmos de clasificación se hizo frente a un problema intrínseco de los datos, el desbalanceo de clases. Este desbalanceo se produce por el hecho de que a lo largo del año se aplican una media de ocho tratamientos, frente al resto de días en que no se da dicho tratamiento. Este problema ha de tratarse ya que nos interesa que nuestros algoritmos de aprendizaje clasifiquen correctamente la clase minoritaria, y un desbalanceo generalmente provoca que el clasificador favorezca a la clase mayoritaria. Para solucionar este problema se ha recurrido a una técnica de preprocesamiento bien conocida, el *oversampling*. Esta técnica consiste en la creación de un superconjunto del conjunto de datos original mediante la realización de algunas muestras, o la creación de muestras nuevas a partir de las originales. En este proyecto se utiliza SMOTE (*Synthetic Minority Over-sampling TEchnique*) [13], que viene incorporado en la herramienta Weka. Esta técnica de *oversampling* se centra en crear muestras sintéticas a partir de las ya existentes. Para ello introduce muestras nuevas en las líneas que unen las que se corresponden a clases minoritarias en el espacio de características.

4.2. Métodos de clasificación

En esta sección se describirán los principales algoritmos de aprendizaje máquina empleados en este proyecto: C4.5, Naïve Bayes, IB1, SVM y Regresión logística. Además se expondrán los resultados obtenidos sobre el conjunto de datos históricos. Todos estos esquemas de aprendizaje son ampliamente utilizados en gran variedad de ámbitos que requieran resolver un problema de clasificación. En el caso de este proyecto el problema de clasificación involucra dos clases, una que hace referencia a días en los que no se ha aplicado un tratamiento, y otra para cuando sí se ha aplicado. Sin embargo, estos algoritmos son extensibles a problemas que no sean de carácter binario, es decir, con más de dos clases. El objetivo de la clasificación es obtener el menor error posible a la hora de clasificar una muestra. A continuación se describen cada uno de los algoritmos

utilizados en este proyecto y sus resultados.

4.2.1. C4.5

El algoritmo C4.5 [7] fue desarrollado por Quinlan en 1993 y forma parte de una familia de algoritmos basados en árboles y reglas de decisión. Este algoritmo añade numerosas mejoras a uno previo, ID3 (Iterative Dichotomiser 3). El árbol de decisión permite representar los procesos de decisión involucrados en las tareas de clasificación de manera que una muestra es filtrada de forma descendente por sus características hasta obtener un valor de clase en uno de los nodos hoja.

El objetivo del algoritmo consiste en generar el árbol de decisión óptimo que proporcione una clasificación con el menor error posible. Para ello C4.5 ejecuta un proceso repetitivo en el que selecciona una característica y posteriormente se reparten las muestras de entrenamiento en las ramas, dependiendo del valor correspondiente a la característica que discrimina el nodo padre. La elección de la característica que ocupa cada nodo se realiza utilizando el ratio de ganancia (GR), que a diferencia de ID3, que utiliza la ganancia de información, penaliza los atributos con muchos valores y muchos uniformemente distribuidos:

$$GR = \frac{G(A_i)}{H(A_i)}$$

donde A_i denota una característica, $G(A_i)$ la ganancia de esa característica respecto al conjunto de patrones y $H(A_i)$ la entropía respecto a los valores de la característica A_i , que se formula de la siguiente manera:

$$H(A_i) = - \sum_j \frac{n_{ij}}{n} \times \log \frac{n_{ij}}{n}$$

donde n_{ij} número de muestras que en la característica A_i toman el valor $j \in V_i$, siendo V_i el conjunto de valores de A_i y n es número total de muestras

De esta manera se consigue generar el menor número de subconjuntos, ayudando así a conseguir un árbol óptimo. Para evitar el sobreajuste que producía ID3, C4.5 aplica

un proceso de poda, una vez se haya construido el árbol inicial⁷. Este proceso comienza desde abajo, examinando los subárboles de los nodos no terminales iterativamente, y escogiendo el nodo que mayor mejora implique, hasta que no haya mejora.

Con la intención de poder manejar atributos con valores perdidos C4.5 opta por asignarle un valor de frecuencia de acuerdo con el resto de ejemplos, de manera que la ganancia adquiere un nuevo criterio:

$$G(A_i) = F \times (H - H(A_i))$$

donde H es la entropía de información del conjunto total de muestras, $H(A_i)$ es la entropía que corresponde a la característica A_i y F es el nuevo valor de frecuencia que mide el porcentaje de valores conocidos dentro de todo el conjunto de datos.

Los resultados obtenidos por este algoritmo y expresados en la matriz de confusión 4.1 reflejan que el algoritmo clasifica adecuadamente la clase que se corresponde a días donde no hay que dar tratamiento. Por el contrario no es capaz de clasificar correctamente los días de tratamiento. Debido a este desbalanceo en el conjunto de datos la precisión que nos proporciona el algoritmo no es buena.

		Predicción del sistema	
		NO TRATAR	TRATAR
Caso real	NO TRATAR	120	16
	TRATAR	6	2

Tabla 4.1: Matriz de confusión de C4.5

4.2.2. Naïve Bayes

Este algoritmo de clasificación [11] [12] se basa en dos suposiciones derivadas de la teoría de decisión bayesiana, que son en primer lugar que el problema de decisión se puede describir en términos probabilísticos, y en segundo lugar que todos los valores de

⁷A este proceso se le denomina *post-pruning*

las probabilidades del problema son conocidas. Para ello se basa en el conocido teorema de Bayes:

$$P(h|D) = \frac{P(D|h) \times P(h)}{P(D)}$$

donde:

- $P(h|D)$ es la probabilidad de observar el dato D , cuando se sabe que se cumple la hipótesis h . Se denomina *Probabilidad a posteriori de la observación D* .
- $P(D|h)$ es la probabilidad de que se cumpla la hipótesis h , una vez se haya obtenido el dato D . Se denomina *Probabilidad a posteriori de la hipótesis h* .
- $P(h)$ es la probabilidad de que la hipótesis h sea cierta. Se denomina *Probabilidad a priori de la hipótesis h* .
- $P(D)$ es la probabilidad de que realicemos la observación D . Se denomina *Probabilidad a priori de la observación D* .

La mayoría de los esquemas basados en el teorema de Bayes buscan hallar la hipótesis más probable a partir de los datos observados. A esta hipótesis se le denomina hipótesis MAP (*Maximo A Posteriori*). Por lo que

$$h_{MAP} = \underset{h \in H}{\operatorname{argmax}} P(h|D) = \underset{h \in H}{\operatorname{argmax}} \frac{P(D|h) \times P(h)}{P(D)}$$

El clasificador de Naïve Bayes además supone independencia condicional entre los datos, por lo que el cálculo de la hipótesis MAP se simplifica

$$V_{NB} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \prod_i P(a_i|v_j)$$

dando lugar a un ahorro computacional.

Una de las principales dificultades de este método es que se requiere el conocimiento inicial de muchas probabilidades, pero en la mayoría casos se solventa con estimaciones estadísticas basadas en el conjunto de patrones.

La matriz de confusión 4.2, resultado de evaluar el algoritmo, refleja que el algoritmo no es capaz de al clasificar correctamente ambas clases, ya que los niveles de falsos positivos y negativos son altos.

		Predicción del sistema	
		NO TRATAR	TRATAR
Caso real	NO TRATAR	75	61
	TRATAR	4	4

Tabla 4.2: Matriz de confusión de Naïve Bayes

4.2.3. IB1

El algoritmo IB1 [1] forma parte de la familia de algoritmos llamada IBL (Instance Based Learning), los cuales tratan de clasificar un nueva instancia en base a las ya conocidas. De esta manera no existe un entrenamiento del sistema, sino que al intentar clasificar una nueva muestra se extrae un conjunto de datos similares, procedentes de todo el conjunto de datos, y se utilizan para clasificar el nuevo dato. El algoritmo IB1 es prácticamente idéntico al NN (Nearest Neighbor) donde se utiliza la distancia euclídea como medida para encontrar el vecino más cercano.

$$d(I_1, I_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

donde I_1 e I_2 son dos muestras en un espacio bidimensional y (x_1, y_1) (x_2, y_2) son sus coordenadas.

Además IB1 normaliza los rangos de los atributos para poder tratar con diferentes

rangos, procesa las instancias de manera incremental, lo que supone que el clasificador puede ir cambiando a lo largo del tiempo, y además permite tratar valores de atributos perdidos, tal y como ocurría en C4.5.

Este algoritmo al ser de tipo perezoso, tiene la ventaja de poder ir adaptándose con cada nueva muestra, y la descripción de esta puede ser tan compleja como se quiera. Sin embargo al realizar todo el cómputo en tiempo de clasificación, el coste computacional es alto y atributos irrelevantes pueden afectar a la medida de similitud, por lo que conjuntos de datos que no hayan pasado por una selección de características pueden verse seriamente comprometidos. Este algoritmo nos proporciona los resultados reflejados en la matriz de confusión 4.3 y que nos muestran un resultado similar al C4.5, donde la clase que se corresponde a días donde no hay que dar tratamiento está clasificada correctamente y la clase que representa los días de tratamiento no se identifica adecuadamente.

		Predicción del sistema	
		NO TRATAR	TRATAR
Caso real	NO TRATAR	111	25
	TRATAR	6	2

Tabla 4.3: Matriz de confusión de IB1

4.2.4. SVM

Las Máquinas de Soporte Vectorial, en inglés SVM (Support Vector Machine) [10] fueron presentadas en 1992 y se derivan de la teoría de aprendizaje estadístico postulada por Vapnik y Chervonekis. Este tipo de esquema de aprendizaje ha ido ganando popularidad debido a su firme base matemática y a los buenos resultados obtenidos en una amplia variedad de campos. La base de este algoritmo es la construcción de un hiperplano que permita separar las clases con un margen máximo entre vectores de soporte, que son las muestras más cercanas al hiperplano que maximizan la distancia del margen.

La búsqueda del hiperplano que maximice el margen se realiza mediante la optimización cuadrática y puede resolverse mediante un problema dual introduciendo

multiplicadores de Lagrange. Así la solución del hiperplano se puede escribir en función de los vectores de soporte.

Cuando el problema al que nos enfrentamos es demasiado complejo, encontrar una solución óptima es muy complicado dentro del espacio de datos de entrada, por eso SVM proyecta este en uno de mayor dimensión por medio de una función Kernel, permitiendo así transformar algoritmos lineales en no lineales, siempre que estos dependan solamente del producto escalar de dos vectores.

Las funciones kernel deben ser continuas, simétricas y, preferiblemente, tener una matriz de Grann semidefinida positiva. Con la última condición se cumpliría el teorema de Mercer con lo que se garantiza que el problema convergerá y la solución será única. Estas condiciones se cumplen en numerosos casos por lo que tenemos una amplia gama de kernels disponibles para poder enfrentar este esquema a muchos tipos de problemas. Entre los kernels más utilizados encontramos el kernel lineal, polinómico o laplaciano, entre otros.

Este tipo de métodos tienen numerosas ventajas sobre métodos tradicionales como las redes de neuronas, haciendo que el entrenamiento sea más fácil, aceptando datos de diferentes tipos o eliminando los óptimos locales. Sin embargo encontrar una función de kernel que se ajuste bien o conseguir ajustar los parámetros hacen que haya un proceso laborioso previo.

Una vez evaluado el algoritmo, la matriz de confusión 4.4 muestra un resultado similar al C4.5 e IB1, donde la clase que se corresponde a días donde no hay que dar tratamiento se detecta correctamente, mientras la clase que representa los días de tratamiento no se clasifica adecuadamente.

		Predicción del sistema	
		NO TRATAR	TRATAR
Caso real	NO TRATAR	104	32
	TRATAR	5	3

Tabla 4.4: Matriz de confusión de SVM

4.2.5. Regresión Logística

La regresión logística [12] [1] es un modelo probabilístico desarrollado por D.R. Cox en 1958. Este modelo, al igual que la regresión lineal ordinaria, permite relacionar una variable dependiente con varias variables independientes, dándonos una medida de conformidad según se ajuste más o menos el modelo. Sin embargo, mientras la regresión lineal ajusta la línea por medio de mínimos cuadrados, la regresión logística utiliza modelos de máxima verosimilitud. La regresión logística está pensada para explicar el comportamiento de la variable dependiente cuando esta es de naturaleza binaria.

En este caso la matriz de confusión 4.5 nos indica que los resultados siguen la misma tendencia que los anteriores.

		Predicción del sistema	
		NO TRATAR	TRATAR
Caso real	NO TRATAR	102	34
	TRATAR	5	3

Tabla 4.5: Matriz de confusión de Regresión Logística

4.3. Resultados del modelo de Goidanich

En esta sección presentamos los resultados de aplicar el modelo de Goidanich a los datos históricos. Para ello se ha utilizado un umbral del 70 % hasta el 90 % de desarrollo para determinar el momento en que se debería de dar el tratamiento, tal y como ha indicado el experto, ya que a partir del 90 % la planta ya habría sufrido daños. Para esto se ha utilizado como conjunto de test cada año por separado y después se ha hecho la media. Podemos ver los resultados en la matriz 4.6, que nos muestran que esta técnica ha dado resultados positivos al hacer una predicción de ambas clases.

		Predicción del sistema	
		NO TRATAR	TRATAR
Caso real	NO TRATAR	135	2
	TRATAR	1	7

Tabla 4.6: Matriz de confusión del modelo de Goidanich

4.4. Comparación de resultados

En esta sección se pretende comparar los resultados de los algoritmos anteriores y los del modelo de Goinadich. Para ello se hará uso de las métricas presentadas en la sección 2.2, las cuales determinan la eficacia de las diferentes técnicas.

La tabla 4.7 muestra los valores de sensibilidad, especificad y precisión para cada una de las diferentes técnicas.

Método	Sensibilidad	Especificidad	Precisión
C4.5	0.882	0.250	0.847
Naïve Bayes	0.551	0.500	0.549
IB1	0.816	0.250	0.785
SVM	0.765	0.375	0.743
Regresión logística	0.750	0.375	0.730
Modelo de Goidanich	0.985	0.875	0.979

Tabla 4.7: Comparación de los principales métodos

En la tabla comprobamos que hay un método que destaca sobre los demás, el modelo de Goidanich. Este modelo consigue una tasa alta de acierto a la hora de predecir la fecha de un tratamiento y no se equivoca a la hora de determinar cuando no se debe tratar. El resto de métodos, como ya habíamos ido avanzando en la sección 4.2, no consiguen que el resultado sea fiable a la hora de determinar la fecha de un tratamiento. Con el modelo de Goidanich podemos tener la certeza de que al menos el resultado sea

una predicción acorde con el criterio del experto.

Desarrollo del sistema

Dentro de los objetivos de este proyecto estaba crear un sistema con una interfaz amigable, que mostrara aspectos esenciales para el seguimiento del hongo mildiu, y al que se incorporara uno de los métodos evaluados para crear una predicción del riesgo de infección. Para el diseño de la interfaz se ha contado con las recomendaciones de los expertos, que han dado sus opiniones sobre los parámetros más importantes, la manera de visualizarlos y otros aspectos importantes a la hora de tomar decisiones. El modelo que se ha implementado e incorporado al sistema ha sido el que ha dado mejores resultados en los experimentos y que, como podemos ver en la sección 4.4, ha sido el modelo de Goidanich.

Previamente al diseño de la interfaz se establecieron los criterios que esta debía de seguir para cumplir el objetivo final de ayuda a la toma de decisiones.

- **Requisitos funcionales** Aquí describiremos las funcionalidad básicas que debe tener la aplicación.
 - Mantener actualizados los datos meteorológicos.
 - Establecer fechas de tratamientos y tipo de tratamiento asociado, que se irán guardando en un fichero.
 - Visualizar los datos de cada una de las estaciones meteorológicas.
 - Calcular el desarrollo del mildiu desde el último tratamiento mediante el modelo Goidanich.
- **Requisitos no funcionales** Aquí describiremos como debe presentar la aplicación aspectos relevantes de la misma.
 - Visualizar cada parámetro por separado.

- Establecer una alerta visual, utilizando un código de colores. Rojo cuando el riesgo es alto y se debe aplicar un tratamiento, amarillo cuando el riesgo es medio y verde cuando el riesgo es bajo.
- Establecer dos gráficas que permitan comparar la evolución de parámetros diferentes a la vez.
- La visualización de los datos abarcará los últimos 10 días.
- Utilizar dos métodos para establecer la fecha de tratamiento:
 - Uno que permita determinar la fecha de tratamiento en el día previo.
 - Otro método que permita seleccionar en un calendario la fecha del último tratamiento aplicado.
- Poder revertir los cambios.
- Visualizar el histórico de tratamientos.⁸

Con el objetivo de mantener los datos actualizados de manera automática se ha implementado un cliente Restful que se conecta a un servicio, creado por la empresa *Monet Viticultura*, que proporcionará los datos actualizados de las mediciones hechas por las estaciones meteorológicas. Para ello se ha utilizado el lenguaje Python junto con librerías especializadas. También se ha implementado con este lenguaje un parser JSON para el tratamiento de los datos descargados en este formato y generar ficheros con formato *csv*, el cual es fácilmente reconocido por Matlab. Se muestra a continuación, en la figura 5.1, el diagrama de clases correspondiente a esta parte del sistema.

⁸Este aspecto es importante para el experto, pues es necesario saber que tipos de productos han sido utilizados para determinar cuál será el siguiente.

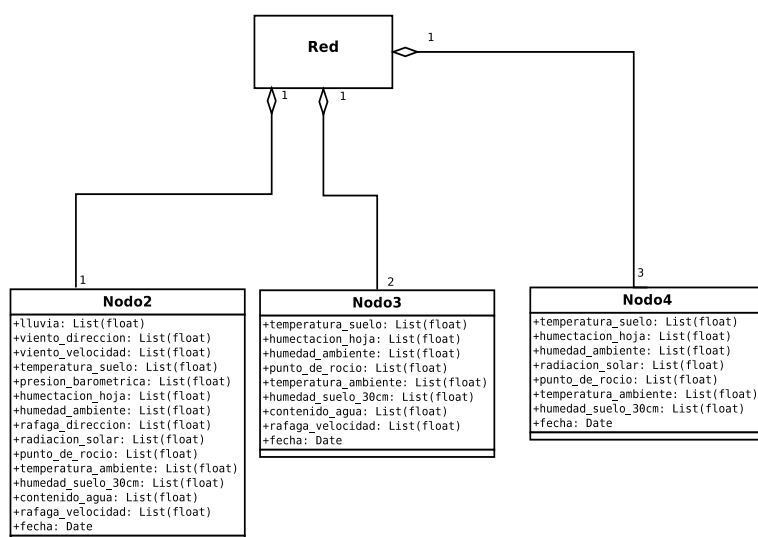


Figura 5.1: Diagrama de clases de cliente Resful.

A continuación presentaremos en la figura 5.2 el diagrama de casos de uso que se han tenido en cuenta a la hora de desarrollar el sistema.

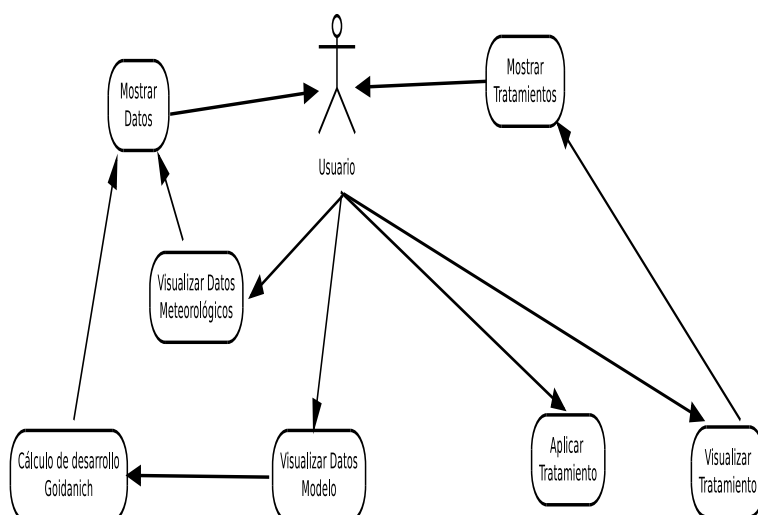


Figura 5.2: Diagrama de casos de uso.

5.1. Interfaz

Para el diseño de la interfaz se ha utilizado la herramienta Guide, que está dentro del paquete de herramientas de Matlab. En la imagen 5.3 se puede ver una captura con todos los elementos numerados. A continuación se describirá cada uno de esos elementos.

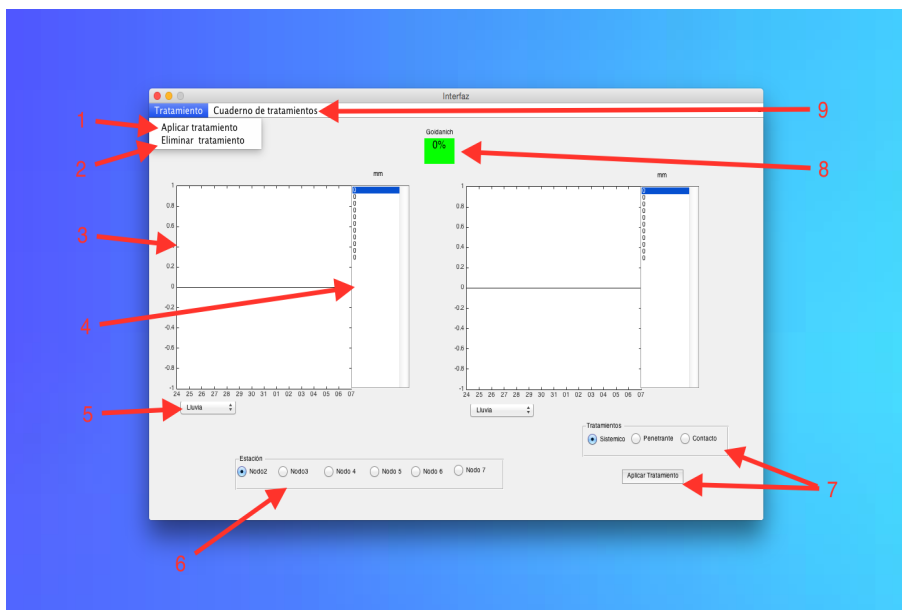


Figura 5.3: Interfaz de usuario

1. Este botón del submenú nos permite establecer la fecha del último tratamiento en un calendario.
2. Este botón hace que eliminemos el último tratamiento, pudiendo así revertir el único cambio permanente que el usuario puede hacer sobre la interfaz.
3. Esta es una de las dos gráficas gemelas. En el eje de las abscisas siempre nos encontraremos con las fechas a las que se corresponde cada muestra, representando el día del mes, mientras que en el eje de las ordenadas tendremos escala correspondiente a cada parámetro.
4. Esta lista, y su gemela, nos presentan los valores numéricos de cada una de las muestras representada en la gráfica correspondiente. En la parte superior tenemos la métrica que emplea cada una de ellas.

5. Este botón desplegable nos sirve para seleccionar el tipo de parámetro que se mostrara en la gráfica superior. El botón desplegable que hay a la derecha hace lo mismo con su correspondiente gráfica. Los parámetros que contiene cambian con respecto a la estación seleccionada, debido a que algunas estaciones cuentan con sensores con los que otras no cuentan.
6. Cada uno de estos botones seleccionables representa una estación meteorológica. Mientras uno esté marcado, los datos cargados en la aplicación se corresponderán con esa estación.
7. Estos dos elementos permiten establecer que se ha dado un tratamiento en el día previo. En la parte superior se indica el tipo de tratamiento y en la inferior el botón que hace el cambio.
8. Este elemento muestra el desarrollo del hongo según el modelo de Goidanich. Para poder determinar el riesgo de una manera más visual se han marcado tres umbrales que determinarán el código de color, como se puede ver en la figura 5.4. Entre 0 y 30 se usa el color verde (riesgo bajo), entre 30 y 60 se usa el color amarillo (riesgo medio) y > 60 se usa el color rojo (riesgo alto).

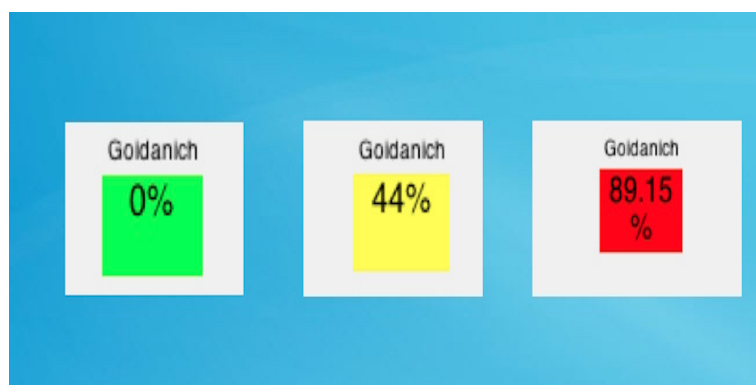


Figura 5.4: Código de colores

9. Este botón del menú nos abrirá el cuaderno de tratamientos históricos, que podemos ver en la figura 5.5, donde se pueden visualizar todos los tratamientos dados hasta la fecha, junto con el tipo de tratamiento que se ha dado.

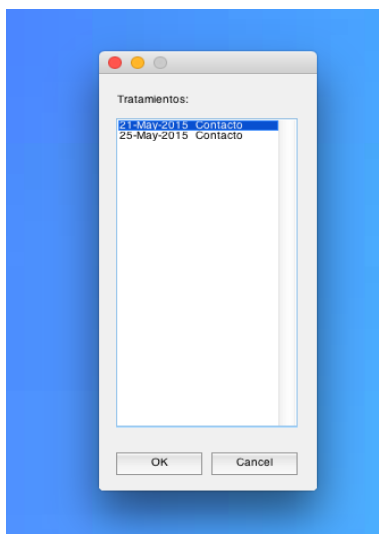


Figura 5.5: Cuaderno de tratamientos.

La interfaz cuenta además con una ventana de carga que nos mantiene en espera hasta que se termine la descarga de datos inicial. Solo aparece al inicio de la ejecución.

A continuación se describirán se describirán dos ejemplos de casos de uso:

- **Caso1.** Este ejemplo trata la tarea de establecer una fecha de tratamiento en un día, que no necesariamente sea el día previo. Es útil por ejemplo para introducir casos históricos que no estaban previamente en la base de datos. Para ello nos presentará un calendario en el que podremos seleccionar el día de tratamiento. Una vez hecho esto se abrirá un cuadro de diálogo donde podremos seleccionar el tipo de tratamiento (penetrante, contacto o sistémico) y, una vez seleccionado, nuestro fichero de tratamientos se actualizará con los valores deseados. El proceso está representado en la figura 5.6.

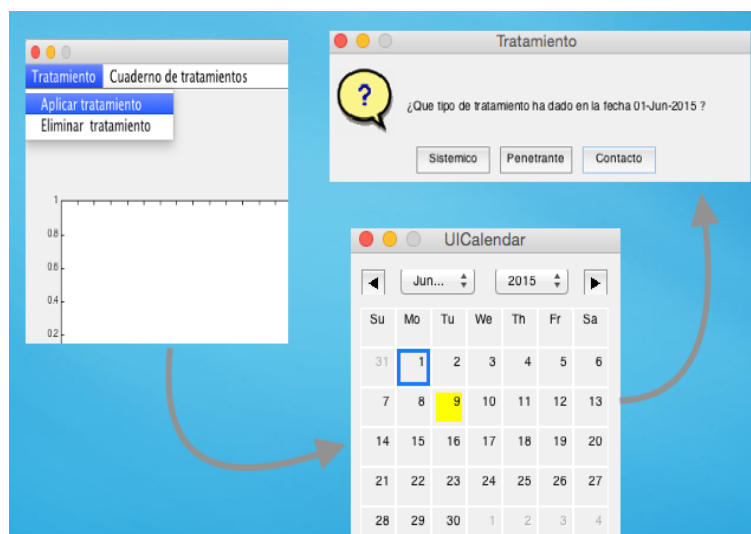


Figura 5.6: Caso de uso tratamiento.

- **Caso2.** Este ejemplo trata la tarea de cambiar el parámetro que se quiere visualizar. Para ello abrimos el menú desplegable y seleccionamos el nuevo parámetro. El resultado es una gráfica y una lista con nuevos valores, correspondientes al nuevo parámetro. El proceso está representado en la figura 5.7.

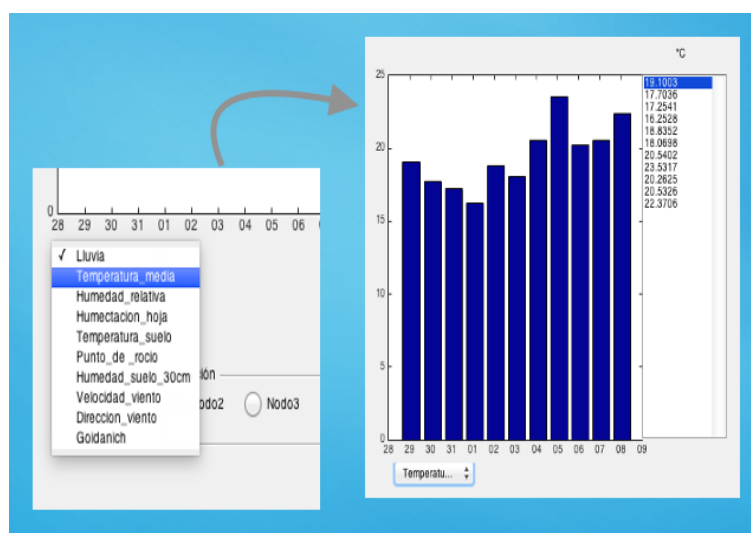


Figura 5.7: Caso de uso parámetro.

Después de describir los elementos y características de este sistema, hay que destacar que el objetivo de este es ser una herramienta para la monitorización de la enfermedad

fúngica, por lo que se ha prestado especial atención a elementos de visualización de datos y se ha apostado por la sencillez a la hora de hacer alguna acción, con un máximo de tres pasos para la realización de la más compleja.

Conclusiones y trabajo futuro

Los objetivos marcados para este proyecto respondían a dos enfoques diferentes. Por una parte se pretendía evaluar un modelo de los propuestos, que pudiera dar una predicción lo más fiable posible del estado de desarrollo del hongo mildiu y así poder evaluar el riesgo. Por otro lado se evaluaron diferentes esquemas de aprendizaje máquina, con la intención de conseguir igualar, y si fuese posible, superar, los resultados de los modelos comúnmente utilizados en el campo. Una vez se evaluaran las diferentes opciones se implementaría la mejor opción y se incorporaría a la interfaz. El marco en el cual han sido comparados estos dos enfoques ha sido el conjunto de datos históricos, el cual disponía de un parámetro binario que marcaba la fecha del tratamiento. Este parámetro ha sido extraído del cuaderno de campo, que ha sido elaborado por el experto y, que en algunos casos, se ha apoyado en el cálculo manual del modelo de Goidanich para determinar una fecha de tratamiento.

Los resultados han sido adecuados en el caso del modelo de Goidanich mientras que en el caso de los esquemas de aprendizaje no se han obtenido resultados de clasificación similares a los del modelo anterior. Esto se debe principalmente a dos factores, por una parte un reducido conjunto de datos históricos que dificulta la tarea de aprendizaje y, por otro lado, la flexibilidad del modelo de Goidanich que permite superar las carencias que sufren estos datos.

Los resultados obtenidos en el proyecto han corroborado los comentarios de los expertos, que aseguraban que es muy complicado obtener una solución completamente automática que haga una predicción exacta de la fecha en que se debe de tratar la planta, ya que la decisión no sólo responde a parámetros meteorológicos, sino que es necesario conocer el estado de las plantas de primera mano atendiendo a posibles manchas. Extraer el conocimiento de los expertos sobre cómo tomar esta decisión es complicado, por lo que el fin último de este proyecto ha sido ayudar a tomar esta decisión de manera que sea posible minimizar el número de tratamientos.

6.1. Trabajo futuro

Finalizado este trabajo se pueden destacar varios aspectos que se pueden mejorar o extender:

- **Ampliación del área de acción:** Durante el último mes de desarrollo de este proyecto se han puesto en marcha más estaciones de recogida de datos. Aprovechar estos datos podría ser una ampliación importante para la mejor comprensión de esta enfermedad y de cómo se relaciona con datos climáticos. Se daría un paso más hacia la mejora en el Manejo Integrado de Plagas y Enfermedades. Además la actualización sería sencilla dada la escalabilidad del código.
- **Nuevos datos históricos:** Unido a la mejora anterior sería posible reunir nuevos datos históricos y de más calidad que podrían volver a alimentar los diferentes algoritmos de clasificación y poder mejorar el rendimiento de estos.
- **Nuevas características:** Tal y como se ha comentado antes, los datos meteorológicos no son todo lo necesario para tomar la decisión con respecto al hongo, sino que hay otros aspectos como la aparición de manchas en las hojas. Este aspecto está siendo estudiado hoy en día y ya existen estrategias automáticas que consiguen determinar la aparición de estas manchas. Para ello en diversos lugares se están utilizando drones que vuelan a baja altura sobre las viñas y que detectan las manchas con diferentes técnicas de visión artificial que implican análisis de colores y texturas.
- **Utilizar base de datos de productos químicos:** En este proyecto se ha simplificado el tipo de productos que se pueden dar en un tratamiento, diferenciando tres. En realidad existen infinidad de productos con sus características, por lo que en un trabajo futuro se podría extraer la información de estos productos de la base de datos que posee el Ministerio de Agricultura para así poder refinar más el resultado del modelo.

I.1. Implementar petición al servidor

```
#Generar string fechas x días pasados desde hoy
today=datetime.datetime.now()+ relativedelta(days=-x)
month_ago = datetime.datetime.now()+ \
relativedelta(days=-(x+1))
date_format='%d-%m-%Y'
fecha_inicial=month_ago.strftime(date_format)+' 00:00'
fecha_final=today.strftime(date_format)+' 00:00'

urllib3.contrib.pyopenssl.inject_into_urllib3()

#Acceso a la hora del servidor para
# la construcción del token
url = 'https://clientes.h2210457.stratoserver.net' \
'/index.php?r=consultas/SYNC_time'
response = urllib2.urlopen(url).read()

#Construcción del token
sub1=response[1:15]
m=md5.new('Vcosteira_json'+sub1+'xxxxxxxxxx')
token=m.hexdigest()

#Contruir petición
url='https://clientes.h2210457.stratoserver.net' \
'/index.php?login=Vcosteira_json&r=consultas/JSON_data'
```

```
params = {'token':token, 'fecha_inicial':fecha_inicial, \
'fecha_final':fecha_final}
url_parts = list(urlparse.urlparse(url))
query = dict(urlparse.parse_qs(url_parts[4]))
query.update(params)
url_parts[4] = urllib.urlencode(query)
url= urlparse.urlunparse(url_parts)
```

```
#Enviar petición
```

```
response = urllib.urlopen(url).read()
json_response = json.loads(response)
```



Instalación

II.1. Requisitos software

Para la correcta ejecución de este programa se debe de disponer del siguiente software instalado:

- Python 2.7, disponible en <https://www.python.org/downloads/>, junto a los siguientes paquetes:
 - urllib, urllib2 y urllib3
 - json
 - md5
 - urlparse
 - requests
 - datetime
 - calendar
- MATLAB Compiler Runtime v84, disponible en <http://es.mathworks.com/products/compiler/mcr/>

II.2. Instalación

Se describe a continuación el proceso de instalación en entornos Linux y OS X una vez haya sido instalado el software requerido.

- Ir a la carpeta SAPM donde se encuentra el script `run_SAPM.sh`
- Ejecutar el comando `sudo chmod +x`
- Ejecutar el comando `./run_SAPM.sh <v84_dir>`, donde `<v84_dir>` es el directorio donde se encuentra el MATLAB Compiler Runtime v84 instalado.

Una vez realizado estos dos pasos el proceso puede tardar un poco debido a la descarga de los datos de los últimos treinta días. Una vez se abra la pantalla principal el sistema está listo para su uso.

Bibliografía

- [1] Witten,I.H y E.Frank y M.A. Hall: *Data Mining: Practical machine learning tools and techniques* , Morgan Kaufmann, San Francisco, 3rd Edition, 2011.
- [2] Waikato, Machine Learning Group University of: *WEKA - The Waikato Environment for Knowledge Analysis*, http://weka.sourceforge.net/wekadoc/index.php/Main_page. Último acceso: Mayo 2015.
- [3] Mathworks, <http://www.mathworks.com/>. Último acceso Mayo 2015
- [4] *Matlab, Creating Graphical User Interfaces, R2014b* , The MathWorks Inc, SNatick, 2014.
- [5] Python Documentation, <https://www.python.org>. Último acceso Mayo 2015
- [6] Perl Documentation, <https://www.perl.org>. Último acceso Mayo 2015
- [7] Quinlan, J. R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [8] Ethem Alpaydin : *Introduction to Machine Learning*. MIT Press, 2004.
- [9] Russell y Norvig: *Inteligencia Artificial, un enfoque moderno*. Pearson, 2004.
- [10] Boser, B. Guyon, I., and V. Vapnik: *A training algorithm for optimal margin classifiers. Fifth Annual Workshop on Computational Learning Theory*. ACM Press, Pittsburgh, 1992.
- [11] Mitchell, T . *Machine Learning*, McGraw Hill 1997.
- [12] Ng, A.Y. and Jordan, M. I. *On Discriminative vs. Generative Classifiers: A comparison of Logistic Regression and Naive Bayes*, *Neural Information Processing Systems*, Ng, A.Y., and Jordan, M. 2002.
- [13] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, W. Philip Kegelmeyer, *SMOTE: Synthetic Minority Over-sampling Technique*, *Journal of Artificial Intelligence Research* 16:321-357, 2002.

- [14] G.Goidanich: *Manuale di Patologia Vegetale*. Edizioni Agricole, Bologna, 1959
- [15] G. Barrios , R. Coscolla , A.L. Espadas, J.J. Pérez , J. L. Pérez, J. Toledo: *Los parásitos de la vid, estrategias de protección razonada*, Ministerio de Agricultura, pesca y alimentación, Ediciones Mundi-Prensa, 4ª Edición, 1998.
- [16] Alain Reynier: *Manual de viticultura*, Ediciones Mundi-Prensa, 6º Edición, 2005.
- [17] Pérez Marín, J.L: *Modelización de las enfermedades mildiu, podredumbre gris y odio de la vid: casos prácticos y situación actual en España*, Pythoma nº 114: 78-79, 1998
- [18] P. Spencer-Philips, M. Jeger: *Advances in downy mildew research*, Volume 2, Kluwer Academic Publishers, 2004.
- [19] IPM, Statewide Integrated Pest Management Program, <http://www.ipm.ucdavis.edu/PMG/r302100311.html>. Último acceso Mayo 2015.
- [20] G. Medina, J. Grageda, J. A. Ruiz, A. D. Báez, *Uso de estaciones meteorológicas en la agricultura*, Instituto Nacional de Investigaciones Agrícolas y Pecuarias, Folleto informativo nº 50, Mayo de 2008.
- [21] V. Rossi, T. Caffi, A. Cossu, F. Fronteddu, *Empirical vs Mechanistic models for primary infections of Plasmopara viticola*, Wagenigem, 17 Octubre 2006
- [22] Pérez Marín, J.L: *Sintomatología y sistemas de control del mildiu de la vid*, Pythoma nº 83: 114-116, 1998.
- [23] C. Iglesias, A. Martínez, J.P. Mansilla, M. Vilas, R. Pérez, J. Fernandes: *Creación de Redes de Videovigilancia e Prevención Agrícola como medio de ayuda da posta en marcha da Producción Integrada*, Xunta de Galicia, Consellería do Medio Rural, 1ª Edición, 2008.
- [24] G. Barrios, J. Reyes: *Modelización del mildiu en la vid*, Phytoma nº 164, Diciembre 2004.
- [25] L. Frías, Ramón, R. de Acuña, F. Trujillo, E. Perera: *Ensayo de validación del modelo predictivo de mildiu en viña Goidanich en cuatro comarcas vitícolas de tenerife..*
- [26] *El mildiu y su modelo de predicción*, Diputación foral de alava, Servicio de viticultura y enología.
- [27] P. Mansilla, Pérez, C. Pintos: *Ficha 10/99 (P. viticola)*. Estación Fitopatológica do Arriero. Servicio Agrario, Diputación Provincial de Pontevedra.