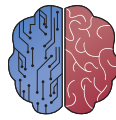




UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería de la Salud



INGENIERÍA
DE LA SALUD

**TFG del Grado en Ingeniería de la
Salud**

**Minería de datos y
aprendizaje automático
aplicado a la predicción de
incidencia de párkinson
basado en la biometereología.**

Presentado por Lorena Calvo Pérez
en Universidad de Burgos

4 de abril de 2025

Tutores: Antonio Canepa Oneto – Tutor 2



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería de la Salud



D. Tutor 1, profesor del departamento de departamento, área de área.

Expone:

Que el alumno D. Pepe Pérez, con DNI 123456A, ha realizado el Trabajo final de Grado en Ingeniería de la Salud titulado título del trabajo.

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, 4 de abril de 2025

Vº. Bº. del Tutor:

Vº. Bº. del Tutor:

D. Tutor 1

D. Tutor 2

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 |
| Objetivos | 3 |
| Conceptos teóricos | 5 |
| 3.1. Enfermedad del Parkinson | 5 |
| 3.2. La Biometeorología y su Relación con la Salud | 7 |
| 3.3. Minería de datos | 8 |
| 3.4. Estado del arte y trabajos relacionados. | 10 |
| Metodología | 11 |
| 4.1. Descripción de los datos. | 11 |
| 4.2. Procesamiento de datos. | 12 |
| 4.3. Técnicas y herramientas. | 14 |
| Resultados | 15 |
| 5.1. Resumen de resultados. | 15 |
| 5.2. Discusión. | 15 |
| Conclusiones | 17 |
| 6.1. Aspectos relevantes. | 17 |

| | |
|----------------------------------|-----------|
| Lineas de trabajo futuras | 19 |
| Bibliografía | 21 |

Índice de figuras

| | |
|--|---|
| 3.1. Pie de la figura de la figura bla bla bla | 9 |
|--|---|

Índice de tablas

3.1. 9

Introducción

Descripción del contenido del trabajo y de la estructura de la memoria y del resto de materiales entregados.

Objetivos

Objetivos principales del trabajo realizado.

Este apartado explica de forma precisa y concisa cuales son los objetivos que se persiguen con la realización del proyecto. Se puede distinguir entre:

1. Los objetivos marcados por los requisitos del software/hardware/análisis a desarrollar.
2. Los objetivos de carácter técnico, relativos a la calidad de los resultados, velocidad de ejecución, fiabilidad o similares.
3. Los objetivos de aprendizaje, relativos a aprender técnicas o herramientas de interés.

Conceptos teóricos

Explicación de los conceptos teóricos básicos necesarios para que cualquier miembro del tribunal pueda entender el trabajo realizado.

Esta sección puede contener el número de subsecciones que sean necesarias.

3.1. Enfermedad del Parkinson

La enfermedad del Parkinson es un trastorno neurodegenerativo del SNC (sistema nervioso central), el desorden de movimiento más común y la segunda enfermedad neurodegenerativa más frecuente tras el Alzheimer.

La prevalencia de la enfermedad es significativa, afectando a un 1-2 % de la población mayor de 65 años. La edad en la que suele manifestarse la enfermedad es entre los 65 y 70 años, pero pueden aparecer en mayores de 50 e incluso en los adolescentes.

Cabe destacar que hay una mayor incidencia en hombres que en mujeres.

La enfermedad de Parkinson se origina cuando ciertas neuronas en el cerebro dejan de funcionar adecuadamente. Estas células son responsables de producir dopamina, una sustancia que transmite señales a la parte del cerebro encargada de controlar el movimiento y la coordinación del cuerpo.

Las neuronas afectadas se encuentran en una región llamada sustancia negra. El Parkinson se desarrolla cuando estas células empiezan a morir o deteriorarse, lo que ocurre debido a alteraciones en su metabolismo.

Un aspecto clave de la enfermedad de Parkinson es la acumulación de una proteína llamada α -sinucleína. Esta proteína, en condiciones normales,

tiene una función en el cerebro, pero en los pacientes con Parkinson, se pliega de manera anormal, formando agregados que se acumulan en las neuronas, lo que contribuye al daño celular. Estos depósitos de α -sinucleína forman estructuras conocidas como cuerpos de Lewy, que son característicos de la enfermedad.[Zhang et al., 2018]

La producción insuficiente de dopamina desencadena los principales síntomas de la enfermedad, como temblores, lentitud de movimiento, rigidez y problemas de equilibrio.[Poewe et al., 2017]

El diagnóstico de la enfermedad de Parkinson (EP) se basa principalmente en la evaluación clínica. Los criterios más reconocidos fueron establecidos por la UK Parkinson Disease Society - Brain Bank e incluyen cuatro signos clave: bradicinesia o acinesia, temblor en reposo, rigidez e inestabilidad postural.[Marín et al., 2018]

El tratamiento de los síntomas motores de la enfermedad de Parkinson se basa en la terapia de reemplazo de la dopamina o la utilización de agonistas dopaminérgicos. No obstante, la dopamina no puede atravesar la barrera hematoencefálica, por lo que el tratamiento de referencia es la administración de levodopa (L-Dopa), su precursor, que se convierte en dopamina en el cerebro por la acción de la enzima di-hidroxi-fenilalanina descarboxilasa.

El uso de L-Dopa resulta ser efectivo para mejorar la calidad de vida de los pacientes al aliviar los síntomas motores causados por la pérdida de dopamina. Sin embargo, con el paso del tiempo su eficacia se ve disminuida, posiblemente debido a la progresión de la enfermedad y la muerte continua de neuronas dopaminérgicas.

Por esta ello, es necesario aumentar de forma progresiva la dosis, pudiendo dar lugar a la aparición de efectos secundarios, especialmente disquinesias (movimientos involuntarios anormales), que suelen manifestarse tras 5 a 10 años de tratamiento crónico. Estas disquinesias pueden afectar de forma significativa la vida diaria del paciente, dificultando actividades laborales y cotidianas.

Debido a estas complicaciones, la levodopa no siempre es recomendada como primera opción de tratamiento, salvo que sea estrictamente necesario. Por lo que, se prefiere el uso de los agonistas dopaminérgicos, ya que pueden retrasar la necesidad de L-Dopa y disminuir el riesgo de desarrollo de disquinesias tempranas.

En muchas ocasiones, para extender temporalmente los beneficios de la L-Dopa y retrasar la aparición de disquinesias, se pueden utilizar tratamientos combinados, como:

- Inhibidores de la catecol-O-metiltransferasa (COMT) y de la monoaminooxidasa (MAO): ayudan a prolongar la acción de la dopamina en el cerebro.
- Amantadina: actúa como antagonista de los receptores NMDA y puede reducir la severidad de las disquinesias.

[Hurtado et al., 2016]

3.2. La Biometeorología y su Relación con la Salud

La Biometeorología es la rama de la ciencia que trata las relaciones entre los procesos atmosféricos y los seres vivos.[Ramos, 2014] Por otro lado, la biometeorología médica, estudia cómo los fenómenos meteorológicos repercuten en el cuerpo humano y cómo los cambios del clima a lo largo de un año provocan variaciones importantes en la salud.[RTVE, 2017] Por ende, los cambios climáticos y las variables ambientales tienen un efecto importante en la salud humana, influyendo en el bienestar tanto físico como mental. Estos factores pueden contribuir al desarrollo de diferentes enfermedades. Los principales componentes considerados que influyen en la salud de las personas son:

- Calidad del aire: Entre los contaminantes que representan un grave riesgo para la salud pública se encuentran las partículas en suspensión, el monóxido de carbono, el ozono, el dióxido de nitrógeno y el dióxido de azufre. La contaminación del aire, tanto en interiores como en exteriores, causa enfermedades respiratorias y de otro tipo, y es una fuente importante de morbilidad y mortalidad[World Health Organization, 2025]
- Agua y saneamiento: El acceso a agua limpia y un saneamiento adecuado son esenciales para prevenir enfermedades transmitidas por el agua. La contaminación de fuentes de agua, por desechos industriales y urbanos, puede generar graves problemas de salud.
- Alimentos y seguridad alimentaria: Los contaminantes ambientales como pesticidas y metales pesados afectan la cadena alimentaria y pueden causar enfermedades tanto agudas como crónicas.
- Suelos y contaminación: La contaminación del suelo por productos químicos y desechos peligrosos afecta la salud humana a través del contacto directo o por la ingestión de alimentos contaminados.

- Cambio climático: Las alteraciones en el clima aumentan los riesgos para la salud, como la propagación de enfermedades, el estrés térmico y los desastres naturales, además de causar desplazamientos poblacionales.[Díaz Cordero, 2012] [Ambientum, 2025]

La Organización Panamericana de la Salud (OPS) resalta que el cambio climático representa un riesgo significativo para la salud y el bienestar.[(PAHO), 2025]

3.3. Minería de datos

La minería de datos es una disciplina que se centra en el desarrollo de métodos y algoritmos diseñados para extraer automáticamente información relevante, lo que facilita la identificación de patrones ocultos en grandes volúmenes de datos. Además, uno de los objetivos de la minería de datos es garantizar que la información obtenida tenga capacidad predictiva, optimizando así, el proceso de análisis de datos.[Martinez, 2001]

Tanto la minería de datos como el aprendizaje automático se han vuelto fundamentales en el campo de la salud, ya que permiten procesar y analizar grandes cantidades de datos clínicos y biométricos, lo que facilita la detección de patrones, la predicción de enfermedades y el apoyo en la toma de decisiones médicas.[Raul et al., 2016]

En conclusión, la combinación de la minería de datos y el aprendizaje automático en la predicción de la incidencia del Parkinson, teniendo en cuenta la biometeorología, permite identificar relaciones entre los variables ambientales y la salud de los pacientes. Contribuyendo a una mejor comprensión de la interacción del entorno y la salud.

Sub Subsección

En esta sección y el resto de secciones de la memoria puede ser necesario incluir listas de items.

- item1
- item2
- item3
- item4

| a | b | c |
|---|---|---|
| 1 | 2 | 3 |
| 4 | 5 | 6 |

Tabla 3.1:

Listas enumeradas.

1. item1
2. item2
3. item3

Figuras, como la figura 3.1 que aparece en la página 9.

Puedes aprender más de las figuras en la dirección https://es.overleaf.com/learn/latex/Inserting_Images



Figura 3.1: Pie de la figura de la figura bla bla bla

También se pueden insertar tablas como 3.1, que ha sido generada con <https://www.tablesgenerator.com/>.

Es necesario que todas las figuras y tablas aparezca referenciadas en el texto, como estos ejemplos.

Todos los conceptos teóricos deben de estar correctamente referenciados en la bibliografía. Por ejemplo, aquí estoy citando la página de L^AT_EX de Wikipedia [[Wikipedia, 2015](#)].

También puede ser necesario utilizar notas al pie ¹, para aclarar algunos conceptos.

¹como por ejemplo esta

3.4. Estado del arte y trabajos relacionados.

Revisión bibliografica de que se está haciendo en la industria o la academia relativo al problema que se está tratando.

Enumeración y resumen de todos los trabajos relacionados de interés.

Metodología

4.1. Descripción de los datos.

Los datos utilizados en este trabajo provienen de la plataforma Our World in Data (OWID), una fuente de datos abiertos sobre diversas temáticas globales. Para obtener los datos necesarios, primero se exploraron los conjuntos de datos disponibles en su web, seleccionando aquellos que eran más relevantes para el trabajo.

4.1.1. Obtención de datos.

Una vez identificados los datasets de interés, se procedió a localizar los endpoints de la API correspondientes, lo que facilitó la automatización de la obtención de los datos. Estos endpoints permiten acceder a los datos en tiempo real, lo que asegura que la información utilizada se mantenga actualizada.

El proceso consistió en obtener las URLs de los endpoints específicos, los cuales proporcionan los datos en formato JSON, un formato adecuado para su posterior manipulación y análisis. En algunos casos, al acceder a los metadatos, los datos proporcionados inicialmente incluían solo información descriptiva (como años y países), por lo que fue necesario cambiar la extensión del endpoint de "metadata.json" a "data.json" para obtener los valores reales de los indicadores, organizados por país y año. Este proceso permitió integrar los datos automáticamente en el análisis.

4.1.2. Limitaciones de Acceso y Obtención Manual de Datos

. Algunos conjuntos de datos presentan restricciones de redistribución que impiden su acceso directo a través de la API. En estos casos, al intentar acceder a los endpoints correspondientes, se muestra el mensaje: "Los datos de este gráfico no están disponibles para descargar debido a restricciones de licencia". Esta limitación impide la obtención automática de los datos en ciertos casos.

Ante esta restricción, se optó por un enfoque manual utilizando herramientas del navegador, como la función Control + Shift + I, para extraer la información directamente desde la página web de Our World in Data (OWID). A través de este proceso, se pudieron obtener los valores requeridos y generar los datasets necesarios para el análisis.datos.

4.2. Procesamiento de datos.

Una vez obtenidos los datos desde la plataforma Our World in Data (OWID) fue necesario estructurarlos y procesarlos para su análisis.

Sin embargo, la información necesaria no se encontraba en un único archivo JSON, sino que estaba distribuida en dos fuentes de datos distintas dentro de la misma plataforma:

- Fuente de datos principal: Contiene los valores de los indicadores analizados, organizados en una estructura donde los datos están asociados a identificadores numéricos de países y años.
- Fuente de metadata: Proporciona información adicional necesaria para interpretar correctamente los datos, como la correspondencia entre identificadores numéricos y nombres de países, así como la lista de años disponibles en el conjunto de datos.

Dado que cada fuente presenta una estructura diferente, se tuvo que diseñar un proceso de integración para combinar ambas y construir un conjunto de datos estructurado en forma de DataFrame, donde cada fila representa un país en un año determinado con su valor correspondiente del indicador analizado.

Para llevar a cabo esta integración, primero se cargaron los datos desde las API correspondientes, asegurando que la información estuviera correctamente obtenida antes de su procesamiento.

4.2.1. Carga de Datos

Para obtener los datos se implementó una función que se encarga de realizar peticiones a las direcciones URL de los datos y la metadata, obteniendo así la información en formato JSON. Esto permitió estructurar los datos de manera uniforme para su posterior manipulación y análisis.

4.2.2. Estructuración de Datos

El procesamiento de los datos es un paso esencial para estructurar la información de manera homogénea. Primero, se extrajeron los valores numéricos de los indicadores desde los datos obtenidos. Luego, se identificaron los años disponibles en el archivo metadata, ya que estos definen la dimensión temporal del análisis. Finalmente, se recuperó la lista de países o entidades presentes en el dataset, lo que permitió organizar la información de manera estructurada.

Si alguno de estos elementos no está presente en los datos obtenidos, se han establecido condiciones para detectar y gestionar estos errores, garantizando la coherencia de la información utilizada.

4.2.3. Conversión a DataFrame

Para facilitar el análisis, los datos procesados fueron organizados en un formato tabular utilizando un DataFrame. Se estructuraron en filas donde cada una representa una combinación de país y año, asignándole el valor correspondiente del indicador. En los casos en los que no existió un valor disponible para una combinación específica, se dejó un espacio vacío para mantener la coherencia del conjunto de datos.

Gracias a este procesamiento, se logró una organización clara y accesible de los datos, lo que permitió su integración en el análisis posterior de manera eficiente.

4.2.4. Conclusión

El diseño de este proceso permitió estructurar los datos de manera eficiente, independientemente de la fuente de la que provinieran. Gracias a las funciones implementadas, se garantizó una extracción adecuada de la información, su procesamiento y su conversión a un formato tabular que facilitó el análisis. Esto aseguró que los datos estuvieran correctamente organizados por país y año, permitiendo realizar estudios y obtener conclusiones a partir de ellos.

4.3. Técnicas y herramientas.

Esta parte de la memoria tiene como objetivo presentar las técnicas metodológicas y las herramientas de desarrollo que se han utilizado para llevar a cabo el proyecto. Si se han estudiado diferentes alternativas de metodologías, herramientas, bibliotecas se puede hacer un resumen de los aspectos más destacados de cada alternativa, incluyendo comparativas entre las distintas opciones y una justificación de las elecciones realizadas. No se pretende que este apartado se convierta en un capítulo de un libro dedicado a cada una de las alternativas, sino comentar los aspectos más destacados de cada opción, con un repaso somero a los fundamentos esenciales y referencias bibliográficas para que el lector pueda ampliar su conocimiento sobre el tema.

Resultados

5.1. Resumen de resultados.

Breve resumen de los resultados. En caso de ser un trabajo muy experimental, los resultados completos pueden aparecer en su anexo correspondiente.

Debería haber una correspondencia entre los objetivos y los resultados explicados en esta sección

5.2. Discusión.

Discusión y análisis de los resultados obtenidos.

Conclusiones

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.

6.1. Aspectos relevantes.

Este apartado pretende recoger los aspectos más interesantes del **desarrollo del proyecto**, comentados por los autores del mismo.

Debe incluir los detalles más relevantes en cada fase del desarrollo, justificando los caminos tomados, especialmente aquellos que no sean triviales.

Puede ser el lugar más adecuado para documentar los aspectos más interesantes del proyecto y también los resultados negativos obtenidos por soluciones previas a la solución entregada.

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.

Lineas de trabajo futuras

Este capítulo debería ser informe crítico indicando cómo se puede mejorar el proyecto, o cómo se puede continuar trabajando en la línea del proyecto realizado.

Bibliografía

- [Ambientum, 2025] Ambientum (2025). Enfermedades emergentes: El rol crucial de la salud ambiental. https://www.ambientum.com/ambientum/cambio-climatico/enfermedades-emergentes-el-rol-crucial-de-la-salud-ambiental.asp?utm_source=chatgpt.com. Accedido: 2025-03-26.
- [Díaz Cordero, 2012] Díaz Cordero, G. (2012). El cambio climático. *Ciencia y sociedad*.
- [Hurtado et al., 2016] Hurtado, F., N Cárdenas, M. A., Cardenas, F., and León, L. A. (2016). La enfermedad de parkinson: Etiología, tratamientos y factores preventivos. *Universitas Psychologica*, 15(SPE5):1–26.
- [Marín et al., 2018] Marín, D. S., Carmona, H., Ibarra, M., and Gámez, M. (2018). Enfermedad de parkinson: fisiopatología, diagnóstico y tratamiento. *Revista de La Universidad Industrial de Santander. Salud*, 50(1):79–92.
- [Martinez, 2001] Martinez, B. B. (2001). Minería de datos. *Cómo hallar una aguja en un pajar. Ingenierías*, 14(53):53–66.
- [(PAHO), 2025] (PAHO), P. A. H. O. (2025). Cambio climático y salud. https://www.paho.org/es/temas/cambio-climatico-salud?utm_source=chatgpt.com. Accedido: 2025-03-26.
- [Poewe et al., 2017] Poewe, W., Seppi, K., Tanner, C. M., Halliday, G. M., Brundin, P., Volkman, J., Schrag, A.-E., and Lang, A. E. (2017). Parkinson disease. *Nature reviews Disease primers*, 3(1):1–21.
- [Ramos, 2014] Ramos, M. B. (2014). Biometeorología humana en la ciudad de punta alta.

- [Raul et al., 2016] Raul, A., Patil, A., Raheja, P., and Sawant, R. (2016). Knowledge discovery, analysis and prediction in healthcare using data mining and analytics. In *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, pages 475–478. IEEE.
- [RTVE, 2017] RTVE (2017). ¿cómo afecta el clima a nuestra salud?
- [Wikipedia, 2015] Wikipedia (2015). Latex — wikipedia, la enciclopedia libre. [Internet; descargado 30-septiembre-2015].
- [World Health Organization, 2025] World Health Organization (2025). Air pollution. https://www.who.int/health-topics/air-pollution#tab=tab_1. Accedido: 2025-03-26.
- [Zhang et al., 2018] Zhang, G., Xia, Y., Wan, F., Ma, K., Guo, X., Kou, L., Yin, S., Han, C., Liu, L., Huang, J., et al. (2018). New perspectives on roles of alpha-synuclein in parkinson’s disease. *Frontiers in aging neuroscience*, 10:370.