Inteligencia Artificial Explicable para estimar la depresión y esquizofrenia en pacientes basadas en datos de sensores de IoT

Explainable Artificial Intelligence to estimate depression and schizophrenia in patients based on IoT sensor data

Trabajo de Fin de Máster

Curso 2022-2023

Autora

Ana Jiménez Arévalo

Director

Iván García-Magariño García

Máster en Internet de las Cosas

Facultad de Informática

Universidad Complutense de Madrid

Inteligencia Artificial Explicable para estimar la depresión y esquizofrenia en pacientes basadas en datos de sensores de IoT

Explainable Artificial Intelligence to estimate depression and schizophrenia in patients based on IoT sensor data

Trabajo de Fin de Máster en Internet de las Cosas

Departamento de Ingeniería de Software e Inteligencia Artificial

Autora

Ana Jiménez Arévalo

Director

Iván García-Magariño García

Convocatoria: Junio/Julio 2023

Calificación: 6.2

Máster en Internet de las Cosas

Facultad de Informática

Universidad Complutense de Madrid

Agradecimientos

A mi familia, por el apoyo que me habéis dado durante todos estos años. Gracias por confiar en mi siempre, por regalarme siempre los mejores consejos, en los buenos y en los malos momentos, pero sobre todo, por quererme como lo hacéis día a día.

Resumen

Inteligencia Artificial Explicable para estimar la depresión y esquizofrenia en pacientes basadas en datos de sensores de IoT

Actualmente, la Inteligencia Artificial puede lograr distintas tareas como la toma de decisiones y la resolución de problemas. La Inteligencia Artificial Explicable (XAI) es un conjunto de procesos y métodos que permite a los humanos confiar y entender la salida de los algoritmos de aprendizaje automático.

En el presente trabajo se realizan distintos métodos de clasificación y predicción a partir de dos conjuntos de datos que fueron recogidos por sensores incorporados en un reloj actigráfico en pacientes con trastorno depresivo o esquizofrénico. Posteriormente, estos modelos son explicados mediante las dos librerías de XAI más conocidas, SHAP y LIME. Además, se realiza una comparación con otros artículos escritos a partir de estos mismos datasets.

Palabras clave

Inteligencia Artificial Explicable (XAI), depresión, esquizofrenia, actividad motora, actigrafía, series temporales, clasificación, predicción, SHAP y LIME.

Abstract

Explainable Artificial Intelligence to estimate depression and schizophrenia in patients based on IoT sensor data

Nowadays, Artificial Intelligence can accomplish different tasks such as making decisions and solving problems. Explainable Artificial Intelligence (XAI) is a set of processes and methods that allow humans to trust and understand the output of machine learning algorithms.

In the work, different classification and prediction methods are performed on two datasets that were collected by sensors incorporated in an actigraphic clock in patients with depressive or schizophrenic disorder. Subsequently, these models are explained using the two most well-known XAI libraries, SHAP and LIME. In addition, a comparison is made with other articles written from these same datasets.

Keywords

Explainable Artificial Intelligence (XAI), depression, schizophrenia, motor activity, actigraphy, time series, classification, prediction, SHAP and LIME.

Índice general

Índice de figuras

Capítulo 1

Introducción

Para comenzar, en este capítulo se expone una breve introducción y motivación del contenido del trabajo. Además se explican unos objetivos iniciales que se plantearon y que deben ser cumplidos al finalizar la lectura. Y por último, una descripción de la estructura de la memoria.

1.1. Motivación

El Internet de las Cosas describe la red de objetos que llevan incorporados sensores, software y otros mecanismos cuyo objetivo es conectarse entre si e intercambiarse información y datos con otros dispositivos y sistemas a través de Internet. En los últimos años se ha convertido en una de las tecnologías más importantes del siglo. Ahora, es posible conectar todo tipo de objetos a Internet a través de dispositivos integrados y, con esto, la comunicación se hace más sencilla. Dentro de estos instrumentos, cabe destacar los sensores portátiles, que permiten la monitorización ambulatoria de diversos datos y, que cada vez son más comunes dentro de muchos ámbitos, como la salud. Al recolectar información a largo plazo, pueden adelantar el diagnóstico de una enfermedad. Hoy en día son muy usados en el campo de la salud mental. Pueden aportar información sobre distintas variables fisiológicas, como la presión arterial, la frecuencia cardíaca o la actividad física.

La depresión o trastorno depresivo mayor es una enfermedad emocional que causa sentimiento de tristeza constante y una pérdida de interés a la hora de realizar ciertas actividades. Afecta los sentimientos, pensamientos y comportamientos de una persona. Por otro lado, la esquizofrenia es una enfermedad mental más grave que afecta la forma en que el sujeto piensa, siente y se comporta. Estos pacientes pueden parecer que han

perdido el contacto con la realidad.

Este trabajo ha sido realizado para dar a conocer a los lectores el concepto de Inteligencia Artificial Explicable. Hoy en día no es muy conocido, sin embargo, es de gran utilidad para la realización de modelos de machine learning y, métodos de clasificación y predicción, ya que aporta explicaciones a los resultados obtenidos y ayuda a entenderlos. A su vez, también se trabaja con dos enfermedades comunes e importantes en el mundo y se dan a conocer un poco más a los lectores. Y, por último, quise encontrar y comparar distintos métodos para aplicar a un mismo conjunto de datos y con ellos poder llegar a unas conclusiones similares. Con la descripción de los dos artículos, se pretende mostrar esta faceta de la Inteligencia Artificial.

1.2. Breve descripción del trabajo

Los sensores están incorporados en unos relojes actigráficos que llevan pacientes con episodio depresivo mayor, tanto monopolar como bipolar, y sujetos con esquizofrenia, tanto paranoide como no. En concreto, 23 pacientes deprimidos, 32 controles sanos y 22 personas con esquizofrenia llevaron el reloj aproximadamente durante dos semanas (más o menos días dependiendo del paciente), y se estuvo midiendo la actividad motora por cada minuto a lo largo de ese tiempo. En especial, los valores recogidos de la actividad motora se han empleado como una serie temporal, ya que estas se definen como sucesiones de datos medidos en determinados momentos ordenados cronológicamente. Tratándolos de tal forma, ha sido posible diferenciar, hacer filtrados, buscar a partir de fechas concretas y, lo más importante, estudiar el comportamiento futuro. A partir de un día de un paciente de cada enfermedad se han intentado predecir, con la actividad motora por cada minuto, las 24 horas siguientes y, posteriormente, comparar con los valores iniciales.

Los datos iniciales de estas dos enfermedades han sido algo modificados. Lo primero, es que los sujetos fueron divididos por cada día de tratamiento, de manera que para la depresión se pasó de tener 23 observaciones a 291, para la esquizofrenia, de 22 a 285 observaciones y, las de los controles sanos en lugar de las 32 iniciales, se ha trabajado con 402. Por otro lado, las variables de cada uno de ellos también sufrieron algún cambio que será explicado de manera más exhaustiva en la sección correspondiente dentro del Capítulo 3.

La Inteligencia Artificial, definida como una simulación de la inteligencia humana que crea algoritmos y sistemas informáticos capaces de ejecutar tareas simples y complejas, permite que se hayan realizado a lo largo del trabajo distintos métodos de clasificación y de predicción para ver si se podía distinguir, a partir de distintas

propiedades de un paciente o de únicamente la actividad motora, el tipo de trastorno que padecía. Una vez realizados estos modelos, pueden surgir cuestiones como ¿cuál de las propiedades de los pacientes influye más en el tipo de enfermedad que este tiene?, ¿se tienen los mismos resultados para cada persona? o, ¿qué pasa dentro de un modelo? La mayoría de ellas son posibles resolverlas con Inteligencia Artificial Explicable (XAI), un conjunto de técnicas, procesos y estrategias que aporta explicaciones para las predicciones, recomendaciones y decisiones de sistemas inteligentes.

Hay muchas formas de clasificar la XAI y distintos métodos para aplicarla, pero en este trabajo se usan SHAP y LIME. SHAP usa cálculos del campo de la teoría de juegos para averiguar qué variables tienen más influencia en las predicciones de las técnicas de machine learning. Por otro lado, LIME utiliza el método de la caja negra para determinar el mínimo número de variables que generan la máxima probabilidad de acierto, y así explicar el por qué fue clasificada una instancia determinada con la clase asignada.

La cantidad de estudios, comparaciones y búsquedas que se pueden hacer a partir de un mismo conjunto de datos es inmensa. Este es otro beneficio que nos aporta el aprendizaje automático y la Inteligencia Artificial, la posibilidad de llegar a unos mismos resultados a partir de diversos métodos y, comparar cada uno de ellos. Se seleccionaron dos trabajos para hablar de ellos y ver otras posibles formas de estudiar estos datos. Uno procedente de *Google Scholar*, que hace una comparación entre los tres tipos de pacientes a partir del cálculo de tres medidas no paramétricas, la estabilidad interdiaria, la variabilidad intradiaria y la amplitud relativa. El otro, de *ScienceDirect*, lo hace mediante la representación de UMAPs, un método no supervisado y no lineal de reducción de la dimensionalidad.

1.3. Objetivos

El principal objetivo a la hora de proceder con este trabajo fue intentar encontrar un patrón en la actividad motora de cada uno de los sujetos. Es decir, ver relaciones a ciertas horas del día en distintos pacientes de una misma enfermedad y, diferenciar un paciente con trastorno esquizofrénico de uno con trastorno depresivo únicamente por su actividad motora. Además, también se buscó aprender y entender los modelos mediante Inteligencia Artificial Explicable. Al ser esta una rama que cada vez es más conocida pero no existe todavía mucha información, se pretendió conocer a detalle los dos métodos (SHAP y LIME) y qué aporta cada uno. Y, por último, ver y aprender diversos estudios que se pueden hacer sobre un mismo conjunto de datos.

1.4. Estructura del trabajo

La estructura de la memoria está dividida en cuatro partes. La primera, formada por el Capítulo 2, detalla el marco teórico necesario para comprender de forma correcta todo el trabajo. Explica de manera breve, qué es la Inteligencia Artificial Explicable, así como la actigrafía y las series temporales.

La segunda parte, formada por el Capítulo 3 y el Capítulo 4, contiene la parte principal. Aclara los dos trastornos de los que se habla, la procedencia de los datos con los que se trabaja y la creación de los nuevos. Además, también se presentan los métodos de clasificación, que para el trastorno depresivo diferencia entre bipolar o monopolar, y para el esquizofrénico si la enfermedad es paranoide o no. Como parte final del Capítulo 3, se hacen cuatro comparaciones entre la variedad de los pacientes a partir de la actividad motora media por cada hora del día. Por último, la aplicación de la XAI a estos modelos. A las primeras clasificaciones se les aplica tanto SHAP como LIME, dos librerías muy útiles y conocidas que permiten detallar las variables más importantes e influyentes del modelo, así como una explicación de los resultados obtenidos.

El Capítulo 5 compone la tercera parte. En esta, se realiza un método de predicción para cada conjunto de datos de las dos enfermedades. Tomando estos como series temporales, ya que miden la actividad motora por cada minuto, se estudia el comportamiento futuro y una posible predicción de días posteriores.

El último integrante es el Capítulo **??**, en el que se hace una pequeña comparación con otros artículos. En esta parte, se detalla como se pueden hacer distintos estudios a partir de un mismo dataset, pero a su vez llegando a una conclusión similar.

Para terminar, está escrito un breve resumen de las conclusiones extraídas del trabajo así como un pequeño plan de futuro.

Capítulo 2

Estado del arte

2.1. Inteligencia Artificial Explicable aplicada al IoT

A lo largo de este trabajo se va a hacer referencia al Internet de las Cosas (IoT) y a la Inteligencia Artificial Explicable (XAI). Antes de aclarar estos dos conceptos importantes, es necesario saber qué es la Inteligencia Artificial. La IA está cada vez más presente en nuestras vidas. Se podría definir como la combinación de algoritmos que intentan simular acciones de los humanos o, a su vez, ir más allá de nosotros, de la inteligencia humana. Se aplica a cualquier campo o sector como la salud, las finanzas, el transporte o la educación y, sin saberlo, la usamos siempre en nuestro día a día. La detección facial de los móviles, los asistentes virtuales de voz, los *bots* o algunas aplicaciones móviles son algunos ejemplos. El objetivo principal de todos ellos es hacer más fácil la vida de las personas.

Se pueden reconocer distintos tipos de Inteligencia Artificial como los sistemas que piensan como humanos o que actúan como ellos. Los primeros son capaces de automatizar actividades como la toma de decisiones, la resolución de problemas y el aprendizaje. Un ejemplo de estos son las redes neuronales artificiales. Los segundos, sin embargo, son computadores que realizan tareas de forma similar a la de las personas, por ejemplo los robots. Hace unos años estos se podían ver como una idea prácticamente imaginaria que en algún momento podría aparecer en nuestras vidas. Actualmente, es posible encontrarse con estos sistemas en un restaurante, tienda, o cualquier lugar accesible. También existen sistemas que piensan racionalmente. Estos intentan emular el pensamiento lógico racional de los humanos, es decir, se investiga la manera de lograr que las máquinas puedan percibir o razonar. Por último, aquellos sistemas que actúan racionalmente imitando de manera lógica el comportamiento

humano.

El uso del Big Data para trabajar con grandes cantidades de datos ayuda a proporcionar ventajas comunicacionales, comerciales y empresariales [23].

2.1.1. Internet de las Cosas

El término Internet de las Cosas surgió en el año 1999 por Kevin Ashton, un profesor británico del MIT, que estaba trabajando en el campo de la tecnología de la identificación por radiofrecuencia. Sin embargo, los primeros conceptos sobre la creación de una red de dispositivos inteligentes se discutieron en 1982. La evolución y el concepto de IoT se relacionaron con la incorporación de sensores y conexiones en cualquier tipo de dispositivo que pudiera admitirlo, incorporando así inteligencia a estos. Fue con la entrada del año milenio cuando ya comenzó su investigación en el ámbito académico, y el desarrollo y aparición de nuevos dispositivos inteligentes fue aumentando con el tiempo. Tan solo en 9 años la cantidad de elementos conectados a la red ya superaba la población mundial (aproximadamente 12.500 millones). Desde entonces, el crecimiento de Internet de las Cosas es exponencial y la tecnología cada vez seguirá progresando más.

Una vez contada su historia, se podría definir el Internet de Cosas como la interconexión en red de todos los objetos cotidianos que, a menudo, están equipados con algún tipo de inteligencia [21]. Estos objetos o dispositivos están conectados entre sí mediante redes cableadas o inalámbricas y se intercambian distintos datos. Varias conexiones unidas forman un ecosistema de IoT que puede ser de menor (dentro de una empresa) o mayor (en comunidades cerradas o ciudades) escala. Para poder llegar a entender mejor el concepto, se puede dividir en tres capas: la capa física, formada por los artilugios, sensores y controladores, es decir, cualquier aparato que se pueda conectar a una red cableada o inalámbrica puede formar parte; la capa informática, referida al almacenamiento y procesamientos de datos; y, la capa de aplicación que se trata del conjunto de servicios integrados a la nube para obtener información. Existen tres disciplinas sin las cuales no habría sido posible el progreso: el Big Data, el Cloud Computing y la Inteligencia Artificial.

2.1.2. Inteligencia Artificial Explicable

A medida que la Inteligencia Artificial avanza y está más integrada en la vida de las personas, cada vez es más importante comprender cómo y por qué se toman ciertas decisiones. Sin embargo, este proceso no es fácil. Además, el crecimiento de la IA provoca que los modelos de machine learning sean más complejos e incomprensibles.

Estos potentes modelos, difíciles de entender, se denominan 'caja negra'. Es por esto que el concepto de **Inteligencia Artificial Explicable** (XAI) atrae la atención de los investigadores para crear sistemas de IA explicables e interpretables. Esta es una rama de la IA que aporta explicaciones para las predicciones, recomendaciones y decisiones de sistemas inteligentes. La XAI es un conjunto de técnicas, procesos y estrategias que las bibliotecas deberían adoptar y defender para garantizar que el machine learning esté al servicio. Además, ayuda a comprender los modelos complejos de 'caja negra' aportando cierta claridad, de manera que se muestran correlaciones entre variables. Así, no se obtiene únicamente el resultado de un modelo aplicado a un conjunto de datos sin saber lo que ocurre en el trasfondo, si no que, se resuelven las dudas que cualquier desarrollador o lector puede tener, como ¿por qué el porcentaje de predicción es ese? o ¿qué influencia tiene cada variable en el modelo?

La dificultad de entender los modelos de investigación científica no es un concepto que haya surgido en los últimos años, en 1970 ya comenzaron estas dudas. Sin embargo, con el avance de la IA es cuando, en 2016, la interpretación y explicabilidad empezaron a ser más conocidas. También está presente en el mundo de la programación, el concepto de IML (*Interpretable Machine Learning*), un subconjunto de XAI que se centra en la lógica de los algoritmos. Se podría decir que este nuevo término se centra más en la interpretabilidad que en la explicabilidad como la XAI. La explicabilidad se define como la capacidad de proporcionar información sobre el funcionamiento interno del modelo y, la interpretabilidad como la capacidad de ampliar un modelo o sus predicciones comprensibles por el ser humano. Además, otros conceptos importantes son la comprensibilidad, definida como la capacidad de expresar la información aprendida para hacerla comprensible al ser humano y, la fidelidad, expresada como la capacidad de ser coherente de elegir las características verdaderamente relevantes.

La clasificación de la Inteligencia Artificial Explicable se puede hacer según tres formas distintas [11]. Teniendo en cuenta el cuándo y cómo se produce la explicación del modelo se distingue la clasificación ante-hoc y post-hoc. La explicación de un modelo se puede hacer antes, durante o después del entrenamiento. Los métodos ante-hoc interpretan externamente antes de la fase de entrenamiento o internamente durante ella y así, al acabarla, el modelo ya es explicable. Los post-hoc sin embargo, se aplican externamente tras el entrenamiento. Otra clasificación surge dependiendo de la clase del modelo al que se le aplica la XAI, ya sean explicaciones específicas o agnósticas. Los métodos específicos de modelo suelen diseñarse para un tipo de modelo como las redes neuronales profundas, pero tienen un inconveniente, ya que existe una limitación al determinar el modelo cuando la necesidad del tipo particular es explicación. Los métodos agnósticos pueden aplicarse a cualquier tipo de modelo sin limitar las clases de este. Por último, según el alcance del modelo, las explicaciones

se pueden hacer local o globalmente. Las primeras describen por qué y cómo pueden generarse determinadas predicciones a nivel local, mientras que el objetivo de las segundas es caracterizar el modelo en su totalidad.

Al estar en pleno crecimiento, existen muchos tipos de métodos para aplicar XAI. Se pueden usar para conseguir una explicación visual (como el *Partial Dependence Plot* o el *Individual Conditional Explanation*), para explicar las características del modelo (como *SHAP* o *LIME*), aplicar ejemplos (como el *Contrastive Explanation Method*), para basarse en perturbaciones, es decir, explican el modelo de la caja negra probando iterativamente con diferentes entradas (como por ejemplo el *Random Input Sampling for Explanations*), o basarse en una retropropagación, ya que se centran en el flujo de información (por ejemplo *Saliency Maps*).

Tres grandes beneficios que aporta la Inteligencia Artificial Explicable son:

- Genera confianza en la Inteligencia Artificial garantizando la interpretabilidad y la explicabilidad de sus modelos. Además, simplifica el proceso de evaluación e incrementa la transparencia de estos.
- Optimiza los resultados evaluando de forma continua y mejorando el rendimiento de los modelos.
- Gestiona los requisitos normativos, de cumplimineto y de riesgos.

Antes de comenzar con la parte práctica del trabajo y con la información del dataset, se presenta una breve explicación las dos grandes librerías que se han usado para aplicar toda la información de esta sección. Se han aplicado SHAP y LIME, dos métodos de explicación de características del modelo.

SHAP

Lloyd Shapley fue un matemático y economista que introdujo el **valor de Shapley** en 1953. Este es un método de distribución de riquezas usado en la teoría de juegos cooperativos. Para cada uno de estos juegos, se asigna un único reparto del beneficio generado entre todos los jugadores que participen. Dado que algunos jugadores pueden contribuir más o menos a la coalición, dependiendo de ciertas características, el valor de Shapley aporta una respuesta a la pregunta: *¿qué importancia tiene cada jugador para la cooperación global?*

SHAP (SHapley Additive exPlanations) es una de las librerías de explicabilidad de modelos más conocida y utilizada. Usa cálculos del campo de la teoría de juegos para averiguar qué variables tienen más influencia en las predicciones de las técnicas

de machine learning. Conecta la asignación óptima de créditos con las explicaciones locales utilizando los valores de Shapley y sus extensiones relacionadas. En este caso, los jugadores son las variables empleadas en el modelo. Ayuda a saber cuáles de ellas son más influyentes analizando el impacto de cada una. A partir de estos valores, la explicación se representa como la suma de las contribuciones. Es posible instalarla a partir de PyPI o con conda-forge como indica la página web [13].

Los resultados se representaron mediante tres gráficas distintas.

- El beeswarm clasifica las características por la suma de las magnitudes de los valores Shap en todas las muestras y los usa para mostrar el impacto que tiene cada una en la salida del modelo.

- El gráfico de barras muestra la magnitud con la que una variable impacta sobre la predicción.

- El diagrama de fuerza se usa cuando se quiere describir el de una instancia específica. Así, es posible explicar a un paciente cómo llegó su modelo a la predicción que hizo.

LIME

LIME (Local Interpretable Model-Agnostic Explanations) es otro método para aplicar Inteligencia Artificial Explicable. Es un marco de referencia que explica el por qué un modelo predice de la forma en que lo hace. Utiliza la metáfora de la caja negra, en donde las variables de entrada se encuentran a un lado y las de salida en el otro, sin que se tenga acceso a la caja per-se. Actualmente, LIME admite explicaciones para modelos tabulares, clasificadores de texto y clasificadores de imágenes. Es posible instalarlo en Python o a partir del propio terminal [20].

Su principal objetivo es determinar el mínimo número de variables que generan la máxima probabilidad de acierto y con esto explicar el por qué fue clasificada una instancia determinada con la clase asignada. Lo hace creando datos *fake* cercanos para cada una de estas instancias. A partir de estos, calcula la distancia entre estos datos creados y los originales, y se hacen predicciones a estos nuevos datos.

La gráfica de esta biblioteca está dividida en tres partes distintas. La primera de ellas muestra los porcentajes de predicción de la observación en concreto, la segunda la influencia de cada variable tanto positiva como negativamente y, la tercera y última, los valores reales de cada característica dentro de esta instancia.

La aplicación de la Inteligencia Artificial Explicable sobre Internet de las Cosas es muy amplia. En sistemas autónomos y robots, la IA aporta la habilidad de hacer tareas que los humanos son incapaces o, actividades que necesitan rapidez y destreza como pintar, planchar y soldar. Resultados experimentales demuestran que las explicaciones de los robots pueden mejorar la confianza y transparencia sobre ellos. Los sistemas de decisión energética basados en IA se utilizan en ciudades inteligentes, redes inteligentes y aplicaciones domésticas inteligentes. La XAI ayuda a interpretar ciertas predicciones críticas sobre el uso de la energía y las emisiones. También se aplica a otros servicios dentro de las *smart cities* como la monitorización del aire, la gestión del agua o el control de la temperatura. Al ser aspectos de la vida cotidiana y que afectan al ser humano, son necesarios modelos interpretables, que ayuden a evitar peligros. Por ejemplo, se han creado a lo largo de los años varias plataformas para la investigación que destacan las capacidades en Inteligencia Artificial Explicable para el procesamiento de eventos en un entorno de terreno urbano denso [11]. Estos y muchos más sectores como sistemas financieros, sanidad, el ámbito industrial, la seguridad y privacidad y, la agricultura necesitan algoritmos de Inteligencia Artificial Explicable dentro de sus dispositivos o servicios de IoT para poder tanto explicar los hechos, como evitar problemas futuros.

Los datos recolectados por sensores de IoT contienen información incompleta o, en ocasiones, errónea que puede provocar resoluciones inexactas. La XAI ayuda a descubrir y reparar cualquier fallo o punto débil inesperado, garantizando así un proceso de toma de decisiones fluido. Permite descubrir hechos e información ocultos, así como nuevas perspectivas sobre problemas de machine learning. Si se puede interpretar el algoritmo, se puede descubrir información adicional y patrones que no se habían detectado antes.

2.2. Actigrafía

Los sensores son herramientas que responden a información del entorno físico. En la actualidad, la investigación biomédica utiliza muchos tipos de sensores para detectar procesos biológicos, químicos o físicos y, que luego mandan esta información a profesionales o usuarios individuales. Si estos sensores están integrados en algún dispositivo o tecnología, se denominan sensores portátiles. Estos están muy integrados en el día a día de las personas, ya que los datos que crean y/o controlan son esenciales; como por ejemplo pasos diarios, calorías quemadas, registros continuos de la frecuencia cardíaca o niveles de actividad. Ayudan a los investigadores a prevenir casos graves de salud como derrame cerebral o ataque cardíaco y casos de salud mental.

La actigrafía es un método que permite, mediante la colocación de un sensor

(actígrafo), saber el nivel de actividad del cuerpo, en general para medir la cantidad de sueño de una persona [7]. Este dispositivo, llamado actígrafo, está formado por un acelerómetro y una memoria donde se guardan las medidas de actividad. La actigrafía se usa en la investigación de ciertos parámetros del sueño, así como en estudios cronobiológicos y en cuantificaciones de la actividad física de las personas. Cuando se observa el sueño del paciente, se valora la latencia de este, el tiempo total de sueño y su eficiencia. A la hora de hacer un estudio cronobiológico, se establece el ritmo circadiano de una persona para poder observar si lo tiene bien estructurado y regular. Por último, la medición de la actividad física permite saber la cantidad de movimiento de un sujeto.

En los datos que se recogieron, tanto para el estudio del trastorno depresivo como de la esquizofrenia y, el de los controles sanos, el actígrafo usado fue un reloj, en concreto un modelo llamado AW4, desarrollado por Cambridge Neurotechnology Ltd.

Reloj actigrafo AW4

El AW4 es un acelerómetro de muñeca con dimensiones 37x29x10 mm que pesa 16g. Esta tecnología detecta la amplitud máxima de la aceleración del movimiento y genera un señal transitoria proporcional a la tasa de aceleración. Los recuentos de la actividad se realizan mediante algoritmos matemáticos de integración seleccionándose el recuento máximo para cada segundo individual. Estos recuentos máximos se recogen durante intervalos de tiempo específico que oscila entre 2 segundos y 15 minutos. La frecuencia de muestreo es de 32 Hz y se registran los movimientos superiores a 0,05g. Se produce un voltaje correspondiente que se almacena como recuento de actividad en la unidad de memoria del reloj actigráfico. El número de recuentos es proporcional a la intensidad del movimiento [9].

El MotionWatch es la nueva generación de *CamNtech*, inventores del Actiwatch 4 y 7. El MotionWatch 8 que se muestra en la Figura 2.1, con un acelerómetro digital triaxial, gráficos de actividad y un software especializado, se utiliza también para cuantificar la intensidad y duración de la actividad física diaria.

Alguna de las características de este actígrafo de muñeca son, que es resistente al agua, tiene una transferencia de datos mediantes USB, un sensor de luz y marcador de eventos incluidos de serie, batería sustituible por el usuario y que registra hasta 180 días con una época de un minuto. Por último, el software incluye la función NPCRA, donde todos los datos sin procesar y analizados pueden exportarse a programas de terceros.

Figura 2.1: MotionWatch 8

2.3. Series temporales

Una **serie temporal** es una secuencia de N observaciones ordenadas y equidistantes cronológicamente sobre una o varias características.

Una serie temporal univariante (o escalar) se representa de la forma

$$y_1, y_2, ..., y_N; (y_t)_{t=1}^{N}; (y_t : t = 1, ..., N),$$

donde y_t es la observación $t(1 \leq t \leq N)$ de la serie y N es la longitud de la serie.

Una serie temporal multivariante (o vectorial), se escribe como

$$y_1, y_2, ..., y_N; (y_t)_{t=1}^{N}; (y_t : t = 1, ..., N),$$

donde $y_t = [y_{t1}, y_{t2}, ..., y_{tM}]' (M \geq 2)$ es la observación $t, (1 \leq t \leq N)$ de la serie, N es el número de observaciones y M las características.

A partir de este momento se hablará de series temporales univariantes, ya que es el caso de este trabajo. Los valores de una serie temporal van ligados a instantes de tiempo (con periodos desde horarios hasta anuales), de manera que su uso implica un estudio de dos variables, la previamente dicha como característica principal y el tiempo. El análisis de series temporales presenta un conjunto de técnicas estadísticas que permiten estudiar y modelizar el comportamiento de un fenómeno que evoluciona a lo largo del tiempo, así como realizar previsiones de los valores futuros [6].

La forma más sencilla de comenzar con un estudio práctico de una serie temporal es mediante una representación gráfica. Con esta, es posible conocer ciertas carac-

terísticas importantes como su tendencia, la existencia de ciclos o la presencia de valores atípicos. El gráfico muestra el paso del tiempo periódicamente en el eje x y los distintos valores de la variable característica en el eje y. Otra forma es un diagrama de cajas en el que se podrían ver los valores atípicos y así poder estudiarlos, descartarlos o aprender de ellos.

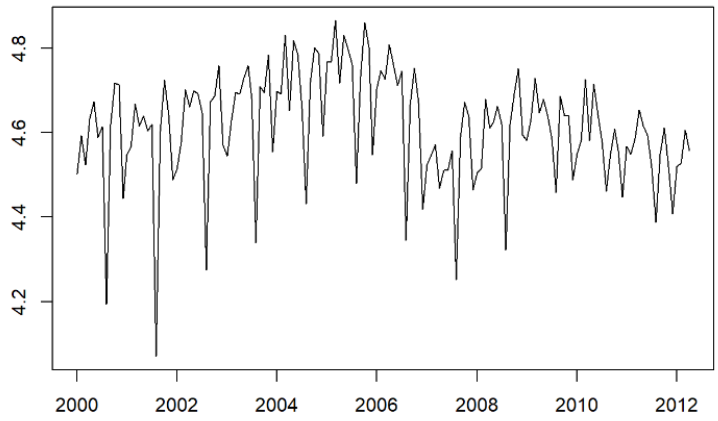

Figura 2.2: Serie temporal (índice de producción industrial de Cantabria.Log.)

La componente más relevante de una serie es la tendencia, ya que en función de si la presenta o no, se podrá clasificar en serie estacionaria o no. La tendencia es una propiedad determinista definida como un cambio a largo plazo que se produce respecto a la media. Las **series estacionarias** son aquellas en las que se observa un patrón periódico. Estas son estables, es decir, la media y la variabilidad son constantes a lo largo del tiempo. Son fáciles de estimar y predecir un nuevo valor. Cuando esto no ocurre, se habla de una **serie no estacionaria**, donde la media y/o la variabilidad si que cambian en el tiempo. Una vez estudiada la estacionalidad y haberla eliminado, pueden persistir unos valores que son aleatorios.

Otra cuestión a determinar a la hora de estudiar una serie temporal es si existe dependencia entre la variabilidad y el nivel. La variabilidad es la dispersión de los valores de una variable, y el nivel, es una medida de tendencia central como la media, por ejemplo. Si no dependen una de la otra, significa que el incremento debido a la estacionalidad siempre es el mismo, aunque exista tendencia creciente o decreciente. Por el contrario, si existe esta dependencia, los elementos de la serie se combinan de forma multiplicativa, es decir, el incremento debido a la estacionalidad aumenta o disminuye conforme a la tendencia.

2.4. Trabajos relacionados

En esta última sección del capítulo se hablará sobre dos artículos que usan los mismos datasets pero aplican dos métodos distintos. Las alteraciones dentro de la actividad motora en personas con depresión y esquizofrenia son reales, sin embargo, pocas veces se han estudiado objetivamente y obtenido explicaciones de por qué. Es posible hacer distintas comparaciones, medidas, métodos de clasificación y predicción para los diversos objetivos de cada estudio. El objetivo de esta sección es demostrar que a partir de las diversas opciones de un estudio de aprendizaje automático, es posible llegar a un mismo resultado con una buena conclusión.

El primero [3], titulado '*Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia then in major depression*', es un breve informe cuyo propósito era estudiar la complejidad de los patrones de actividad motora en estos pacientes utilizando la actigrafía.

Como ya sabemos, esta medida se registró mediante un reloj de muñeca durante periodos de entre una y dos semanas para pacientes con trastorno depresivo, esquizofrénico y, personas sanas. En el artículo se calcularon tres variables no paramétricas a lo largo de todo el conjunto de días de cada paciente. Estas son la estabilidad interdiaria (IS), la variabilidad intradiaria (IV) y la amplitud relativa (RA) y, mediante ellas, hace comparaciones para cada uno de los sujetos. La primera cuantifica la variabilidad entre los días, es decir, la fuerza de acoplamiento del ritmo a factores ambientales supuestamente estables. La segunda indica la fragmentación del ritmo, es decir, la frecuencia y el alcance de las transiciones entre el reposo y la actividad. Y por último, la tercera se calcula a partir de los datos del periodo más activo de 10 horas y del periodo menos activo de 5 horas diarias. Los resultados para cada conjunto de datos son los recogidos en la Tabla 2.1, procedentes del propio informe.

	Control	Esquizofrenia	Depresión
Estabilidad interdiaria	0,446 ± 0,113	0,526 ± 0,154	0,428 ± 0,129
Variabilidad intradiaria	0,901 ± 0,168	0,742 ± 0,190	0,825 ± 0,282
Amplitud relativa	0,837 ± 0,118	0,801 ± 0,141	0,836 ± 0,121

Cuadro 2.1: Resultados del cálculo de las variables no paramétricas

De la misma forma que en este trabajo, se dividió la actividad motora en intervalos de 24 horas. Comparando con las personas sanas, tanto los pacientes con depresión como los de esquizofrenia, mostraron reducciones en la actividad total diaria. Además, los sujetos con trastorno esquizofrénico, teniendo en cuenta únicamente la actividad nocturna, tuvieron una reducción más pronunciada respecto a los depresivos. Los va-

lores de la IS, que oscilan entre 0 (mínima regularidad) y 1 (máxima regularidad), son un 18 % más altos teniendo esquizofrenia que ninguno de los dos, y un 4 % más bajo para los que tienen depresión. Esto quiere decir que los pacientes con trastorno esquizofrénico tienen algo más de regularidad del ritmo diario que el resto. La variabilidad intradiaria muestra que las personas con más fragmentación del ritmo son los controles, ya que los valores de esta variable se mueven entre 0 (mínima fragmentación, por lo que se ajusta a una onda cosenoidal) y 2 (máxima fragmentación).

Dentro del conjunto esquizofrénico, es necesario hacer referencia a los medicamentos. Se obtuvieron distintos resultados para aquellos que ingirieron clozapina a otro antipsicótico. Como se pudo observar en la parte de Inteligencia Artificial Explicable de la clasificación de las personas con este trastorno, paranoide o no paranoide, el más influyente era el neuroléptico. A pesar de esto, nueve pacientes tomaron clozapina, el más conocido, por lo que el artículo hace referencia a las tres medidas de las variables no paramétricas para esta característica. Se llegó a la conclusión de que el grupo de clozapina tuvo una reducción en la actividad motora nocturna mucho más pronunciada que el otro grupo. Además, muestran un aumento de la estabilidad interdiaria (38 %) y, una reducción de la variabilidad intradiaria (24 %) respecto a los pacientes sanos. Al igual que en los resultados obtenidos a la hora de la predicción, el autor del informe comentó que hay alguna comparación insignificante dentro de los sujetos con esquizofrenia paranoide y no paranoide, ya que estos últimos únicamente eran 5 de los 23 pacientes totales.

Como se pudo ver en este trabajo, existen claras diferencias entre los dos tipos de pacientes. En la Figura 3.1.3, el paciente con más alteraciones dentro del trastorno depresivo llega al máximo peso de la actividad motora, siendo este aproximadamente 1700. Sin embargo, la Figura 3.5 muestra que este valor es 1200. En el artículo, llega también a esta conclusión, a que los pacientes con trastorno depresivo muestran un ritmo menos estructurado, con más altibajos. Esto es compatible con los resultados de un incremento de la IS y una reducción de la IV dentro del otro conjunto (esquizofrenia). "*Los pacientes esquizofrénicos mostraban un patrón de actividad caracterizado tanto por una baja actividad total como por una menor actividad nocturna, y además un ciclo de actividad de descanso más regular*"[3]. En ciertos momentos del artículo, estas diferencias de más o menos nivel de la actividad lo justifica por la hospitalización o no de los pacientes, y por su trabajo. Las personas sanas estaban a lo largo del estudio trabajando y realizando sus actividades diarias normales, mientras que los pacientes no. Esto, claramente, influye en unos valores más altos para los *controls* (personas que no tienen ni trastorno depresivo ni esquizofrenia). También existe una diferencia a la hora estar ingresado o no en un hospital.

La suma de la actividad de cada grupo dependiendo de las distintas horas del día

se muestran en la Tabla 2.2. Las horas seleccionadas para el estudio, que son las diez intermedias, coinciden con el rango de más actividad.

Hora del día	Control	Esquizofrenia	Depresión
00:00-06:00	14799	12062	12687
06:00-12:00	119528	67122	74223
12:00-18:00	164809	84698	112324
18:00-24:00	118102	48985	75582

Cuadro 2.2: Actividad motora media a cada hora del día

Para concluir con los resultados y comparaciones de este primer artículo, comentar que los propios autores de este afirman la dificultad de distinguir entre paciente depresivo, esquizofrénico o sano a partir de la actividad motora. "*Hasta donde sabemos, este es el primer estudio basado en actigrafía que aborda las diferencias en el patrón de actividad motora entre pacientes con esquizofrenia y pacientes con depresión*". Por ello, el presente trabajo se centró más en una clasificación dentro de cada una de las enfermedades y posteriormente una posible predicción, en lugar de establecer una mayor comparación entre los tres tipos que se distinguen.

El segundo artículo, procedente de *ScienceDirect* y que se titula '*An unsupervised machine learning approach using passive movement data to understand depression and schizophrenia*' presenta un trabajo algo diferente ya que se centra en la representación de varios UMAPs (Uniform Manifold Approximation and Projection).

El UMAP es un método no supervisado y no lineal de reducción de la dimensionalidad capaz de establecer y organizar conglomerados informativos a partir de datos de alta dimensión [18]. Mediante este, los datos actigráficos por cada minuto (divididos en semanas) se redujeron a dos dimensiones, se visualizaron y se etiquetaron en función de la enfermedad para ver las relaciones entre la actividad, así como las tendencias de cada grupo. Al usar este método, hay que fijarse en la distancia relativa entre los puntos, ya que las proyecciones no pueden interpretarse directamente. La distancia euclídea entre cada punto representa el patrón de movimiento de un individuo dentro de un diagnóstico. Por último, los datos se volvieron a reducir a una única dimensión para observar e interrogar la influencia relativa del comportamiento de movimiento mediante SHAP. Sin embargo, en este caso, este método de XAI se usó para identificar qué horas del día eran más influyentes en el movimiento semanal de los pacientes. Al expresar la variación del espacio bidimensional en una sola dimensión, se pudo construir los valores de forma única por persona y por característica para representar cómo el modelo informaba el espacio latente. Estos valores se visualizaron en

intervalos promediados de dos horas frente a los datos actigráficos del individuo para asignar las características influyentes a los patrones de movimiento correspondientes [18].

El informe comienza afirmando que desde hace mucho tiempo ha habido perturbaciones en el movimiento de personas tanto con un trastorno depresivo mayor como con esquizofrenia. Una lenta psicomotricidad es uno de los síntomas de estas enfermedades, y muchas investigaciones han demostrado que estos dos grupos de pacientes difieren, respecto a la actividad motora, de personas sin ningún trastorno. Además, relacionándolo con la actigrafía, esta es una herramienta muy usada y útil dentro de la medicina para poder caracterizar alteraciones en el movimiento y a la hora de dormir. Uno de los principales objetivos de los autores era determinar si un método de machine learning no supervisado podría diferenciar entre un trastorno depresivo mayor, un trastorno esquizofrénico o, un control sin desorden. Por otro lado, una vez encontrada la validez del método, se buscaba caracterizar los distintos fenotipos de movimiento asociados a los tres casos. Y por último, al igual que se ha hecho en este trabajo, mediante Inteligencia Artificial Explicable, entender los patrones del movimiento en sujetos con depresión y esquizofrenia.

En la Figura 2.3 se observa el UMAP de los grupos de datos, control, esquizofrenia y depresión. Cada punto de la gráfica muestra un individuo, siendo sus coordenadas bidimensionales (X_i, Y_i) la representación del movimiento de ese sujeto durante una semana. Cuanto más cerca estén los puntos entre si, más parecida será la actividad motora. Además de detectar la similitud de movimiento dentro de los sujetos de cada grupo, con estas gráficas también se pueden ver los participantes con actividad parecida, independientemente de sus diagnósticos. En una imagen donde se muestra el conjunto de individuos total, se puede ver como los de una misma enfermedad tienden a agruparse, especialmente los clasificados con esquizofrenia y como sanos. Esto complica la posibilidad de diferenciar unos de los otros. Los pacientes depresivos tienden a rodear a los otros dos y, situarse más en la periferia. Para confirmar que la interpretación del UMAP era correcta, se calculó la distancia euclídea media:

	Distancia euclídea
Control	1,41 ± 0,70
Esquizofrenia	1,56 ± 0,83
Depresión	2,03 ± 1,02

Cuadro 2.3: Distancia (media ± desviación) entre los individuos de cada grupo

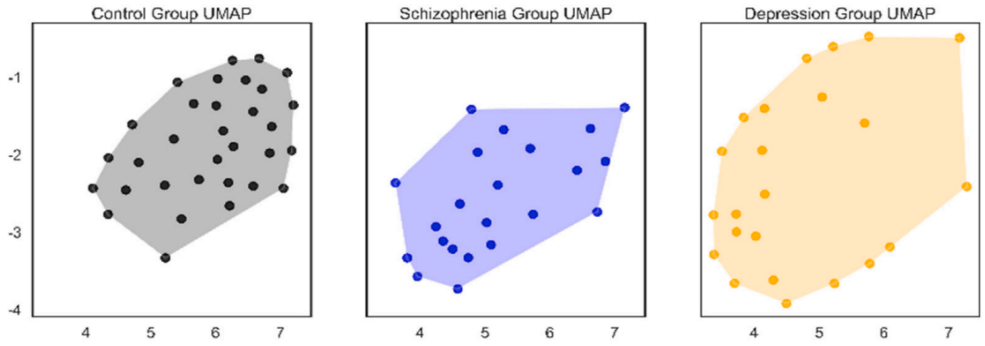

Figura 2.3: Representación UMAP de los datos actigráficos de cada grupo

Los resultados que se obtuvieron en este segundo documento son similares tanto a los del primero, como a los de este trabajo. Para los pacientes sanos, los valores de SHAP indicaron un patrón diurno consistente sobre las 9:00 de la mañana. Para el segundo grupo, los pacientes con esquizofrenia, la amplitud global de la salida representativa de la actigrafía era inferior en comparación con las del primer conjunto. Por último, los pacientes con trastorno depresivo mostraron más irregularidad en su actividad, muy poca estabilidad interdiaria y pautas de sueño volátiles, ya que había horas de despertar inusuales. En general, los participantes representativos del grupo de depresión mostraron bajos niveles de rutina e intensidad en sus patrones de movimiento, y se diferenciaron en gran medida del grupo de control [18]. Todas las horas representativas en la actividad motora de cada grupo están incluidas en el intervalo escogido y usado de este trabajo. En conclusión, respecto al análisis cualitativo, los participantes de control eran más propensos a tener horarios regulares de sueño y actividad como se demostró en el método aplicado. Por el contrario, los depresivos (no se distingue entre bipolares o monopolares), como ya se ha comentado tenían horarios diarios más irregulares. Por último, se podría intuir en un principio una actividad similar entre los pacientes depresivos y esquizofrénicos, sin embargo, las personas con el segundo trastorno tenían horarios de sueño y vigilia más regulares, más parecidos a los controles. Al obtener este resultado es normal pensar que pudo haber algún error, pero no, coinciden con el trabajo original que encontró un patrón de comportamiento más estructural en la esquizofrenia que en la depresión.

A pesar de tener cierto éxito y obtener conclusiones razonables, los autores de este artículo confirman de nuevo la dificultad de trabajar con estos datos. El tamaño de las características, el número de pacientes y la variabilidad entre ellos, incluso dentro de un mismo grupo, provocaron que la tarea no fuera sencilla. Por ello, para abordar este reto, se usó un algoritmo de aprendizaje automático no supervisado que redujo

el tamaño, en este caso de 10080 observaciones a dos valores representados en un sistema de coordenadas cartesianas. Y, para terminar, se aplicó el método de SHAP para comprender mejor la importancia relativa de las características de los datos y determinar el efecto de las perturbaciones de los puntos de actividad.

Capítulo 3

Métodos de Clasificación

3.1. Depresión

La depresión es una enfermedad que se caracteriza por una tristeza constante y por la pérdida de interés en las actividades con las que se suele disfrutar, así como por la incapacidad para llevar a cabo las actividades cotidianas, durante al menos dos semanas. Tiende a ser más frecuente en las personas de mayor edad, aquellos que padecen enfermedades somáticas crónicas o graves y en las mujeres (en razón de 2-3 mujeres por cada hombre). Actualmente cada vez está más detectada y se puede observar que influye en la vida cotidiana de las personas. Es causada por una combinación de factores genéticos, biológicos, ambientales y psicológicos.

Dentro de los tipos de depresión que se pueden encontrar, el más grave es el episodio depresivo mayor que tiene un inicio claro y debe perdurar a lo menos dos semanas. Este es fácilmente distinguible del carácter de quien lo padece. Se define como un estado patológico persistente de ánimo deprimido asociado a otros síntomas psíquicos y/o físicos, que afecta significativamente el funcionamiento global del sujeto [12]. La distimia o trastorno depresivo persistente es un tipo de depresión más suave, con pocos síntomas excluyendo algunos de los más graves como la ideación suicida y, que dura como mucho dos años. Tiende a confundirse con el estilo característico del sujeto, lo que lleva a hacer difícil precisar su comienzo. Es frecuente que en los pacientes con distimia, que es una depresión crónica, se complique con un episodio depresivo mayor, más agudo e intenso. En estos casos, se habla de depresión doble. Por último, la melancolía se considera un subtipo de la depresión haciéndose distinción entre depresión melancólica y la no melancólica. Las personas con depresión melancólica suelen sentirse muy desesperadas y culpable y no tienen el más mínimo

ápice de felicidad. Estas depresiones son las más difíciles de tratar.

No todas las personas con enfermedades depresivas experimentan los mismos síntomas. La gravedad, frecuencia y duración de los síntomas varían dependiendo de la persona y su enfermedad en particular. Algunos indicios que se pueden destacar son alteraciones emocionales como angustia, alteraciones del pensamiento como fallos en la concentración o indecisión, alteraciones somáticas como insomnio o anorexia, alteraciones de los ritmos vitales y alteraciones de la conducta.

Las causas son múltiples. Es posible encontrar una relación biológica, como alteraciones hormonales o bioquímicas y herencia, y una disposición biográfica como un estilo de personalidad melancólico, que puedan desencadenarse en situaciones ambientales y de conflicto, fracaso o pérdidas.

En este trabajo, como se ha comentado previamente, se habla sobre personas depresivas bipolares y monopolares. El trastorno bipolar es una enfermedad crónica y recurrente que se manifiesta principalmente por episodios depresivos y periodos de exaltación del humor e incremento de la vitalidad (episodios maníacos o hipomaníacos) [14].

Existen varios tipos de trastornos bipolares. El episodio maníaco desarrolla síntomas variados en un mismo tiempo como humor eufórico e irritable junto con ideas de grandiosidad o hiperactividad. Este trastorno puede unirse a su vez con síntomas psicóticos o una conducta de riesgo. El trastorno bipolar de tipo I alterna episodios maníacos, depresivos y mixtos. El trastorno bipolar de tipo II se caracteriza por síntomas depresivos e hipomaníacos. Por último, la ciclotimia es un tipo de bipolaridad más leve.

Los síntomas más habituales de un trastorno bipolar son episodios maníacos como por ejemplo euforia, aceleración del habla y cierta hiperactividad. y episodios depresivos. En un episodio mixto coexisten ambos síntomas. Las causas están muy ligadas a la genética, combinadas con situaciones vitales adversas.

La depresión bipolar I se refiere al episodio depresivo mayor en un paciente con trastorno bipolar de tipo I. La depresión bipolar II afecta a aquellas personas con trastorno bipolar de tipo II y un episodio depresivo mayor y, la depresión monopolar es en la que no existe una presencia de episodios de bipolaridad.

Es necesario comentar, antes de empezar con el tratamiento de los datos, varias diferencias entre los casos. Se ha estudiado que un diagnóstico incorrecto prolonga el sufrimiento asociado con una peor calidad de vida, por ello los indicadores clínicos alertan a los pacientes con un estado depresivo mayor sobre la posibilidad de estar en presencia de una depresión bipolar. Estos pacientes tienden más a una actitud

de suicidio y de consumo de alcohol y drogas, así como de mayor trastorno ansioso y de la personalidad. Respecto al tratamiento, cuando un paciente con depresión bipolar es tratado como si sufriera de una monopolar se obtiene solo una recuperación parcial. Las principales características clínicas sugerentes de riesgo de bipolaridad en un paciente con un episodio depresivo mayor son antecedentes familiares, el curso de la enfermedad, los síntomas de esta y la respuesta al tratamiento con antidepresivos.

A continuación se define el cuestionario o escala MADRS, un valor encontrado en el conjunto inicial de los datos que sirve para evaluar la gravedad y síntomas de la depresión.

3.1.1. Cuestionario MADRS

La escala MADRS (Montgomery Asberg Depression Rating Scale) es un instrumento cuyo objetivo es la evaluación de los síntomas y gravedad de la depresión. Consta de 10 ítems relativos a distintos indicios. Cada uno de estos se puntúa del 0 al 6, siendo el más bajo la ausencia del síntoma y el más alto el máximo nivel de gravedad de este.

La tristeza observada, la declarada por el paciente, la tensión interna, el sueño reducido, el apetito reducido, las dificultades para concentrarse, la lasitud, la incapacidad para sentir, los pensamientos pesimistas y los suicidas son las diez características que este cuestionario mide.

La puntuación total de este cuestionario se determina con la suma de cada uno de los 10 apartados, por lo que en total puede llegar a 60 puntos. Se declara una depresión moderada a partir de los 20 puntos, y a partir de 35 se puede considerar grave. El cuestionario está incluido en la bibliografía [1].

3.1.2. Dataset: *The Depresion Dataset*

Este dataset contiene grabaciones de la actividad motora de 23 pacientes depresivos, tanto bipolares como monopolares, y de 32 pacientes sin ningún trastorno psicológico que contribuyeron al estudio. En concreto, dentro de los afectados, hay 15 personas con trastorno depresivo monopolar, 1 persona con trastorno depresivo bipolar I y 7 personas con trastorno depresivo bipolar II. Todos ellos fueron medidos y observados a lo largo de distinta cantidad de días. El dataset está introducido como referencia de la bibliografía [8].

El primer conjunto de datos contiene un archivo de cada uno de los 55 sujetos de su actividad motora a lo largo de cada minuto de los días controlados. Esta actividad

motora fue medida por el reloj actigráfico que se les colocó en la muñeca derecha. Cada archivo contiene tres columnas:

- '*timestamp*': intervalos por cada minuto. Indica el día exacto y la hora de la medida (2003-05-07 12:04:00).
- '*date*': establece únicamente el día (2003-05-07).
- '*activity*': la medida de la actividad motora en ese preciso momento.

El segundo conjunto de datos tiene características de cada persona, que son:

- '*number*': identificador del paciente.
- '*days*': número de días de mediciones. En torno a dos semanas duró el tratamiento.
- '*gender*': género del paciente, mujer (1) u hombre (2). Participaron 13 hombres y 10 mujeres.
- '*age*': rango de la edad del paciente. Desde los 20 años hasta los 69.
- '*afftype*': tipo de trastorno del paciente. Si el trastorno es depresivo bipolar II (1), monopolar (2) o depresivo bipolar I (3). Un único paciente tenía trastorno depresivo bipolar I, 7 pacientes bipolar II y, el resto, depresivo monopolar.
- '*melanch*': si la depresión es melancólica (1) o no (2). Solo un paciente tenía esta melancolía y, 3 de ellos no se llegó a saber.
- '*inpatient*': si el paciente fue hospitalizado (1) o llevado al ambulatorio (2), que fue un total de 18.
- '*edu*': educación agrupada en años.
- '*marriage*': si el paciente está casado o conviviendo con alguien formaba parte de los 11 del total (1) o, si está soltero (2).
- '*work*': si el paciente está trabajando o estudiando (1), o por el contrario está en el paro o sin estudiar (2). Únicamente trabajaban 3 de ellos.
- '*madrs1*': medida del MADRS una vez empezado el tratamiento.
- '*madrs2*': medida del MADRS una vez acabado el tratamiento.

A la hora de trabajar con todos estos datos, se ha realizado alguna modificación o ajuste para un mejor resultado que se explicará a continuación.

3.1.3. Métodos de Clasificación

Como se ha explicado tanto en la teoría como en la descripción del dataset, en el estudio participaron personas con trastorno depresivo monopolar y bipolar (I y II). El fin de este apartado del trabajo es, mediante distintos modelos de clasificación, conseguir saber qué tipo de depresión tiene cada paciente, distinguiendo sólo entre los casos de trastorno depresivo monopolar o bipolar. Es decir, al haber únicamente un paciente con depresión bipolar I, se ha hecho una clasificación binaria entre monopolar o bipolar.

Durante el proyecto, los datos originales medidos en intervalos de un minuto a lo largo de distinta cantidad de días se han tratado como una serie temporal. Al tener miles de datos por cada paciente, lo primero que se planteó fue observar por separado los días medidos de cada paciente, pasando así de tener 23 observaciones a 291. Una vez hecho esto, la columna '*timestamp*' se estableció como índice para cada dato y así poder hacer filtrados por fechas directamente. Y, más tarde, mediante funciones aplicadas a series temporales, se fue calculando la actividad media por cada hora.

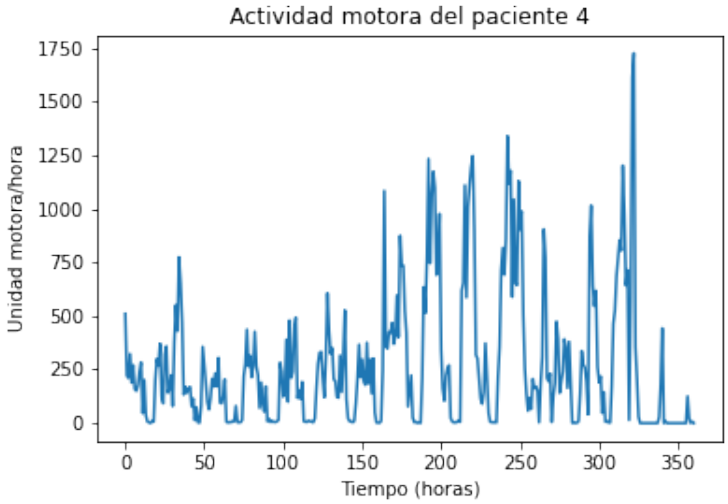

Figura 3.1: Paciente con trastorno depresivo

Para tener una idea inicial, antes de realizar cualquier cálculo, se representó en una señal de tiempo la variación de la actividad de cada persona a lo largo del tiempo. Analizando cada paciente por separado y sus cambios diarios, el cuarto fue de los que más altibajos mostró y el que llegó al punto máximo de actividad, en concreto en el día 12, como se puede observar en la siguiente imagen.

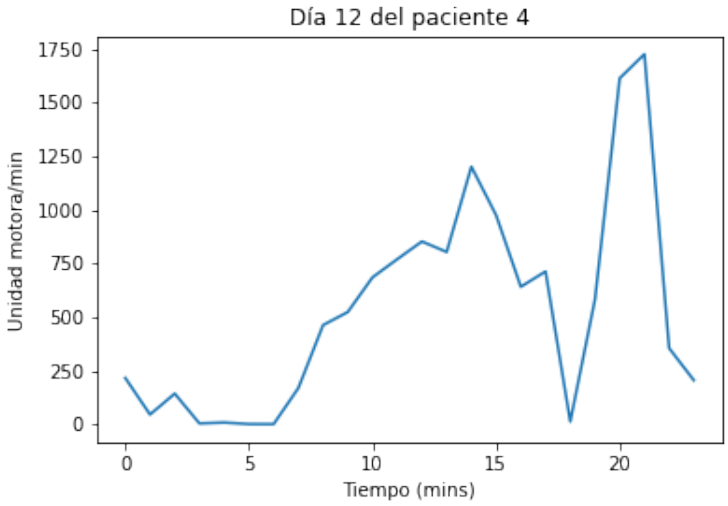

Figura 3.2: Día 12 del paciente

Viendo cada señal, las horas medias del día en las que se pueden ver mejores resultados son de la quinta a la decimoquinta, a pesar de haber algún valor alto fuera de este rango. Por ello, se creó un nuevo dataset a partir del original en el cada fila representa un único día del paciente correspondiente, y las nuevas variables son:

- '*Tipo*': el tipo de trastorno depresivo del paciente, si es trastorno depresivo bipolar (0) o trastorno depresivo monopolar (1). 196 observaciones del total eran del tipo 1, y 95 del tipo 0.
- '*Melancolía*': la aparición de melancolía en la enfermedad. (0) Si esta existe, en concreto en 13 casos, (1) si no, en 249 observaciones, y (2) si este dato no se sabe, el resto.
- '*Edad*': la edad media del rango de edad del paciente.
- '*MADRS1*': la medida MADRS al inicio del tratamiento.
- '*MADRS2*': la medida MADRS al final del tratamiento.
- '*Hosp*': la hospitalización del paciente. (1) Si el sujeto fue llevado al ambulatorio y (0) si al hospital. 222 observaciones del total son del tipo 1.
- '*Día*': la actividad motora media de ese día.
- '*Hora5*': la actividad motora media a la quinta hora del mismo día.
- '*Hora6*': la actividad motora media a la sexta hora del mismo día.

- '*Hora7*': la actividad motora media a la séptima hora del mismo día.
- '*Hora8*': la actividad motora media a la octava hora del mismo día.
- '*Hora9*': la actividad motora media a la novena hora del mismo día.
- '*Hora10*': la actividad motora media a la décima hora del mismo día.
- '*Hora11*': la actividad motora media a la décimo primera hora del mismo día.
- '*Hora12*': la actividad motora media a la décimo segunda hora del mismo día.
- '*Hora13*': la actividad motora media a la décimo tercera hora del mismo día.
- '*Hora14*': la actividad motora media a la décimo cuarta hora del mismo día.
- '*Hora15*': la actividad motora media a la décimo quinta hora del mismo día.

Tipo	Melancolía	Edad	MADRS1	MADRS2	Hosp	Dia	Hora5	Hora6	Hora7	Hora8
1	1	37	19	19	1	164.070516	209.850000	129.116667	60.466667	99.950000
1	1	37	19	19	1	110.545196	12.650000	4.516667	6.433333	5.666667
1	1	37	19	19	1	130.902982	184.883333	206.700000	192.216667	447.583333
1	1	37	19	19	1	226.821107	12.450000	5.250000	9.816667	9.683333
1	1	37	19	19	1	159.088027	119.783333	170.333333	80.250000	284.633333
...	...	...	...	...	...	...	...	...	...	...
1	1	32	29	23	0	349.058435	494.600000	757.466667	520.700000	404.833333
1	1	32	29	23	0	130.404472	11.383333	45.000000	5.083333	15.183333
1	1	32	29	23	0	1.012703	0.000000	0.000000	2.150000	3.716667
1	1	32	29	23	0	0.549797	0.000000	0.000000	1.200000	0.566667
1	1	32	29	23	0	4.233079	0.000000	0.000000	0.000000	0.550000

Figura 3.3: Datos depresión

Predicción a partir de la actividad motora media por hora

El objetivo inicial de este primer método fue intentar predecir la depresión de cada paciente a partir de su actividad motora únicamente. En este método se usarán como variables las calculadas a partir de la media de cada hora, de la quinta a la decimoquinta, de todos los días totales.

Antes de aplicar cualquier método, se dividieron las observaciones en un conjunto de entrenamiento (70 %) y de test (30 %). Como se ha comentado, las variables que participan en este modelo son las calculadas como la media de la actividad motora en una de las diez horas intermedias del paciente en un día determinado. La variable

a predecir será *'afftype'* en el dataset original o *'tipo'* en el creado para este proceso. Se crearon tres modelos distintos para hacer una comparación y detectar cuál predice mejor en este caso.

Como primer método se aplicó **Regresión Logística**. Este es un modelo estadístico que se utiliza para determinar la probabilidad de que ocurra un evento. Muestra la relación entre características y luego calcula la probabilidad de un resultado determinado. En Machine Learning se utiliza para ayudar a crear predicciones precisas. En este caso se usa Regresión Logística binaria ya que solo existen dos resultados posibles para la respuesta categórica.

El modelo obtuvo un resultado de predicción de un 70,936 % de los casos. Este número no es suficientemente alto para poder determinar el trastorno de una persona depresiva.

Como segundo modelo se usaron **Redes Neuronales**, que son un modelo inspirado en el funcionamiento del cerebro humano. Está formado por un conjunto de nodos conocidos como neuronas artificiales que están conectadas entre sí y se transmiten señales desde la entrada hasta la salida. Los valores de entrada llegan a cada neurona, y estas lo modifican y trasmiten a las próximas redes. En el final de la red, se obtiene una salida que es la predicción calculada por la red. En este caso, se obtuvo un 97,537 % de aciertos.

Por último, se hizo un estudio con **Keras**, una biblioteca de código abierto escrita en Python que acelera la creación de redes neuronales. Funciona como una API que permite acceder a varios frameworks de aprendizaje automático y desarrollarlos.

Primero, se creó un modelo vació de tipo *Sequential*, es decir, una serie de capas de neuronas secuenciales. Se agregaron cuatro capas *Dense*, cada una con distintas unidades y capas de activación. Se hicieron diversas pruebas con varias combinaciones, ya que no existe un modelo concreto correcto, y este creado obtuvo unos resultados excelentes. Además, se añadieron más *epochs*. Su porcentaje de predicción, aproximadamente un 99,50 %.

Como conclusión, se puede observar la aproximación de los resultados obtenidos en forma de resumen en la siguiente tabla:

	Predicción	Tiempo de ejecución
Regresión Logística	70,936 %	0,011 s
Redes Neuronales	97,537 %	0,143 s
Keras	99,50 %	3,406 s

Cuadro 3.1: Predicción del trastorno depresivo a partir de la actividad motora media

Predicción a partir de distintas propiedades

Además de la clasificación anterior, se hizo otra a partir de algunas medidas de los pacientes, en concreto, la edad de cada uno, si estuvo hospitalizado o no el sujeto, si la depresión es melancólica y las medidas MADRS tanto al iniciar el tratamiento como al finalizarlo. A estas se añaden las actividades motoras medias tanto por cada día como a la hora 12, calculadas de nuevo tratando los datos como una serie temporal. De nuevo, con los conjuntos de *train* y *test*, se realizó una comparación entre los distintos métodos: Regresión Logística, Redes Neuronales y Keras. El dataset usado en esta sección es el mismo de la parte anterior, que se tuvo que modificar para poder tener más observaciones y estudiar cada día de un paciente por separado.

Se obtuvieron unos resultados algo más bajos, esto se debe a que la actividad motora es más influyente en esta clasificación del trastorno depresivo monopolar o bipolar. Además, es cierto que el dataset no es suficientemente grande como para poder determinar la enfermedad de un paciente con estas pocas características. A pesar de ello, los pocentajes de predicción no son descartables.

Para la **Regresión Logística** se añadieron más iteraciones en el modelo y se obtuvo un mejor resultado, un 83,25 % de los casos.

Por último, el porcentaje de aciertos de predicción para los otros dos métodos se redujeron en esta sección. Mediante **Redes Neuronales**, el *score* obtenido fue de nuevo un 83,25 %.

El método de **Keras** era casi insignificante si se usaba una única capa *Dense*, ya que no llegaba a predecir ni un 64 % de los sucesos. Sin embargo, se observó que añadiendo más *epochs* al modelo, y más capas como en el creado en la predicción anterior, este porcentaje aumentaba. Los *epochs* hacen referencia al número de iteraciones que se usan durante el modelo.

Se pueden observar los resultados que se obtuvieron en todos los métodos y sus tiempos de procesado en la siguiente tabla:

	Predicción	Tiempo de ejecución
Regresión Logística	83,25 %	0,024 s
Redes Neuronales	83,25 %	0,109 s
Keras	86,21 %	2,176 s

Cuadro 3.2: Predicción del trastorno depresivo a partir de varias propiedades

3.2. Esquizofrenia

La esquizofrenia es un trastorno mental grave por el cual las personas interpretan la realidad de manera anormal [22]. La población suele tener un concepto erróneo de esta enfermedad, ya que la esquizofrenia no significa personalidad doble, ni que el sujeto sea más peligroso o violento, sino un trastorno cerebral crónico que no está muy extendido en la actualidad. La investigación ha demostrado que afecta tanto a hombres como mujeres por igual aunque suele aparecer antes en los primeros. En general, los pacientes con esquizofrenia fallecen más jóvenes que el resto de personas. A pesar de que se desconoce la causa exacta de la aparición de esta enfermedad, existen algunos factores que aumentan su desarrollo como antecedentes familiares, complicaciones durante el embarazo o nacimiento y consumo de drogas que alteran la mente durante la adolescencia.

Esta enfermedad se ha clasificado en cinco tipos que son paranoide, desorganizada, catatónica, indiferenciada y residual. La paranoide se caracteriza por la preocupación excesiva de una o más ideas delirantes o bien por alucinaciones auditivas frecuentes. La esquizofrenia desorganizada muestra en el paciente un comportamiento alterado ya que no tiene orden ni afectividad y no responde a los estímulos externos de manera adecuada. Los sujetos con esquizofrenia catatónica presentan inmovilidad (catalepsia), negativismo, mutismo, adopción de posturas extrañas o, por el contrario, una actividad motora excesiva y limitación de las palabras. La indiferenciada y la residual no tienen síntomas concretos como las anteriores. La esquizofrenia indiferenciada se diagnostica cuando los indicios son como los descritos en los tres tipos anteriores pero en su conjunto no se puede clasificar como paranoide, desorganizada o catatónica. Por último, en la residual se dan manifestaciones leves de síntomas tanto positivos como negativos pero de menor magnitud.

Algunos de los signos que se suelen identificar en alguien con esquizofrenia, sin tener en cuenta los tipos con sus características explicados anteriormente, son alucinaciones, ya que por lo general ven o escuchan cosas que no existen, fantasías, síntomas negativos referidos a la capacidad limitada para vivir de manera normal, y un comportamiento motor extremadamente anormal. Los estudios de imágenes cerebrales muestran diferencias en la estructura del cerebro y el sistema nervioso central de las personas con esquizofrenia y han demostrados que ciertos químicos como la dopamina o el glutamato pueden contribuir a su evolución.

Si no se trata este trastorno, se pueden producir graves problemas que afectan a todos los ámbitos de la vida, como por ejemplo intentos de suicidio, trastornos de ansiedad y trastornos obsesivos compulsivos, depresión, abuso de alcohol y drogas, problemas financieros y falta de vivienda, aislamiento social o más problemas de

salud y médicos. Un medicamento que suele ofrecerse a personas que han intentado suicidarse, hacerse daño físico, o que no han recibido otros fármacos, es la clozapina. Es un antipsicótico atípico indicado para tratar síntomas de esquizofrenia y que su acción consiste en cambiar la actividad de ciertas sustancias naturales en el cerebro. Otro tipo de antipsicótico son los neurolépticos que se emplean para tratar los síntomas de la psicosis, tales como alucinaciones, delirios y demencia. La mayoría impiden la acción de ciertas sustancias químicas en el sistema nervioso. Por último, para tratar fuertes oscilaciones del humor y energía, un medicamento psiquiátrico que se suele recetar es un estabilizador del estado de ánimo. El trastorno más común al que se aplica es el bipolar, en el cual los estabilizadores suprimen los balanceos entre episodios maníacos y depresivos. También sirven para tratar el trastorno límite de la personalidad.

A continuación se introduce el cuestionario o escala BPRS, un valor encontrado en el conjunto inicial de los datos que sirve para evaluar la gravedad y síntomas de la esquizofrenia.

3.2.1. Cuestionario BPRS

La escala BPRS (Brief Psychiatric Rating Scale) fue diseñada por Overall y Gorham y usada para valorar la respuesta al tratamiento farmacológico en pacientes psicóticos así como para la clasificación de este trastorno como positivo o negativo. Existen múltiples versiones de este cuestionario, cada uno con distintas formas de puntuación (0-6, 1-7, 0-4) y número de ítems (10, 14, 16, 18, 20, 22,). La versión más común consta de 18 apartados a los que se suman dos finales, uno de gravedad y otro de mejoría general, puntuado cada uno del 1 (ausencia del síntoma) al 7 (extremadamente grave), por lo que la puntuación total oscila entre los 18 y 126 puntos y dependiendo de esta se indica la aproximación del nivel de esquizofrenia del paciente [15].

Dentro de las características que estudia este cuestionario se encuentran algunas como la preocupación somática, la ansiedad psíquica, el aislamiento emocional, las alucinaciones y el humor depresivo.

Los síntomas negativos de la esquizofrenia se valoran en los ítems 3 (retraimiento emocional), 13 (enlentecimiento motor), 16 (aplanamiento efectivo) y 18 (desorientación). Los positivos sin embargo, en el 4 (desorganización conceptual), 11 (suspicacia), 12 (alucinaciones) y 15 (contenidos del pensamiento inusuales).

3.2.2. Dataset: *A Motor Activity Database of Patients with Schizophrenia*

Los datos recogidos para este caso continúan la estructura de los explicados para el trastorno depresivo. 22 pacientes con esquizofrenia, 17 de ellos paranoide y 5 no, y los mismos 32 controles sin trastorno psicológico que también contribuyeron en el estudio de la depresión, fueron medidos a lo largo de distintos días. De nuevo, mediante un reloj actigráfico, se registró la actividad motora por cada minuto y distintas propiedades de cada sujeto. El dataset está introducido como referencia de la bibliografía [10].

El primer conjunto del dataset contiene un archivo de estas medidas de actividad para cada persona:

- '*timestamp*': intervalos por cada minuto. Indica el día exacto y la hora de la medida (2003-05-07 12:04:00).
- '*date*': establece únicamente el día (2003-05-07).
- '*activity*': la medida de la actividad motora en ese preciso momento.

El segundo, datos de distintas cualidades de los pacientes esquizofrénicos. Estas son:

- '*number*': identificador del paciente.
- '*gender*': género del paciente. Hay 19 hombres frente a 3 mujeres.
- '*age*': rango de la edad del paciente. Pacientes desde los 25 años hasta los 69.
- '*days*': número de días de mediciones. Alrededor de dos semanas.
- '*schtype*': tipo de trastorno del paciente. Si el trastorno esquizofrénico es paranoide (17 personas) o no (5 personas).
- '*migraine*': si además tenía migrañas. 18 del total de los pacientes no tenían.
- '*bprs*': medida del BPRS.
- '*cloz*': consumo de clozapina (8 pacientes).
- '*trad*': neuroléptico tradicional o moderno. Existen dos tipos de este fármaco, el tradicional que fue consumido por 6 pacientes y el no tradicional o modernno, por 16.

- '*moodst*': estabilizador del estado de ánimo. 5 de los sujetos tomaron este estabilizador.

- '*agehosp*': edad en la que fue hospitalizado.

Por último, hay dos archivos más. Uno que contiene el identificador de los 54 registros y el número de días del estudio de cada uno, y otro que contiene la media, desviación típica y proporción de ceros de la actividad motora de cada persona.

3.2.3. Métodos de Clasificación

Al igual que en el estudio del trastorno depresivo, se realizaron dos predicciones distintas para la esquizofrenia. Una basada en únicamente la actividad motora a lo largo de las diez horas intermedias del día, y otra basada en distintas propiedades de los pacientes, como puede ser su medida BPRS o, si tomó clozapina. Para ello, las 22 observaciones se transformaron en 285, una para cada día de los pacientes.

Tratando los datos como una serie temporal se pudo hacer un pequeño análisis dependiendo de las fechas (días y horas). Es importante recordar que cada valor viene dado en intervalos de un minuto, sin embargo, a partir de estos, se fue calculando la media por cada hora. Observando el cambio de la actividad motora a lo largo de los días e identificando los sujetos, se pudo llegar a la conclusión de que el primer paciente tiene unos altibajos más impactantes como se observa en la Figura 3.4.

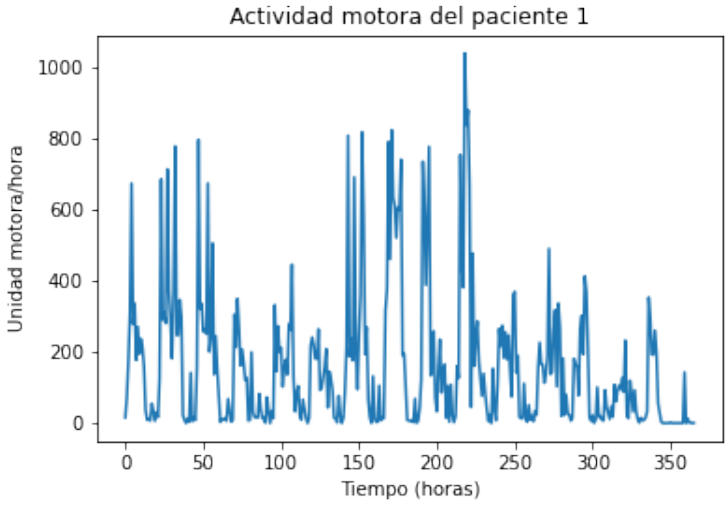

Figura 3.4: Paciente 1 con esquizofrenia (paranoide)

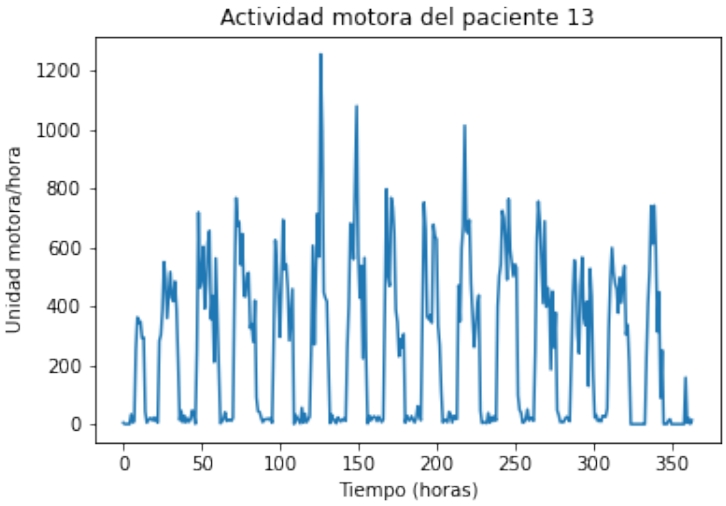

Figura 3.5: Paciente 13 con esquizofrenia (paranoide)

Sin embargo, el paciente número 13 (con esquizofrenia paranoide también) fue el que llegó al máximo valor, en el día 5. En la Figura 3.6 se muestra su actividad motora.

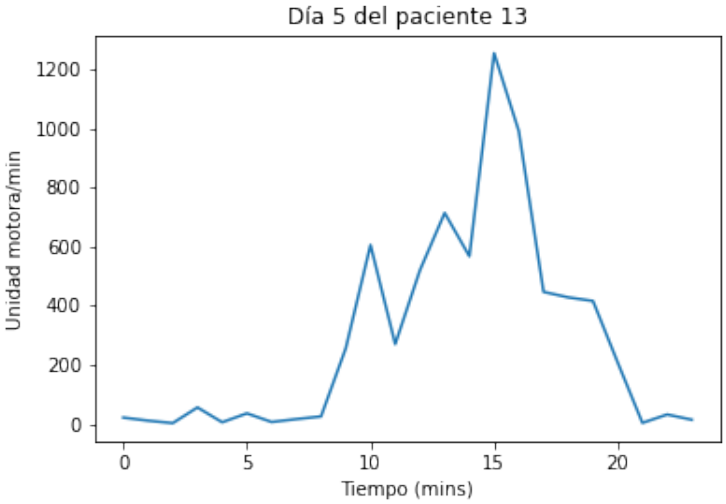

Figura 3.6: Día 5 del paciente 13

Algunas de las variables dadas en la página web [10] fueron descartadas a la hora de hacer ambos métodos, ya que realizando diversas pruebas, estas no mostraban

un mejor resultados de aciertos, al contrario, lo disminuían. Además, las diez horas diarias intermedias volvían a demostrar que eran las más relevantes. Por todo esto, que el dataset creado, a partir del original, está formado por:

- '*Tipo*': el tipo de trastorno esquizofrénico del paciente, si es paranoide (1) o no (0). Se ha pasado a tener 220 observaciones con trastorno esquizofrénico paranoide, y 65 con no paranoide.
- '*Migraña*': la aparición de migrañas en la enfermedad. (0) Si estas aparecen y (1) si no. Ahora, 52 observaciones del total tenían migraña.
- '*Edad*': la edad media del rango de edad del paciente.
- '*BPRS*': la medida BPRS.
- '*Clozapina*': el consumo de clozapina (1) por 103 observaciones, o (0) si no se tomó este medicamento, 182.
- '*Neuroléptico*': o tradicional (0) o moderno (1). 207 de ellas era moderno.
- '*Estabilizador*': no se medicó un estabilizador del estado de ánimo (0) o si (1). 221 observaciones del total no fueron recetadas.
- '*Día*': la actividad motora media de ese día.
- '*Hora5*': la actividad motora media a la quinta hora del mismo día.
- '*Hora6*': la actividad motora media a la sexta hora del mismo día.
- '*Hora7*': la actividad motora media a la séptima hora del mismo día.
- '*Hora8*': la actividad motora media a la octava hora del mismo día.
- '*Hora9*': la actividad motora media a la novena hora del mismo día.
- '*Hora10*': la actividad motora media a la décima hora del mismo día.
- '*Hora11*': la actividad motora media a la décimo primera hora del mismo día.
- '*Hora12*': la actividad motora media a la décimo segunda hora del mismo día.
- '*Hora13*': la actividad motora media a la décimo tercera hora del mismo día.
- '*Hora14*': la actividad motora media a la décimo cuarta hora del mismo día.
- '*Hora15*': la actividad motora media a la décimo quinta hora del mismo día.

Tipo	Migraña	Edad	BPRS	Clozapina	Neuroléptico	Estabilizador	Dia	Hora5	Hora6	Hora7
1	1	47.0	48	1	1	0	205.832839	278.050000	336.783333	177.383333
1	1	47.0	48	1	1	0	249.150563	247.100000	247.683333	346.400000
1	1	47.0	48	1	1	0	115.515116	3.216667	10.200000	9.500000
1	1	47.0	48	1	1	0	121.451689	15.166667	0.883333	72.516667
1	1	47.0	48	1	1	0	95.580913	0.000000	13.933333	217.183333
...	...	...	...	...	...	...	...	...	...	...
1	1	52.0	50	0	0	0	267.421210	446.333333	313.066667	371.316667
1	1	52.0	50	0	0	0	185.010186	6.433333	4.166667	6.083333
1	1	52.0	50	0	0	0	135.222148	3.166667	9.166667	8.550000
1	1	52.0	50	0	0	0	179.597364	42.733333	348.016667	293.500000
1	1	52.0	50	0	0	0	106.606950	403.300000	346.300000	138.566667

Figura 3.7: Datos esquizofrenia

Predicción a partir de la actividad motora media por hora

Siguiendo la estructura del trabajo, al igual que en la sección de *Depresión*, se investigó mediante dos formas la manera de clasificar a los pacientes esquizofrénicos según el tipo de su enfermedad, paranoide o no.

Esta primera manera, utiliza los datos de la actividad motora de todos los sujetos. Una vez divididos todos ellos según su número de días, se fue observando uno a uno las horas con más variaciones o con puntos más altos y a su vez los más bajos. De nuevo, la diez horas diarias intermedias de la quinta a la décimo quinta, fueron las que presentaban más desigualdades. Cogiendo los intervalos de 60 en 60 (debido a que están recogidos por minuto), se calculó la media de esas horas, así como la diaria.

Teniendo en cuenta las 285 observaciones totales de este caso, lo primero que hubo que hacer fue dividir los datos en un conjunto de entrenamiento (70 %) y otro de test (30 %). Los tres modelos creados independientemente obtuvieron resultados similares a los de la sección anterior mediante estas mismas variables, siendo Regresión Logística algo más bajo que los otros dos.

Para la **Regresión Logística**, el porcentaje ya fue relativamente considerable al acertar el 79,40 % de los casos. Sin embargo, teniendo en cuenta lo sucedido con los otros modelos, se esperaba conseguir un mayor acierto con los siguientes.

Efectivamente, el *score* mejoró. Mediante **Redes Neuronales**, se obtuvo un 96,48 % de aciertos en la clasificación de la esquizofrenia en un tiempo de 0,17 segundos. Este resultados es bastante bueno para poder determinar si un paciente tiene esquizofrenia paranoide o no a partir de su actividad motora diaria.

El último método, **Keras**, mejoró estos valores. Añadiendo un total de 100 *epochs* en lugar de 80, se logró mejorar el resultado de un 64,82 % a un 97,98 %.

Se pueden observar los resultados que se obtuvieron en todos los métodos y sus tiempos de procesado en la siguiente tabla:

	Predicción	Tiempo de ejecución
Regresión Logística	79,40 %	0,014 s
Redes Neuronales	96,48 %	0,168 s
Keras	97,98 %	3,211 s

Cuadro 3.3: Predicción de la esquizofrenia a partir de la actividad motora media

Predicción a partir de distintas propiedades

Estos métodos de clasificación fueron para determinar si la esquizofrenia de los pacientes era paranoide o no, por segunda vez, pero a partir de varios rangos característicos: la existencia de migrañas, la edad de la persona, su medida BPRS, si tomó alguno de los medicamentos antipsicóticos (clozapina, neuroléptico o estabilizador del estado de ánimo) y por último, dos actividades motoras medias, la diaria y la de mediodía.

El código, como todos los anteriores, se pueden observar en el GitHub definitivo establecido en la bibliografía [24].

Como resultados generales de los tres métodos, se obtuvieron unos muy considerables porcentajes de aciertos de predicción. El modelo creado a partir de **Regresión Logística** mejoró respecto al método anterior (actividad motora media por hora) a este, siendo su *score* de 90,95 %.

El segundo método, **Redes Neuronales**, repitió número de éxito, un 96,48 % de los casos consiguió saber si el paciente padece esquizofrenia paranoide o no. Y, por último, en un tiempo de casi 3 segundos, el modelo creado mediante **Keras**, obtuvo un 93,97 %.

	Predicción	Tiempo de ejecución
Regresión Logística	90,95 %	0,080 s
Redes Neuronales	96,48 %	0,118 s
Keras	93,97 %	2,598 s

Cuadro 3.4: Predicción de la esquizofrenia a partir de varias propiedades

3.3. Clasificación a partir de la actividad motora

Como métodos de clasificación conjuntos y, que únicamente se hablarán en esta sección, se realizaron experimentos con cuatro nuevos datasets de los tres tipos de pacientes. Estos vienen insertados con el resto de los códigos y archivos [24].

A partir de la actividad motora media por cada hora del día, se hizo una primera tabla de valores que contenía el tipo de trastorno que padecía el paciente, sin tener en cuenta los distintos subtipos dentro de la depresión y de la esquizofrenia, (0: control, 1: depresión, 2: esquizofrenia) así como los datos del movimiento de la hora 1 a la 24. De la misma forma, se crearon tres archivos nuevos con cada combinación, esquizofrenia(2)-control(0), depresión(1)-control(0) y esquizofrenia(2)-depresión(1). A continuación, se muestra una tabla de los resultados obtenidos con los distintos modelos, y después, se hablará brevemente de cada método, aplicándoles *Regresión Logística* y *Redes Neuronales*.

	0 - 1 - 2	0 - 1	0 - 2	1 - 2
Regresión Logística	46,80 %	58,30 %	66,48 %	62,61 %
Redes Neuronales	92,20 %	97,11 %	96,17 %	98,04 %

Cuadro 3.5: Clasificación entre los tres tipos a partir de la actividad motora

El primero de ellos, la diferenciación entre el tipo 0, 1 y 2, fue la que peor valores de predicción obtuvo. Tras las distintas observaciones y experimentaciones de los datos originales, esto puede deberse a las extrañas relaciones que hay entre ellos. Como bien afirma el artículo [18], la distancia relativa de la actividad motora entre los pacientes con esquizofrenia y los controles es relativamente parecida, a diferencia de los pacientes con depresión, que muestran unos movimientos muy dispersos. La similitud de los primeros pudo provocar confusiones a la hora de realizar la predicción. En la Figura 5.1, se observa la media de cada hora de los tres grupos. Es posible detectar que las personas sanas tienen valores más altos, sin embargo, estos valores y los de los esquizofrénicos son más similares entre si mismos que los de los depresivos. Dentro de este último grupo, la información muestra más altibajos. Mediante SHAP, se logró una respuesta a la pregunta *¿qué horas afectan más a la hora de categorizar la enfermedad o trastorno?* Las horas más influyentes según este método de Inteligencia Artificial Explicable, son desde las 23:00 hasta las 8:00 aproximadamente, estando en primer lugar la hora 24. Estas fueron las horas de dormir de los pacientes y son en las que más comportamientos diferentes presentaron.

La última clasificación, depresión-esquizofrenia, es la que mejor resultados consiguió. De nuevo, al ser el grupo depresivo el que más se diferencia, respecto a la

desviación entre cada hora, distinguir entre el tipo 1 y el tipo 2 fue medianamente sencillo. Dentro de las otras dos, existe una leve disminución en la predicción del tercer modelo, con un 96,17 %.

	Control	Depresión	Esquizofrenia
0	0	1	2
1	212.337852	179.679267	129.739825
2	216.994942	176.734364	127.324854
3	206.757172	174.824284	134.606784
4	228.98039	165.465865	136.879649
5	228.570854	165.465865	140.671287
6	248.153648	166.546793	136.609123
7	227.00199	164.532532	135.808187
8	233.888433	160.134307	126.04193
9	215.386774	149.715693	130.357895
10	204.515091	165.642841	128.342573
11	203.671208	157.391065	128.948947
12	210.33724	150.181271	132.566082
13	208.356385	138.54433	119.683041
14	221.429063	140.819817	119.426901
15	206.284577	145.914891	131.329649
16	198.478151	155.409336	130.731813
17	209.335323	163.388488	130.077895
18	225.744735	166.891581	132.60462
19	213.565216	172.738717	132.979474
20	204.99602	163.072566	129.161111
21	212.453566	179.858419	123.019415
22	214.828897	155.915464	136.262281
23	207.714345	207.714345	136.440175
24	213.06563	162.27142	138.879123

Figura 3.8: Actividad motora media horaria

Capítulo 4

Experimentación con XAI: SHAP y LIME

4.1. Explicación de la clasificación depresiva

Como se comentó en la sección 2.1.2, la Inteligencia Artificial Explicable es necesaria para poder comprender cómo y por qué se obtienen ciertos resultados en modelos de machine learning. ¿Cuáles son las variables que ayudaron en este trabajo a predecir el tipo de trastorno depresivo de los pacientes (monopolar o bipolar)?, ¿qué variable fue más influyente? o ¿por qué se obtuvieron estos resultados?, son algunas de las muchas preguntas que al realizar el trabajo surgieron o se plantea cualquier lector. Mediante los dos métodos más comunes, SHAP y LIME, se investigó y se resolvieron estas dudas.

Para ver y obtener información de las distintas características del dataset, se usó el modelo de Regresión Logística para predecir a partir de distintas propiedades. Se estudió este modelo debido a que en la página web [13] es uno de los más explicados y de los que más artículos se pueden encontrar en internet. Además, se obtuvo un porcentaje de predicción del 83,25 % en un tiempo muy bajo. Las variables afectadas eran: la melancolía, la edad, las medidas MADRS, la hospitalización del paciente y, las actividades motoras medias tanto diaria como al mediodía. Por último, recordar que la medida a predecir se transformó en si el trastorno depresivo era monopolar (1) o bipolar (0).

SHAP

Una vez instalada la librería *shap*, lo primero a lo que se procedió fue a calcular los valores de Shapley. Para obtenerlos, se necesitó el modelo entrenado y los datos de *train*, que se usaron para calcular las expectativas condicionales. Cuando se trata de un modelo de Regresión Logística, shap utiliza su *Linear Explainer* que devuelve un objeto de tipo *explanation* = *shap_values* que permite dibujar gráficas de forma sencilla.

Summary plot o beeswarm

El primer gráfico que se realizó fue el *beeswarm* o *summary plot* cuya función es clasificar las características por la suma de las magnitudes de los valores Shap en todas las muestras, y los usa para mostrar el impacto que tiene cada una en la salida del modelo. El eje x determina el valor del matemático, y los datos se recogen para mostrar su densidad. El color representa el valor original de cada propiedad. En el eje y es posible ver las variables ordenadas de mayor a menor influencia en el modelo.

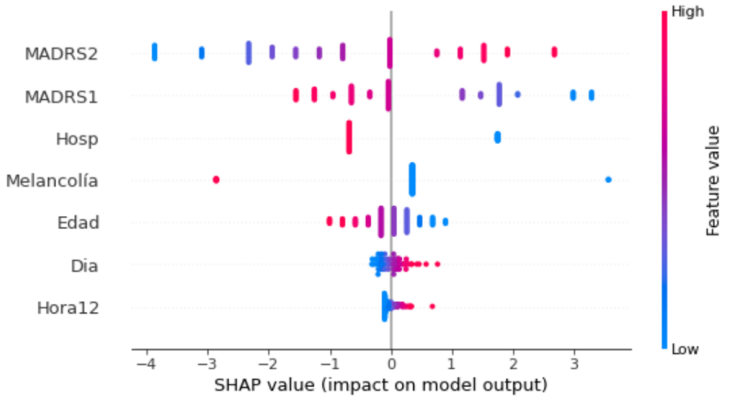

Figura 4.1: Depresión: Beeswarm SHAP

La Figura 4.1 fue la obtenida en esta primera explicación. La primera variable y por tanto la que más afectó a la hora de predecir fue 'MADRS2'. Cuánto más alto es el valor original de esta, más posibilidad hay de que la persona tenga un trastorno depresivo monopolar, ya que tiende el color rojo hacia la derecha de la gráfica. Con 'MADRS1' sucede al revés, un alto valor de esta medida influye más en la predicción del trastorno depresivo bipolar. Esto debió ser porque al iniciar el tratamiento, los pacientes con depresión bipolar alcanzaron una medida más alta, pero sin embargo a la hora de finalizarlo consiguieron reducirla más. Calculando las diferencias en valor absoluto entre estos valores ($|MADRS1 - MADRS2|$), no se percibió gran diferencia.

Los resultados para la hospitalización de los pacientes tuvieron sentido, ya que al referirse el 1 al ambulatorio, el color rojo, y el 0 al hospital, el azul, si una persona fue llevada al ambulatorio ayudó a saber que hay más probabilidad de que esta padeciera un trastorno depresivo bipolar. Respecto a la melancolía fue difícil fiarse, ya que se establecieron como valor 2 aquellos que eran *NaN*. Por esto, nos indica que es más probable no encontrar melancolía en los bipolares, y encontrarla en el monopolar. Con la edad existe una relación que establece que cuanto más mayor, más probabilidad de que el trastorno fuera bipolar. Por último, las actividades motoras indicaban que los pacientes con trastorno depresivo monopolar solían tenerla más alta.

Gráfico de barras

En este caso, fue más sencillo interpretar el resultado. Este gráfico únicamente muestra la magnitud con la que una variable impacta sobre la predicción. Es decir, no importa si suma o resta, si no el valor absoluto de SHAP. Se puede ver como el orden de la izquierda es el mismo que antes.

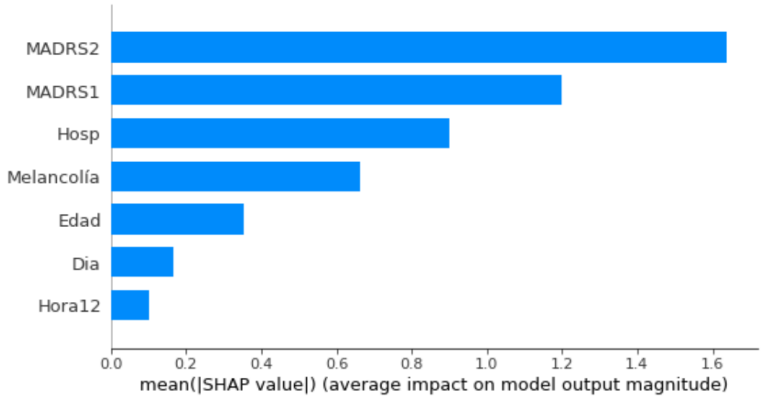

Figura 4.2: Depresión: Gráfico de barras SHAP

Diagrama de fuerza

Para terminar, el diagrama de fuerza se usa cuando, en lugar de observar el efecto de todas las muestras, se quiere describir el de una instancia específica. Así, es posible explicar a un paciente cómo llegó su modelo a la predicción que hizo. Para interpretar los resultados, el número en negrita indica el *score* final del modelo para esta observación y el *base value* representa la predicción que se obtendría si el resto de las variables no aportaran información. Las barras rojas suman y las azules restan a la predicción y, su magnitud viene determinada por el tamaño de estas. En este ejemplo se seleccionó la quinta fila. Este paciente padecía depresión monopolar, y como indica el diagrama, las propiedades más dominantes fueron la medida 'MADRS2'

y el traslado al ambulatorio.

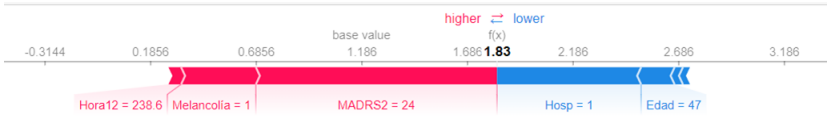

Figura 4.3: Depresión: Diagrama de fuerza SHAP

LIME

Al igual que el diagrama de fuerzas del método anterior, LIME estudia predicciones individuales, obteniendo la contribución de cada variable sobre la predicción de una observación en concreto. A partir de esta observación que se quiere interpretar, LIME genera nuevos datos permutados modificados a partir del original tomando valores de una distribución de media y varianza respectivos al predictor. Sobre estas nuevas predicciones entrena un modelo lineal interpretable que debe aportar una buena aproximación de la predicción de manera local. En este caso estamos sobre un modelo de clasificación, por tanto se necesitan las probabilidades de la variable a predecir. A la hora de visualizar los gráficos, en SHAP es más sencillo interpretar los resultados.

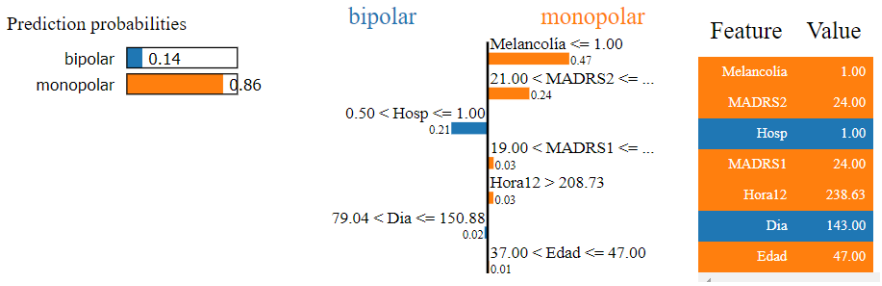

Figura 4.4: Depresión: Gráfica LIME

El ejemplo de la Figura 4.4 muestra los resultados que se obtuvieron en la predicción de un paciente con trastorno depresivo monopolar.

La tabla de la derecha es el valor real de cada propiedad del paciente. La de la izquierda, establece que el modelo supo que se trataba de una trastorno depresivo monopolar con un 86 % de seguridad. Se observó como el color azul influía positivamente en pronosticar la bipolaridad, y el naranja negativamente. Es decir, había que fijarse en el segundo para detectar qué variables intervinieron en el resultado, y con

qué magnitud, ya que este paciente sufría el trastorno depresivo monopolar. La importancia de cada una se ve en el largo de las barras de la tabla del medio. Respecto al valor de la melancolía, al no tener, afectó de manera positiva al aumento de la predicción, así como la medida 'MADRS2' que es más cercana al máximo encontrado en los datos (28) que al mínimo (11). Por el contrario, que la variable 'Hosp' fuera un 1, quiere decir que el paciente fue llevado al ambulatorio. Esto redujo el porcentaje ya que tiende a ser más visto en pacientes con trastorno depresivo bipolar.

4.2. Explicación de la clasificación esquizofrénica

A la hora de hacer el modelo de clasificación para la esquizofrenia paranoide o no paranoide, surgieron nuevas dudas sobre los resultados obtenidos. De nuevo, la Inteligencia Artificial Explicable resuelve estas y muchas otras dudas, y da explicaciones de la predicción. Mediante tres gráficas de SHAP y el método de LIME, fue posible entender el proceso.

El método de Regresión Logística de predicción de la esquizofrenia a partir de distintas propiedades obtuvo un porcentaje de aciertos del 90,95 %. Se introdujo este modelo para explicarlo. Las variables que se usaban en él y que se irán nombrando a lo largo de la sección son: si tenían o no migraña, la edad del paciente, su medida BPRS, los tres medicamentos psiquiátricos (clozapina, neuroléptico y estabilizador del estado de animo) y las actividades motoras medias tanto diaria como al medio día. La variable a predecir 'Tipo', tomaba valor 1 si la esquizofrenia era paranoide y 0 si no lo era.

SHAP

Para esta parte, se procedió de la misma forma que en la sección anterior para la depresión. Se entrenó el modelo, y una vez divididos los conjuntos de entrenamiento y validación, se usaron los primeros para el cálculo de las expectativas condicionales. Mediante el objeto de tipo *explanation*, fue posible representar gráficamente las gráficas y diagramas siguientes. El *summary plot*, el diagrama de barras y el de fuerza son tres de los muchos gráficos que aporta esta librería. Con ellos es posible conseguir el objetivo de este trabajo, la explicabilidad del modelo creado. Como ya fue explicado cada uno anteriormente, se explicarán más los conocimientos que aportan.

Summary plot o beeswarm

Hay que recordar que los colores (azul,rosa y rojo) se refieren a los valores originales de las características, de menor a mayor, que las variables de la izquierda están de

arriba a abajo ordenadas por la influencia sobre el modelo a menor, y que el eje x indica los valores SHAP de cada una de ellas.

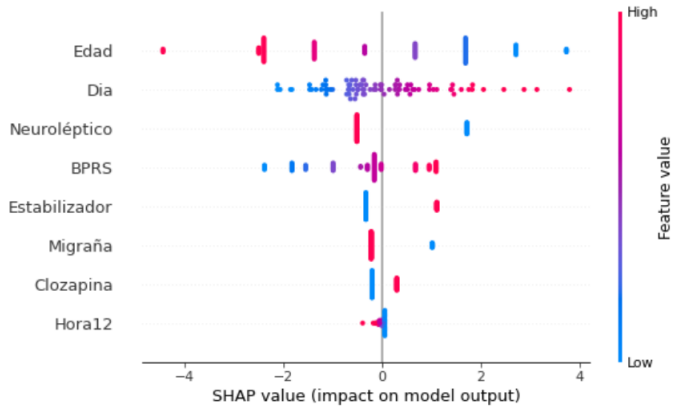

Figura 4.5: Esquizofrenia: Beeswarm SHAP

Empezando por la edad, se pudo llegar a la conclusión que los pacientes con mayor edad tendían más a padecer esquizofrenia no paranoide. La segunda variable siguiendo el orden de influencia sobre el modelo fue la actividad motora diaria. Cuánto más alta esta medida, más probabilidad de que el trastorno fuera paranoide. También se pudo ver que los pacientes con esquizofrenia no paranoide ingerían neuroléptico tradicional en lugar de moderno y, que solían tener una medida más baja del BPRS.

Gráfico de barras

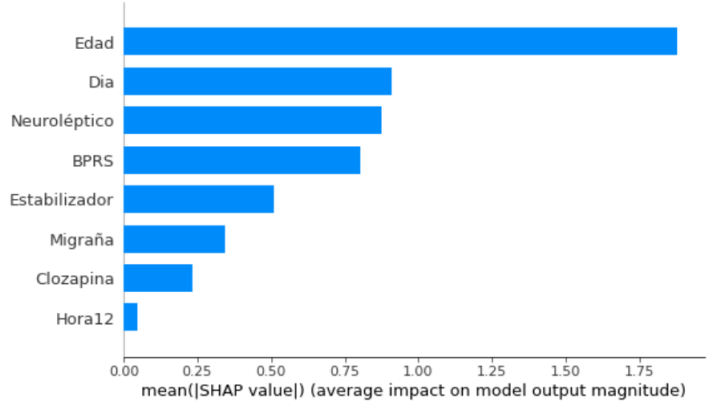

Figura 4.6: Esquizofrenia: Gráfico de barras SHAP

En un experimento de clasificación binaria, uno de los gráficos de mayor impor-

tancia de SHAP es un gráfico de barras, que muestra el valor de Shapley absoluto medio para cada una de las características. Esto simboliza la influencia en el resultado previsto del objetivo, independientemente de si suma o resta a este.

La actividad motora media al medio día es la propiedad que menos influye a la hora de predecir el tipo de enfermedad de las personas, seguida de la clozapina y de si tuvieron a su vez migrañas o no. Por el otro lado, la edad y la actividad motora media diaria son las dos que más ayudan a saber esta información.

Diagrama de fuerza

El diagrama de fuerza, en lugar de dar una gráfica global de todas las observaciones, explica una a una la instancia que se quiera. En este caso, las variables que más afectan al resultado de predicción de este paciente en concreto pueden no seguir el orden del *beeswarm* o de la gráfica de barras, ya que se habla de una persona individual.

En este caso, en la Figura 4.7 se mostró un paciente con esquizofrenia paranoide. Este sujeto tenía las características que se ven en la imagen. Su edad, 32 años, la actividad motora media del día, 63,35, y su medida BPRS, 59, son las tres cualidades con mayor influencia, ya que las barras son más largas.

Para poder representar este diagrama en el programa, es necesario haber ejecutado en Python el siguiente código: *shap.initjs()*.

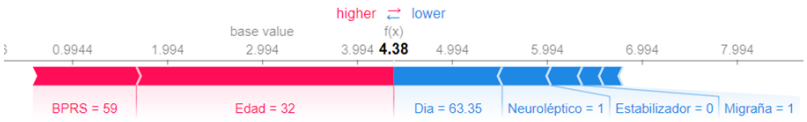

Figura 4.7: Esquizofrenia: Diagrama de fuerza SHAP

LIME

LIME devuelve unos resultados similares al del diagrama de fuerza de la librería de SHAP. Para cada observación que se quiera, devuelve la seguridad que tiene de la posible predicción, los valores originales de las variables estudiadas y la importancia e influencia de cada una. En teoría, la predicción generada debería ser similar en magnitud a la generada por el modelo de predicción original.

En el ejemplo de la Figura 4.7 se observó que para este paciente, lo que más afectaba para predecir el resultado era la actividad motora de ese día positivamente y negativamente el neuroléptico (moderno) y la edad (47 años).

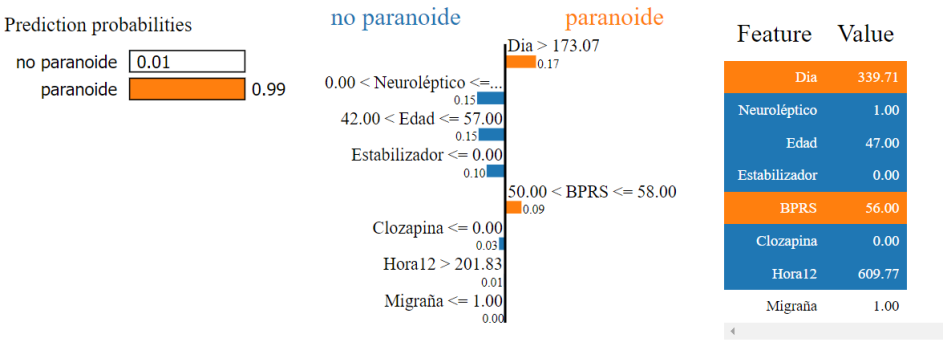

Figura 4.8: Esquizofrenia: Gráfica LIME

4.3. Conclusiones

En conclusión, algunas de las propiedades son más o menos influyentes dependiendo del paciente que se estudia. Dentro de los primero gráficos (*beeswarm* y *diagrama de barras*) es posible ver las variables que podrían o no cambiar el resultado obtenido y, las que en principio, los generan. Esto es debido a que estas dos representaciones muestran el resultado general del peso de todas las variables dentro del modelo. En los otros dos se puede ir analizando paciente a paciente, u observación a observación, ya que indica la prioridad de las propiedades para cada sujeto. SHAP, LIME y las demás opciones de explicación que existen son muy útiles para tanto desarrolladores como investigadores, para así obtener una justificación de la clasificación o predicción.

Dentro de los pacientes depresivos las características que más influyeron en el resultado fueron las medidas MADRS, la inicial y la final. Este orden se puede pensar que es así de importante debido a la diferencia que hay al iniciar el tratamiento y al acabarlo, ya que es complejo atribuir una enfermedad a una persona por el valor de un cuestionario. Las de una persona con esquizofrenia fueron la actividad motora media y la edad del paciente. Observando la actividad motora por cada día de los sujetos con trastorno depresivo, se pudo detectar que no había una relación entre los pacientes bipolares o monopolares. La actividad no mostraba un patrón claro. Sin embargo, observando este movimiento hora por hora, o a lo largo de los minutos si que podía existir más diferenciación entre los distintos tipos. Por eso, la clasificación dependiendo de únicamente la actividad motora, hecha en la sección 3.1.3, obtuvo buenos resultados. Respecto a los pacientes con esquizofrenia, la actividad motora si que mostraba diferencias entre un paciente con trastorno paranoide o no. A pesar de no tener unos valores muy dispersos unos de los otros, esta característica ayudó a clasificar el tipo de enfermedad.

Capítulo 5

Experimentación con Métodos de Predicción con Series Temporales

Lo posiblemente más interesante de las series temporales es poder estudiar su comportamiento, intentar detectar ciertos patrones a lo largo del tiempo y hacer un pronóstico sobre cómo serán en un futuro. Esa es la idea de esta sección. A partir de los datos que, como ya se ha dicho, son las medidas de la actividad motora medidas por intervalos de un minuto, se busca predecir un día futuro del movimiento de un paciente con trastorno depresivo mayor y con esquizofrenia. Una vez obtenidos los resultados, se compararán con los iniciales y se verá si tienen o no sentido.

El código procede de una página web que intenta predecir las ventas futuras de ciertos productos de una empresa a partir de los datos de casi dos años [2]. Este se ha modificado para primero, observar y estudiar ambas series, y posteriormente hacer la predicción.

Para empezar, se estableció la columna *'timestamp'* como índice del dataset y así poder hacer filtrados y búsquedas a partir de las fechas de cada paciente. Una vez hecho esto, se usaron varias funciones para comenzar el estudio, como por ejemplo la función *.describe()*, que aporta la media, desviación, valor mínimo, máximo y más características del archivo deseado. Con esto, se pudo detectar qué paciente tenía la máxima actividad motora, el que más desviación tenía, etc. Respecto al trastorno depresivo, como se comentó en la sección 3.1.3, el cuarto paciente era el más característico, pues con este se realizó la predicción. Para la esquizofrenia, el método de

predicción se creó a partir del primer paciente. Además, es posible acceder a la media diaria u horaria de cada paciente con una única línea de código. Esto ya se usó en la clasificación a partir de la actividad motora media de cada paciente al usar las diez horas intermedias del día. También se representaron gráficamente los datos de la Tabla 5.1.

timestamp	activity
2003-06-03	206.196949
2003-06-04	248.211806
2003-06-05	141.854861
2003-06-06	144.496528
2003-06-07	158.490278
2003-06-08	179.154861
2003-06-09	185.045139
2003-06-10	375.261111
2003-06-11	468.204167
2003-06-12	429.307639
2003-06-13	584.284722
2003-06-14	235.078472
2003-06-15	296.350694
2003-06-16	530.465278
2003-06-17	28.700694
2003-06-18	13.777778

Figura 5.1: Actividad media horaria

El método que se usa en la página de referencia para hacer el pronóstico es una red neuronal con una arquitectura sencilla. El que se empleó en el presente trabajo fue similar. Se aumentó el número de iteraciones y se trasformaron los datos a valores entre -1 y 1 para el método tangencial utilizado.

El problema se transformó en uno de tipo supervisado para poder, mediante un número concreto de columnas, predecir el siguiente entrenándolo con *backpropagation*. Es decir, seleccionando un número de horas, se intentó conocer las siguientes. Esto fue posible convirtiendo los valores de una única columna en varias. El método se observa mejor en la Figura 5.2.

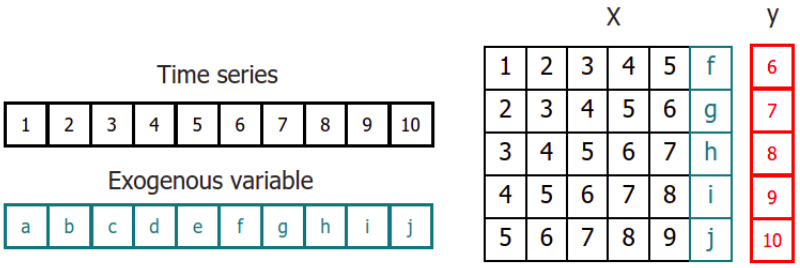

Figura 5.2: Transformación del problema

En concreto, con la función *series_to_supervised()*, la columna de valores se transformó en una tabla de datos de 101 columnas. Estos números fueron transformados en valores entre el -1 y 1 para poder posteriormente usar en el modelo la activación tangencial. A continuación, antes de crear el modelo, se crearon los conjuntos de *test* y *train*. Como dice el ejemplo, en este caso al ser una serie temporal, es importante mantener el orden y no coger valores aleatorios para cada subconjunto. Posteriormente ya se comenzó a elaborar el modelo, entrenarlo y a predecir. Al cabo de aproximadamente 150 *epochs*, se obtuvieron los siguientes resultados de predicción de la actividad motora para los pacientes con depresión.

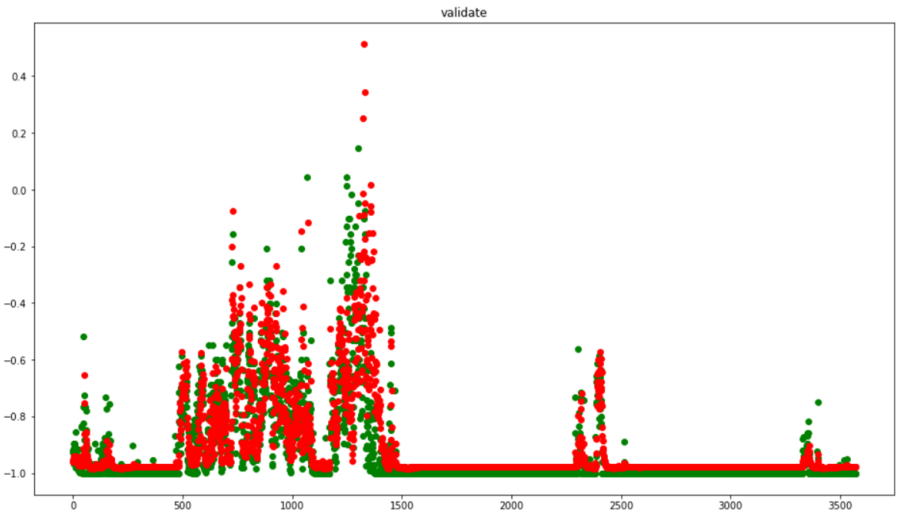

Figura 5.3: Predicción de la actividad motora del trastorno depresivo

Se pueden observar dos colores distintos. Los verdes intentan aproximarse al máximo a los rojos. Así, cuanto más superpuestos estén, mejor será la predicción.

Una vez hecho esto, el ejemplo realiza una prueba introduciendo valores de una semana e intenta predecir la siguiente. En este caso, se procedió de la misma forma. Para el cuarto paciente se seleccionó uno de sus días intermedios para intentar adivinar la actividad motora del siguiente. De nuevo se transformaron los datos pero sin incluir los valores de salida en la tabla. Los resultados obtenidos se compararon con los existentes y se llegó a la conclusión de que la actividad se aproximaba razonablemente a los reales, como se puede observar en la gráfica.

Para la esquizofrenia se realizó el mismo ejercicio. Los resultados se muestran en la Figura 5.4, y de nuevo los puntos verdes se aproximan a los rojos. Para este caso, los datos del primer paciente, con trastorno esquizofrénico paranoide, fueron los empleados.

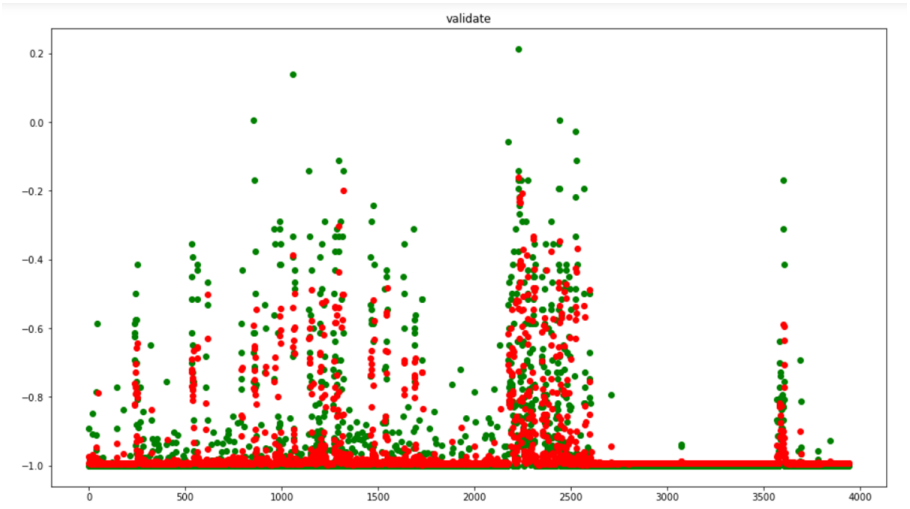

Figura 5.4: Predicción de la actividad motora del trastorno esquizofrénico

Siguiendo los mismos pasos de transformación de datos, división de subconjuntos (entrenamiento y validación), y creación del modelo, se obtuvieron de nuevos unos datos semejantes a los de un día de una persona con esquizofrenia.

Capítulo 6

Conclusiones y trabajo futuro

Para terminar, en este último apartado se comentarán ciertas conclusiones y algunas observaciones para mejorar en un posible futuro trabajo o que han podido dificultar el proceso.

Primero de todo la cantidad de observaciones. Un total de 77 pacientes son pocos para poder aplicar métodos de aprendizaje automático. La división de estos valores según el número de sus respectivos días de tratamiento fue una buena solución para poder obtener mejores resultados. El cambio a casi 300 valores por enfermedad fue suficiente para proceder con la parte práctica.

A la hora de desempeñar las dos clasificaciones principales (entre el trastorno depresivo monopolar y bipolar, y el esquizofrénico paranoide y no paranoide), se alcanzaron porcentajes muy parecidos. En 3 de los 4 métodos, el modelo que mejores resultados obtuvo fue el de keras. La regresión logística sí que mostró diferencias significativas cuando se trabajó con la actividad motora de los trastornos por separado, llegando a diferenciarse casi un 20 % en la predicción con las redes neuronales y keras. Con las distintas propiedades (e.g., edad, actividad motora media diaria, MADRS, BPRS) no existió apenas variedad. Por ello, fue el que se introdujo en la aplicación de las librerías SHAP y LIME. Esto es debido a la cantidad de información y ejemplos que había sobre este modelo.

Una vez obtenidas las explicaciones de los resultados, se llegó a conclusiones similares a las del segundo artículo [18]. En este, gracias a los UMAPs, se observaba que la actividad de los pacientes con trastorno depresivo era más dispersa y la de los pacientes con esquizofrenia más cercana. Los gráficos representados por SHAP y LIME indican que la actividad motora media diaria ('*día*') es muy influyente en los resultados de la categorización de la esquizofrenia paranoide o no paranoide, y casi

nada importante en la depresión. Esto confirma un patrón en el comportamiento de los pacientes con esquizofrenia más estructurado que en las personas con trastorno depresivo.

Respecto a la predicción futura de la actividad de un paciente, hay poco que añadir más allá de lo comentado en el Capítulo 5. Las gráficas señalan como en ambos casos los puntos verdes se aproximan considerablemente a los rojos.

Para un trabajo futuro se propone estudiar pacientes actuales. Si se pudieran recolectar nuevos datos de la actividad motora de personas con esquizofrenia y depresión, se podría realizar una comparación con los valores usados en el presente trabajo. Como los dos conjuntos de datos fueron recogidos antes del 2010, una nueva recolección de estos podría ayudar a sacar otras conclusiones u observar un patrón diferente. El tamaño de observaciones debería ser más grande ya que fue difícil encontrar diferencias entre los pacientes de una misma enfermedad pero distinto tipo, ya sea una esquizofrenia paranoide o no paranoide, o una depresión monopolar, bipolar I o bipolar II.

Con los avances de la tecnología y de la IA, la actigrafía y el reloj actigráfico también han evolucionado. Dentro del reloj *MotionWatch 8* se han hecho mejoras y han aparecido nuevos dispositivos en el mercado. Por ejemplo, el *PRO-Diary* es un nuevo reloj de muñeca que combina la actigrafía con un diario electrónico del paciente. Además de recolectar los datos del movimiento del sujeto con un acelerómetro triaxial, también hace preguntas a este que puede responder de manera táctil. El uso de este aparato, podría obtener unos resultados más precisos o nuevos valores que ayuden a un estudio algo diferente.

Capítulo 7

Introduction

This chapter provides a brief introduction and motivation for the content of the work. It also explains some initial objectives that were set and that must be fulfilled at the end of the reading. And finally, a description of the memory structure.

7.1. Motivation

The Internet of Things describes the network of objects with embedded sensors, software and other mechanisms whose purpose is to connect to each other and exchange information and data with other devices and systems over the Internet. In recent years it has become one of the most important technologies of the century. It is now possible to connect all kinds of objects to the Internet through integrated devices, making communication easier. Among these instruments, the most important ones are portable sensors, which allow ambulatory monitoring of various data and are becoming increasingly common in many fields, such as health. By collecting long-term information, they can advance the diagnosis of a disease. Today they are widely used in the field of mental health. They can provide information on different physiological variables, such as blood pressure, heart rate or physical activity.

Depression or major depressive disorder is an emotional illness that causes feelings of constant sadness and a loss of interest in certain activities. It affects a person's feelings, thoughts and behaviors. On the other hand, schizophrenia is a more serious mental illness that affects the way the subject thinks, feels and behaves. These patients may appear to have lost touch with reality.

This work has been done to introduce readers to the concept of Explainable Artifi-

cial Intelligence. Nowadays it is not very well known, however, it is very useful for the realization of machine learning models, classification and prediction methods, since it provides explanations to the results obtained and helps to understand them. In turn, it also works with two common and important diseases in the world and makes them a little more known to the readers. And finally, I wanted to find and compare different methods to apply to the same data set and with them to reach similar conclusions. With the description of the two articles, it is intended to show this facet of Artificial Intelligence.

7.2. Brief description of the work

The sensors are incorporated in actigraphic watches worn by patients with major depressive episode, both monopolar and bipolar, and subjects with schizophrenia, both paranoid and nonparanoid. Specifically, 23 depressed patients, 32 healthy controls and 22 people with schizophrenia wore the watch for approximately two weeks (more or less days depending on the patient), and motor activity was measured for each minute over that time. In particular, the collected values of motor activity have been used as a time series, as these are defined as successions of data measured at certain times in chronological order. By treating them in such a way, it has been possible to differentiate, filter, search from specific dates and, most importantly, study future behavior. From one day of a patient of each disease, we have tried to predict, with the motor activity per minute, the following 24 hours and, subsequently, to compare with the initial values.

The initial data for these two diseases have been somewhat modified. First, the subjects were divided by each day of treatment, so that for depression we went from having 23 observations to 291, for schizophrenia, from 22 to 285 observations, and for the healthy controls, instead of the initial 32, we worked with 402. On the other hand, the variables for each of them also underwent some changes that will be explained more exhaustively in the corresponding section in Chapter 3.

Artificial Intelligence, defined as a simulation of human intelligence that creates algorithms and computer systems capable of executing simple and complex tasks, allows different classification and prediction methods to be carried out throughout the work to see if it was possible to distinguish, based on different properties of a patient or on motor activity, the type of disorder the patient was suffering from. Once these models have been made, questions may arise such as: which of the patient's properties most influences the type of disease he/she has?, are the same results for each person? or, what happens within a model? Most of them can be solved with Explainable Artificial Intelligence (XAI), a set of techniques, processes and strate-

gies that provide explanations for the predictions, recommendations and decisions of intelligent systems.

There are many ways to classify XAI and different methods to apply it, but in this work SHAP and LIME are used. SHAP uses calculations from the field of game theory to find out which variables have the most influence on the predictions of machine learning techniques. On the other hand, LIME uses the black box method to determine the minimum number of variables that generate the maximum hit probability, and thus explain why a given instance was classified with the assigned class.

The number of studies, comparisons and searches that can be made from the same set of data is immense. This is another benefit provided by machine learning or AI, the possibility of arriving at the same results from different methods and comparing each one. Two reports were selected to talk about them and to see other possible ways to study this data. One from *Google Scholar*, which makes a comparison between the three types of patients based on the calculation of three non-parametric measures, inter-day stability, intra-day variability and relative amplitude. The other, from *ScienceDirect*, is based on representing UMAPs, an unsupervised, nonlinear method of dimensionality reduction.

7.3. Objectives

The main objective in proceeding with this work was to try to find a pattern in the motor activity of each of the subjects, i.e., to see relationships at certain times of the day in different patients with the same disease, and to differentiate a patient with schizophrenic disorder from one with depressive disorder only by their motor activity. In addition, we also tried to learn and understand the models through XAI. As this is a branch that is becoming more and more known but there is still not much information available, the aim was to know in detail the two methods (SHAP and LIME) and what each one contributes. And finally, to see and learn about the various studies that can be done on the same data set.

7.4. Structure

The structure of the report is divided into four parts. The first one, formed by Chapter 2, details the theoretical framework necessary to correctly understand the whole work. It explains in brief, what is Explainable Artificial Intelligence, as well as actigraphy and time series.

The second part, consisting of Chapter 3 and Chapter 4, contains the main part. It clarifies the two disorders being discussed, the origin of the data being worked with and the creation of new ones. In addition, the classification methods are also presented, which for the depressive disorder differentiates between bipolar or monopolar, and for the schizophrenic whether the illness is paranoid or not. As a final part of Chapter 3, four comparisons are made between the variety of patients on the basis of average motor activity for each hour of the day. Finally, the application of XAI to these models. Both SHAP and LIME are applied to the first classifications, two very useful and well-known libraries that allow detailing the most important and influential variables of the model, as well as an explanation of the results obtained.

Chapter 5 composes the third part. In this, a prediction method is performed for each data set of the two diseases. Taking these as time series, since they measure motor activity per minute, the future behavior and a possible prediction of subsequent days are studied.

The last component is the Chapter **??**, in which a small comparison with other articles is made. In this part, it is detailed how different studies can be made from the same dataset, but at the same time reaching a similar conclusion.

Finally, a brief summary of the conclusions drawn from the work and a small plan for the future is written.

Capítulo 8

Conclusions and future work

Finally, in this last section we will talk about certain observations or conclusions to be improved in a possible future work or that may have hindered the process. First of all, the number of observations. A total of 77 patients is too few to be able to apply machine learning methods. The division of these values according to the number of their respective treatment days was a good solution in order to obtain better results. The change to almost 300 values per disease was sufficient to proceed with the practical part.

When performing the two classifications, very similar percentages were achieved. In 3 of the 4 methods, the model with the best results was the keras model. Logistic regression did show significant differences when working with the motor activity of the disorders, reaching a difference of almost 20 % in the prediction with the neural networks and keras. With the different properties (e.g., age, average daily motor activity, MADRS, BPRS) there was hardly any variety. Therefore, it was the one that was introduced in the application of the SHAP and LIME libraries. This is due to the amount of information and examples that were available on this model.

Once the explanations of the results were obtained, similar conclusions were reached to those of the second article [18]. In this one, thanks to the UMAPs, it was observed that the activity of patients with depressive disorder was more dispersed and that of patients with schizophrenia was closer. The graphs plotted by SHAP and LIME indicate that average daily motor activity ('*day*') is very influential in the results of paranoid or non-paranoid schizophrenia categorization, and almost not at all important in depression. This confirms a more structured pattern in the behavior of patients with schizophrenia than in people with depressive disorder.

Regarding the future prediction of a patient's activity, there is little to add beyond

what was discussed in Chapter 5. The graphs show how in both cases the green points are considerably closer to the red ones.

For future work it is proposed to study current patients. If new motor activity data could be collected from people with schizophrenia and depression, a comparison could be made with the values used in the present work. As the two datasets were collected before 2010, a new collection might help to draw other conclusions or observe a different pattern. The size of observations should be larger as it was difficult to find differences between patients of the same illness but different type, whether paranoid or non-paranoid schizophrenia, or monopolar, bipolar I or bipolar II depression.

With advances in technology and AI, actigraphy and actigraphic watch have also evolved. Within the *MotionWatch 8* improvements have been made and new devices have appeared on the market. For example, the *PRO-Diary* is a new wristwatch that combines actigraphy with an electronic patient diary. In addition to collecting the subject's movement data with a triaxial accelerometer, it also asks the subject questions that it can answer tactilely. The use of this device could lead to more accurate results or new values that would help in a somewhat different study.

Bibliografía

[1] Asberg M. (1978). Montgomery and Asberg Depression Rating Scale (Madrs). Academic Department of Psychiatry. https://www.veale.co.uk/wp-content/uploads/2010/10/MADRS.pdf

[2] Bagnato J.I. (2019). Pronóstico de Series Temporales con Redes Neuronales en Python. *Aprende machine learning.* Aprende Machine Learning. https://www.aprendemachinelearning.com/pronostico-de-series-temporales-con-redes-neuronales-en-python/

[3] Berle J.O., Hauge E.R., Oedegaard K.J., Holsten F., Fasmer O.B. (2010). Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia then in major depression. *National Library of Medicine.* BMC Research Notes. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2890507/

[4] CamNtech (2022). MotionWatch 8. *Actigraphy for Sleep, Chronobiology and Physical Activity.* CamNtech. https://www.camntech.com/motionwatch-8/

[5] Cañadas Bustos D. (2021). Cinco tipos de esquizofrenia y sus causas. *Enfermedades Psiquiátricas.* Salud Blogs Mapfre. https://www.salud.mapfre.es/enfermedades/psiquiatricas/los-tipos-de-esquizofrenia/#:~:text=La%20esquizofrenia%20es%20una%20enfermedad,%2C%20catat%C3%B3nica%2C%20indiferenciada%20y%20residual

[6] Esparza Catalán C. Series temporales. Laboratorio de Estadística CSIC. http://humanidades.cchs.csic.es/cchs/web_UAE/tutoriales/PDF/SeriesTemporales.pdf

[7] Ferré A. (2022). ¿Cómo cuantificamos el sueño o el ritmo de sueño? ¿Qué es la Actigrafía? *Pruebas del sueño.* Medicina del sueño. https://doctorferre.com/pruebas-del-sueno/actigrafia/

[8] Garcia-Ceja E., Riegler M.A., Jakobsen P., Torresen J. Nordgreen T., Oedegaard K.J., Fasmer O.B. (2018). The Depresjon Dataset. *Depresjon*. Simula. https://datasets.simula.no/depresjon/

[9] Hum Kinet J. (2012). Intra- and Inter-Instrument Reliability of the Actiwatch 4 Accelerometer in a Mechanical Laboratory Setting. *PubMed Central*. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3588665/

[10] Jakobsen P., Garcia-Ceja E., Stabell L.A., Oedegaard K.J., Berle J.O., Thambawita V., Hicks S.A., Halvorsen P., Fasmer O.B., Riegler M.A. (2020). A Motor Activity Database of Patients with Schizophrenia.*Psykose*. Simula. https://datasets.simula.no/psykose/

[11] Kök I., Yildirimi Okay F., Muyanli Ö., Özdemir S. (2022). Explainable Aritificial Intelligence (XAI) for Internet of Things: A Survey. *Cornell University*. Arxiv. https://arxiv.org/pdf/2206.04800.pdf

[12] Leyton A. F., Barrera A. (2010). El diagnóstico diferencial entre la Depresión Bipolar y la Depresión Monopolar en la práctica clínica. *Revista médica de Chile*. Scielo. https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0034-98872010000600017

[13] Lundberg S. (2018). Welcome to the SHAP documentation. *SHAP*. https://shap.readthedocs.io/en/latest/index.html

[14] Ortuño Sánchez-Pedreño F. (2022). Trastorno bipolar. Clínica Universidad de Navarra. https://www.cun.es/enfermedades-tratamientos/enfermedades/trastorno-bipolar#:~:text=El%20trastorno%20bipolar%20es%20trastorno,(episodios%20man%C3%ADacos%20o%20hipoman%C3%ADacos)

[15] O., G. Brief Psychiatric Rating Scale (BPRS). https://cdn.sanity.io/files/0vv8moc6/psychtimes/685845afc5dcf7058b340eea897eed10cc4f0b1c.pdf/bprsform.pdf

[16] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay, E. (2011). Logistic Regression. *Machine Learning in Python*. Scikit-Learn. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

[17] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay, E. (2011). MLPClassifier. *Machine Learning in Python*. Scikit-Learn. https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

[18] Price G.D., Heinz M.V., Zhao D., Nemesure M., Ruan F., Jacobson N.C. (2022). An unsupervised machine learning approach using passive movement data to understand depression and schizophrenia. *Journal of Affective Disorders*. Elsevier. https://www.sciencedirect.com/science/article/pii/S0165032722008631?ref=pdf_download&fr=RR-2&rr=7cd479fb3924ff08

[19] Radecic D. (2020). LIME vs. SHAP: Which is Better for Explaining Machine Learning Models?. *Two of the most popular Explainers compared*. Towards Data Science. https://towardsdatascience.com

[20] Ribeiero Tulio M. (2016). Local Interpretable Model-Agnostic Explanations (lime). *lime*. https://lime-ml.readthedocs.io/en/latest/index.html

[21] Salazar J., Silvestre S. (2017). Internet de las cosas. *TechPedia*. Erasmus+. https://psm.fei.stuba.sk/pages/95/LM08_F_ES.pdf

[22] Torres F. (2020). What is Schizophrenia? *Schizophrenia*. American Psychiatric Association. https://www.psychiatry.org/patients-families/schizophrenia/what-is-schizophrenia#:~:text=Schizophrenia%20is%20a%20chronic%20brain,thinking%20and%20lack%20of%20motivation

[23] ¿Qué es la Inteligencia Artificial? (2023). *Innovación*. Iberdrola. https://www.iberdrola.com/innovacion/que-es-inteligencia-artificial#:~:text=La%20Inteligencia%20Artificial%20(IA)%20es,a%20d%C3%ADa%20a%20todas%20horas.

[24] Github: https://github.com/anajim13/TFM-XAI.git