



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



**TFG del Grado en Ingeniería
Informática**

**Análisis de datos de
temperatura y humedad de
suelo procedentes de sensores
IoT desplegados en un viñedo**



Presentado por Gabriel Hernández Vallejo
en Universidad de Burgos — 16 de abril
de 2023

Tutores: Rubén Ruiz González, Alejandro
Merino Gómez



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



D. Rubén Ruiz González y D. Alejandro Merino Gómez, profesores del departamento de Ingeniería Electromecánica, área de Ingeniería de Sistemas y Automática.

Exponen:

Que el alumno D. Gabriel Hernández Vallejo, con DNI 71709111-X, ha realizado el Trabajo final de Grado en Ingeniería Informática titulado *"Análisis de datos de temperatura y humedad de suelo procedentes de sensores IoT desplegados en un viñedo"*.

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Burgos, 16 de abril de 2023

Vº. Bº. del Tutor:

Vº. Bº. del co-tutor:

D. nombre tutor

D. nombre co-tutor

Resumen

En este primer apartado se hace una **breve** presentación del tema que se aborda en el proyecto.

Descriptores

Palabras separadas por comas que identifiquen el contenido del proyecto Ej: servidor web, buscador de vuelos, android ...

Abstract

A **brief** presentation of the topic addressed in the project.

Keywords

keywords separated by commas.

Índice general

Índice general	iii
Índice de figuras	v
Índice de tablas	vi
Introducción	1
1.1. Estructura de la memoria	2
1.2. Materiales adjuntos	3
Objetivos del proyecto	5
2.1. Objetivos generales	5
2.2. Objetivos técnicos	5
2.3. Objetivos personales	6
Conceptos teóricos	7
3.1. Pre-procesamiento de datos	7
Técnicas y herramientas	11
4.1. Metodologías	11
4.2. Control de versiones	11
4.3. Alojamiento del repositorio	11
4.4. Gestión del proyecto	12
4.5. Comunicación	12
4.6. Entorno de desarrollo	12
4.7. Librerías	13
4.8. Restful API	13

Aspectos relevantes del desarrollo del proyecto	15
Trabajos relacionados	17
Conclusiones y Líneas de trabajo futuras	19
Bibliografía	21

Índice de figuras

Índice de tablas

Introducción

El cuidado de las tierras de cultivo vitícolas ha supuesto a menudo a lo largo de la historia diferentes retos desde el comienzo de su domesticación entorno al 3500-3100 a.C., continuando en la época del Antiguo Egipto, lugar en el que se cree se comenzó a popularizar la ingesta del producto resultante de la labranza y cosecha de los cultivos: el vino [1].

En la península ibérica el interés por este proceso y su resultante comenzó en el periodo fenicio cerca del 1100 a.C., expandiéndose por todo el territorio peninsular gracias al comercio, lo que dio como resultado que los procesos enológicos fueran ya bien conocidos en los siglos IV-III a.C. [1, 2].

Entre los principales retos a los que se han enfrentado los viticultores en la península ibérica se encuentran numerosas plagas de insectos, ácaros, nematodos, diferentes vertebrados, moluscos, bacterias, etc [3]. Siendo la más devastadora la causada por la histórica palga de filoxera (*Phylloxera vastatrix*) a finales del siglo XIX y comienzos del XX, que obligó a una reestructuración prácticamente completa de la viticultura española [4].

Sin embargo, el principal factor influyente en el cultivo de la vid es el clima, siendo este el causante de sus alteraciones, fisiopatías y la presencia de casi todas las plagas, de manera que estas se encuentran relacionadas con diferentes factores climáticos como la temperatura, las precipitaciones, la humedad y la humectación. De esta forma, por ejemplo, las altas lluvias o humedades y/o bajas temperaturas pueden producir corrimiento del racimo (escasez anormal de bayas en los racimos) o las altas temperaturas pueden generar desequilibrios hídricos que afecten a los injertos y a aquellas cepas infectadas por hongos de la madera [3].

Por otro lado, los datos climatológicos históricos indican que la temperatura global ha aumentado $1,1^{\circ}\text{C}$ desde el inicio de la Revolución Industrial hasta nuestros días, siendo este desequilibrio causado en los últimos años por la actividad humana, produciéndose a valocidad tal que los ecosistemas no tienen tiempo de adaptarse [5]. Las precipitaciones, además, no son uniformes sino que aumentan en zonas donde ya son muy abundantes y disminuyen en las regiones más secas [5].

Ante esta situación se plantea la cuestión de cuáles de estos efectos podríamos ser capaces de predecir y en qué proporción para, si bien no subsanar completamente, si ser capaces de paliar en cierta medida conociendo con anterioridad ciertos datos climatológicos que ayudarán a las diferentes tomas de decisiones de los viticultores.

De esta manera, siendo conocedores a corto o medio plazo la humedad y la temperatura superficial del suelo, entre otras variables, podríamos adaptar los sistemas de riego para adecuarlos a las necesidades de los cultivos y por otro lado, gestionar de forma adecuada los recursos hídricos nacionales (en este caso de la zona hidrográfica del Duero) al mismo tiempo que protegemos a las vides de plagas y trastornos fisiológicos debidos a diferentes factores como la humectación o la temperatura del suelo.

En este proyecto se propone el análisis de los datos de temperatura y humedad del suelo procedentes de sensores IoT (*Internet of Things*) desplegados en un viñedo situado en la zona sur de la provincia de Burgos, para, de esta forma, obtener modelos predictivos que nos permitan conocer el comportamiento de ciertas variables en función de otras y anticipar su evolución futura a corto plazo.

1.1. Estructura de la memoria

La memoria del proyecto cuenta con la siguiente estructura:

- Introducción: descripción del problema que el proyecto pretende resolver junto con la estructuración de la memoria y los materiales adjuntos.
- Objetivos del proyecto: descripción de los objetivos a cumplir con el desarrollo del proyecto.
- Conceptos teóricos: explicación de los conceptos teóricos necesarios para la comprensión del proyecto, el problema a abordar y la solución propuesta.

- Técnicas y herramientas: listado de técnicas y herramientas empleadas durante el desarrollo del proyecto.
- Aspectos relevantes del desarrollo: aspectos a destacar durante la realización del proyecto.
- Trabajos relacionados: trabajos relacionados con el problema que aborda el proyecto.
- Conclusiones y líneas de trabajo futuras: conclusiones obtenidas de la realización del proyecto y posibilidades de ampliarlo o de introducir mejoras.

Además de la memoria, se incluyen una serie de anexos:

- Plan de proyecto software: planificación temporal y estudio de viabilidad del proyecto.
- Especificación de los requisitos del software: descripción de la fase de análisis (objetivos y requisitos funcionales y no funcionales).
- Especificación del diseño: descripción de la fase del diseño.
- Manual del programador: aspectos relevantes del código fuente del proyecto.
- Manual de usuario: guía de usuario para el manejo de la aplicación asociada al proyecto

1.2. Materiales adjuntos

El proyecto incluye el siguiente contenido:

- Python Notebook con el pre-procesamiento de datos y las diferentes expliciones.
- Python Notebook con el modelado de la Red Neuronal Artificial empleada para la regresión de humedad y temperatura y las explicaciones correspondientes.
- Conjunto de datos de los sensores desplegados en el viñedo.
- Conjunto de datos del pluviómetro desplegado en el viñedo.

- Script de Python para mostrar las gráficas de los datos (de sensores y pluviómetro) sin procesar.
- Script de Python para mostrar las gráficas de los datos (de sensores y pluviómetro) procesados
- Script de Linux para la instalación del entorno virtual.
- Script de Windows PowerShell para la instalación del entorno virtual.
- Script de Windows CMD para la instalación del entorno virtual.

Los recursos mencionados se encuentran disponibles en GitHub [\[6\]](#).

Objetivos del proyecto

El proyecto aborda diferentes objetivos. Podemos hacer una distinción entre objetivos generales, objetivos técnicos y objetivos personales.

2.1. Objetivos generales

- Realizar un análisis de datos para ser capaces de predecir la temperatura y humedad del suelo en un viñedo.
- Facilitar la comprensión de los datos recogidos mediante representaciones gráficas.
- Buscar las correlaciones entre las diferentes medidas obtenidas y las diferentes variables ambientales y/o de características del suelo.
- Encontrar modelos predictivos que permitan explicar el comportamiento de ciertas variables en función de las otras y anticipar la evolución futura a corto plazo.

2.2. Objetivos técnicos

- Realizar visualizaciones de los datos recogidos por los sensores IoT (*Internet of Things*) desplegados en el viñedo.
- Realizar un pre-procesamiento de datos mediante librerías de manipulación de datos de Python como Pandas.

- Modelar una Red Neuronal Artificial con Keras que permita predecir la humedad y temperatura del suelo gracias a los datos proporcionados.
- Emplear Git como sistema de control de versiones distribuido mediante la plataforma GitHub.
- Emplear ZenHub para la gestión de proyectos mediante las metodologías ágiles.
- Aplicar en la medida de lo posible las metodologías ágiles mediante la técnica Scrum aprendida durante el desarrollo del grado (ciertos aspectos como las "*daily*" no pueden aplicarse debido al carácter del TFG).

2.3. Objetivos personales

- Emplear los conocimientos y técnicas adquiridas durante el desarrollo de los diferentes cursos del Grado en Ingeniería Informática.
- Profundizar en el uso de la inteligencia artificial para la resolución de problemas cotidianos que tendrá un reflejo en un sistema real (como es el campo del cultivo de la vid).
- Profundizar en la utilización de un lenguaje de programación tan versátil como es Python para el análisis de datos, empleando diferentes librerías nativas y de terceros, así como crear nuevos módulos.
- Profundizar en la librería de Tensorflow que se encuentra tan extendida en los diferentes ámbitos de la vida cotidiana (mundo laboral e investigador).

Conceptos teóricos

El desarrollo del proyecto cuenta con diferentes fases desde el pre-procesamiento de los muestreos de los sensores y el pluviómetro, continuando con la selección del modelo de Red Neuronal Artificial que se empleará en la predicción a corto plazo, así como su modelado para ser empleada en este problema concreto y su validación.

En esta sección se presentarán los conceptos teóricos en cada etapa para facilitar su comprensión.

3.1. Pre-procesamiento de datos

En el pre-procesamiento de datos se pretende realizar la integración y limpieza de datos, de forma que disminuyan los posibles problemas de calidad que puedan surgir en los diferentes sistemas de información.

Como norma general el **proceso de integración** debe ser realizado durante la fase de recopilación de los datos. La limpieza permite la detección y corrección de los problemas no resueltos en la fase anterior como los valores anómalos (*Outliers*) o faltantes [7].

Tras la integración de los datos de las diferentes fuentes (v.g. bases de datos), se pueden realizar un resumen de características de atributos, en la que se mostrarán las características generales de estos como medias, mínimos, máximos y valores posibles.

En esta tabla que proporciona diferente información podríamos obtener información trascendental para proceso de análisis de datos, sobre todo para atributos categóricos. En el caso de atributos numéricos un mecanismo visual que es especialmente útil es la gráfica de dispersión, que es la técnica

de visualización de datos que se empleará mayoritariamente durante el desarrollo del proyecto [7].

En el conjunto de datos podemos encontrar **valores faltantes**, perdidos o ausentes que pueden ser reemplazados por diferentes razones. Una de ellas es que el modelo que empleemos puede no tratar bien estos valores o que utilice un mecanismo de tratamiento que no sea adecuado al contexto. Un problema asociado a la detección de los valores faltantes es que estos no estén representados como nulos, lo que puede introducir sesgo en el conocimiento extraído [7].

Ante esta situación se puede actuar de diferentes maneras:

- Ignorarlos (ciertos algoritmos son tolerantes a los valores faltantes).
- Eliminar la columna que contiene valores faltantes.
- Filtrar las filas: eliminar las filas afectadas, lo que introduce cierto sesgo en muchas ocasiones.
- Reemplazar el valor por otro que preserve la media y la varianza del conjunto de datos en caso de atributos numéricos y la moda en atributos categóricos. Una forma de reemplazar los valores faltantes es la imputación de datos perdidos, que consiste en predecirlos a partir de otros ejemplos. Existen también algoritmos que se emplean tradicionalmente para este fin.
- Segmentar: se segmentan las filas por los valores disponibles y se obtiene un modelo por cada uno de los segmentos y se combinan.
- Modificar la política de calidad de datos y esperar a que los faltantes estén disponibles.

Además de las situaciones anteriores, es posible que el conjunto de datos cuente con **valores erróneos** que se deben detectar y tratar. La detección de estos campos puede realizarse de diferentes maneras dependiendo de su formato y origen.

En el caso que nos atañe se deben buscar los valores extremos, que no significa que estos sean erróneos, sino que estadísticamente se clasifican como anómalos, aunque representen un estado genuino de la realidad. Con todo y con eso, estos valores pueden suponer un problema para algunos métodos que se basan en el ajuste de pesos como las redes neuronales. En

otras ocasiones pueden haber datos erróneos que caen en la normalidad, por lo que no pueden ser detectados [7].

La falta de detección de estos valores puede resultar en problemas si posteriormente se normalizan los datos, puesto que la mayoría de datos estarían en un rango muy pequeño y puede haber poca precisión en algunos modelos ante estas situaciones.

Técnicas y herramientas

4.1. Metodologías

Scrum

La metodología *Scrum* es un marco de trabajo de desarrollo ágil de software iterativo (dividido en líneas de tiempo denominados *sprints*) e incremental, que surge como antítesis a la gestión de proyectos predictiva (centrada en la planificación, presupuestos y los plazos de entrega) [8, 9].

4.2. Control de versiones

El control de versiones se realizará mediante **Git** puesto que a diferencia de otras herramientas de control de versiones, provee de una administración del repositorio distribuida en lugar de centralizada, lo que permite a los desarrolladores tener un clon del repositorio en su equipo local [10].

4.3. Alojamiento del repositorio

Para el *hosting* del repositorio se ha barajado emplear: **GitHub**, **Bitbucket**, **GitLab**.

Se ha optado por emplear **GitHub** debido a la facilidad de uso que ofrece y las múltiples funcionalidades disponibles, sumado a la destreza adquirida durante el desarrollo del grado con esta plataforma de alojamiento de repositorios.

4.4. Gestión del proyecto

Para la gestión del proyecto se ha considerado emplear [ZenHub](#), por su integración completa con GitHub.

Esta herramienta proporciona un *canvas* que permite gestionar las tareas (*issues*) en el *sprint* ordenándolas en columnas según sea su estado y priorizándolas según se encuentren en esta, además de permitir una estimación de su duración, asignar su desarrollo a uno o más desarrolladores e indicar el *milestone* y el *sprint* al que pertenecen las tareas.

4.5. Comunicación

La comunicación durante desarrollo del proyecto se realiza via e-mail, debido a la rapidez, versatilidad y facilidad de la comunicación, además de la familiaridad de las partes con el medio de comunicación.

Para las tutorías telemáticas se escogió emplear [Microsoft Teams](#) por la existencia de forma predefinida de una cuenta de usuario de la Suite Office de Microsoft proporcionada por la Universidad.

4.6. Entorno de desarrollo

Debido a la flexibilidad que aporta Visual Studio Code (editor de texto bajo licencia MIT desarrollado por Microsoft [11]) gracias a el entorno proporcionado de forma nativa y a las extensiones disponibles de Microsoft y terceros, se ha optado por emplearlo para el desarrollo de los scripts de Python, los Python Notebooks, la documentación del repositorio en Markdown y la realización de la memoria en LaTeX.

En cuanto al intérprete Python se ha empleado [Python 3.9.13](#) mediante un entorno virtual que permite instalar las dependencias sin modificar el entorno base.

En lo referente a la compilación de los ficheros LaTeX, se ha optado por emplear MikTeX debido a la facilidad de actualización e instalación de dependencias. Para la integración del compilador de LaTeX con Visual Studio Code se emplea [LaTeX Workshop](#), puesto que permite la compilación automática al guardar los cambios en los ficheros sin realizar ninguna configuración explícita.

4.7. Librerías

Pandas

Pandas es una herramienta de análisis de datos escrita en Python que ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales [12].

Tensorflow

Tensorflow es una librería de código abierto para aprendizaje automático desarrollada por Google [13].

En concreto se ha optado por emplear su *API*: *Keras*, debido a su modularidad y extensibilidad [14].

Matplotlib

Matplotlib es una librería para generar gráficos bidimensionales en Python a partir de diferentes estructuras de datos como listas [15].

Json

El módulo JSON de Python proporciona una *API* (interfaz de programación de aplicaciones) para manejar objetos en formato JSON (*JavaScript Object Notation*) [16].

Os

El módulo OS de Python proporciona una manera portable de emplear la funcionalidad del sistema operativo [17].

4.8. Restful API

Para obtener los datos meteorológicos de la zona en la que se situa el viñedo y poder realizar la comparación con las muestras obtenidas en el pluviómetro desplegado se han estudiado diferentes opciones:

- *OpenWeatherMap*
- *Weatherbit*

- AEMET OpenData

Todas las opciones requieren de una petición al servidor mediante HTTPS o HTTP, sin embargo, en el primer caso a pesar de que se muestran los datos en una frecuencia de una hora, el limitante se encuentra en que solo permite recuperar hasta un año atrás. En el segundo caso los requerimientos se cumplen al completo, pero tras contactar con el soporte de la *API* no hay respuestas que permitan el acceso al *endpoint* correspondiente.

Finalmente la última opción proporciona datos históricos sin límite, la contraparte es que la estación meteorológica consultada se encuentra en Aranda de Duero y se tratan de datos diarios.

Aspectos relevantes del desarrollo del proyecto

Este apartado pretende recoger los aspectos más interesantes del desarrollo del proyecto, comentados por los autores del mismo. Debe incluir desde la exposición del ciclo de vida utilizado, hasta los detalles de mayor relevancia de las fases de análisis, diseño e implementación. Se busca que no sea una mera operación de copiar y pegar diagramas y extractos del código fuente, sino que realmente se justifiquen los caminos de solución que se han tomado, especialmente aquellos que no sean triviales. Puede ser el lugar más adecuado para documentar los aspectos más interesantes del diseño y de la implementación, con un mayor hincapié en aspectos tales como el tipo de arquitectura elegido, los índices de las tablas de la base de datos, normalización y desnormalización, distribución en ficheros³, reglas de negocio dentro de las bases de datos (EDVHV GH GDWRV DFWLYDV), aspectos de desarrollo relacionados con el WWW... Este apartado, debe convertirse en el resumen de la experiencia práctica del proyecto, y por sí mismo justifica que la memoria se convierta en un documento útil, fuente de referencia para los autores, los tutores y futuros alumnos.

Trabajos relacionados

Este apartado sería parecido a un estado del arte de una tesis o tesina. En un trabajo final grado no parece obligada su presencia, aunque se puede dejar a juicio del tutor el incluir un pequeño resumen comentado de los trabajos y proyectos ya realizados en el campo del proyecto en curso.

Conclusiones y Líneas de trabajo futuras

Todo proyecto debe incluir las conclusiones que se derivan de su desarrollo. Éstas pueden ser de diferente índole, dependiendo de la tipología del proyecto, pero normalmente van a estar presentes un conjunto de conclusiones relacionadas con los resultados del proyecto y un conjunto de conclusiones técnicas. Además, resulta muy útil realizar un informe crítico indicando cómo se puede mejorar el proyecto, o cómo se puede continuar trabajando en la línea del proyecto realizado.

Bibliografía

- [1] J. Piqueras, *Historia de la vid y el vino en España edades Antigua y Media*. Valencia: Univeritat de València, 2014, ch. 1.
- [2] Ministerio de Agricultura, Pesca y Alimentación, “El cultivo de la vid en España,” [Online; Accedido 20-marzo-2023]. [Online]. Available: <https://www.mapa.gob.es/es/ministerio/archivos-biblioteca-mediateca/archivos/Cultivo.aspx>
- [3] Mundi-Prensa, *Los parásitos de la vid estrategias de protección razonada*, 5th ed. Madrid: Mundi-Prensa, 2004, ch. 1, pp. 17–18.
- [4] —, *Los parásitos de la vid estrategias de protección razonada*, 5th ed. Madrid: Mundi-Prensa, 2004, ch. 3, p. 66.
- [5] I. Z. L. Zúñiga, *Meteorología y climatología*, Ed. digital act.: sept. de 2021 ed., ser. Grado. Madrid: Universidad Nacional de Educación a Distancia, 2021, ch. 11, pp. 194, 216.
- [6] G. Hernández, “Repositorio del Trabajo Fin de Grado ‘Análisis de datos de temperatura y humedad de suelo procedentes de sensores IoT desplegados en un viñedo’ por Gabriel Hernández Vallejo.” [Online]. Available: <https://github.com/GabiHV/TFG22-23>
- [7] J. Hernández Orallo, *Introducción a la minería de datos*. Madrid: Pearson Educación, 2004, ch. 4.
- [8] Wikipedia contributors, “Scrum (software development) — Wikipedia, the free encyclopedia,” 2023, [Online; accessed 24-March-2023]. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Scrum_\(software_development\)&oldid=1144400875](https://en.wikipedia.org/w/index.php?title=Scrum_(software_development)&oldid=1144400875)

- [9] M. Palacio, *Scrum Master*. Iubaris Info 4 Media SL, 2021, ch. 1, p. 11.
- [10] S. F. Conservancy, “About - git,” 2023. [Online]. Available: <https://git-scm.com/about/distributed>
- [11] Wikipedia contributors, “Visual studio code — Wikipedia, the free encyclopedia,” 2023, [Online; accessed 24-March-2023]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Visual_Studio_Code&oldid=1145496466
- [12] Wikipedia, “Pandas (software) — wikipedia, la enciclopedia libre,” 2023, [Internet; descargado 24-marzo-2023]. [Online]. Available: [https://es.wikipedia.org/w/index.php?title=Pandas_\(software\)&oldid=148434974](https://es.wikipedia.org/w/index.php?title=Pandas_(software)&oldid=148434974)
- [13] —, “Tensorflow — wikipedia, la enciclopedia libre,” 2021, [Internet; descargado 24-marzo-2023]. [Online]. Available: <https://es.wikipedia.org/w/index.php?title=TensorFlow&oldid=139184038>
- [14] —, “Keras — wikipedia, la enciclopedia libre,” 2022, [Internet; descargado 25-marzo-2023]. [Online]. Available: <https://es.wikipedia.org/w/index.php?title=Keras&oldid=148156962>
- [15] —, “Matplotlib — wikipedia, la enciclopedia libre,” 2022, [Internet; descargado 24-marzo-2023]. [Online]. Available: <https://es.wikipedia.org/w/index.php?title=Matplotlib&oldid=146868590>
- [16] P. S. Foundation, “json — json encoder and decoder,” 2023. [Online]. Available: <https://docs.python.org/3/library/json.html#module-json>
- [17] —, “os — miscellaneous operating system interfaces,” 2023. [Online]. Available: <https://docs.python.org/3/library/os.html?highlight=os#module-os>