



ugr | Universidad
de **Granada**

TRABAJO FIN DE GRADO

GRADO EN INGENIERÍA INFORMÁTICA

POSTCOVID-AI analytics

modeling human behaviour at population level

Autor

Álvaro Santana Sánchez

Directores

Oresti Baños Legrán

Claudia Villalonga Palliser



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

Granada, Noviembre de 2023

POSTCOVID-AI analytics: modeling human behaviour at population level

Álvaro Santana Sánchez

Palabras clave: predicción, forecasting, análisis de series temporales, deep learning

Resumen

El siguiente trabajo aborda el problema de la predicción de actividad física de una población de individuos. Los datos de dicha población, fueron recopilados a través de dispositivos smartphone durante un periodo de aproximadamente un mes. Durante este periodo de tiempo, fue utilizada ActivityRecognitionApi de Google para obtener la actividad realizada por los participantes con una frecuencia de 1 minuto. Posteriormente los valores de estas actividades fueron pasados a su equivalente en una medida metabólica que representa las calorías que quema una actividad de la vida cotidiana, los METs, y dispuestos de manera longitudinal en series temporales a nivel individual. Mediante un modelo de memoria a corto y largo plazo LSTM se comparan diferentes *pipelines* en el proceso de predicción en pro de obtener la mejor alternativa para la predicción de los valores de actividad poblacional. Los datos son presentados en series temporales agregadas, en series temporales individuales que son agregadas a posteriori de la salida del modelo, y en matrices de datos que incluyen en cada fila un individuo de la población. Las ventajas y desventajas de los modelos son expuestas durante el trabajo así como el desempeño de los mismos en la predicción de actividad física. Algunos de estos cuentan con mayor información poblacional, otros individual e incluso toda la información teniendo por ende un rango diferente de entradas más y menos sencillas que contienen una mayor o menor información. Además, se trabaja con distintos horizontes de predicción para ver como estos interactúan con los modelos utilizados. Concretamente se realizará una predicción sobre de intervalo de 2 horas, la cuál será realizada tanto en esta escala como en su correspondiente en minutos obteniendo también predicciones de 120 puntos.

Los resultados indican que la predicción de METs es posible mediante un modelo LSTM para cada una de las alternativas. El modelo con series temporales agregadas es el que mejor comportamiento expuso respecto al resto, siendo el de mayor capacidad para detectar tendencias específicas para intervalos de tiempo más pequeños así como las tendencias generales en la línea de actividad de la población. Los modelos individual y matricial también presentan resultados positivos en la predicción a nivel global, exponiendo menor facilidad para detectar cambios bruscos o repentinos en

los valores de salida. Debido a ello, estos modelos pueden exponer algún outliers y mayor variabilidad en sus respuestas, haciendo de ellos modelos menos sólidos. Para el total de METs consumidos durante todo el intervalo de tiempo el modelo Agregado obtuvo un error medio absoluto de 188,69 en el horizonte en horas y 190,68 en el horizonte en minutos. El modelo individual 248,55 y 358,41 para estos mismos horizontes y finalmente el matricial 307,83 y 234,56. Todos ellos respecto a una media de 4027,03 METs. También han sido detectados otros valores que inclinan la balanza hacia el uso de predicciones más genéricas utilizando horizontes compuestos por un número menor de puntos. Mediante este trabajo se refuerza la posibilidad de realizar predicciones sobre la actividad física y se llega a la conclusión de que el uso de agregación a nivel poblacional y la predicción de salidas de menor número de valores son los dos mejores enfoques para predecir valores de actividad.

POSTCOVID-AI analytics: modeling human behaviour at population level

Álvaro Santana Sánchez

Keywords: predicción, forecasting, análisis de series temporales, deep learning

Abstract

This investigation delves into the challenge of predicting physical activity within a demographic using data collected from smartphone devices. The data, spanning approximately one month, was acquired through the utilization of the Google ActivityRecognitionApi, capturing minute-by-minute activities. These activities were subsequently converted into metabolic equivalent of task (MET) values and organized into individual-level time series datasets. Employing Long Short-Term Memory (LSTM) models, diverse prediction methodologies were compared to ascertain the most effective means of forecasting population-wide activity levels. The data presentation comprised aggregated time series, individual time series subsequently aggregated, and matrices of data, reflecting varying degrees of intricacy and information incorporation. Multiple prediction horizons were investigated, encompassing two-hour intervals and their corresponding minute-level predictions.

The outcomes suggest that forecasting MET values using LSTM models is viable across all methodologies. The aggregated time series model exhibited optimal performance, excelling in discerning specific trends within shorter time intervals and general activity patterns within the population. The individual and matrix-based models also generated positive results, though they demonstrated reduced sensitivity to abrupt variations in output values, displaying some degree of variability and the occasional outlier. In terms of the total METs expended over the entire target timeframe, the aggregated model yielded mean absolute errors of 188.69 for hourly predictions and 190.68 for minute-level forecasts. The individual model reported errors of 248.55 and 358.41, while the matrix-based model displayed errors of 307.83 and 234.56, relative to an average of 4027.03 METs. These findings suggest that, in specific scenarios, favoring more generalized predictions with fewer data points may be a prudent approach in the forecasting process. This research contributes insights into the capabilities of LSTM models in predicting physical activity and underscores the effectiveness of an aggregated time series approach in capturing overarching trends within the population.

Yo, **Álvaro Santana Sánchez**, alumno de la titulación de Ingeniería Informática de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI 77770603J, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Álvaro Santana Sánchez

Granada a Noviembre de 2023.

D. Oresti Baños Legrán, Profesor del Departamento de Ingeniería de Computadores, Automática y Robótica de la Universidad de Granada.

D. Claudia Villalonga Palliser, Profesor del Departamento Departamento de Ingeniería de Computadores, Automática y Robótica de la Universidad de Granada.

Informan:

Que el presente trabajo, titulado ***POSTCOVID-AI analytics: modeling human behaviour at population level***, ha sido realizado bajo su supervisión por **Álvaro Santana Sánchez**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a Noviembre de 2023.

Los directores:

Oresti Baños Legrán Claudia Villalonga Palliser

Agradecimientos

Primero de todo quiero expresar mi gratitud a los tutores del trabajo, tanto doctor Oresti Baños Legrán como la doctora Claudia Villalonga Palliser, que me han guiado y acompañado durante la realización del mismo. Así de la misma manera a Don Ciro Rodríguez León cuya supervisión ha seguido de cerca mi trabajo y sido de gran utilidad en los avances del mismo. Sin su orientación experta y consejos de la materia este trabajo no podría haber sido posible.

También agradecer a la Universidad de Granada y todo su equipo por las enseñanzas proporcionadas durante los años de estudio que me han brindado la posibilidad de llevar a cabo este trabajo de fin de grado.

Agradecer a todas las personas que participaron en el estudio de cuyos datos nace la posibilidad de este mismo.

Finalmente, agradecer a mis seres queridos por el acompañamiento durante todo el proceso de realización así como el apoyo moral aportado.

Índice general

1. Introducción	1
1.1. Contexto	1
1.2. Planteamiento del problema	4
1.3. Motivación	5
1.4. Solución propuesta	6
1.5. Objetivos	7
2. Enfoques y evolución en el campo de la predicción	8
2.1. Predicción de series temporales	8
2.2. Modelos estadísticos	9
2.2.1. Auto-Regresivos	10
2.2.2. Media-Móvil	10
2.2.3. Integrado	10
2.3. Modelos Computacionales	12
2.3.1. Machine Learning: dificultades y bondades	12
2.3.2. Deep Learning	13
3. Estado del arte	15
4. Fundamentos teóricos	19
4.1. RNN	19
4.1.1. Problemas en los gradientes	20
4.1.2. Long Short Term Memory	20
5. Gestión del Proyecto	25
5.1. Metodología de desarrollo	25
5.2. Planificación	26
5.2.1. Hitos	26
5.3. Presupuesto	28
6. Datos	30
6.1. Utilización de METs	31
6.1.1. Reducción del dataset	32

7. Metodología	34
7.1. Diseño	35
7.1.1. Alternativa 1: Entrada individual de elementos 1xM con agregación al final	35
7.1.2. Alternativa 2: entrada matricial de elementos NxM . .	36
7.1.3. Alternativa 3: entrada agregada de elementos 1XM previamente agregados	37
7.1.4. M: Ventana de tiempo	38
7.2. P: Horizonte de predicción	38
7.2.1. Horizonte 1: Predicción en minutos	39
7.2.2. Horizonte 2: Predicción en horas	39
7.3. Implementación	40
7.3.1. Aprendizaje	40
7.3.2. Test: 9-fold Cross-Validation	46
8. Evaluación	49
8.1. Resultados en las predicciones de test	51
8.1.1. Resultados punto a punto	51
8.1.2. Resultados para el valor total	53
8.1.3. Comportamiento del modelo en largo plazo	54
8.1.4. Peores y mejores resultados por split	67
8.2. Discusión	72
8.2.1. Diferencias entre los modelos	72
8.2.2. Comportamiento en bloque	74
9. Conclusiones y trabajos futuros	77
9.1. Conclusiones	77
9.2. Trabajos Futuros	79

Índice de figuras

1.1. Estimación del INE de defunciones semanales 2019-2023. Fuente: [1].	2
1.2. Tasa de variación anual de la ocupación INE durante los primeros años de pandemia. Fuente: [2].	3
2.1. Framework de construcción de las series temporales. Fuente: [3].	11
2.2. Histograma de publicaciones de modelos DL en el campo financiero. Fuente: [4].	14
3.1. Resultados en la predicción de Actividad Física. Fuente: [5].	16
3.2. Arquitectura BeWell24Plus. Fuente: [6].	17
4.1. RNN básica. Fuente: [7].	19
4.2. Arquitectura LSTM. Fuente: [7].	21
4.3. Forget Gate. Fuente: [7].	22
4.4. Input Gate. Fuente: [7].	23
4.5. Output Gate. Fuente: [7].	24
5.1. Ejemplo de metodología en cascada. Fuente: [8].	25
5.2. Diagrama de Gannt con la planificación del proyecto	27
6.1. Ejemplo de filas del dataframe de datos procesado.	32
6.2. Consumo de METs poblacional de los primeros 24 días para 100 participantes.	33
6.3. Consumo de METs poblacional de los primeros 24 días para 25 participantes.	33
7.1. Framework con entrada de datos a nivel individual.	35
7.2. Framework con entrada de datos a nivel matricial	36
7.3. Framework con entrada de datos agregados al inicio	37
7.4. Ejemplo de horizonte predicho con formato en minutos.	39
7.5. Ejemplo de horizonte predicho con formato en horas.	40
7.6. Definición del modelo con entradas matriciales.	41
7.7. Definición del modelo con entradas matriciales.	43

7.8.	Funcionamiento de EarlyStopping. Fuente: [9].	44
7.9.	9 fold cross validation.	47
7.10.	Representación Fold 1.	48
8.1.	Representación por serie temporal de los valores punto a punto y totales.	50
8.2.	Representación del calculo de errores punto a punto.	51
8.3.	Representación del calculo de errores totales.	53
8.4.	Representación de la extracción del experimento a largo plazo para el split 1.	55
8.5.	Predicción en horas de dos días para el split 1 modelo Agregado.	57
8.6.	Recta de Mínimos Cuadrados Ordinarios para el modelo de entradas agregadas con horizonte en horas.	57
8.7.	Predicción en horas de dos días para el split 1 modelo Individual.	58
8.8.	Recta de Mínimos Cuadrados Ordinarios para el modelo de entradas individuales con horizonte en horas.	58
8.9.	Predicción en horas de dos días para el split 1 modelo Matricial.	59
8.10.	Recta de Mínimos Cuadrados Ordinarios para el modelo de entradas matriciales con horizonte en horas.	59
8.11.	Predicción en minutos de dos días para el split 1 modelo Agregado.	60
8.12.	Recta de Mínimos Cuadrados Ordinarios para el modelo de entradas agregadas con horizonte en minutos.	60
8.13.	Predicción en minutos de dos días para el split 1 para el modelo Individual.	61
8.14.	Recta de Mínimos Cuadrados Ordinarios entradas individuales.	61
8.15.	Predicción en minutos de dos días para el split 1 modelo Matricial con el horizonte en minutos.	62
8.16.	Recta de Mínimos Cuadrados Ordinarios para el modelo de entradas matriciales con el horizonte en minutos.	62
8.17.	Predicción de un solo valor en cada intervalo predicho sobre el modelo de entradas agregadas para el horizonte en minutos.	64
8.18.	Recta de mínimos cuadrados para la predicción sumada de entradas agregadas.	64
8.19.	Predicción de un solo valor en cada intervalo predicho sobre el modelo de entradas individuales en minutos.	65
8.20.	Recta de mínimos cuadrados para la predicción sumada de entradas individuales.	65
8.21.	Predicción de un solo valor en cada intervalo predicho para el modelo de entradas matriciales en minutos.	66
8.22.	Recta de mínimos cuadrados para la predicción sumada de entradas matriciales.	66
8.23.	Mejor MAE para el modelo agregado en el split 1.	68
8.24.	Mejor MAE para el modelo individual en el split 1.	69

8.25. Mejor MAE para el modelo matricial en el split 1.	69
8.26. Peor MAE para el modelo agregado en el split 1.	71
8.27. Peor MAE para el modelo individual en el split 1.	71
8.28. Peor MAE para el modelo matricial en el split 1.	72
8.29. Serie temporal agregada a nivel poblacional durante el día 2.	73
8.30. Serie temporal correspondiente al participante 0 durante el día 2.	73

Índice de tablas

5.1.	Estimación precio de recursos humanos.	28
5.2.	Estimación precio de recursos informáticos.	29
6.1.	Equivalencias entre actividad física y METs.	32
8.1.	Valores medios a predecir horizonte en horas	50
8.2.	Valores medios a predecir horizonte en minutos	50
8.3.	Resultados punto a punto para el horizonte en horas.	51
8.4.	Resultados punto a punto para el horizonte en minutos.	52
8.5.	Resultados para el horizonte en horas	53
8.6.	Resultados para el horizonte en minutos	54
8.7.	Comportamiento a largo plazo horizonte en horas	55
8.8.	Comportamiento a largo plazo horizonte en minutos	56
8.9.	Mejores resultados por split horizonte en horas.	67
8.10.	Mejores resultados por split horizonte en minutos.	68
8.11.	Peores resultados horizonte en horas.	70
8.12.	Peores resultados horizonte en minutos.	70
1.	Valores medios a predecir horizonte en horas.	84
2.	Valores medios a predecir horizonte en minutos.	84
3.	Resultados punto a punto para el horizonte en horas.	85
4.	Resultados punto a punto para el horizonte en minutos.	86
5.	Resultados totales para el horizonte en horas.	87
6.	Resultados totales para el horizonte en minutos.	88
7.	Comportamiento a largo plazo para el horizonte en horas.	88
8.	Comportamiento a largo plazo para el horizonte en minutos.	89

Capítulo 1

Introducción

1.1. Contexto

El 31 de enero de 2020 es diagnosticado el primer caso español de COVID-19 (SARS-CoV-2) en San Sebastián de la Gomera. La alta capacidad contagiosa del virus unido al desconocimiento en la prevención y trata del mismo, llevaron a una rápida expansión de la enfermedad provocando un alto grado de preocupación en la población. Este continuó creciendo hasta que finalmente el gobierno declaró el confinamiento de la población el 14 de marzo de 2020. Dicho periodo se extendió durante 98 largos días y fue seguido de otros dos Estados de Alarma [10]. Durante este tiempo los ciudadanos españoles tuvieron que enfrentarse a un nuevo estilo de vida más restrictivo, donde las actividades cotidianas se veían limitadas al índole del hogar, el aislamiento y soledad estaban a la orden del día y el estrés producido por la preocupación hacia los más vulnerables era una realidad dentro de la sociedad.

Se estima que en España a finales de 2020 la pandemia ya había causado más de 80.000 defunciones y que durante el año 2022 más de 92.591 españoles fallecieron con una prueba PCR positiva [10]. Para conocer realmente como de devastadores fueron estos datos en la población, basta con mirar el crecimiento en los casos de defunción dentro del país a raíz de la pandemia y compararlos con los casos registrados en 2019. Estos pueden ser encontrados en la estimación de defunciones del Instituto Nacional de Estadística para los años de pandemia [1]. Observe el pico de mortalidad alcanzado entorno a la semana 14 en la Figura 1.1 para el año 2020.

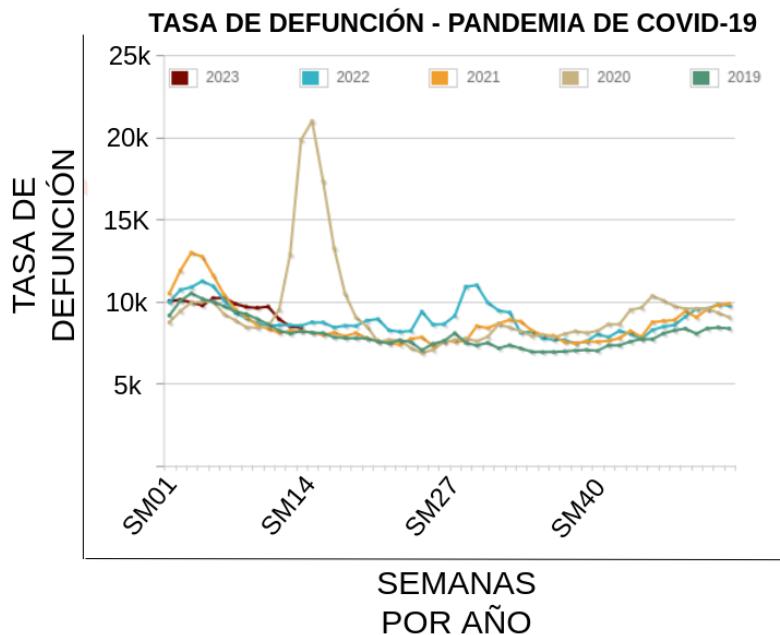


Figura 1.1: Estimación del INE de defunciones semanales 2019-2023. Fuente: [1].

Todas estas dificultades junto a otras como el deterioro económico y la pérdida de empleo, han sido sin lugar a dudas el detonante de un cambio sustancial en la población. Actualmente, en 2023, la sociedad continúa enfrentando las consecuencias que la pandemia ha traído.

Muchos de estos valores son sencillos de cuantificar. Véase cómo aumenta o disminuye la tasa de paro en los diferentes cuatrimestres del año o el número de ERTE expedidos por el gobierno. Un simple vistazo a una gráfica temporal de estos valores aporta una gran información sobre el comportamiento y la tendencias de los mismos. Obsérvese como la Figura 1.2 muestra una reducción drástica de la tasa de ocupación al inicio de la pandemia, la cual va reduciéndose hasta volver a mejorar en 2021.

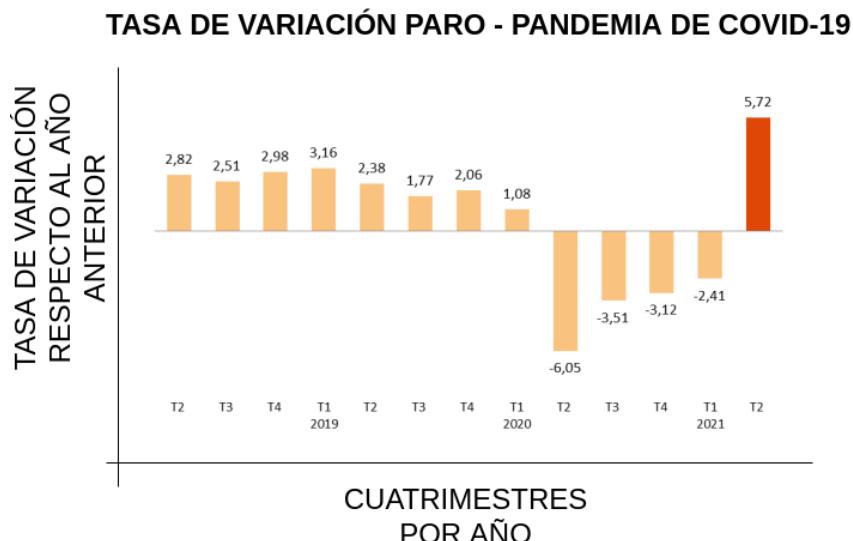


Figura 1.2: Tasa de variación anual de la ocupación INE durante los primeros años de pandemia. Fuente: [2].

Sin embargo, otros factores como el bienestar o el impacto emocional no son tan sencillos de medir. En este contexto aparece POSTCOVID-AI, un proyecto de investigación financiado por la Convocatoria de Investigación Social de la Fundación "la Caixa" 2020 en el que investigadores de la Universidad de Granada buscan contestar a la pregunta: ¿Qué factores influencian el bienestar de la población durante la pandemia y en el periodo posterior? Este es un gran proyecto con diversas vertientes y líneas de investigación, el cual planteó cuatro objetivos generales que son citados a continuación [11]:

1. Crear un conjunto de datos público, grande y anónimo de datos longitudinales sociales, conductuales y emocionales en el contexto posterior al COVID-19.
2. Usar algoritmos de Inteligencia Artificial para implementar un framework inteligente que nos ayude a comprender los factores de resiliencia relacionados con el COVID.
3. Contribuir al desarrollo de nuevas directrices, políticas y recomendaciones informadas por la evidencia generada a través del framework metodológico.
4. Avanzar en el campo científico de las ciencias computacionales psicológicas, conductuales y sociales a través de la monitorización y análisis del contexto de la vida diaria.

Este trabajo de fin de grado busca la contribución al desarrollo de los puntos anteriormente mencionados y esta especialmente relacionado con el segundo de ellos. La aportación se realizará mediante el enfoque de la resolución de un problema concreto, y que además pueda aportar un cierto valor en el campo de las ciencias computacionales y los sistemas inteligentes.

1.2. Planteamiento del problema

Se aborda el problema de estimación de actividad física de una población durante un periodo de tiempo futuro, a partir de datos de actividad recaudados para dicha población. Esto requiere de análisis objetivo poblacional de la actividad física, que permita posteriormente realizar predicciones de la misma. Algunas de las problemáticas relacionadas con dicho análisis son muy similares o vendrán derivados a los problemas metodológicos que podemos encontrar al plantear la recopilación de estos mismos datos de actividad física así como su procesado. El análisis poblacional debe tratar los datos de manera que consiga una representación fiel a los valores de actividad física medidos y permitir posteriormente realizar un análisis objetivo a nivel poblacional. Los datos a nivel individual por ende deben disponerse de manera que permitan posteriormente su puesta en común a nivel de población. Algunos de los puntos principales que pueden tener una influencia directa en el análisis de actividad física poblacional y su predicción son los siguientes:

- ¿Cómo es obtenido el indicador de actividad física utilizado? Debe hacerse de una manera que permita extender su medición a un grupo de personas representativo de cualquier población, de tal forma que sea lo menos intensivo posible para los individuos de la misma. Medidas como las tomadas en las pruebas realizadas a los futbolistas en los equipos de primer nivel como electrocardiogramas o epiometrías no son realmente escalables a una población grande de individuos. Esto se debe a factores como el precio o la cantidad de equipamiento y recursos que son necesarios para tomar estos exámenes. Por tanto, el análisis requerirá lidiar con datos que probablemente no sean todo lo indicativos que se pudiera desear en este ámbito. Lo cuál llevaría al siguiente punto.
- ¿Cómo de representativo es el indicador obtenido? La medición de una actividad física no es sencilla. Una persona puede pasar de hacer sentadillas, a realizar un trote suave o un sprint, o desde montar en bicicleta a realizar yoga o un deporte de contacto... Dependiendo de la actividad en cuestión se activan algunos atributos corporales u otros (actividades cardiovasculares, de fuerza, agilidad, flexibilidad...), por lo que se debe buscar un indicador que sea común a una mayoría de

estos y representativo de la actividad física de una persona y tener en cuenta estás cuestiones a la hora de realizar un análisis de población que posteriormente permita la predicción de estos mismo valores.

- ¿Que factores podrían alterar este indicador? ¿Es posible realizar la toma de datos de manera precisa? Se debe tener en cuenta el grado de precisión de los sensores de los dispositivos con los que van a ser tomadas las mediciones. Por ejemplo, un sensor de detección de pasos que utiliza cambios en el acelerómetro o giroscopio del teléfono, puede verse alterado por cambios bruscos de movimiento de la mano y es bien sabido que la medición de estos sensores no es realmente fiel a la realidad.

Posterior a un análisis satisfactorio de la actividad física y procesados los datos para una correcta representación, estos son utilizados para abordar el problema principal de estimación o predicción de valores de actividad. Los valores resultantes de la misma a su vez deben tener en cuenta los siguientes puntos:

- Debe ser un valor útil y representativo de la población.
- Deberán de ser precisos, de forma que los valores resultantes describen fielmente el comportamiento poblacional.
- Como problema de predicción deberá de enfrentarse a factores exógenos que puedan alterar el valor de los resultados.
- Aún no siendo la predicción un valor que deba ser calculado en tiempo real, debe poder realizarse en un periodo de tiempo lo suficientemente corto para que los resultados presenten valor práctico.

1.3. Motivación

La salud de un individuo está medida por muchos factores tanto sociales como personales que pueden variar a lo largo del tiempo. En concreto, la actividad física está ligada estrechamente al bienestar mental de las personas provocando diversos beneficios en los individuos que practican deporte de manera habitual. Entre otros debido a la liberación de endorfinas que producen una sensación de felicidad y euforia así como la reducción de la ansiedad y el estrés. Según la Organización Mundial de la Salud (OMS): *"Las personas que son insuficientemente activas tienen un 20 % a 30 % más de riesgo de muerte en comparación con las personas que son suficientemente activas."* [12].

Durante la pandemia, la ciudadanía tuvo que enfrentar una gran limitación en la realización de actividades físicas, siendo el sedentarismo y los problemas ligados del mismo mucho más comunes de lo habitual. Se estima que durante el estado de confinamiento con una confianza del 95 %, el sedentarismo en la población global aumentó unas 2,4 horas diarias respecto a la situación anterior [13]. Finalizado el confinamiento, la incorporación de la actividad física se trató de realizar gradualmente. Por ejemplo, el *"running"* fue una de las primeras actividades recuperadas, debido a que su realización al aire libre y de manera individual permite su ejecución de manera más segura que el resto de deportes. Pero finalmente los gobiernos tuvieron que abordar una cantidad abundante de corredores que salían a la calle a recuperar su hábito o incorporar uno nuevo. En esta situación, una estimación de aquellas horas que mayor actividad física iban a presentar podría haber sido una herramienta muy útil de cara a la prevención de riesgos. Gracias a ello se podrían haber tomado medidas que permitieran una mayor seguridad y control, facilitando así a la población la realización de este deporte.

Como presenta el tercero de los puntos del proyecto POSTCOVID-AI, estos datos de predicción de actividad física contienen la posibilidad de informar las decisiones en organizaciones gubernamentales o sin ánimo de lucro. Estas podrían utilizarlos para realizar campañas de fomentación de la actividad física en aquellos lugares o sectores de la población donde sea más necesario y especialmente en los momentos donde se predice una menor cantidad de actividad. También permitiría facilitar la infraestructura necesaria en aquellos lugares donde se estima un crecimiento de actividad física, mediante la construcción de carriles bici, la creación parques o en un plazo de tiempo más corto para que pistas deportivas puedan ser abiertas a cierta hora o aumentar el nivel de seguridad y primeros auxilios en zonas que vayan a experimentar un nivel de actividad alto.

De cara a un futuro, lo ideal es que las tendencias en actividad física pudieran analizarse en un alto periodo de tiempo, localizar los factores que están influyendo en ello y poner un plan de actuación que evite una tendencia al sedentarismo de la población cuando esta sea detectada.

1.4. Solución propuesta

A raíz de los datos longitudinales de actividad física del proyecto POSTCOVID-AI recogidos como series temporales, se busca la creación de un framework mediante técnicas de inteligencia artificial y aprendizaje profundo que permita la predicción a futuro en un determinado horizonte temporal de estos valores. Para poder alcanzar un resultado satisfactorio con los objetivos, se realizará una experimentación mediante diferentes metodologías para la predicción. Estas técnicas serán analizadas y comparadas

así como su actuación sobre los datos de cara a alcanzar un producto que permita realizar las predicciones con la mayor exactitud posible.

1.5. Objetivos

El fin de este proyecto es la creación de un framework que mediante modelos de aprendizaje automático, logre realizar predicciones sobre la actividad física de una población en un horizonte de tiempo que permita sacar partido a esta información. Para ello se establecen los siguientes objetivos específicos:

- Entendimiento de los fundamentos bajo los valores recopilados de actividad física, así como su preprocesamiento de cara a obtener medidas que permitan realizar predicciones y sean representativas de la población.
- Investigación sobre los modelos y herramientas de inteligencia artificial utilizados en el campo de la predicción de series temporales, así como los más populares en el contexto actual para la predicción de actividad física.
- Establecer diferentes líneas de investigación para realizar la predicción de valores poblaciones y establecimiento de una metodología para toma y comparación de resultados.
- Evaluar los resultados y comprobar que la metodología final permite llevar a cabo la predicción de manera precisa y cuyos horizontes de predicción pueden ser utilizados en un entorno real.

Capítulo 2

Enfoques y evolución en el campo de la predicción

A continuación se realiza un repaso completo sobre el desarrollo de modelos y técnicas utilizados para la predicción de valores, abordando las diferentes técnicas y enfoques que son utilizados para el mismo. La información incorporada en este capítulo tratará el campo de la predicción y no el problema concreto de la estimación de valores de actividad física el cuál se comenta en el capítulo del estado del arte (Capítulo 3).

2.1. Predicción de series temporales

A pesar de los nuevos métodos de computación e inteligencia artificial que hacen de este un campo de actualidad, con un desarrollo y mejora de los modelos y técnicas conocidos prácticamente constante, la predicción de valores futuros siempre ha sido uno de los mayores intereses del ser humano debido a su gran potencial. Desde las predicciones sobre el clima, así como la medicina y la economía, las distintas aplicaciones y problemas planteados fueron los impulsores de un desarrollo formal de esta área.

El inicio del análisis de series temporales conocido actualmente se estima en 1927 con las publicaciones de G.U Yule [14], ya que postulan la posibilidad de que las series temporales sean composiciones provenientes de procesos estocásticos frente a los enfoques anteriores caracterizados por la idea de un mundo determinista [15]. Además este, junto algunos de sus contemporáneos como Gilbert Walker introdujeron los conceptos de modelos auto regresivos y media móvil.

No es hasta 40 años después, en la década de 1970 que aparecería *Time Series Analysis: Forecasting and Control* [16], trabajo de los coautores Box y Jenkins que introducen un enfoque sistemático para el trabajo de series temporales, cimentando así los principios del campo. Con esta publicación Box y Jenkins integran todo el conocimiento desarrollado hasta el momento, y proponen un ciclo iterativo de tres fases para la identificación, estimación y verificación de series temporales ampliamente utilizado en la literatura [17].

Actualmente los modelos de inteligencia artificial y Deep Learning han tomado una gran popularidad dentro del campo de predicción de series temporales y producido un cambio completo en la manera de tratarlas. Por ello, podemos distinguir entre dos tendencias o enfoque de cara al tratamiento de estas series:

- La primera consistiría en este enfoque clásico y basado en estadística, que utiliza las propiedades matemáticas para analizar y consecuentemente predecir valores a lo largo de estas series.
- La segunda nace junto a esta llegada de la Inteligencia Artificial y el alto poder de computación que permite el manejo de grandes cantidades de datos [18].

2.2. Modelos estadísticos

El análisis de series temporales permite reconocer patrones, tendencias y relaciones. Estas propiedades son utilizadas por los modelos estadísticos para la predicción de valores en el tiempo [19]. Existen muchos modelos estadísticos con enfoques diferentes cuyas debilidades y ventajas dependen, en esencia, de la naturaleza de la serie temporal en estudio. Aun así, podemos considerar que las propiedades recogidas por el modelo ARIMA *autoregressive integrated moving average*, constituyen las características principales en el estudio de series temporales, y sin duda dan base para el resto de técnicas y modelos utilizados para la predicción. Tras la publicación de Box y Jenkins y junto a la popularización de la computación, este modelo comienza a ser ampliamente utilizado y a raíz de él han sido desarrolladas una gran cantidad de variaciones, como podría ser SARIMA, un modelo similar a ARIMA cuya letra "S" hace referencia a la consideración de la estacionariedad. Como indican sus siglas, ARIMA nace de la unión de tres propiedades incluidas en los siguientes tipos de modelos:

2.2.1. Auto-Regresivos

Los modelos auto-regresivos o AR(p) asumen que el proceso de generar los datos es una combinación lineal de observaciones pasadas, y por tanto utilizan los datos recogidos para realizar las predicciones futuras. El orden de estos modelos depende del número de observaciones.

$$\hat{y}_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$

Siendo:

- y_t valor de la serie en el instante t
- c una constante
- ϵ_t ruido blanco
- $\phi_1..p$ son los parámetros del modelo

2.2.2. Media-Móvil

El modelo de media móvil MA(q), es común en el modelado de series temporales univariadas. Utiliza el error en predicciones pasadas para mejorar la predicción realizada. Este error se asume como independiente y normalmente distribuido (ruido blanco):

$$\hat{y}_t = c + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

En este caso se añade θ son los parámetros del modelo

2.2.3. Integrado

En este modelo I(d), la diferencia se toma una cantidad d de veces para así poder hacer la serie estacionaria.

$$\hat{y}_t = (1 - B)^d y_t$$

donde $B y_t = y_t - 1$

Finalmente el modelo ARIMA(p,d,q) queda de la siguiente manera:

$$\hat{y}_t = c + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$

Pasapitch Chuai, Nittaya Kerdprasop y Kittisak Kerdprasop analizaron la capacidad de predicción de los modelos ARIMA y ARMA del consumo de electricidad en un hogar en los años comprendidos entre 2006 y 2010 [3], en los cuales se tomaron medidas de consumo con una frecuencia de 1 minuto. La realización del mismo se hizo siguiendo la metodología de Box y Jenkins. Uno de los detalles más interesantes de este estudio reside en que se experimentó con diferentes unidades de tiempo y horizontes de predicción 2.1.

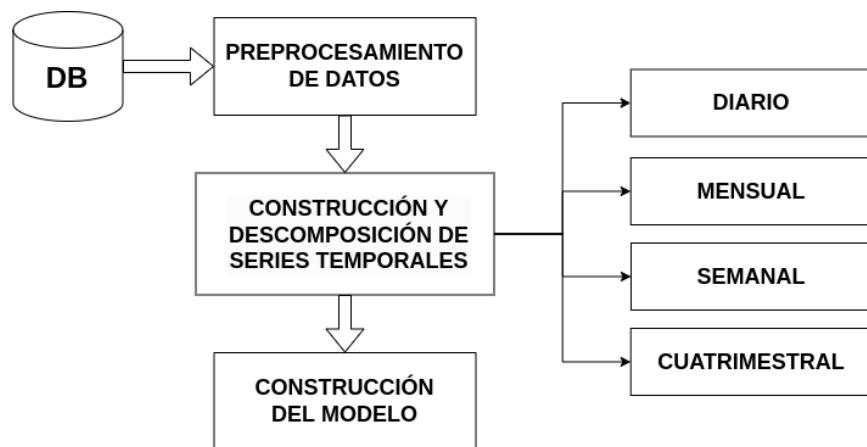


Figura 2.1: Framework de construcción de las series temporales. Fuente: [3].

Como resultado el modelo ARIMA era capaz de realizar mejores predicciones en períodos de tiempo mensuales y cuatrimestrales, encontrando los mejores valores con horizontes de predicción de 6 meses a dos cuatrimestres. Por otro lado, el modelo ARMA capturaba mejor las series temporales diarias y semanales, encontrando su pico de mejores actuaciones entre los horizontes de predicción de 28 a 5 semanas.

Otro proyecto donde ARIMA expone un buena actuación sobre la tarea de pronóstico es la predicción de consumo eléctrico en la zona oeste de Arabia Saudí entre Agosto de 1992 y Julio de 1993 [20]. Este estudio realizado por el departamento de investigación energética saudí, utiliza datos de los 5 años anteriores y fue anteriormente realizado para modelos de regresión lineal y abductive networks. En este caso, ARIMA consiguió mejores resultados en términos de error medio absoluto, a pesar de que sus competidores utilizaban data multivariante donde existía información del tiempo, la economía e información demográfica. Concretamente el mejor de los modelos ARIMA obtuvo un error del 3.8 % en sus predicciones respecto del 8.1 % y

5.6 % en los mejores resultados de sus contrincantes.

La propuesta realizada por los modelos estadísticos es capaz de representar las dinámicas de una gran cantidad de series temporales. Sin embargo, se ha encontrado que este enfoque es insuficiente para la predicción de muchas series temporales de la vida real cuyo comportamiento contiene una gran no-linealidad. Debido a ello, los modelos de Machine Learning y especialmente los diferentes tipos de *Artificial neuronal Networks* ANN han tomado gran relevancia en el campo [21]. Estos modelos si bien son capaces de captar dinámicas complejas dentro de las series temporales, sabemos que necesitan de un conjunto de datos grande para el entrenamiento y su uso, que han provocado que su entrada haya sido más complicada que en otros campos y sus resultados más cuestionados.

2.3. Modelos Computacionales

2.3.1. Machine Learning: dificultades y bondades

En la historia de las competiciones de la predicción se series temporales [22] las *M-Competition* creadas por Spyros Makridakis y Michèle Hibon en 1982 y que actualmente en 2023 siguen activas, comparan gran parte de los modelos de cada época en la que se realizan para un conjunto de datos que iba creciendo en cada competición. Para la *M3-competition* algunos modelos utilizaron redes neuronales, pero debido a la escasez, el profesor Sven Crone de la universidad de Hamburgo organizó una nueva competición con 111 series temporales mensuales de la *M3-competition* y entorno a 60 nuevos algoritmos fueron añadidos a la competición.

Como resultado ninguno de ellos mejoró respecto a los participantes de la *M3-competition*, reforzando la idea de que las redes neuronales y en general modelos no-líneales no se ajustan bien a series temporales cortas, ya que en esta competición la mayor de ellas estaba compuesta de 126 observaciones. *M4-competition* fue una de las últimas competiciones realizada finalizando en Marzo de 2018. Si bien mantiene unos resultados pobres por parte de los modelos puramente de Machine Learning, sorprendió al concluir cómo los modelos híbridos que combinan ambos enfoques habían sido los mejores tanto en exactitud como en precisión. [23].

Otra de las consecuencias importantes de M4-competition recae en que los modelos estadísticos no producen mejoras considerables al aumentar el tamaño de los datos. Tomando en cuenta los dos factores nombrados en

contra de estos modelos (no-linealidad y tamaño de los datos), comienza a adquirir sentido al uso de Machine Learning, ya que permite para otros problemas dentro del campo que este obtenga mejoras respecto de los modelos tradicionales, especialmente debido al aumento de los datasets utilizados así como la complejidad de los mismos. Algunos de los estudios más citados en la defensa del Machine Learning provienen de la mano de Paul J. Werbos [24][25], o de Guoqiang Zhang, B. Eddy Patuwo y Michael Y. Hu [26].

2.3.2. Deep Learning

Los modelos Deep Learning se caracterizan por obtener buenas características de los datos de entrada utilizando un procedimiento de propósito general y han demostrado un gran rendimiento en múltiples tareas como compresión de lenguaje natural, clasificación de sentimientos o contestando preguntas. Esto se debe al uso de múltiples capas de procesamiento que permiten una gran abstracción sobre los datos.

Algunos de los modelos Deep learning más estudiados en la literatura son las Recurrent Neuronal Network (RNN), Long Short Term Memory (LSTM), Convolutional Neuronal Network (CNN)... Para estos modelos son múltiples los estudios comparativos donde se obtienen resultados que opacan a sus competidores y se han posicionado como una de las opciones esenciales a la hora de realizar predicción de series temporales. *“Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms”* [27] es un ejemplo de estos estudios comparativos, concretamente para el problema de predicción del precio de la luz. El rendimiento de los modelos Deep Learning en este mejora considerablemente respecto de los diferentes modelos más consolidados en el enfoque estadístico. DNN (Deep Neuronal Network), GRU (Gated Recurrent Unit) y LSTM superan a sus competidores obteniendo los mejores resultados, seguidos de modelos híbridos como SVR-ARIMA, y muy lejos de los modelos puramente estadísticos como AR.

Por otro lado, la predicción en el ámbito financiero, en la cual se incluyen un abanico amplio de problemas como la predicción de tendencias en el mercado de valores o criptomonedas, al ser sin duda uno de los campos que más interés han concentrado, nos puede dar una buena visión sobre como ha crecido la popularidad del uso de modelos Deep Learning. La Figura [4] muestra el número de publicaciones de modelos DL en el campo financiero contadas por 2.2.

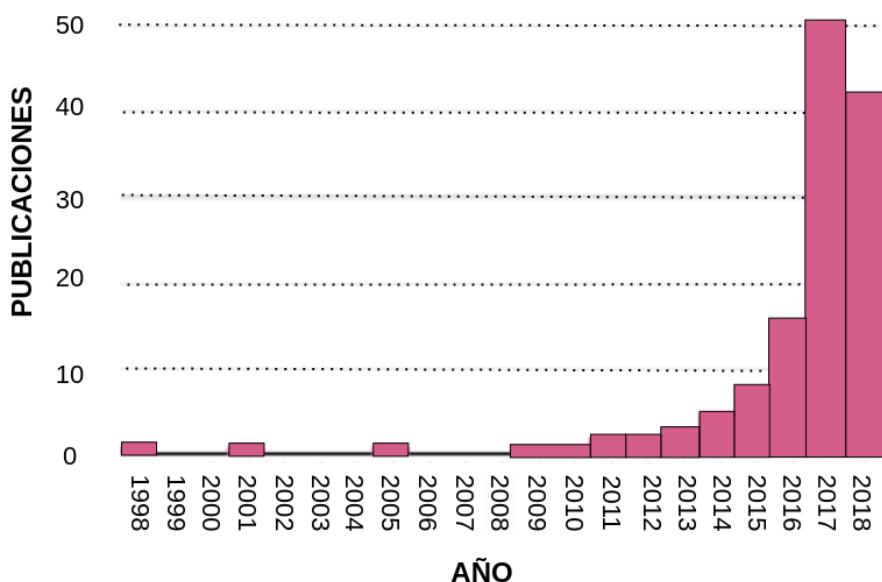


Figura 2.2: Histograma de publicaciones de modelos DL en el campo financiero. Fuente: [4].

Dentro de estas publicaciones además existe una clara apuesta por los modelos recursivos derivados de RNN, siendo más del 50 % de las apariciones de esta familia, y dentro de ella el modelo estrella es LSTM suponiendo el 60,4 % de las menciones referentes a modelos RNN.

Capítulo 3

Estado del arte

La capacidad de predecir y comprender el comportamiento relacionado con la actividad física ha tomado relevancia en la ciencia, la medicina y la tecnología en los últimos años. El estado del arte abarca una amplia gama de enfoques multidisciplinarios, que van desde la monitorización de la actividad física a través de dispositivos portátiles, su representación, o el uso de inteligencia artificial para el análisis de datos biológicos y comportamentales, así como su predicción. Este capítulo realiza un repaso sobre los estudios más recientes en referencia al punto de la predicción y expone las metodología y técnicas en auge o más utilizadas actualmente dentro de este ámbito.

"Physical activity forecasting with time series data using smartphone"[5] es un estudio muy similar a la problemática que se trata durante este trabajo, a pesar de que no se utiliza el concepto de METs como medida de actividad física ni se busca un objetivo a nivel poblacional. De manera similar al trabajo de predicción de METs, los datos utilizados fueron recopilados mediante ActivityRecognitionAPI con una frecuencia de un minuto, así como valores de consumo calórico y pasos recopilados mediante Google Fit Platform cada hora. Posteriormente estos son utilizados para realizar predicciones de los períodos de tiempo que se realizará cada una de las actividades, es decir, cuantos minutos el individuo del estudio estará parado, corriendo o cualquier otra acción, así como del número de calorías y pasos dados por el participante. Las predicciones son realizadas tanto de manera diaria como por horas utilizando BiLSTM, así como otras variaciones de LSTM. Por ejemplo, los valores utilizados con un horizonte en horas indican que entre las 00:00 y las 01:00 un individuo estará 30 minutos parado, 15 andando y 15 montando en bicicleta, posteriormente prediciendo cuantos minutos de actividad físicas serán realizados por un sujeto.

Las diferentes variaciones de la red LSTM, dieron buenos resultados a la hora de estimar el tiempo de actividad física empleados por los participantes cada hora durante el último día de estudio, siendo capaces de captar las tendencias reales en la actividad. Para LSTM, BiLSTM1, BiLSTM2, y CNN BiLSTM el error medio absoluto fue inferior a 9 minutos en todos los casos.

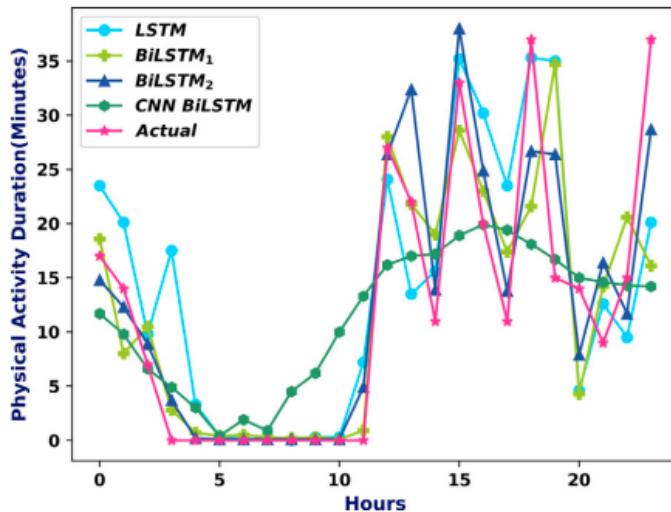


Figura 3.1: Resultados en la predicción de Actividad Física. Fuente: [5].

Otro de los trabajos encontrados sobre la materia es un paper titulado "Prediction of Physical Activity Times Using Deep Learning Method" [28], el cual se encuentra en el libro *International Telecommunications Conference* que recopila una de las reuniones científicas más importantes del mundo sobre telecomunicaciones, procesamiento de señales, redes de datos, seguridad y optimización.

Este estudio trabaja con los datos de *StudentLife dataset* y son provenientes de sensores recopilados a través de teléfonos inteligentes de 49 estudiantes de Dartmouth College por un período de 10 semanas. Cuentan con una gran recopilación de mediciones del acelerómetro cada 2 s que van siendo tomadas durante 1 minuto de cada 4. Posteriormente estos instantes son clasificados en las categorías quieto, andando, corriendo y desconocido. El estudio trata de predecir el ratio de sedentarismo a través de la proporción de tiempo en el que los estudiantes se encontraban parados. Así, si existen 1200 medidas etiquetadas como quieto durante un período de 1800 instantes, se considera un porcentaje de sedentarismo del 0.67 %.

Posteriormente, se predicen porcentajes de sedentarismo durante períodos de 6 horas en los 49 participantes utilizando una RNN. Estos son comparados con las predicciones realizadas por un modelo de referencia, que

utiliza los valores del instante $t-1$ como predicciones para el instante t . Los resultados indican que para 40 de los 49 estudiantes, el resultado mejoró gracias a la red RNN. Concluyeron con la idea de que era posible el uso de modelos Deep Learning para la predicción de actividad física y que éstas podrían ser utilizadas para sugerir mejores horarios de actividad según las rutinas de los individuos.

Finalmente podemos encontrar enfoques más amplios como el propuesto por *"Multimodal Time-Series Activity Forecasting for Adaptive Lifestyle Intervention Design"* [6] donde se combina tanto información de la actividad física realizada por el usuario como información de la interacción del usuario con la aplicación, obteniendo el modelo multimodal observable en la Figura 3.2.

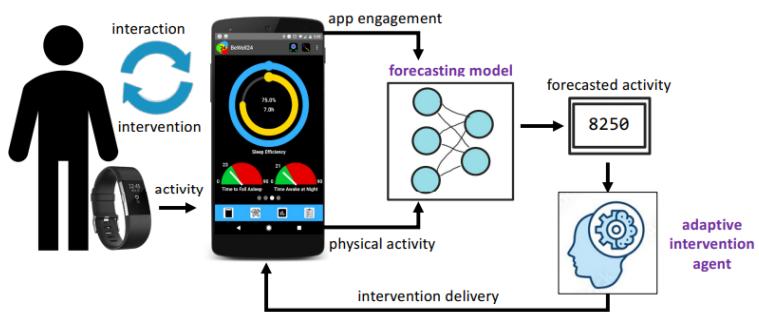


Figura 3.2: Arquitectura BeWell24Plus. Fuente: [6].

Los datos provienen de 58 personas con condiciones pre-diabéticas en un estudio clínico de 9 meses. Se utiliza un LSTM con 100 neuronas, seguido de una capa densa de 50 neuronas. Los resultados de este estudio indican una mejora en los resultados al utilizar un método combinado que incorpora más información, obteniendo un *mean absolute error* (MAE) de 2081 para el modo multimodal, 2372 en el caso de uso de actividad de la aplicación y 2090 para la actividad física.

Conclusiones parciales

Las principales diferenciaciones en el campo nacen en el tratamiento de los datos, la obtención de los valores de actividad física, así como la creación de una medida representativa de esta que permita trabajar correctamente con los modelos utilizados. Por su parte, LSTM y sus variaciones suponen claramente un estándar en los modelos de predicción de actividad física y prácticamente en cualquier problema de predicción, hasta el punto que son

pocos los estudios en los que no se utiliza este o en su defecto algún otro modelo de la familia de las RNN.

Su éxito se debe en gran parte a su sistema de memoria, el cual permite recordar valores de las entradas durante un largo periodo de tiempo y por tanto utilizar ventanas de tiempo mayores en la predicción e incorporar una mayor cantidad de información pasada. Gracias a ello es capaz de detectar patrones complejos en las series temporales de entrada y dependencias en el largo plazo. Además, para manejar la información almacenada en las celdas de memoria se utiliza un sistema de tres puertas que actúan de manera similar a una función de activación. Estas son, puerta de entrada (input gate), puerta de olvido (forget gate) y puerta de salida (output gate) y la combinación de ellas permite al modelo aprender qué entradas usar y qué importancia tendrán estas en cada momento. Estos conceptos son desarrollados en la sección 4.

En definitiva, el uso de LSTM es altamente recomendado para problemas de predicción debido a que su buen funcionamiento ha quedado demostrado en diversidad de campos y todo tipo de problemas basados en series temporales.

Capítulo 4

Fundamentos teóricos

Esta sección aborda los fundamentos teóricos detrás de las redes neuronales recurrentes, especialmente de las LSTM al ser escogidas como modelo para la predicción. Además, se tratarán los problemas más conocidos en el ajuste de estas redes y su tratamiento.

4.1. RNN

Las RNN o redes neuronales recurrentes fueron popularizadas por su capacidad de modelar datos secuenciales y/o dependientes del tiempo, como la generación de texto o la predicción de valores temporales.

Una red neuronal recurrente en su estado más básico, cuenta con una sola capa oculta y una sola secuencia de salida. El input consistirá en una serie de vectores de entrada secuenciales $[x(1), x(2) \dots x(n)]$, donde en el caso de una serie temporal cada una de estas entradas corresponde a un instante en el tiempo [29].

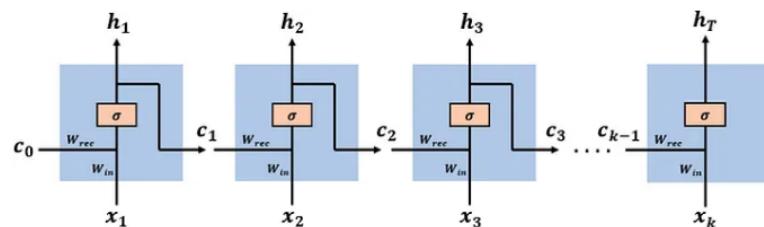


Figura 4.1: RNN básica. Fuente: [7].

La representación básica de un RNN puede ver en la figura 4.1. Cada una de las instancias de la entrada, estará asociada a una salida o vector de estado que contiene la información pasada aprendida por la red $[c(0), c(1) \dots c(n-1)]$. La predicción para un instante t vendrá dada por el vector de entrada $x(t)$ y la información recogida hasta el momento en el vector de estado $c(t-1)$.

Los parámetros de la red están formados por dos matrices W_{in} y W_{rec} , las cuales conectan los dos componentes responsables de la salida, el vector de entrada y el vector de estado. Finalmente, los datos pasan por una función de activación dentro de la capa oculta y el modelo produce una salida.

4.1.1. Problemas en los gradientes

Para ajustar el resultado del modelo y reducir la función de error, los parámetros son ajustados mediante *Back Propagation*:

$$W \leftarrow W - \alpha \frac{\partial E}{\partial W}$$

Para el instante k el gradiente del error vendrá dado por:

$$\frac{\partial E_K}{\partial W} = \frac{\partial E_k}{\partial h_k} \frac{\partial h_k}{\partial C_k} \left(\prod_{t=2}^k \sigma'(W_{rec} \cdot c_{t-1} + W_{in} \cdot x_t) W_{rec} \right) \frac{\partial c_1}{\partial W}$$

El problema del desvanecimiento de gradientes surge cuando la derivada de la función de activación es menor que uno, provocando que el producto sea cada vez menor conforme k aumenta y el gradiente tienda rápidamente a cero [7]. Por otro lado, el efecto contrario conocido como explosión de gradientes ocurrirá cuando los pesos sean muy superiores a la propia derivada de la función de activación y el producto provoque una tendencia hacia infinito de manera exponencial.

4.1.2. Long Short Term Memory

En solución a los problemas en el gradiente que experimentan las redes recurrentes básicas, LSTM o Long Short Term Memory utilizan diferentes capas de agregación que permiten el uso de un volumen mayor de datos [30].

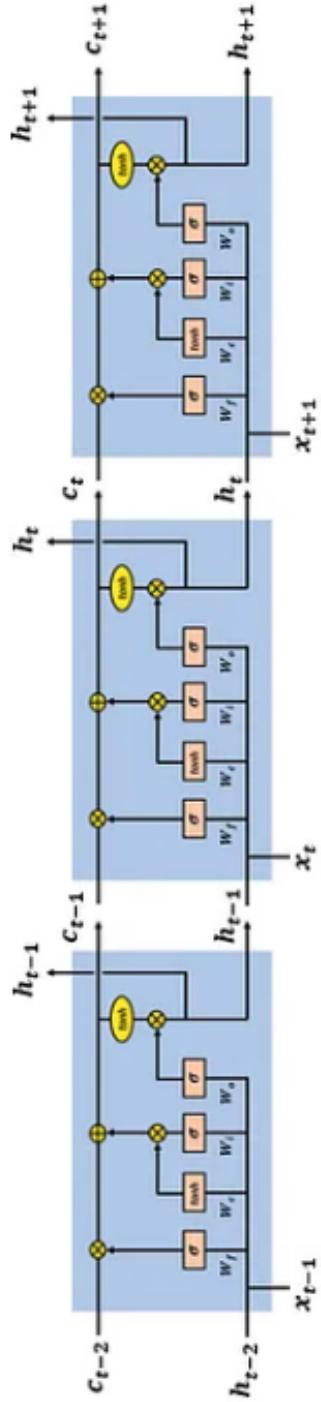


Figura 4.2: Arquitectura LSTM. Fuente: [7].

Estas redes recurrentes representadas en la figura 4.2 están formadas por tres tipos de puertas diferentes, las cuales en un instante t son:

- **Forget gate:** se encarga de controlar que parte de la información pasada almacenada en el estado actual sigue siendo relevante para el resto de la red como se puede apreciar en la figura 4.3.

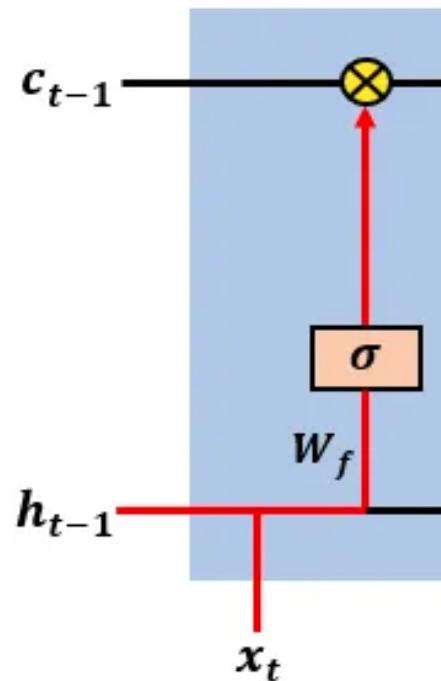


Figura 4.3: Forget Gate. Fuente: [7].

La salida de esta viene dada por el producto de un vector de pesos W_f y la entrada del modelo formada por $[h_{t-1}, X_t]$. Este producto es pasado por una función de activación (en este caso la sigmoidal) obteniendo valores entre 0-1 que posteriormente son multiplicados elemento a elemento con los valores aprendidos hasta el momento c_{t-1} en el vector de estado, indicando la relevancia de estos.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t])$$

- **Input state:** determina que parte de la nueva información es relevante de cara a ser guardada en el vector o celda de estado.

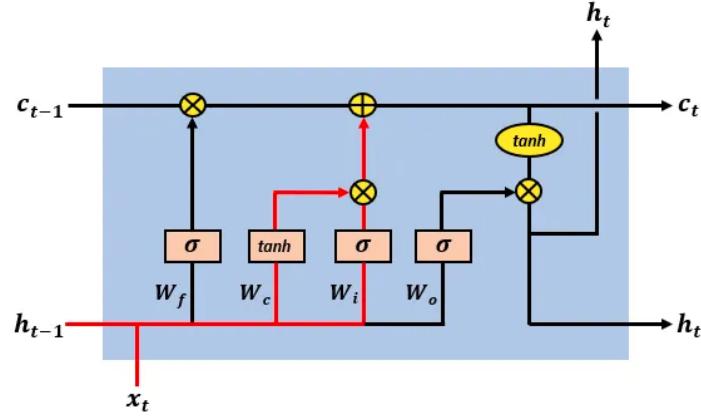


Figura 4.4: Input Gate. Fuente: [7].

Para ello utiliza dos vectores de pesos W_c y W_i , multiplicados por la entrada del modelo y posteriormente combinados mediante el producto elemento a elemento de los mismos. Esta puerta es representada en rojo en la Figura 4.4. La salida de estos es añadida al vector de estado guardando la nueva información que sea relevante.

$$f_{ig} = \tanh(W_c \cdot [h_{t-1}, x_t]) \otimes \sigma(W_i \cdot [h_{t-1}, x_t]) = \hat{c}_t \otimes i_t$$

- **Output Gate:** una vez determinado el vector de estado, se calcula la salida del modelo como una combinación de este y el input, el cuál es previamente multiplicado por un último vector de pesos W_o y una función de activación. Se puede visualizar dicha salida en la Figura 4.5.

$$h_t = o_t \otimes \tanh(c_t)$$

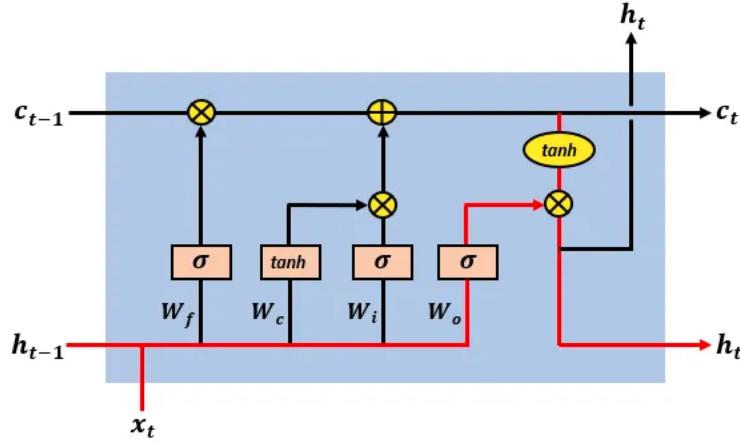


Figura 4.5: Output Gate. Fuente: [7].

Las relaciones y dependencias de largo plazo son codificadas en el vector de estado de tal forma que evitan a los gradientes su desaparición:

$$c_t = c_{t-1} \otimes f_t \oplus \hat{c}_t \otimes i_t$$

Si recordamos el gradiente de la red recurrente:

$$\frac{\partial E_K}{\partial W} = \frac{\partial E_k}{\partial h_k} \frac{\partial h_k}{\partial C_k} \left(\prod_{t=2}^k \frac{\partial c_t}{\partial c_{t-1}} \right) \frac{\partial c_1}{\partial W}$$

Con la nueva definición del vector de estado:

$$\begin{aligned} \frac{\partial c_t}{\partial c_{t-1}} &= \frac{\partial [c_{t-1} \otimes f_t \oplus \hat{c}_t \otimes i_t]}{\partial c_{t-1}} \\ &= \frac{\partial f_t}{\partial c_{t-1}} \cdot c_{t-1} + \frac{\partial c_{t-1}}{\partial c_{t-1}} \cdot f_t + \frac{\partial i_t}{\partial c_{t-1}} \cdot \hat{c}_t + \frac{\partial \hat{c}_t}{\partial c_{t-1}} \cdot i_t \end{aligned}$$

La adición de estos cuatro términos permite la no desavenencia del gradiente ya que se produce un balance entre el valor de los mismos. Esto permitirá el uso de volúmenes de datos mayores y hace de LSTM uno de los modelos más popularizados para la predicción de series temporales.

Capítulo 5

Gestión del Proyecto

En esta sección se describirá la planificación realizada y la estimación de costes tanto monetarios como temporales del mismo.

5.1. Metodología de desarrollo

Las metodologías de desarrollo software permiten realizar los objetivos de manera rápida y eficaz, facilitando el trabajo en equipo y reduciendo el nivel de dificultad general del proyecto. En este caso se ha utilizado una metodología tradicional en cascada [31].

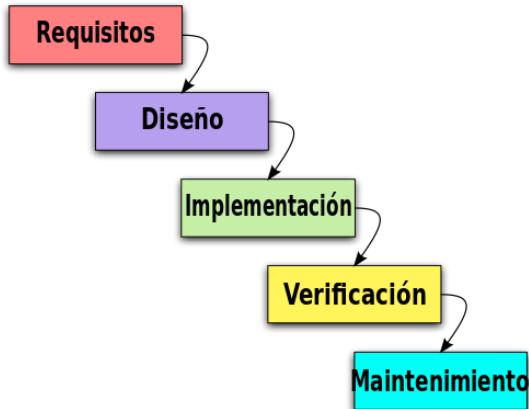


Figura 5.1: Ejemplo de metodología en cascada. Fuente: [8].

La metodología en cascada se basa en un esquema de etapas bien diferenciadas, el cual queda representado en la Figura 5.1. Esta metodología indica que se debe cumplir una serie de propósitos antes de poder pasar a la siguiente etapa.

te y el producto en cada una de estas etapas debe cumplir unos estándares de calidad. Esto hace de la cascada un método lineal que concuerda bien con los pasos a seguir en el desarrollo de un trabajo de ingeniería.

La metodología en cascada es combinada con el método de prototipado, el cuál nos permite crear modelos de prueba para poder detectar errores. Estos son de gran importancia en la elaboración de este proyecto ya que aún marcadas las líneas principales de investigación, son diversas las decisiones y posibilidades a tomar. Se necesita de estos modelos prototípicos para conseguir información sobre cuáles funcionan mejor y poder además obtener la guía de los tutores. Es por ello, que aún manteniendo un proceso estructura lineal en cascada, se lleva a cabo el desarrollo de prototipos que puedan guiar en la corrección de decisiones tomadas durante el proceso de diseño. Por ejemplo, respecto al tamaño del horizonte de predicción, o sobre el tamaño de la ventana utilizada.

5.2. Planificación

Este trabajo ha sido pensado para ser realizado a lo largo del segundo cuatrimestre del curso 2022-2023 de la Escuela Técnica Superior de Ingenierías Informáticas y Telecomunicaciones de la Universidad de Granada, así como durante los meses previos a la convocatoria especial de noviembre del curso 2023-2024. Durante las semanas previas al 20 de Febrero (inicio de este segundo cuatrimestre), el alumno y los docentes se reunieron para plantear el abanico de posibilidades de dicho TFG. Para el mismo Dataset recopilado por el equipo de POSTCOVID-AI, fueron ofrecidas y presentadas al alumno las opciones del forecasting (predicción de series temporales), así como clustering (agrupamiento) y análisis de tendencias, siendo la primera finalmente la escogida.

Establecida la temática del trabajo se programan los siguientes hitos en el desarrollo del mismo cuya planificación puede verse en la Figura 5.2.

5.2.1. Hitos

■ Estado del arte

Investigación de los últimos desarrollos en el campo de la predicción de series temporales y los modelos más utilizados en los principales problemas de la literatura actualmente. Además, aprender sobre las características más relevantes y la teoría que establece las bases de estas series.

- **Fijación de objetivos y establecimiento de la metodología**

Definir a partir de la literatura los modelos y metodologías a poner en práctica. Además, establecimiento de objetivos y resultados que se esperan lograr mediante este trabajo.

- **Análisis y preprocesado de datos**

Estudiar los datos que se utilizarán en el proyecto para ver las fortalezas y debilidades de estos, así como para inquirir la viabilidad de los objetivos previamente fijados. Una vez analizados, realizar el preprocesamiento necesario para un mejor desempeño de los datos.

- **Implementación de programas y ejecutables**

De forma independiente, cada uno de los métodos fijados requiere de un tiempo de programación y desarrollo. Se requiere a su vez de una previa investigación de librerías de inteligencia artificial que permitan el desarrollo completo del pipeline de aprendizaje establecido en la metodología.

- **Toma de marcas**

Realización de la extracción de resultados a través de los programas implementados, siguiendo metodología así como la extracción de gráficas y otras medidas que puedan resultar de interés.

- **Comparación de resultados y extracción de conclusiones**

Obtenidos los resultados realizar el estudio comparativo entre los tres enfoques de predicción y concluir cuál de ellos es mejor a la hora de lograr los objetivos propuestos. Extraer conclusiones y plasmar estas en la memoria del proyecto.

- **Revisión y corrección de errores**

Repaso del trabajo completo y desarrollo de versión final para la entrega.



Figura 5.2: Diagrama de Gannt con la planificación del proyecto

5.3. Presupuesto

El principal de los costes es el destinado al sueldo de la persona encargada de la realización del proyecto. En el desglose de la Tabla 5.1 obtenemos un sueldo por horas en base al salario cobrado por un informático Freelance especializado en Ciencia de Datos.

Recursos Humanos			
Nombre	Precio por unidad	Horas Estimadas	Total
Ingeniero de Datos Freelancer	35€/h	310h	10850€

Tabla 5.1: Estimación precio de recursos humanos.

Respecto de los recursos materiales necesarios para la ejecución del trabajo, es necesario un ordenador compatible con tensorflow 2.12.0 así como el resto de paquetes utilizados en el proyecto y algunas características especiales para el procesamiento de una alta cantidad de datos:

1. **Sistema Operativo:** Ubuntu 16.04 o superior(64-bit). Alternativas: macOS 10.12.6 (Sierra) o superior (64-bit) Windows Native - Windows 7 o superior (64-bit) Windows WSL2 - Windows 10 19044 or higher (64-bit)
2. **GPU:** NVIDIA® GPU card con arquitectura CUDA® 3.5, 5.0, 6.0, 7.0, 7.5, 8.0 o superior (Recomendable).
3. **Memoria RAM:** 16 GB de RAM, debido a la alta cantidad de datos a manejar (la versión ligera del proyecto requiere de una cantidad menor, pero para realizar todas las pruebas desarrolladas durante el proyecto se han necesitado hasta 12 GB de RAM).
4. **Procesador:** preferible un procesador igual o superior a multicore Intel Core i7 o AMD Ryzen 7.
5. **Unidad de Almacenamiento:** Disco duro sólido (SSD) con acceso rápido y entorno a GB de espacio libre.

Finalmente, el precio total de un computador necesario para realizar este trabajo queda reflejado en la Tabla 5.2 al que se añade el precio de una licencia de Pycharm, el IDE empleado para la realización del mismo.

Recursos informáticos		
Nombre	Precio / licencia anual	Acumulado
Computadora	800€	800€
Licencias Py- Charm	249€	1049€

Tabla 5.2: Estimación precio de recursos informáticos.

Por tanto, teniendo en cuenta los recursos estimados en las Tablas 5.1 y 5.2, el costo aproximado del proyecto rondaría los 12200 euros.

Capítulo 6

Datos

POSTCOVID-AI dataset es una base de datos donde fueron registrados los cambios longitudinales, es decir, a lo largo del tiempo de una serie de individuos. Este dataset creado por el equipo del proyecto de mismo nombre, cuenta tanto con información objetiva y subjetiva sobre los participantes introducida por ellos mismos, así como de activación física de los distintos participantes durante aproximadamente un mes. Entre otros, los participantes llenaban encuestas desde sus teléfonos o desde el sitio web del proyecto unas 6 veces al día. Otros de los datos fueron recogidos con una menor frecuencia, como si el paciente o familiar de este había sido infectado, o sobre la situación laboral del mismo. A estos debemos añadir algunos no longitudinales obtenidos mediante un cuestionario inicial como el género, la edad o el código postal y un conjunto de datos obtenidos a través de los sensores del dispositivo móvil. En concreto los valores de actividad física fueron tomados mediante ActivityRecognitionApi [32].

Esta API mediante el uso de algoritmos avanzados para el análisis y procesamiento de datos producidos por sensores como el acelerómetro y el giroscopio de un dispositivo móvil, es capaz de identificar actividades físicas como caminar, correr, andar en bicicleta, conducir o estar en reposo. Los valores recogidos mediante esta API cuentan con las siguientes dificultades:

- **Formato de salida**

La API devuelve un valor que representa el tipo de actividad física que se estaba realizando en el momento. Estos valores requieren de un procesamiento previo a su uso, el cuál transforma las etiquetas por valores numéricos que puedan ser utilizados de cara a entrenar la red. Esta transformación debe de hacerse de manera que los valores sean lógicos y presente relación entre ellos. Por ejemplo, la asignación de valores continuos a cada actividad física estaría aportando una información errónea al concepto de cada actividad. Por ejemplo, si establecemos

andar como un uno, correr como un dos, estar quieto como un tres y conducir como un cuatro. Esto indicaría que correr supone el doble de esfuerzo que andar o la mitad que conducir al disponer estos datos de manera longitudinal en una serie temporal, cuando estas proporciones no tienen por qué ser necesariamente así. Debido a la información aportada por esta API todas las instancias donde las actividades físicas detectadas hayan sido clasificadas de la misma manera deberán llevar un peso similar aunque no necesariamente la intensidad de ambas haya sido la misma.

- **Valores faltantes**

Los valores de actividad física dentro del estudio quedan recogidos con una frecuencia de un minuto. Aunque las series temporales se encuentren completas, no siempre todos los valores recogidos serán realmente representativos de la actividad física realizada en el momento. Esto dependerá de si los sujetos que participaron en el estudio llevaban su teléfono con ellos en el momento de la realización de actividades y si estas realmente se encontraban en el umbral de actividades recogidas por la API.

6.1. Utilización de METs

De cara a solucionar los problemas anteriormente expuestos el equipo de POSTCOVID-AI realizó un preprocesamiento de las etiquetas basado en METs. La palabra MET proviene de "metabolic equivalent of task", la cual es una medida utilizada para representar el nivel de intensidad de una actividad física. Un MET es el nivel de energía utilizada para estar sentado y a raíz de este se atribuyen los niveles para el resto de actividades. Para calcular estos se parte de que las células usan oxígeno para generar la energía con la que se mueven los músculos. Un MET supone aproximadamente 3.5 mililitros de oxígeno por kilogramo cada minuto. Dependiendo del nivel de oxígeno medio utilizado al realizar una actividad en concreto podemos establecer la cantidad de METs que supone dicha actividad.

Por tanto, para que cada una de las acciones tenga un valor que concuerde con el resto, se realiza una conversión del valor de la actividad devuelto por el sensor a la cantidad de METs que esta supone. Estos valores quedan recogidos por el compendio de actividad física para adultos [33]. En este mismo, quedan expuestas las limitaciones del Compendio para Adultos. Como se explica en el compendio, su objetivo no es determinar el costo energético preciso de la actividad física en los individuos, sino más bien proporcionar un sistema de clasificación que estandarice las intensidades MET de las actividades físicas utilizadas en la investigación de encuestas. En el compendio

se registran los valores que calculan el gasto de energía de la actividad física en personas sin tomar en consideración factores como la masa corporal, la adiposidad, la edad, el sexo, la eficiencia del movimiento, y las condiciones geográficas y ambientales en las que se llevan a cabo estas actividades.

Las actividades recogidas mediante ActivityRecognitionAPI y su equivalencia en METs quedan recogidas en la siguiente Figura 6.1.

ACTIVIDAD DETECTADA	METS
IN_VEHICLE	1.3
ON_BICYCLE	7.5
TILTING	1.8
STILL	1.3
ON_FOOT	3.5

Tabla 6.1: Equivalencias entre actividad física y METs.

Finalmente, el dataset tras disponer los valores en su equivalente en METs de manera longitudinal quedaría como en la Figura 6.1. La columna *participant* identifica de manera única a cada uno de los participantes del estudio, *timestamp* recoge el minuto exacto en un formato legible donde estos datos fueron recogidos, *minute* referencia al minuto al que pertenece el valor dentro del estudio y en la columna *METs* se presenta el valor de consumo calórico almacenado en dicho instante.

participant	timestamp	minute	METS
---	---	---	---
str	str	i64	f32
A3FNz	2021-11-17 16:59:00	2459	1.3
A3FNz	2021-11-17 17:00:00	2460	1.3
A3FNz	2021-11-17 17:01:00	2461	1.3
A3FNz	2021-11-17 17:02:00	2462	1.3
A3FNz	2021-11-17 17:03:00	2463	1.3

Figura 6.1: Ejemplo de filas del dataframe de datos procesados.

6.1.1. Reducción del dataset

De cara facilitar el proceso de depuración de código, así como los tiempos en entrenamiento y testeo de modelos, se toma una versión del dataset que se considera representativa y generalizable al modelo completo. Además el uso de el dataset completo trajo diversos problemas de RAM que impedían obtener todos los resultados en su totalidad. Por ello, es utilizado

un subconjunto de los 25 participantes con mayor variabilidad en sus datos. Seleccionar sujetos con la mayor variabilidad, trata de capturar la gama más amplia de comportamientos dentro del conjunto de datos, de cara a comprender mejor la diversidad dentro de la población objetivo así como evitar ruido generado por participantes con una actividad muy monótona. Dicha monotonía proviene normalmente del estado sin movimiento y por ello puede de estar relacionada con una mala praxis del sujeto al no llevar consigo su dispositivo móvil o tenerlo apagado entre otros.

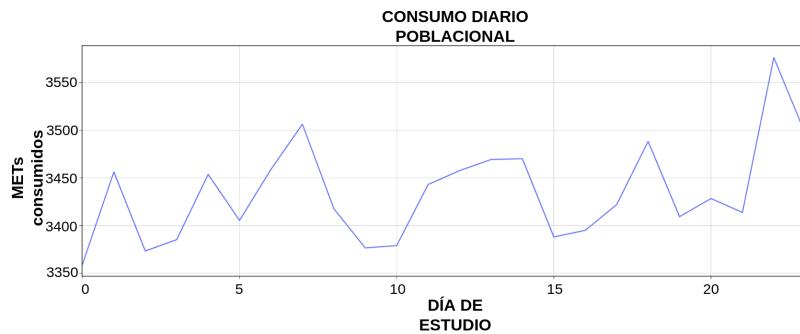


Figura 6.2: Consumo de METs poblacional de los primeros 24 días para 100 participantes.

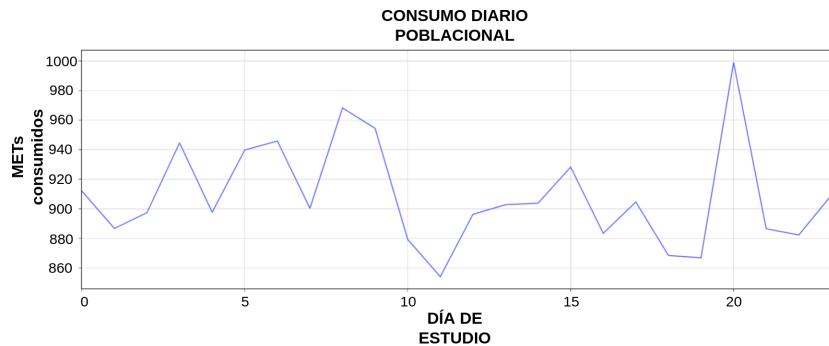


Figura 6.3: Consumo de METs poblacional de los primeros 24 días para 25 participantes.

Las representaciones de las figuras 6.2 referente al dataset completo y 6.3 referente a la versión reducida, muestran como la versión utilizada para el testeo de modelos comparte una correlación grande con la versión original, siendo las tendencias muy similares.

Capítulo 7

Metodología

Este TFG realiza una comparación de diferentes *pipelines* y enfoques, posteriormente valorados según los beneficios o bondades de estos de cara a la predicción de METS a nivel poblacional. Tradicionalmente, los modelos de series temporales utilizan unidades de medida similares para su entrada y salida. Es decir, si un modelo se alimenta de datos recogidos cada minutos, suele realizar predicciones referentes al siguiente minuto. Por ejemplo, véase el siguiente estudio [34] donde se utilizan distintos modelos de Deep Learning para realizar la predicción del precio de mercado de algunas compañías IT en bolsa. En este se utiliza una venta de 100 puntos para realizar una predicción de los siguientes 10 minutos, la cuál puede reutilizarse para realizar un multiple-step forecast y predecir valores más lejanos en el tiempo. ¿Tendría sentido la predicción de valores que no fueran puntos en una posición exacta en la serie temporal? Por ejemplo, ¿Es posible la predicción directa del valor máximo o mínimo de una acción en las próximas 24 horas? ¿Tiene esto mayor potencial que predecir el valor de todos los minutos en las próximas 24 horas y calcular el máximo? ¿Cómo de precisos y/o representativos son los resultados obtenidos de cara al objetivo? A raíz de este estilo de preguntas, es planteada la posibilidad de utilizar diferentes frameworks cuya diferenciación nazca del tratamiento de los datos así como del formato de salida. Debido a la flexibilidad de entradas y salidas ofrecida por los modelos de Deep Learning que se utilizan en la tarea de predicción realizado en este estudio, son muchas las variaciones o posibilidades que pueden ser contempladas. Bajo estas premisas se han propuesto 3 alternativas o variantes, las cuales serán testeadas en las mismas condiciones. Además se comparan también para diferentes horizontes de predicción según se detalla la Sección 7.2.

7.1. Diseño

Durante esta sección son definidos los frameworks utilizados de forma teórica, haciendo hincapié en los inconvenientes y bondades de cada uno de ellos. Para un más correcto entendimiento se identifica como N el número de series temporales registradas, M la ventana de tiempo utilizada en la predicción y P el horizonte de predicción (número de valores a predecir):

7.1.1. Alternativa 1: Entrada individual de elementos 1xM con agregación al final

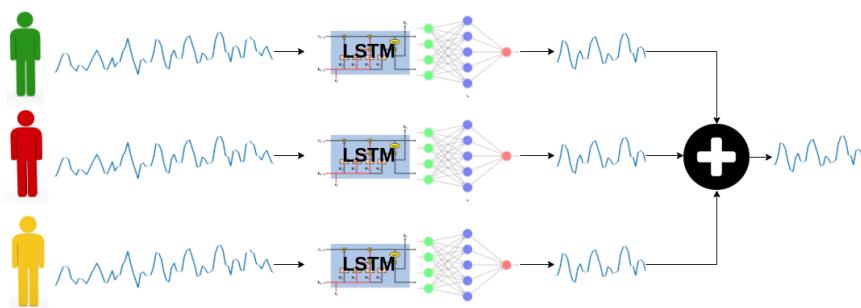


Figura 7.1: Framework con entrada de datos a nivel individual.

En esta primera alternativa se utiliza el enfoque que puede ser considerado más estándar. El modelo es entrenado a partir de las series temporales de cada uno de los participantes, utilizando en cada una de ellas M puntos para realizar una predicción del valor de actividad física del participante durante los P instantes siguientes. Para realizar una predicción sobre el valor de la población, una vez realizadas las predicciones individuales se realizará una agregación sobre los valores obtenidos.

■ Ventajas

Conjunto de entrenamiento grande: al utilizar los datos de cada uno de los participantes por separado, el modelo permite aprovechar toda la información de cada uno de estos, tanto para el entrenamiento como para la validación del mismo.

Flexibilidad: debido a su naturaleza, el modelo una vez entrenado es exportable a diferentes poblaciones. También es posible utilizar la información de distintas poblaciones para entrenar un solo modelo o poner el foco en un grupo o individuo en específico cuyo estudio pueda interesar especialmente. Esto además lo hace más resistente ante valores faltantes o pérdidas de individuos en el sistema.

- **Desventajas**

Ejecuciones múltiples: para obtener el valor agregado de la población se deberá realizar una ejecución para cada uno de los participantes.

Tendencias individuales: el modelo aprenderá sobre series temporales individuales y es posible que muchos detalles sobre tendencias poblacionales, y relaciones entre los diferentes individuos no sean reflejados en el resultado al ser los valores agregados a posteriori.

7.1.2. Alternativa 2: entrada matricial de elementos NxM

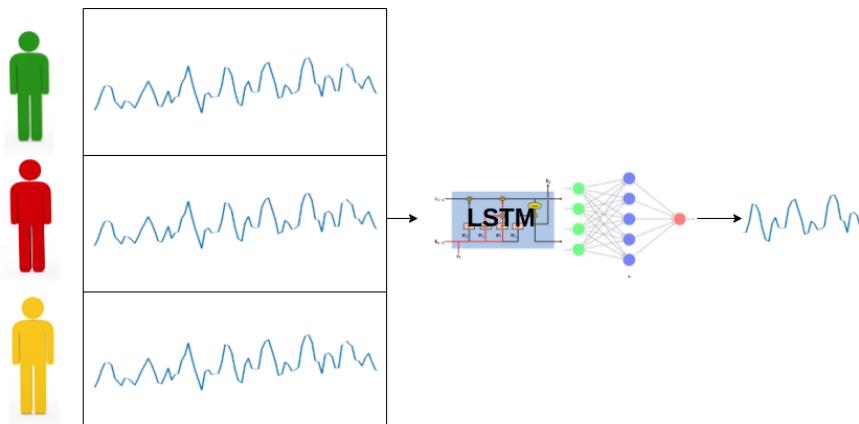


Figura 7.2: Framework con entrada de datos a nivel matricial

Esta segunda alternativa utiliza datos a un mayor nivel de abstracción, donde las entradas incluyen información no solo de los individuos si no de la población en su totalidad. Para ello, se toman los M puntos de cada una de las N series temporales registradas para alimentar el modelo y obtener una predicción de un solo valor. Este valor deberá ser representativo del ejercicio físico realizado por la población completa durante los próximos P instantes de tiempo.

- **Ventajas**

Entrada poblacional: El potencial de este modelo reside en la cantidad de datos que recibe como input. No solo existe información de cada individuo, sino de las relaciones entre estos en cada uno de los instantes. Este factor puede llegar a ser relevante ya que las predicciones se realizan a nivel de la población. Por ende, una entrada que tenga en cuenta todos los individuos al mismo tiempo puede ser más representativa que las series temporales individuales de estos.

- Desventajas

Predicciones complejas: El modelo es más pesado debido a que debe ser entrenado con matrices. Esto hace que la predicción de valores sea menos eficiente, especialmente si el modelo es escalado hacia poblaciones más grandes. A pesar de ello obtenemos un valor para toda la población estudiada en una sola ejecución.

Poca flexibilidad: Partiendo de la suposición de que el modelo pueda recibir entradas de distinto tamaño, su actuación puede verse afectada al cambiar el mismo, aumentando la complicación de reutilizar modelos pre entrenados en poblaciones diferentes o mantener las predicciones cuando los datos no pueden ser recogidos para algunos usuarios.

7.1.3. Alternativa 3: entrada agregada de elementos 1XM previamente agregados

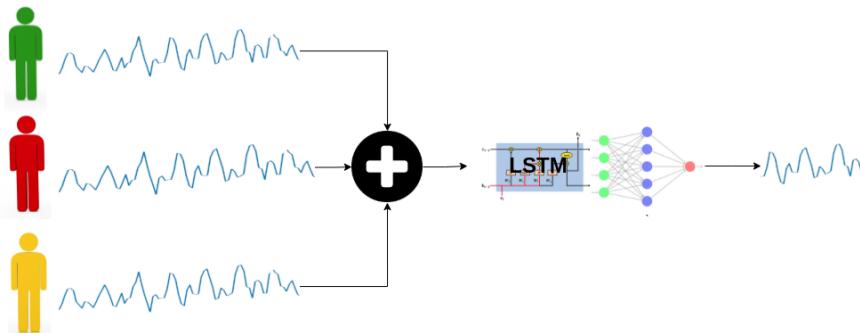


Figura 7.3: Framework con entrada de datos agregados al inicio

Finalmente, un enfoque mixto es propuesto por la última alternativa, que trata de aprovechar las ventajas de ambos modelos. Al igual que la primera alternativa el input de este consistirá en una sola serie temporal de M puntos. La diferencia principal es que esta serie temporal proviene de la agregación de las series temporales de toda la población.

- Ventajas

Información poblacional: existe información de todos los individuos deseados en cada una de las series temporales utilizadas.

Flexibilidad: al ser la entrada una sola serie temporal, el modelo puede generalizarse para distintos tamaños poblacionales ya que estos siempre pasan previamente por la agregación, aunque dependiendo de en qué consista esta la precisión de los resultados puede variar más o menos a la vez que lo hacen las series al agregarse.

- **Desventajas**

Datos de entrenamiento: Al entrenar específicamente con las series agregadas a nivel poblacional obtenemos, a priori, un input más pequeño y con menos información que en la segunda alternativa pero también una cantidad menor que en la primera.

7.1.4. M: Ventana de tiempo

La entrada utilizada siempre contendrá 1440 valores de METS. Estos son valores para cada uno de los minutos del día utilizado en ese momento para realizar la predicción. Estos corresponderán a valores a un determinado individuo o población dependiendo de la alternativa utilizada. El tamaño de esta ventana siempre será el mismo pues la entrada siempre corresponderá con un día de datos con un formato en minutos.

7.2. P: Horizonte de predicción

Además de los tres posibles frameworks o alternativas planteados, para cada uno de estos se llevará a cabo una comparación de diferentes horizontes de predicción. Llamamos horizonte de predicción P al número de valores predichos en una serie temporal:

$$Y_1, Y_2 \dots Y_p$$

. El modelo escogido puede ser entrenado para que produzca salidas en referencia a cualquier concepto deseado. De esta forma, existe la posibilidad de que el modelo logre predecir a partir de una entrada en minutos valores de METs representativos por ejemplo de cuartos de hora. En ese caso concreto, previo al entrenamiento se tomarían estos valores a predecir en todo el conjunto de datos y serían sumados de quince en quince. Al entrenar el modelo con entradas en minutos y salidas representativas de cuartos de hora este aprende a predecir estos valores en lugar de los propios minutos. De esta misma forma es posible trabajar con medias horas y cualquier otra medida de tiempo obtenida a partir de los minutos del conjunto de datos.

La idea final tras cada una de medidas de tiempo escogidas comparte un objetivo común: predecir con la mayor exactitud posible el total de METs poblacionales en un periodo de tiempo establecido. Por tanto, los horizontes de predicción manteniendo la abstracción de lo que representan, cambian respecto al número de valores en su salida (valor al que llamamos P). Por ejemplo, si queremos realizar la predicción de METs consumidos durante un día, el horizonte de predicción cambia si se realiza en minutos o en horas. Para representar dos horas, el horizonte en minutos requiere de 120 puntos.

Por su lado la predicción en horas basta de solo 2 puntos para representar el mismo umbral de tiempo. Lo importante, es que en ambas se mantiene común que en conjunto se predicen dos horas de datos, cambiando únicamente en el horizonte la cantidad de valores de la salida. Bajo estas premisas será comparable como afecta el tamaño de la salida a la exactitud del modelo, pero además será posible explorar las limitaciones de estos a la hora de generar diferentes salidas. Entre otras limitaciones, podremos observar la capacidad del modelo de capturar tendencias reales dentro del umbral de puntos predicho en las salidas de unidades menores (minutos) o cómo puede afectar el cambio de magnitud de las salidas (minutos en las entradas y horas en las salidas) a los resultados del modelo.

7.2.1. Horizonte 1: Predicción en minutos

Para este primer caso el modelo realiza una predicción de 120 valores. Estos son los valores correspondientes a los METs de cada uno de los minutos justamente posteriores a las dos horas continuas a la entrada de 1440 minutos (un día). Dependiendo de la alternativa estudiada, esos minutos representan valores de toda la población o de un individuo en concreto. De esta forma, si los valores de entrada son los METs correspondiente a los minutos 00:00 hasta 23:59 del primer día, la salida contendrá los 120 valores correspondientes a los minutos existentes entre las 00:00 y las 01:59 del segundo día. Posteriormente estos valores son sumados a lo largo del horizonte predicho para obtener el valor total de la población durante todo el periodo estudiado.

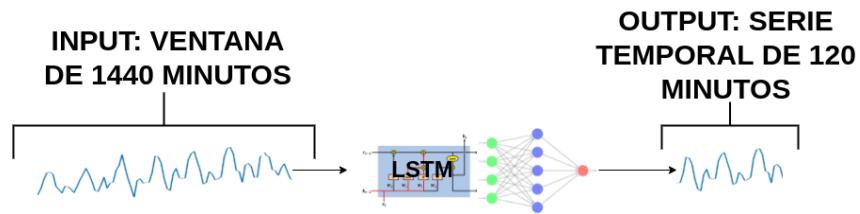


Figura 7.4: Ejemplo de horizonte predicho con formato en minutos.

7.2.2. Horizonte 2: Predicción en horas

El segundo formato de salida a estudiar es el referente a las horas. Dada una entrada de 1440 minutos se predicen los dos valores de METs correspondientes a cada una de las siguientes 2 horas justamente posteriores a dicha entrada. De nuevo, en el caso de que la entrada fuese el primer día desde las 00:00 hasta las 23:59, la salida serían dos valores correspondientes al total de METs consumidos durante la primera hora del segundo día 00:00-00:59

y un segundo con el total de valores desde 01:00 hasta las 01:59. La suma de estos 2 valores corresponde a la predicción de gasto calórico total de la población para el horizonte en horas.

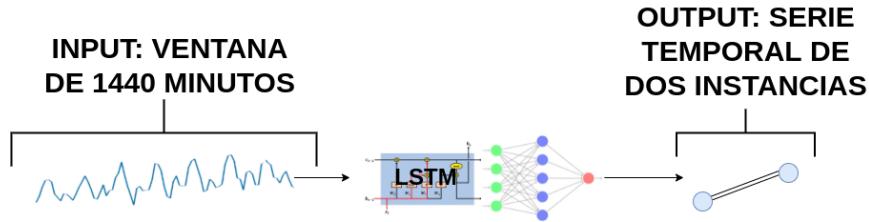


Figura 7.5: Ejemplo de horizonte predicho con formato en horas.

7.3. Implementación

En esta sección serán detallados las estructuras de las diferentes propuestas explicadas durante el apartado 7.1. Todo el cómputo fue realizado mediante el lenguaje de programación Python. Este cuenta con multitud de librerías para el análisis de datos así como de Machine y Deep Learning, haciéndolo una de las mejoras alternativas para la implementación.

7.3.1. Aprendizaje

A continuación se recoge la implementación necesaria para llevar a cabo el proceso de aprendizaje y aquellas definiciones previas al mismo. Como aprendizaje entendemos el proceso por el cual un modelo mejora su capacidad para realizar una tarea específica a medida que se le proporcionan más datos y sus parámetros son ajustados. Por tanto se habla entre otros de modelos y sus capas o técnicas de entrenamiento. Para la secuencia de aprendizaje del trabajo se ha optado por una implementación en *keras* [35]. Esta es una librería de código abierto que ofrece una interfaz sobre la librería *Tensor Flow* [36] para el uso de redes neuronales y otras herramientas de Inteligencia Artificial.

A continuación se presenta la definición python del modelo LSTM utilizado dado un tamaño de ventana y un horizonte de predicción, la cual queda representada en la Figura 7.6. La implementación de este modelo se describe en el Listing 7.1.

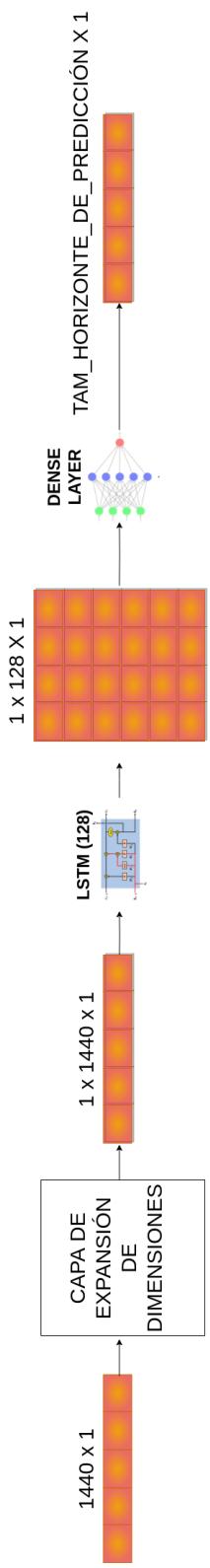


Figura 7.6: Definición del modelo con entradas con entradas matriciales.

```

1
2 import tensorflow as tf
3
4 inputs = tf.keras.layers.Input(shape=(WINDOW_SIZE))
5 # expande la dimension de entrada para ser compatible con LSTM
6 x = tf.keras.layers.Lambda(
7     lambda x: tf.expand_dims(x, axis=1)
8 )(inputs)
9 x = tf.keras.layers.LSTM(128, activation="relu")(x)
10 output = tf.keras.layers.Dense(HORIZON)(x)
11 model_LSTM = tf.keras.Model(
12     inputs=inputs, outputs=output, name="model_5_lstm"
13 )

```

Listing 7.1: Definición de los modelo de entradas 1xM.

El modelo anterior está definido a partir de una capa de entrada de un tamaño similar a la ventana de tiempo utilizada. Posteriormente se añade una capa de expansión para que la entrada sea compatible con la capa LSTM. Esto solo es necesario en los casos donde el input consiste en una única serie temporal, es decir la entrada es de 1xM elementos (alternativas 1 y 3). En el caso de la entrada de NxM (alternativa 2) establecemos el input directamente con el número de participantes y el tamaño de la ventana, como podemos ver en la Figura 7.7. En el Listing 7.2 se puede ver esta implementación en python.

```

1
2 import tensorflow as tf
3
4 inputs = tf.keras.layers.Input(
5     shape=(number_of_participants,WINDOW_SIZE)
6 )
7 x = tf.keras.layers.LSTM(
8     128,return_sequences=True,activation="relu"
9 )(inputs)
10 output = tf.keras.layers.Dense(HORIZON)(x)
11 model_LSTM = tf.keras.Model(
12     inputs=inputs, outputs=output, name="model_5_lstm"
13 )

```

Listing 7.2: Definición del modelo de entradas NxM.

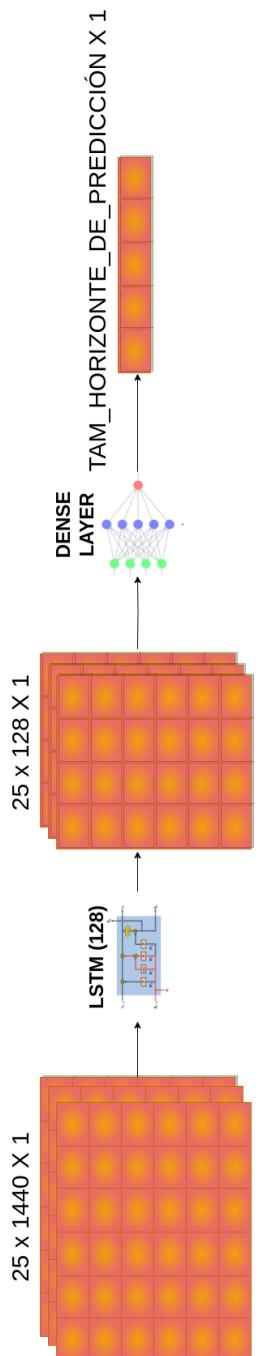


Figura 7.7: Definición del modelo con entradas matriciales.

La capa de LSTM consta de 128 unidades, por lo que se reducirá la dimensionalidad de la entrada, reduciendo la complejidad de los datos y simplificando la extracción de características clave de la secuencia antes de pasar a capas posteriores. Además se introduce una capa activación ReLU (*rectified linear unit*), la cual permite discriminar aquellos valores que realmente han presentado una respuesta a los filtros aplicados de aquellos que no y realmente sus valores actuarían como ruido para el resto. A esta se le añade una capa densa o *Fully Connected* la cuál se encarga de convertir la representación de características de alto nivel en la salida deseada, es decir, un array de valores de un tamaño similar al horizonte de predicción. Con esto queda definido el modelo LSTM pero todavía se requiere de su compilación así como de la definición del resto de objetos de aprendizaje.

Por un lado se añade el uso de *EarlyStopping*. Esta es una técnica de regularización en entrenamiento de modelos de aprendizaje automático ampliamente utilizada para evitar el sobre-ajuste y con ello optimizar el rendimiento del modelo. Esto es logrado mediante la detención del entrenamiento en el punto donde aunque el error de entrenamiento continúa bajando, el error en validación aumenta, lo que indica que el modelo ha alcanzado un punto óptimo en términos de generalización y no debe continuar entrenando [37]. En la siguiente Figura 7.8 puede apreciarse el momento exacto en el que el entrenamiento pararía debido al EarlyStopping.

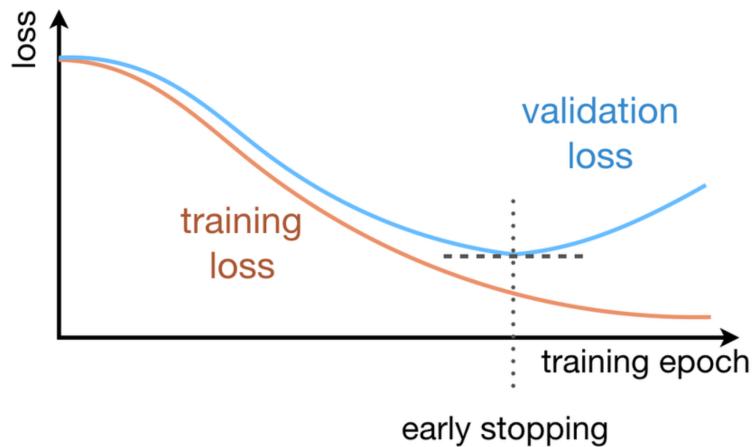


Figura 7.8: Funcionamiento de EarlyStopping. Fuente: [9].

```

1 early_stopping = tf.keras.callbacks.EarlyStopping(
2     monitor="val_loss",
3     patience=4,
4     mode="auto"
5 )

```

Listing 7.3: Definición de early stopping.

El Listing 7.3 muestra como se realiza la implementación del early-stopping. El parámetro *patience* indica la cantidad de épocas adicionales en las que se continuará el entrenamiento cuando el error de validación para de mejorar. Si bien se quiere prevenir el sobre-entrenamiento, debemos evitar que este se detenga prematuramente debido a pequeñas fluctuaciones en las métricas de evaluación conduciendo a modelos subóptimos.

Respecto a la compilación del modelo, es utilizada como métrica de error *Mean Absolute Error*. Esta es una medida que indica la magnitud promedio de los errores entre las predicciones de un modelo y los valores reales de un conjunto de datos. Es intuitiva debido a que se expresa en una medida similar a los valores reales, es decir METS, y penaliza menos los errores debidos a valores atípicos que otras medidas como *Mean Square Error*.

```

1 model_LSTM.compile(loss="mae",
2                     optimizer=tf.keras.optimizers.Adam(),
3                     metrics=["mae"]
4 )

```

Listing 7.4: Compilación del modelo.

Podemos ver en Listing 7.4 el uso de Adam, el cuál es un algoritmo de optimización ampliamente utilizado en el campo del aprendizaje profundo que permite la mejora de la precisión y la velocidad del entrenamiento de modelos de aprendizaje. Esto se debe a su tasa de aprendizaje la cuál va disminuyendo conforme los parámetros confluyen hacia un óptimo en el error, permitiendo que al inicio estos parámetros cambien de una manera más rápida y pero el modelo converja de manera correcta hacia una solución cuando se encuentre cerca de esta. Los valores utilizados son aquellos introducidos por defecto en keras ($\eta = 0.001$).

Por último es realizado el entrenamiento, cuyo código muestra el Listing 7.5. Se establece un número de épocas grandes para que el modelo pueda entrenar hasta que sea detenido por el *EarlyStopping*. El tamaño de batch escogido es un tamaño estándar que permite lotes con una mayor generalización al ser grandes. Esto es importante debido a la alta similitud que puede existir entre los distintos ejemplos.

```

1 model_LSTM.fit(X_train,
2                 y_train,
3                 epochs=100,
4                 verbose=1,
5                 batch_size=256,
6                 validation_data=(X_validation, y_validation),
7                 callbacks=[early_stopping])

```

Listing 7.5: Entrenamiento del modelo.

7.3.2. Test: 9-fold Cross-Validation

La estimación de los resultados ha sido realizado mediante *Cross-Validation* con $k=9$. Esta es una manera robusta de evaluar y seleccionar modelos. La idea detrás de la validación cruzada es dividir el conjunto de datos en varias partes o *folds*. Posteriormente el modelo se entrena y evalúa múltiples veces, utilizando diferentes combinaciones de estos (usualmente $k-1$ son utilizados en el proceso de entrenamiento y el restante durante el proceso de evaluación). A estas combinaciones de folds se le conoce como *split*. Finalmente es realizado un promedio sobre los valores de los diferentes splits, obteniendo una medida del rendimiento del modelo con una buena generalización que evita sesgos en el conjunto de datos utilizados, además de ayudar a detectar sobreajuste en ciertos de estos. Típicamente el número de divisiones en un cross-validation suelen ser un múltiplo de 10, pero en este caso se ha decidido la partición en 9 conjuntos o folds debido a que era una abstracción apropiada en referencia al dataset, ya que 9 representan de manera más natural los datos utilizados. Esto se debe a la división del conjunto de entrenamiento que se ha realizado mediante días, de tal forma que los conjuntos sean totalmente independientes y no existan ejemplos muy similares en particiones diferentes. De otra forma, al generar el conjunto de datos mediante una ventana corrediza de un minuto, sería posible encontrar conjuntos de datos muy similares entre estos *folds* y aún peor, entre los conjuntos de entrenamiento y test afectando al rendimiento del modelo. Podemos observar esto si tomamos por ejemplo la predicción de los minutos del cuarto día entre las 00:00-01:59, para lo que se usarían datos del tercer día entre las 00:00-23:59. El elemento siguiente sería la predicción del cuarto día desde 00:01-02:00 con los datos del tercero desde las 01:00 hasta las 00:00 del cuarto día. Como podemos ver existe un solo minuto de diferencia, haciendo la entrada un 99,93 % similar y la salida un 99,16 %. Por ello, es de gran importancia que los datos de los diferentes splits constituyan días independientes y por ende los elementos de cada splits también lo sean.

Al no existir información de todos los participantes durante los dos primeros días de estudio, estos fueron descartados de los 29 días totales del mismo. En caso de nos descartarlos, las divisiones no podrían realizarse de

manera apropiada debido a que o bien quedarían mezclada información de días similares dentro de cada uno de los splits o bien se perdería una cantidad mayor de información al tener que descartar conjuntos de horas en un mayor número de estos. Realmente, los datos del día 2 comienzan a ser completos para toda la población sobre las 16:00 horas. Pero utilizar este provocaría que más días dentro de los conjuntos empezaran en horas diferentes a las 00:00. Como se ha explicado para no mezclar datos muy similares en los diferentes splits y para descartar el número mínimo de datos los días deben comenzar en su mayoría a las 00:00 dentro de cada conjunto y utilizar como último elemento en ellos predicciones que abarquen hasta las 23:59 de ese mismo día. La otra opción sería mantener un primer conjunto de un tamaño mayor al resto. Esta idea queda descartada ya que se ha considerado que podría ser determinante en los resultados de cada splits el que el tamaño de entrenamiento, test y validación no sea siempre similar.

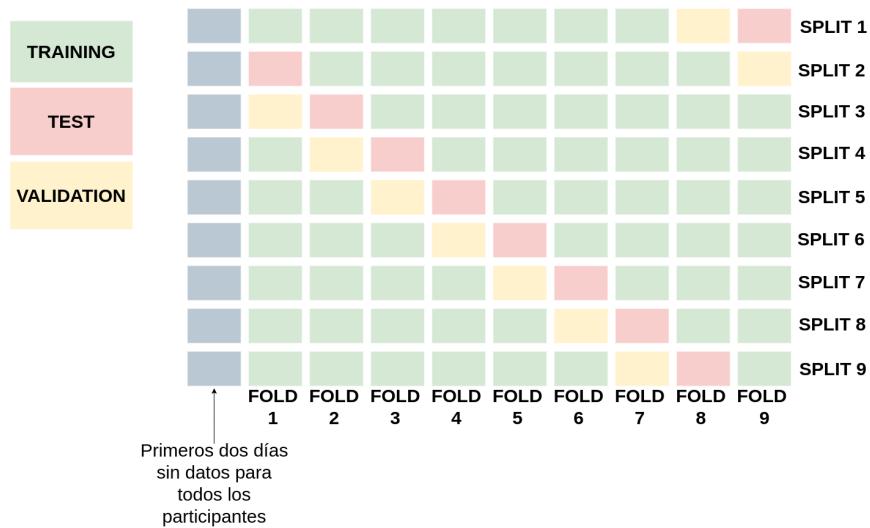


Figura 7.9: 9 fold cross validation.

Para evitar que todo esto pase, las particiones comienzan a realizarse desde el tercer día de estudio el cuál es el primer día con un conjunto completo de datos. Por tanto el estudio queda en 27 días, cuya división más natural se hace si cada partición consta de tres días completos teniendo en cuenta los valores iniciales utilizados en las predicciones y los últimos valores predichos por las mismas. Este efecto es apreciable en la Figura 7.9. De esta forma el primer conjunto inicia desde el tercer día a las 00:00 y dicho ejemplo incluye datos hasta el mismo día a las 23:59. El valor que predice es una ventana de tiempo de dos horas desde el día cuatro a las 00:00 hasta el día cuatro hasta las 01:59. El último día en este mismo conjunto provendrá desde los datos

del día cuatro a las 22:00 hasta el día cinco a las 21:59 que predicen datos del día cinco desde las 22:00 hasta las 23:59. El siguiente conjunto se realiza de manera similar empezando por el día seis a las 00:00. En la Figura 7.10 es observable una mejor representación de este conjunto. Por tanto, algunos datos no se utilizan pero con ello separamos completamente la información incluida en cada uno de los bloques. Por ejemplo, ya que como hemos dicho el fold 2 deberá empezar por el día sexto perderemos datos del quinto día que predicen el día sexto.

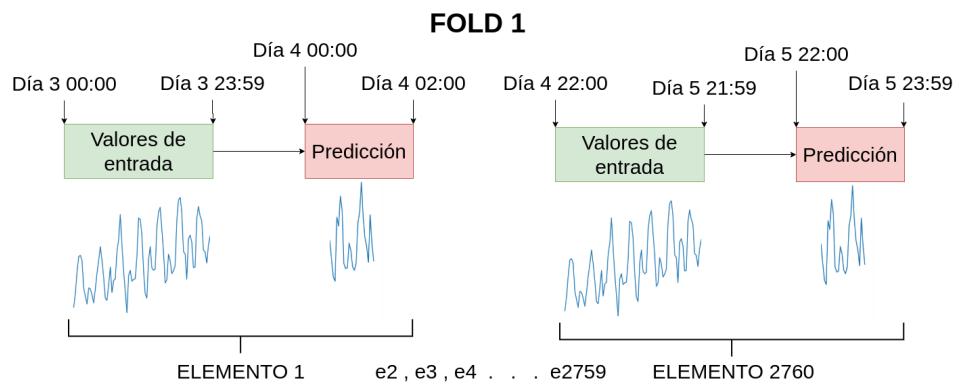


Figura 7.10: Representación Fold 1.

Capítulo 8

Evaluación

En este capítulo se presentan los resultados provenientes de la evaluación de las tres alternativas introducidas en el Capítulo 7 en diferentes test. Durante el capítulo se referencia a las alternativas como: modelo de entradas individuales o framework Individual para Alternativa 1 de la sección 7.1.1, modelo de entradas matriciales o framework Matricial para la Alternativa 2 de la sección 7.1.2 y modelo de entradas agregadas o framework Agregado para la Alternativa 3 de la sección 7.1.3. Los resultados son presentados para los diferentes horizontes y provienen de la realización de predicciones sobre los valores reales de gasto de METs de la población en diferentes instantes de tiempo. En ellos se busca la mayor precisión posible para cada una de los modelos presentados anteriormente. Diferentes medidas y gráficas son expuestas a lo largo del capítulo para entender las debilidades y fortalezas de cada uno de los modelos estudiados.

Para un mejor entendimiento de los resultados, primero se exponen algunos valores medios sobre los datos a predecir. Por un lado tenemos los valores medios punto a punto de las series temporales que se tratan de predecir, así como los valores totales de METs consumidos por la población en el periodo de tiempo predicho. Ambas medidas serán desglosadas a continuación en las secciones 8.1.1 y 8.1.2 referentes a estas medidas. De forma breve comentar que los valores "POINT_TO_POINT" en las tablas 8.1 y 8.2 hacen referencia a los valores individuales dentro de cada una de las series temporales predichas. En el caso de los minutos existen 120 valores que predecir, "POINT_TO_POINT" es el valor medio de cada uno de estos puntos en la predicción. Para el caso de las horas hacen referencia a cual es la media de cada uno de los dos puntos que se predicen. La columna indicada como "ALL_METS" haría referencia al valor promedio de la suma de estos, luego 120 puntos sumados en el caso de los minutos o 2 en el de las horas. Por tanto este último es el valor principal que se trata de predecir, es por ello que en ambas tablas es similar ya que representan lo mismo. A continuación

en la Figura 8.1 se observa como se obtienen dichos valores para cada una de las series temporales. Calculados estos valores por serie, se realiza un promedio con todas las series a predecir en cada split. La media de los valores de todos los split y por ende de los valores a predecir en el conjunto global del trabajo son los valores que se exponen en las tablas 8.1 y 8.2, así como en el resto de tablas del trabajo. Para consultar los valores desglosados por split acuda a la sección resultados completos en el Apéndice.

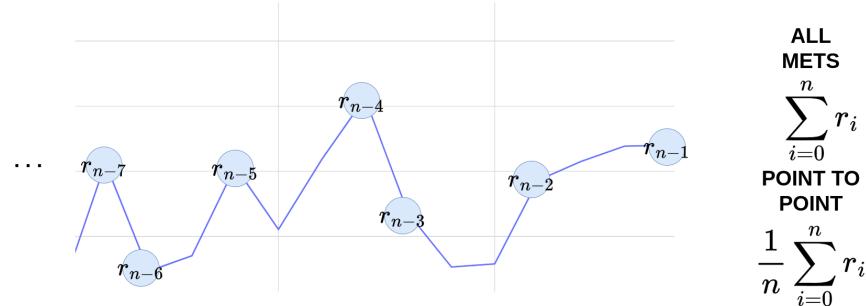


Figura 8.1: Representación por serie temporal de los valores punto a punto y totales.

Tabla 8.1: **Valores medios a predecir horizonte en horas**

Metrica	POINT_TO_POINT	ALL METS
Mean	313.96	4627.93
Std	79.73	159.46

Tabla 8.2: **Valores medios a predecir horizonte en minutos**

Metrica	POINT_TO_POINT	ALL METS
Mean	38.56	4627.93
Std	1.32	159.46

Por tanto, la Tabla 8.1 muestra que en media en las series temporales en horas cada uno de los dos puntos a predecir tiene un valor de 313.96 METs y en horizonte completo 4627.93 METs. Por su lado, el horizonte en minutos en la Tabla 8.2, tiene una media de 38.56 METs en cada uno de los 120 minutos predichos y obviamente un valor similar para el horizonte completo. Los valores de desviación indican en media como varían estos valores en las distintas series temporales. Esto nos indica que en el horizonte en horas entre un ejemplo y otro encontraremos en media hasta 80 METs de diferencia en cada punto, 1.32 METs para los diferentes instantes del horizonte en minutos

y hasta 159.46 METs de diferencia en el total de METs consumidos por la población durante todo el conjunto de dos horas.

8.1. Resultados en las predicciones de test

8.1.1. Resultados punto a punto

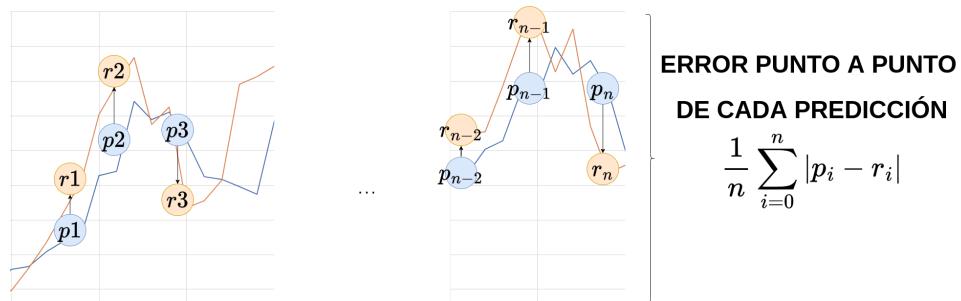


Figura 8.2: Representación del calculo de errores punto a punto.

Los valores punto a punto provienen de la comparación de valores resultantes por la red en cada uno de los instantes de la salida. De tal forma que para el horizonte temporal en minutos cada uno de los 120 valores resultantes es medido respecto del valor real en dicho minuto de tiempo. En el caso de las horas son los dos puntos de salida correspondientes a la predicción los que se comparan con los dos valores reales de METs consumidos en cada una de esas dos horas durante el estudio. Todos los resultados son tratados a nivel poblacional independientemente del modelo utilizado para extraerlos.

Tabla 8.3: **Resultados punto a punto para el horizonte en horas.**

Framework Agregado		
Metrica	MAE Agregado	MSE Agregado
Mean	103.27	19409.44
Std	7.49	2552.54
Framework Individual		
Metrica	MAE Individual	MSE Individual
Mean	129.34	32625.02
Std	13.72	5511.10
Framework Matricial		
Metrica	MAE Matricial	MSE Matricial
Mean	183.53	65314.94
Std	95.90	67535.81

Los resultados de la Tabla 8.3 indican que para el horizonte en horas el modelo con entradas agregadas es el que presenta resultados con menor error, así como una desviación menor entre ellos. Con una media de 2313.96 METs por hora en las instancia de test, el error medio absoluto para los 9 folds es de 103.27 METs, lo que supone para los valores a predichos con respecto a los valores medios reales de la Tabla 8.1 un error 4.46 %. Por otro lado el framework Individual con un MAE medio de 129.3454 METs supone un 5.58 % de error respecto de los valores reales. Finalmente, el modelo de entrada matricial cuenta con el MAE medio más alto de 183.53 METs, es decir un 7.93 %.

De la misma forma, los valores de desviación son mayores para este último framework, llegando a 95.90 METs. Esto es principalmente ocasionado por el split 3 donde se encontró un valor atípico en el error de 425.89 METs, un 370.23 % mayor que el split con menor MAE para el framework (consultar Apéndice para más información). Seguido a este el modelo de entradas individuales con una variación mucho más suave de 13.72 METs y experimentando su menor valor en el caso del framework Agregado donde esta no llega a 7.5 METs.

Tabla 8.4: Resultados punto a punto para el horizonte en minutos.

Framework Agregado		
Metrica	MAE Agregado	MSE Agregado
Mean	2.39	10.55
Std	0.19	1.67
Framework Individual		
Metrica	MAE Individual	MSE Individual
Mean	3.20	20.84
Std	2.53	13.19
Framework Matricial		
Metrica	MAE Matricial	MSE Matricial
Mean	2.53	13.19
Std	0.32	3.12

Respecto al horizonte en minutos (Tabla 8.4), el modelo de entradas agregadas sigue presentando los mejores resultados. Con un MAE medio de 2.39 METs (Tabla 8.2) este supone un 6.20 % respecto al valor real promedio y tan solo una variación de 0.19 METs en el error. Al contrario que el caso anterior, el framework Matricial mejora sus resultados respecto al Individual. Contando este con un MAE promedio 2.53 METs supone un error del 6.57 %, mientras que el framework Individual supone un error del 8.32 %.

8.1.2. Resultados para el valor total

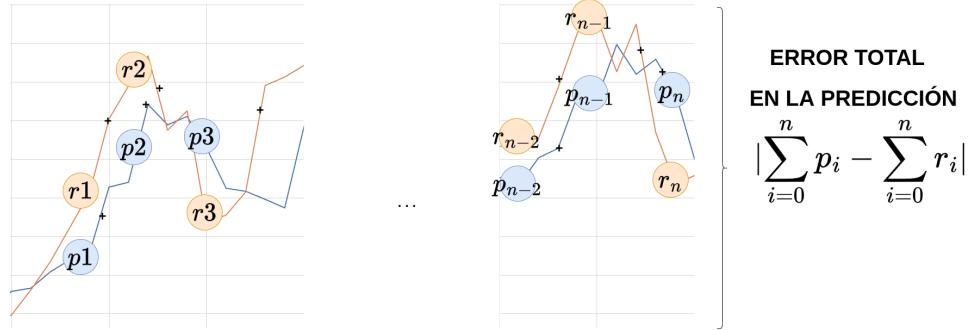


Figura 8.3: Representación del calculo de errores totales.

El valor realmente a predecir es el conjunto de METs consumidos por la población durante el periodo de tiempo establecido. Por lo tanto si estamos prediciendo los valores desde las 00:00 hasta las 01:59 del séptimo día del estudio, el valor total se refiere al total de METs consumidos en este intervalo. Si la salida se realiza en minutos se deben sumar los 120 puntos de la serie temporal de salida. Si la predicción se realiza en horas, se suman los 2 valores de salida. Aunque el análisis es extrapolable, este nos permitirá una mejor comparación de los modelos a posteriori con los diferentes horizontes de predicción.

Tabla 8.5: **Resultados para el horizonte en horas**

Framework Agregado		
Metrica	MAE Agregado	MSE Agregado
Mean	188.69	63807.02
Std	17.06	9295.48
Framework Individual		
Metrica	MAE Individual	MSE Individual
Mean	248.55	114370.189
Std	28.27	20540.39
Framework Matricial		
Metrica	MAE Matricial	MSE Matricial
Mean	307.83	164328.66
Std	74.19	71708.41

Tabla 8.6: Resultados para el horizonte en minutos

Framework Agregado		
Metrica	MAE Agregado	MSE Agregado
Mean	190.68	65028.42
Std	29.14	17716.06
Framework Individual		
Metrica	MAE Individual	MSE Individual
Mean	358.41	217905.15
Std	92.78	86975.99
Framework Matricial		
Metrica	MAE Matricial	MSE Matricial
Mean	245.56	110202.85
Std	45.40	36414.91

Los resultados de las tablas 8.5 y 8.6, al contrario que los resultados valor a valor comparten magnitudes. Esto es debido a que el valor a predecir es similar en ambas (total de METs consumidos por la población en todo el umbral). El framework Agregado mantiene un comportamiento muy similar en ambos casos, aunque la variación con el horizonte en minutos es mayor para el error de los diferentes splits. Respecto al valor medio total de la tabla 8.1 el error para el horizonte en minutos es similar en ambos. En el modelo de entradas agregadas (en torno a un 4%). Para el framework Individual encontramos un error medio del 5.3 % para el horizonte en horas y un 7.7 % en el caso de los minutos. Finalmente el modelo Matricial es el único que empeora al pasar de horas a minutos con porcentajes de error del 6.6 % y 5.3 % respectivamente.

8.1.3. Comportamiento del modelo en largo plazo

Para medir el rendimiento del modelo desde un punto de vista más amplio es propuesto el siguiente experimento. En cada uno de los splits el test está formado por un fold, y por ende por 3 días, donde uno de ellos solo se utiliza como datos de entrada para la predicción. Para los otros dos días, se realizarán las predicciones de 2 horas de aquellos instantes que comienzan en una hora exacta. Es decir, predicción desde 00:00 hasta 01:59, predicción de 02:00 hasta 03:59...

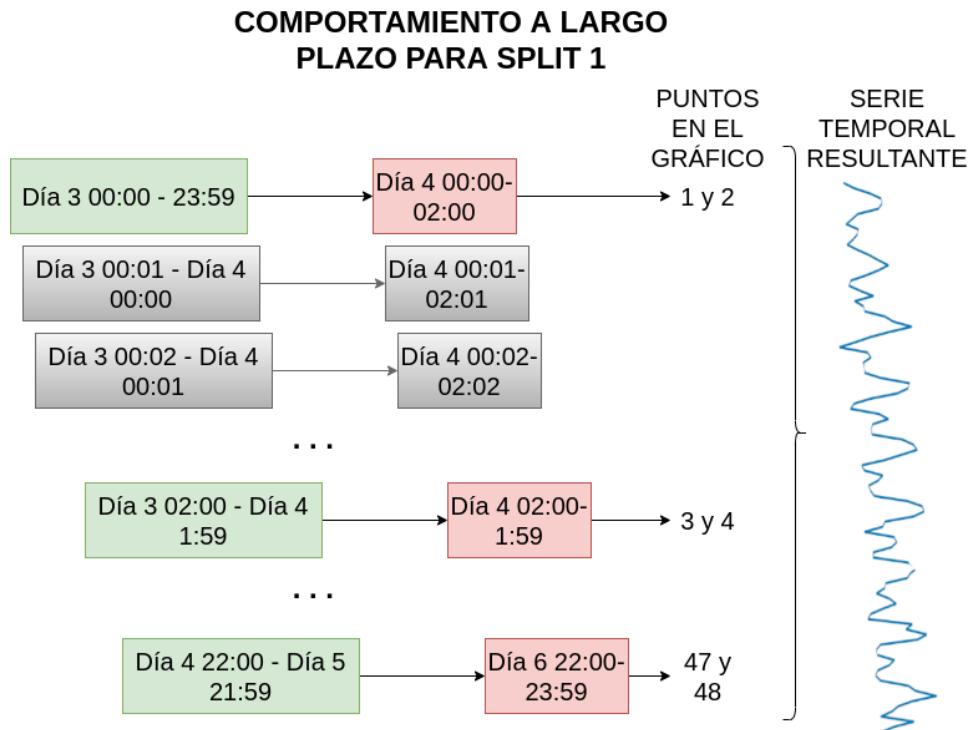


Figura 8.4: Representación de la extracción del experimento a largo plazo para el split 1.

Posteriormente se mide el total de METs que fueron consumidos durante esos dos días y el total de METs que fueron predichos por el modelo. Aunque realmente no se realiza una predicción sobre el consumo de METs en un horizonte de 48 horas, estos resultados permiten observar cómo sería el comportamiento del modelo a lo largo de un periodo de tiempo mayor. Podemos observar mejor la composición de este experimento en la Figura 8.4. Por su parte las tablas 8.8 y 8.7 presentan el valor medio total de METs reales y predichos por cada uno de los frameworks en el experimento. Los resultados muestran que en el peor de los casos (framework Individual con entradas en minutos) el modelo obtiene un resultado 93.03 % similar al valor real.

Tabla 8.7: Comportamiento a largo plazo horizonte en horas

Metrica	Agregado	Individual	Matricial	Real
Mean	105095.22	105739.58	102007.61	106429.81
Std	3436.59	3887.90	6320.319	3613.168

Tabla 8.8: Comportamiento a largo plazo horizonte en minutos

Metrica	Agregado	Individual	Matricial	Real
Mean	105169.41	99018.89	102138.60	106429.81
Std	3436.59	3887.90	6320.319	3613.168

Para una mayor información del comportamiento de los frameworks durante este experimento se presentan las gráficas de valores reales frente a predichos correspondientes al split 1 en cada uno de estos experimentos en el caso de las horas ((Figura 8.5, Figura 8.7 y Figura 8.9) y de los minutos (Figura 8.11, Figura 8.13 y Figura 8.15). De cara poder medir la relación que realmente existe entre las predicciones y los valores reales, se debe utilizar otro método que proporcione información más relevante sobre la correspondencia de los valores que lo que la observación de las figuras de los valores predichos frente a los valores reales puede ofrecer, especialmente en el caso de los valores de los minutos donde cada tramo predicho son 120 puntos diferentes (figuras 8.11, 8.13 y 8.15). Las gráficas de ajuste OLS (Ordinary Least Squares) muestran cómo de bueno es el ajuste realizado por los modelos de diferentes entradas en el experimento a largo plazo. Se toman los valores predichos en el eje y del gráfico, los valores reales en el eje x, y sobre estos se realiza una recta de Mínimos Cuadrados Ordinarios u OLS. Este es un método utilizado en estadística para estimar los parámetros de una relación lineal entre dos o más variables, encontrando la mejor línea recta que se ajusta a un conjunto de datos. Esta técnica utiliza el minimizado de la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por la línea. Como resultado obtenemos una línea que ajusta los datos representada por la ecuación:

$$Y = b + mX$$

cuya idea en este caso será que la pendiente de la recta sea lo más cercana a uno posible así como el término independiente sea cercano a cero. Esto se debe a que la pendiente cercana a uno es significativa de que el modelo es capaz de captar la tendencias de la población y que por ende existe una correspondencia entre los valores reales y los valores predichos, mientras que el término independiente cuanto menor sea, mayor correspondencia existirá entre las magnitudes de los valores predichos. Por ende, este valor es sensible a la magnitud y si la predicción se realiza sobre valores mayores es normal que este valor también sea mayor. Las gráficas correspondientes a este ajuste se exponen inmediatamente después de las gráficas correspondientes al experimento en cuestión.

Valores a largo plazo para el horizonte en horas.

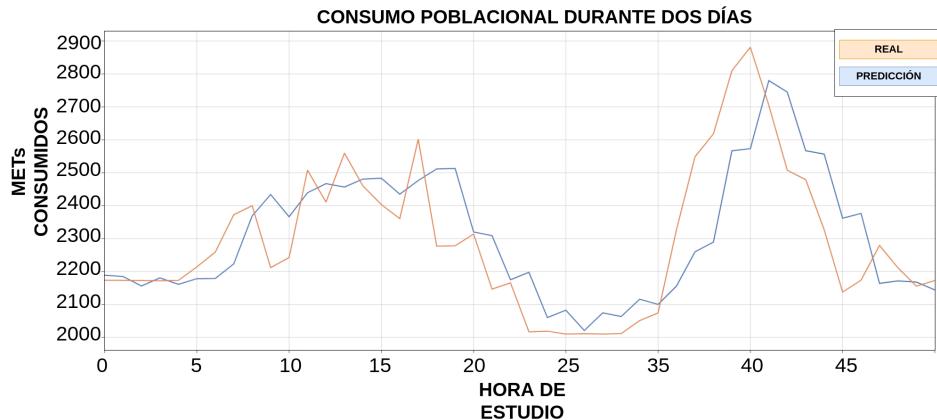


Figura 8.5: Predicción en horas de dos días para el split 1 modelo Agregado.

$$Y = 1300.28374 + 0.7233X$$

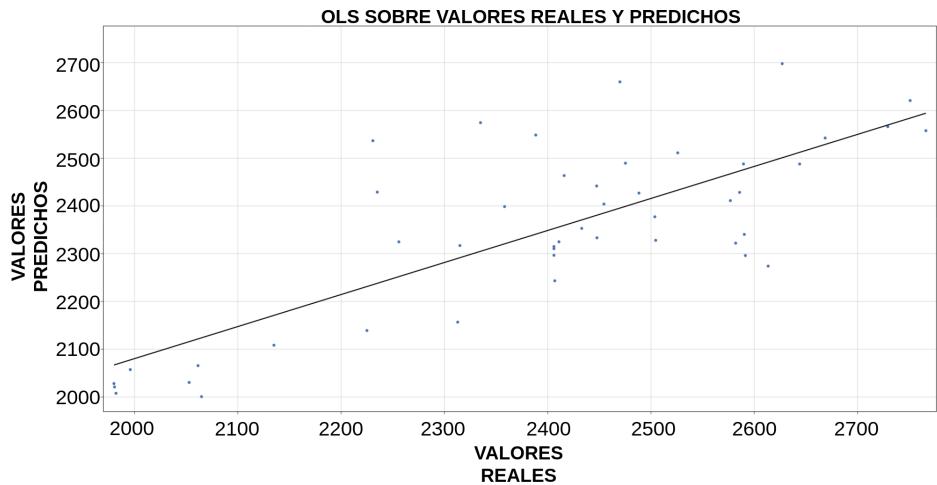


Figura 8.6: Recta de Mínimos Cuadrados Ordinarios para el modelo de entradas agregadas con horizonte en horas.

La Figura 8.5 muestra el resultado de las predicciones hechas por el framework Agregado durante los días de test del split 1 con el horizonte en horas. Posteriormente se presenta la recta OLS (Figura 8.6) correspondiente a dichos valores con una pendiente de 0.73.

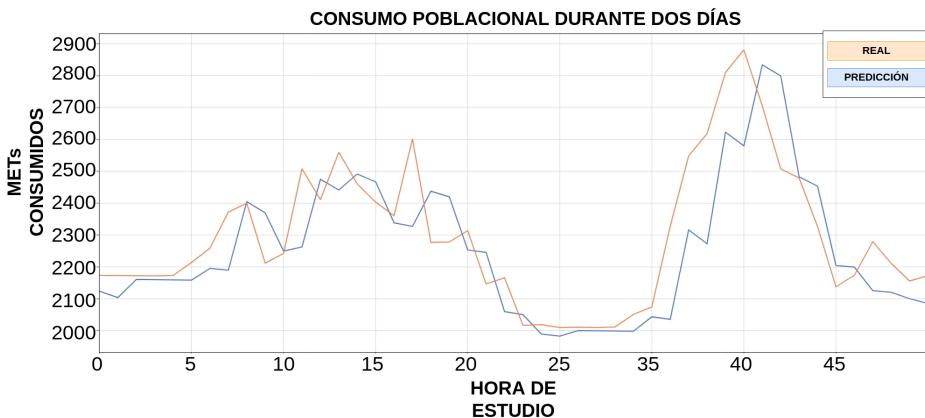


Figura 8.7: Predicción en horas de dos días para el split 1 modelo Individual.

$$Y = 440.809 + 0.7909X$$

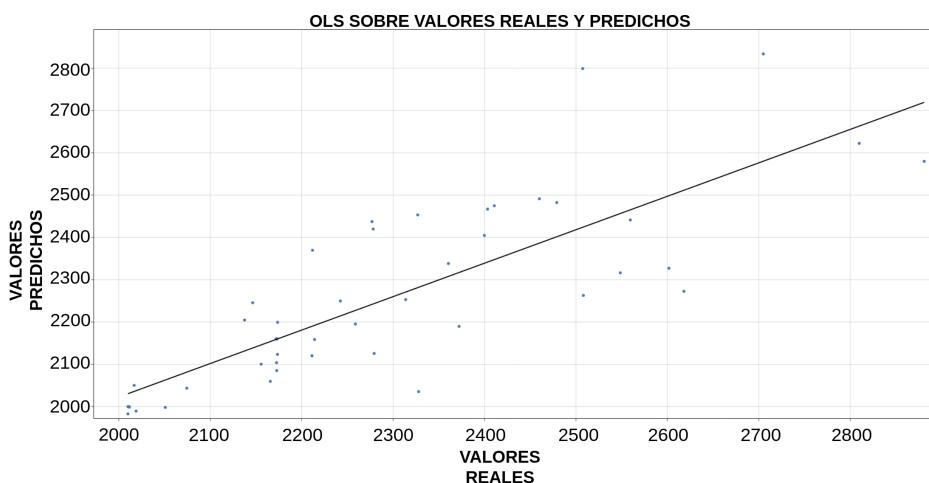


Figura 8.8: Recta de Mínimos Cuadrados Ordinarios para el modelo de entradas individuales con horizonte en horas.

Para el framework Individual la Figura 8.7 muestra el resultado de las predicciones hechas durante los días de test del split 1 con el horizonte en horas. La Figura 8.8 presenta la recta OLS correspondiente a dichos valores con pendiente igual a 0.79.

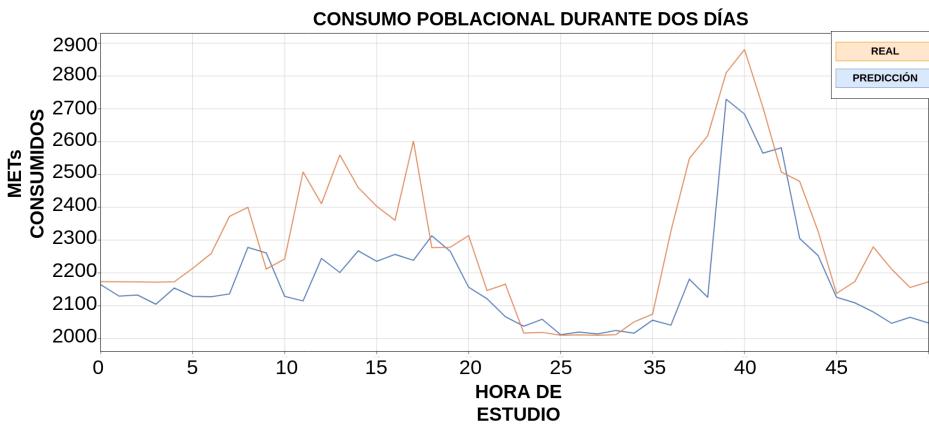


Figura 8.9: Predicción en horas de dos días para el split 1 modelo Matricial.

$$Y = 1212.079 + 0.6869Y$$

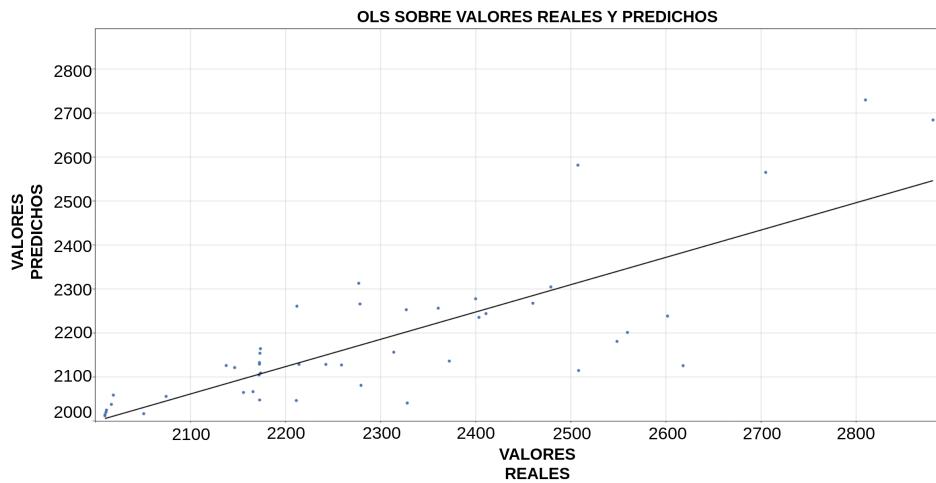


Figura 8.10: Recta de Mínimos Cuadrados Ordinarios para el modelo de entradas matriciales con horizonte en horas.

Por último el valor para el framework Matricial y con el horizonte en horas encontramos la Figura 8.9. Por su parte, la recta OLS de la Figura 8.10 presenta un valor de pendiente 0.79.

Valores a largo plazo para el horizonte en minutos.

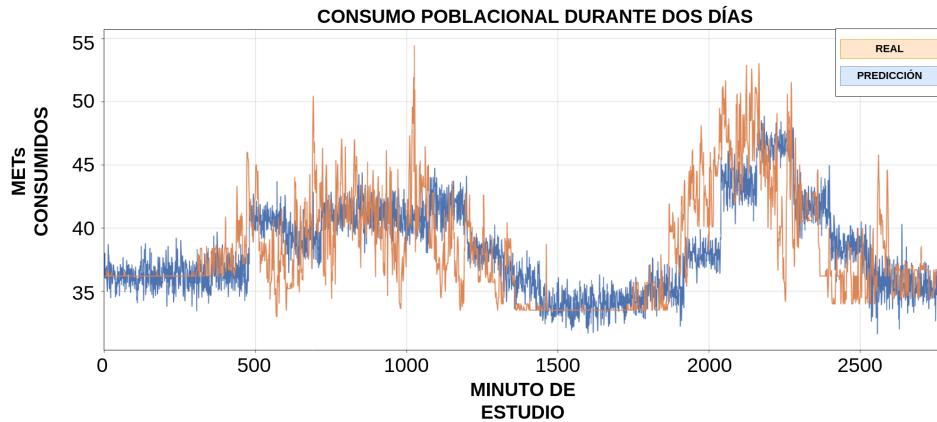


Figura 8.11: Predicción en minutos de dos días para el split 1 modelo Agregado.

$$Y = 18.057038 + 0.529342X$$

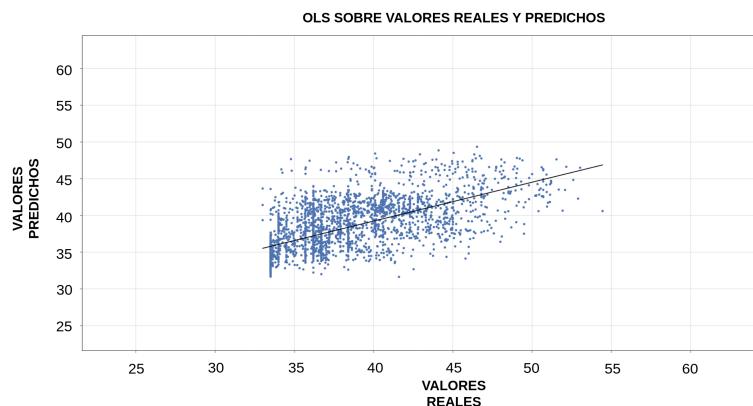


Figura 8.12: Recta de Mínimos Cuadrados Ordinarios para el modelo de entradas agregadas con horizonte en minutos.

La Figura 8.11 muestra el resultado de las predicciones hechas durante los días de test del split 1 con el horizonte en minutos por el framework Agregado y la Figura 8.12 presenta la recta OLS correspondiente a dichos valores con un valor de pendiente 0.52 (respecto al 0.73 del caso en horas).

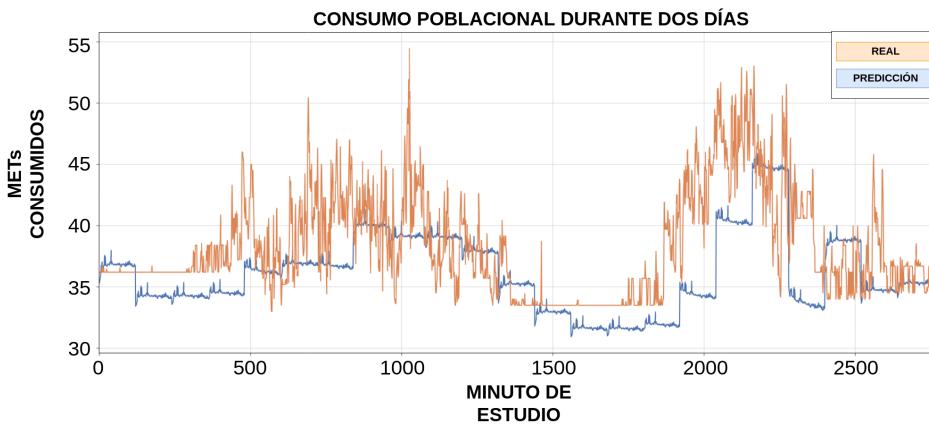


Figura 8.13: Predicción en minutos de dos días para el split 1 para el modelo Individual.

$$Y = 20.79613 + 0.4014X$$

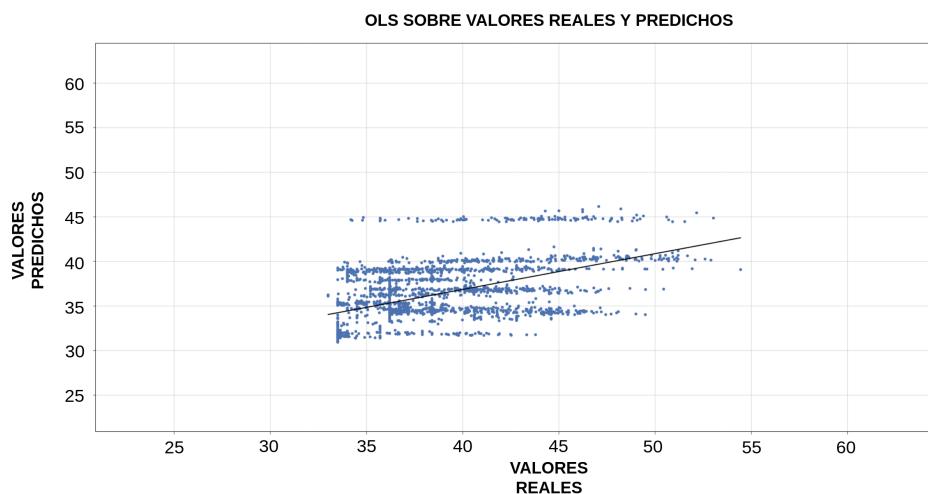


Figura 8.14: Recta de Mínimos Cuadrados Ordinarios entradas individuales.

La Figura 8.13 muestra el resultado de las predicciones hechas durante los días de test del split 1 con el horizonte en minutos por el framework Individual. La Figura 8.14 presenta la recta OLS correspondiente a dichos valores con un valor de pendiente 0.41 (respecto al 0.79 del caso en horas).

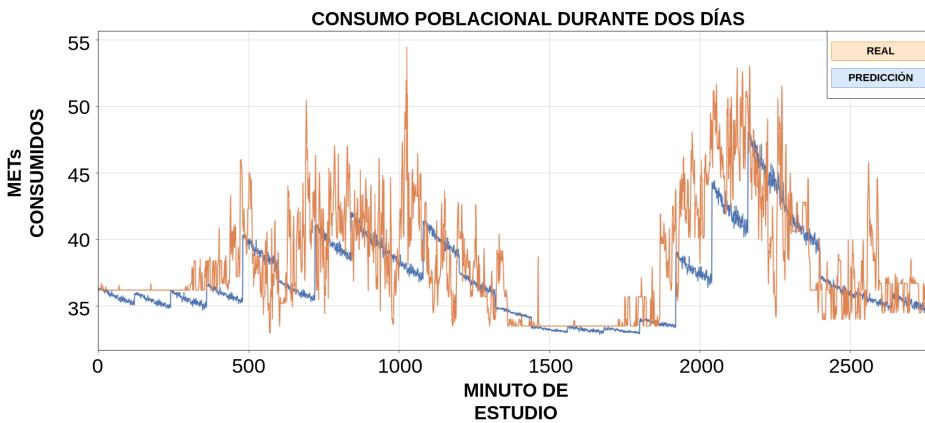


Figura 8.15: Predicción en minutos de dos días para el split 1 modelo Matricial con el horizonte en minutos.

$$Y = 17.52399 + 0.5121X$$

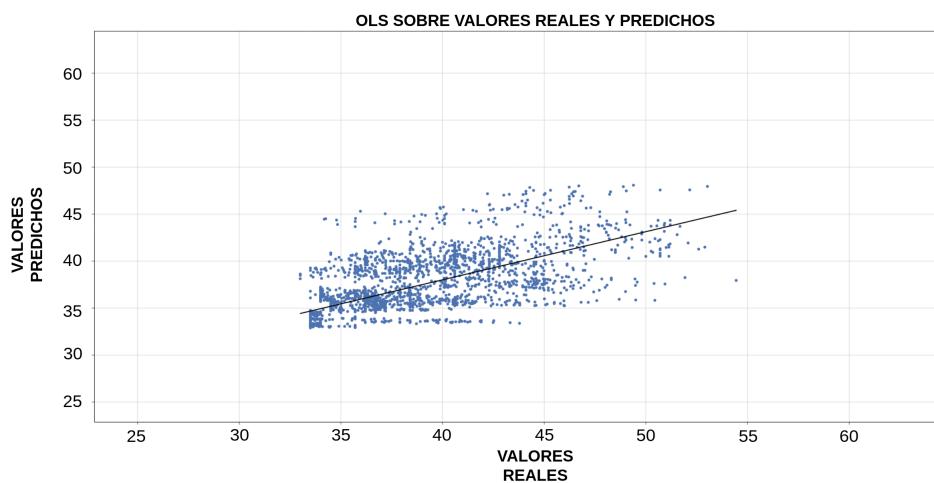


Figura 8.16: Recta de Mínimos Cuadrados Ordinarios para el modelo de entradas matriciales con el horizonte en minutos.

Finalmente, la Figura 8.15 muestra el resultado de las predicciones hechas durante los días de test del split 1 con el horizonte en minutos por el framework Matricial. La recta OLS correspondiente (Figura 8.16) obtiene valor de pendiente 0.51 (respecto al 0.79 del caso en horas).

Experimento a largo plazo en minutos con foco en valores únicos por intervalo.

Si bien las gráficas mostradas en la sección anterior muestran correctamente el funcionamiento del modelo a largo plazo, el objetivo final es la predicción de valores únicos para el umbral de tiempo predicho. En pro de que para el caso de las gráficas referentes a los minutos están pueden verse algo distorsionadas debido a la alta cantidad de datos representada, se repite el experimento centrando el enfoque esta vez en el objetivo de predicción final. Para ello tomamos los valores predichos para las Figuras 8.11, 8.13 y 8.15 y sumamos por cada conjunto de dos horas predicho los 120 valores en un único valor que represente el valor de METs para la población en ese periodo de dos horas. De esta forma obtenemos 23 valores correspondientes a las 23 predicciones realizadas para esas últimas 46 horas (ver figuras 8.17, 8.19 y 8.21).

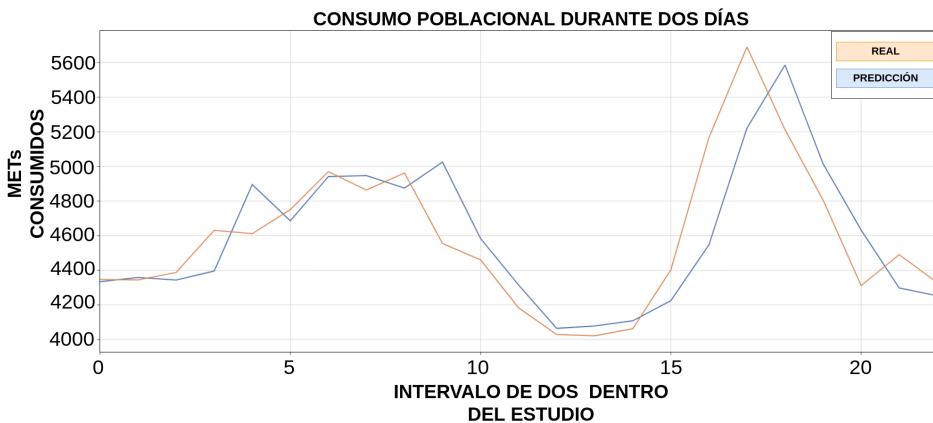


Figura 8.17: Predicción de un solo valor en cada intervalo predicho sobre el modelo de entradas agregadas para el horizonte en minutos.

$$Y = 969.5323 + 0.7901X$$

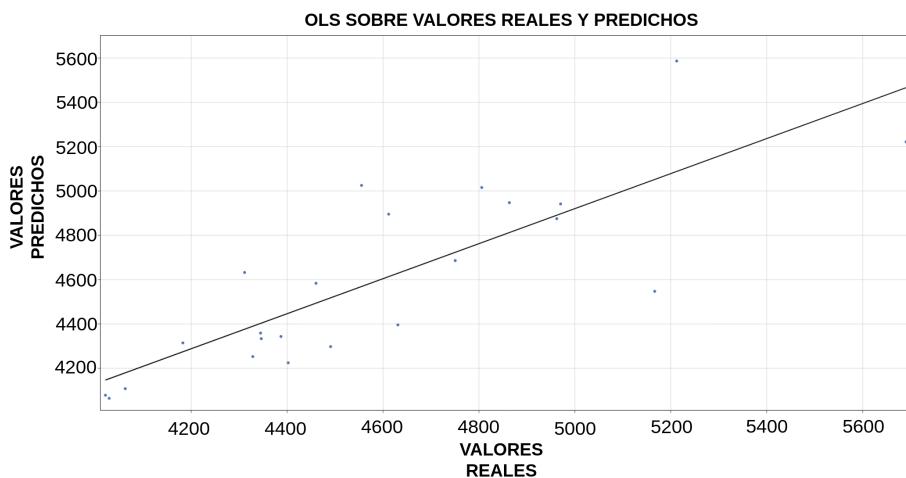


Figura 8.18: Recta de mínimos cuadrados para la predicción sumada de entradas agregadas.

La Figura 8.17 muestra como sería la predicción de la Figura 8.11 al sumar los minutos en cada horizonte predicho en un único valor. La recta OLS correspondiente (Figura 8.18) obtiene valor de pendiente 0.72 (respecto al 0.79 del caso en horas y 0.52 previa suma).

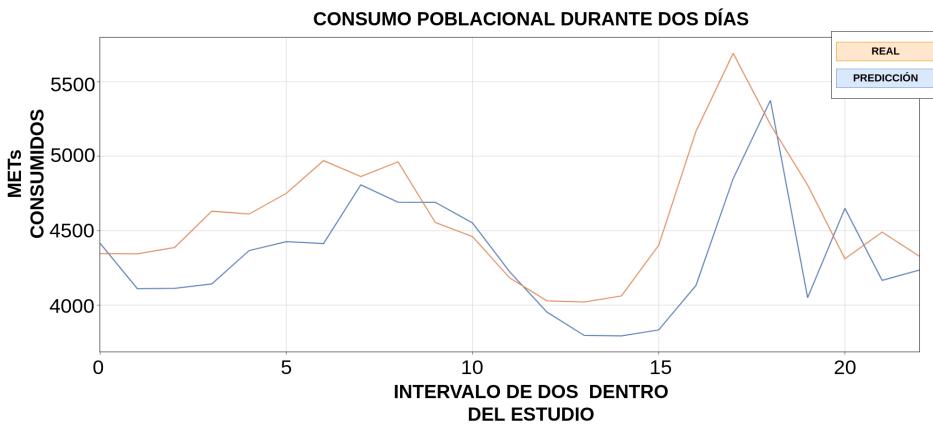


Figura 8.19: Predicción de un solo valor en cada intervalo predicho sobre el modelo de entradas individuales en minutos.

$$Y = 1602.4346 + 0.5960X$$

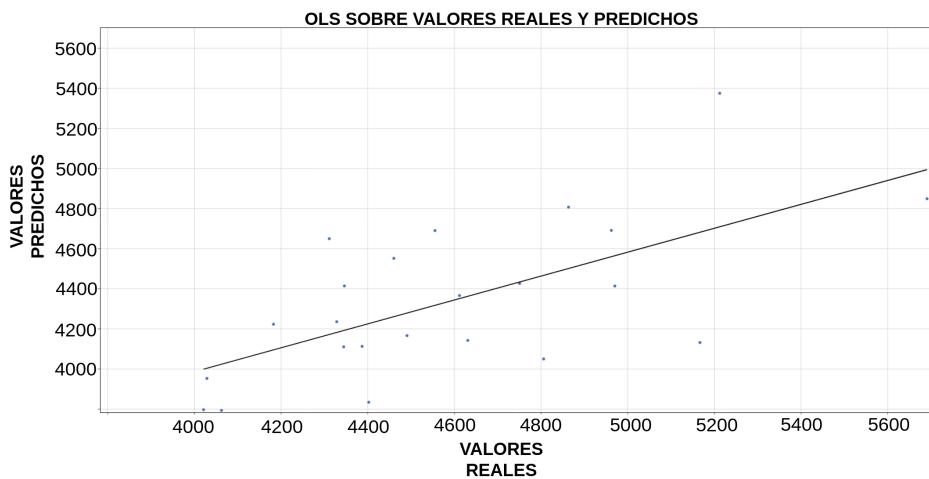


Figura 8.20: Recta de mínimos cuadrados para la predicción sumada de entradas individuales.

La Figura 8.19 muestra como sería la predicción al sumar los minutos en cada horizonte predicho en un único valor de la Figura 8.13. La recta OLS correspondiente (Figura 8.20) obtiene valor de pendiente 0.59 (respecto al 0.79 del caso en horas y 0.4 previa suma).

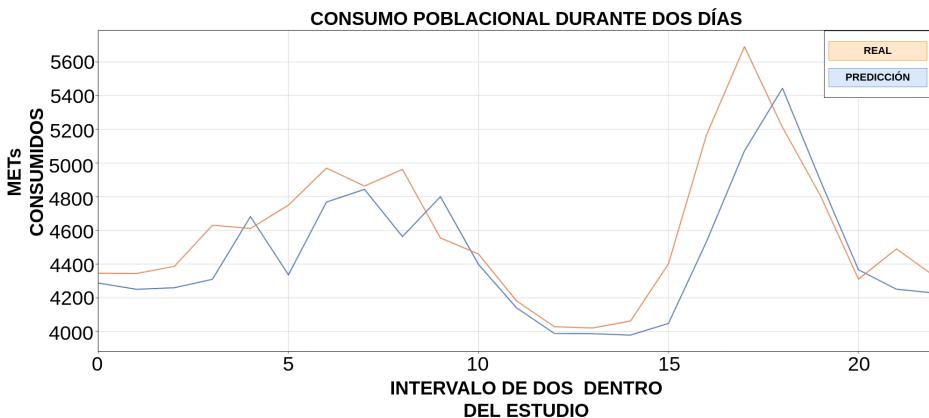


Figura 8.21: Predicción de un solo valor en cada intervalo predicho para el modelo de entradas matriciales en minutos.

$$Y = 972.5857 + 0.7583X$$

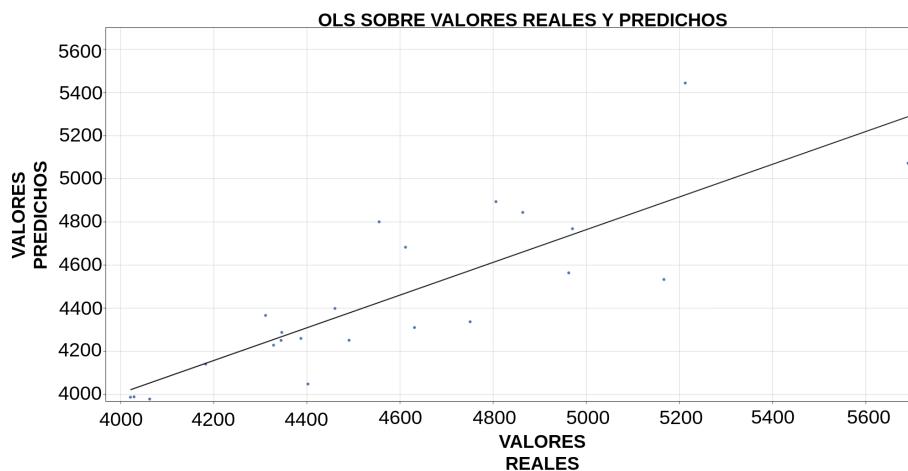


Figura 8.22: Recta de mínimos cuadrados para la predicción sumada de entradas matriciales.

Finalmente, la Figura 8.21 muestra las predicción al sumar los minutos en cada horizonte predicho en un único valor de la Figura 8.15. La recta OLS correspondiente (Figura 8.22) obtiene valor de pendiente 0.75 (respecto al 0.79 del caso en horas y 0.51 previa suma).

8.1.4. Peores y mejores resultados por split

En esta sección se exponen los mejores y peores resultados por split, ya que estos ayudarán a una interpretación más profunda del funcionamiento de los modelos.

Mejor MAE por split

La Tablas 8.9 y 8.10 muestran para los horizontes en horas y minutos cual fue el MAE menor encontrado para todo el split en referencia al total de METs de la población a predecir. En media los mejores resultados para el framework Agregado suponen tan solo 2.16 METs en el caso de las horas y 0.41 METs en el de los minutos. Para el Individual 1.99 y 0.26 METs para horas y minutos respectivamente. El caso Matricial recibe un valor más alto para las horas con 24.76 METs y 0.23 en el caso de los minutos, aunque de nuevo debido al grosso error encontrado en el split 3. Todos estos valores recordemos son respecto a los valores medios punto a predecir de 4627.93 METs (ver Tabla 8.1 o 8.2).

Tabla 8.9: Mejores resultados por split horizonte en horas.

Split	MAE Agregado	MAE Individual	MAE Matricial
1	0.88	0.58	1.94
2	3.38	0.609	1.46
3	0.71	0.50	214.52
4	1.96	5.81	0.53
5	0.64	1.80	1.23
6	5.57	0.70	0.51
7	0.88	2.29	1.35
8	4.12	1.95	0.79
9	1.27	3.69	0.53
Mean	2.16	1.99	24.76
Std	1.78	1.78	71.15

Tabla 8.10: Mejores resultados por split horizonte en minutos.

Split	MAE Agregado	MAE Individual	MAE Matricial
1	0.51	0.38	0.12
2	0.64	0.15	0.34
3	0.34	0.58	0.28
4	0.31	0.23	0.27
5	0.38	0.15	0.16
6	0.45	0.22	0.17
7	0.33	0.21	0.15
8	0.24	0.17	0.22
9	0.49	0.22	0.35
Mean	0.41	0.26	0.23
Std	0.12	0.13	0.08

A continuación se exponen las gráficas de valores predichos y reales con el mejor MAE del primer split (ver Figuras 8.23, 8.24 y 8.25). Los detalles de estas figuras son expuestos en la sección 8.2.

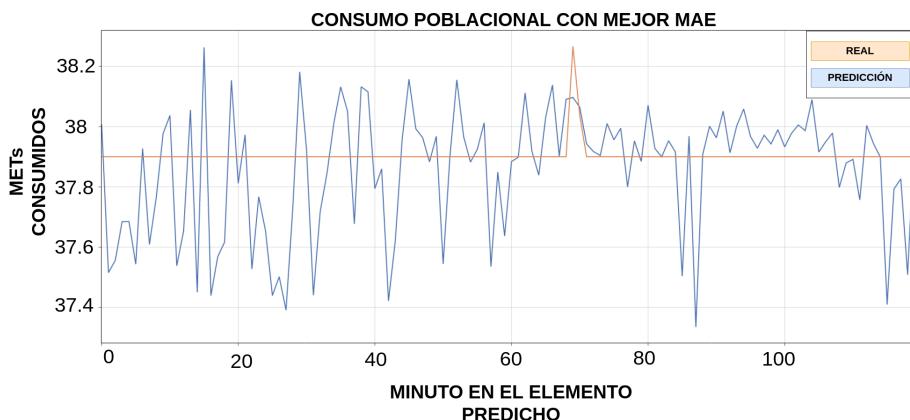


Figura 8.23: Mejor MAE para el modelo agregado en el split 1.

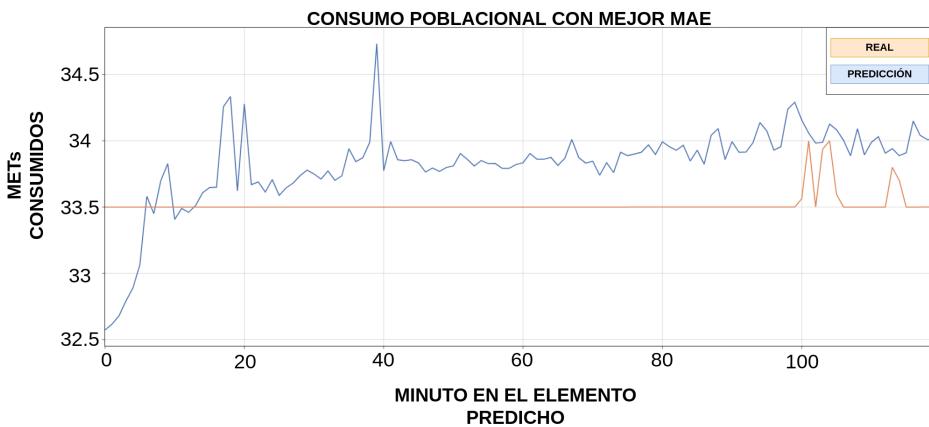


Figura 8.24: Mejor MAE para el modelo individual en el split 1.

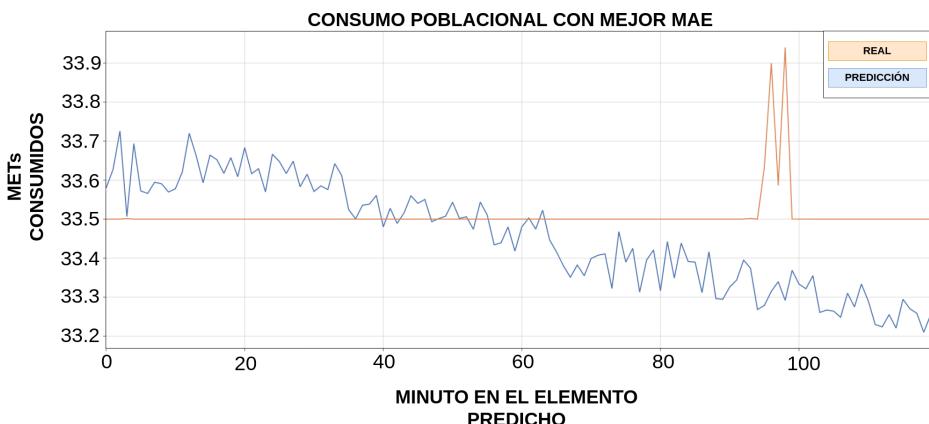


Figura 8.25: Mejor MAE para el modelo matricial en el split 1.

Peor MAE por split

La Tablas 8.11 y 8.12 muestran para los horizontes en horas y minutos cual fue el MAE mayor encontrado para todo el split en referencia al valor total de METs de la población, al igual que se hizo con los mejores resultados. En esta ocasión se pone de manifiesto como todos los modelos sufren carencias fuertes para predecir en algunos ejemplos. El peor error encontrado es para el split 4 del framework Matricial de 707.2 METs respecto de la media 4627.93 METs (error del 15.28 %).

Tabla 8.11: Peores resultados horizonte en horas.

Split	MAE Agregado	MAE Individual	MAE Matricial
1	369.11	523.66	527.05
2	279.95	438.96	395.26
3	357.31	433.066	609.36
4	355.49	469.84	408.10
5	512.82	514.43	707.20
6	468.34	618.07	629.91
7	486.59	492.34	563.24
8	357.38	475.59	683.68
9	321.06	437.35	493.12
Mean	389.789	489.259	557.44
Std	79.98	58.48	111.809

Tabla 8.12: Peores resultados horizonte en minutos.

Split	MAE Agregado	MAE Individual	MAE Matricial
1	0.51	0.38	0.12
2	0.64	0.15	0.34
3	0.34	0.58	0.28
4	0.31	0.23	0.27
5	0.38	0.158	0.16
6	0.45	0.22	0.179
7	0.33	0.21	0.15
8	0.249	0.17	0.22
9	0.49	0.22	0.35
Mean	6.85	10.15	8.11
Std	1.26	0.70	1.51

A continuación se exponen las gráficas de valores predichos y reales con el peor MAE del primer split (ver Figuras 8.26, 8.27 y 8.26). Estas ayudarán a posteriori a una mejor comprensión de los resultados cuya discusión se realiza en la sección 8.2.

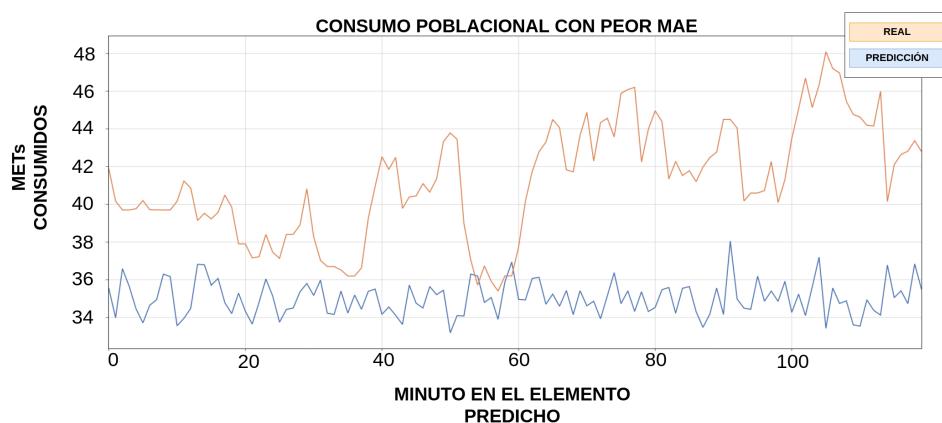


Figura 8.26: Peor MAE para el modelo agregado en el split 1.

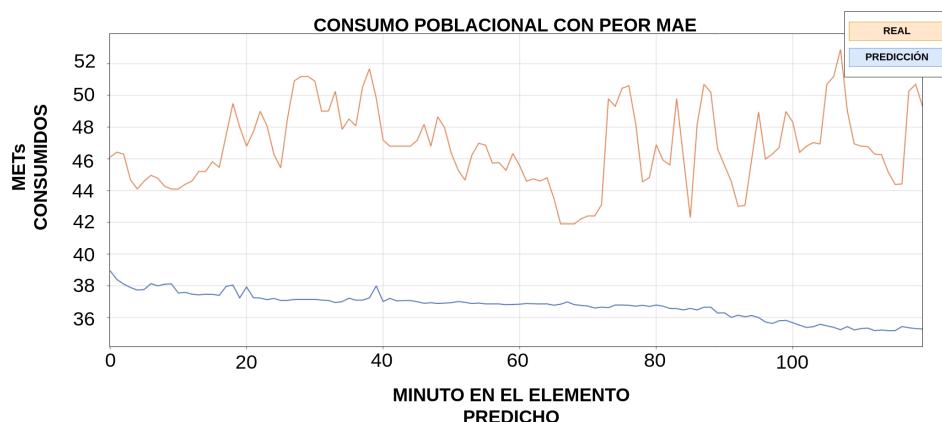


Figura 8.27: Peor MAE para el modelo individual en el split 1.

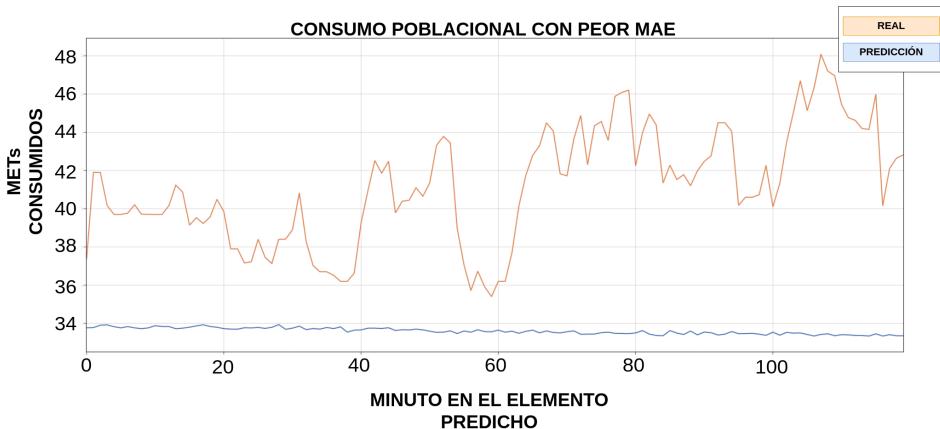


Figura 8.28: Peor MAE para el modelo matricial en el split 1.

8.2. Discusión

8.2.1. Diferencias entre los modelos

Según los datos aportados el mejor modelo de cara a obtener predicciones es el framework Agregado (Alternativa 3 sección 7.1.3). Este modelo es sencillo ya que sus entradas y salidas no son sólo similares sino que ambas tienen un formato parecido a aquello que se trata de predecir. Además estas entradas cuentan con la ventaja de estar incorporando información a nivel poblacional. Aunque el framework individual (Alternativa 2 sección 7.1.1) ponga el foco participante a participante, realmente las series temporales a nivel individual son planas, ya que existe poca variación en los datos y esta se realiza en un rango de valores pequeño. Por tanto estas aportan poca información relevante a nivel poblacional. Al vivir en sociedad el comportamiento que tienen ciertos individuos suele replicarse en otros, ya sea debido a costumbres de está o factores externos. Por ejemplo, conforme empieza a acercarse la noche. Los individuos comienzan a ir a descansar, aunque algunos puedan realizar una actividad al final del día, este momento sigue siendo indicativo de que más y más individuos irán comenzando a acostarse con el paso del tiempo incluidos estos que en este momento se encuentren practicando ejercicio físico. Otros factores como lluvia o mal tiempo repentinos también pueden ser significativos a nivel poblacional y provocar un efecto colectivo en la población.

Si se comparan las entradas de las series temporales de estos dos primeros framework (Agregado e Individual), se recalca cómo las entradas del primero (Agregado) similares a las de la Figura 8.29 reciben cambios mucho más abruptos que en el segundo de los casos representados 8.30.

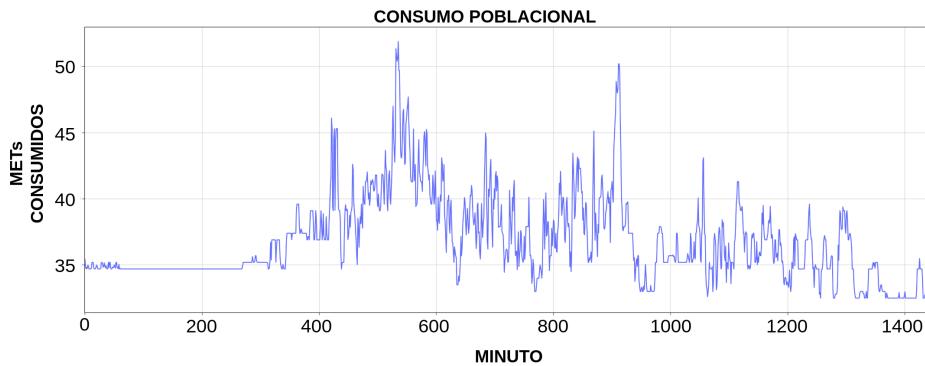


Figura 8.29: Serie temporal agregada a nivel poblacional durante el día 2.

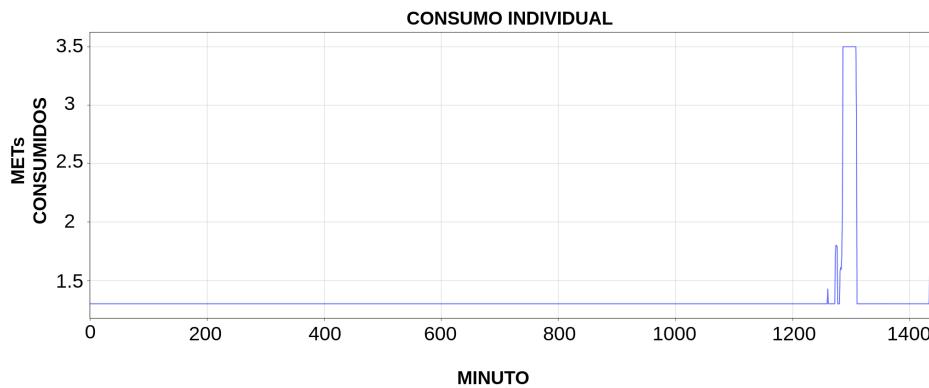


Figura 8.30: Serie temporal correspondiente al participante 0 durante el día 2.

Por otro lado el framework Agregado es un modelo más simple que el Matricial (Alternativa 2 sección 7.1.2), el cual en principio debería ser el que incorporará tanto información a nivel individual como poblacional. Probablemente, para lograr capturar toda la información incorporada en el modelo con entradas matriciales sería necesario modificar el diseño del mismo, de tal forma que el modelo contará con un número mayor de parámetros, así como la cantidad de datos con la que se entrena este. Esto se debe a que el modelo para sacar partido a estas matrices debería ser capaz de detectar las pequeñas variaciones que se van produciendo entre las 25 filas. Esta tarea se dificulta al tener un total de 36000 valores flotantes en su entrada (cada una de las series temporales de 1440 valores de los 25 participantes), respecto de los 1440 que reciben el resto de los modelos.

Un detalle interesante sobre el modelo de entrada individuales es que

según los valores calculados en la Tabla 8.9 con el mejor MAE por split en el horizonte en horas y en la Tabla 8.10 con el mejor MAE en minutos, es que podemos ver como en media el modelo individual es capaz de conseguir el mejor resultado aislado en los splits. Esto quiere decir que en promedio para la mejor de las actuaciones realizadas por los frameworks en cada uno de los conjuntos, el framework Individual suele conseguir la mejor de ellas. De la misma manera, en la tabla de peores resultados para las horas (Tabla 8.11) consigue en promedio el segundo peor valor medio después del conjunto matricial y para el caso de los minutos (Tabla 8.12) el peor valor en promedio. Esto se debe al comportamiento en bloque del mismo unido a predicciones de poca tendencia que hace que prediga muy bien momentos de nula actividad como los nocturnos y pueda sufrir ante cambios muy bruscos en las predicciones. Sobre este efecto trata el siguiente apartado de la discusión.

8.2.2. Comportamiento en bloque

Los modelos claramente sufren en captar tendencias a largo plazo, es decir, cuando los horizontes son mayores. Es por ello, que si nos fijamos en las mejores predicciones realizadas por los modelos para el horizonte en minutos de los diferentes splits, estas predicciones coinciden con momentos donde las series temporales reales son más homogéneas. El comportamiento del modelo parece basarse en buscar en función de la entrada un valor medio que sea representativo de la serie temporal durante los instantes predichos y a partir de este ir realizando pequeñas variaciones e introduciendo ligeras tendencias. Es observable que en las Figuras 8.23, 8.24 y 8.25 correspondientes a las mejores predicciones para el primer split, las series temporales a predecir pertenecen a períodos de tiempo donde la actividad física presenta poca variabilidad. Probablemente perteneciendo estos períodos a las horas de sueño ya que además coinciden con períodos de baja actividad. Las predicciones realizadas varían en un rango pequeño de valores, y por tanto casan bien con los valores reales que como mucho reciben algún pico de actividad, probablemente ligado a que algún paciente se levante momentáneamente. Las predicciones para el peor MAE encontrado en el primer split quedan representadas en las Figuras 8.26, 8.27 y 8.28. Comparando con las mejores anteriormente encontradas, podemos observar que en estos casos se producen cambios bruscos en los valores que se tratan de predecir. Además de ello son valores de actividad física altos, siendo el modelo de entradas agregadas el único que parece ser capaz de captar cierta similitud en las subidas y bajadas de los valores a predecir.

Este fenómeno es observable al tomar el experimento de predicción de dos días explicado en la sección 8.1.3. Concretamente tomando el horizonte en minutos, obtenemos una buena representación de este fenómeno para los diferentes modelos. En el caso de la predicción para el modelo de entradas

agregadas de la Figura 8.11, aunque en la predicción sea apreciable cambios bruscos al terminar un periodo de tiempo de dos horas, estas tienen cierto dinamismo dentro de estos bloques. Por su parte, el modelo de entradas individuales representado en la Figura 8.13 claramente experimenta un comportamiento más monótono, con líneas prácticamente rectas en cada periodo de dos horas que son modificadas bruscamente al terminar estos. De la misma forma el modelo de entradas matriciales 8.15 se ve caracterizado de una manera similar, aunque en lugar de líneas rectas durante los periodos de dos horas sus salidas proponen tendencias descendentes en las franjas de tiempo predichas.

Para entender mejor la calidad de los resultados podemos estudiar el ajuste realizado para cada uno de las rectas de ajuste de mínimos cuadrados en las figuras 8.12, 8.14 y 8.16 de la sección 8.1.3. En ninguno de los tres casos parece presentar relaciones fuertes entre los valores predichos y los valores reales. Especialmente en la Figura 8.14 es observable la presencia de rectas, que indican que un mismo valor fue predicho en varias ocasiones para un umbral amplio de valores reales. Estos se debe al efecto bloque comentado, donde por ejemplo el modelo predice para una ristra de dos horas un solo valor representativo de ese intervalo que varía poco para el conjunto de los 120 valores predichos. A pesar de que las rectas de mínimos cuadrados ordinarios presentados aparentemente no muestran valores de relación altos entre los valores predichos y reales, las predicciones no son malas de cara al objetivo de la predicción de valores de actividad física para la población en su conjunto. Las gráficas correspondientes a los resultados de estos experimentos son Figura 8.17, Figura 8.19 y Figura 8.21 así como sus correspondientes gráficas con ajuste OLS son Figura 8.18, Figura 8.20 y Figura 8.22. Las figuras expuestas con intervalos de tiempo sumados dan una lectura diferente de los resultados que es más acorde con el objetivo del proyecto. Las correlaciones en la recta OLS son mayores y es apreciable que los modelos son bastantes capaces de detectar las tendencias de la población respecto a la actividad física. Es por ello posible concluir que la dificultad o carencia que existe en los modelos al realizar las predicciones en minutos es captar el comportamiento específico de la población a lo largo de los 120 puntos que se tratan de predecir, y tiene por ello sentido apuntar hacia horizontes de predicción con un número de valores menor a pesar de que representan intervalos de tiempo mayores.

En el caso del horizonte en horas, el experimento a largo plazo queda representado en las figuras 8.5, 8.7 y 8.9 para los modelos agregado, individual y matricial respectivamente. En este podemos observar como la tendencia captada por los modelos si es bastante representativa de la realidad, aunque también podemos atribuir a los valores predichos un cierto *delay* sobre los reales. Esto invita a pensar que existe una dependencia alta a los últimos valores de entrada recibidos, aunque esto no supondría necesariamente un

punto negativo ya que si pensamos en la actividad física de una población es habitual que se vea altamente condicionada por los valores de actividad justamente más continuos. El ajuste OLS realizado para estos en las gráficas 8.6, 8.8 y 8.10 representa en todas ellas rectas con valores de Y entorno a 0,7 estableciendo un valor alto de correspondencia.

Capítulo 9

Conclusiones y trabajos futuros

9.1. Conclusiones

La predicción de actividad física es un problema todavía en desarrollo, el cuál ha sido especialmente tratado durante los últimos años debido al surgimiento de teléfonos y otros dispositivos como relojes o pulseras inteligentes que permiten la recopilación de resultados. Este trabajo aborda el problema mediante el enfoque del Deep Learning y utiliza como medida representativa de la actividad física los METs. Durante el estudio se han comparado distintas formas de enfocar el problema y se ha comprobado la viabilidad de la predicción de valores de actividad física a nivel poblacional. También han sido comparados diferentes horizontes de predicción y evaluadas la precisión y viabilidad con la que los modelos pueden predecir para rangos de valores más y menos exigentes en el tiempo.

La predicción de valores de actividad física y otros problemas es posible que mejore con el tiempo debido a nuevos modelos y tecnologías con mayor capacidad de computación. Aún así es necesario entender la complejidad que abarca la realidad humana y la existencia de una gran cantidad de factores externos que pueden alterar el comportamiento de la misma. Por ende los resultados deben ser enfocados desde un punto de vista amplio, entendiendo que de momento los programas que pueden ser desarrollados a raíz de frameworks de trabajo como los propuestos funcionan como una guía y no como una verdad absoluta. Desde este punto de vista los resultados son positivos a la predicción de datos de actividad física. Los resultados del estudio plasman la posibilidad de predecir datos de hasta dos horas a posteriori de actividad, siendo realizadas estas predicciones en el orden de segundos. Se ha visto que las predicciones realizadas son capaces realmente de captar tendencias

poblacionales y como los modelos posterior al entrenamiento son capaces de analizar los comportamientos de las series de entrada para realizar predicciones del comportamiento poblacional. También se ha puesto de manifiesto la dificultad de los mismos para captar picos o movimientos muy específicos dentro de las tendencias generales de la población. El problema es la dificultad que supone ser capaz de predecir que en un periodo del día en concreto que una gran parte de la población va a hacer mucho o poco ejercicio. Por ejemplo, un evento televisivo que acaparase gran atención produciría un descenso claro del consumo de METs durante su duración. También provocaría que aquellos individuos más comprometidos con el ejercicio cambiarán su horario habitual obteniendo picos de actividad. Ambas picos, tanto de subida como de bajada, no quedarían en ningún caso reflejados por los valores de entrada del día anterior donde no existía dicho evento y por tanto los modelos no son capaces de captarlos. Por ello, la bondad de los resultados y por lo que se considera que el modelo realmente capta tendencias es por el resto de las facultades presentadas. Lo interesante es que este es capaz de detectar que el consumo va a ser bajo debido a que los últimos valores entrantes eran bajos ya que los individuos estaban durmiendo, pero también es capaz de detectar que van a comenzar a aumentar debido a que han sido bajos durante un tiempo razonable. Los valores bajos de MAE encontrados así como el bajo porcentaje de error que estos suponen respecto de la media invitan a pensar que realmente estos frameworks captan comportamientos poblacionales.

Finalmente mediante experimentos que ponen en perspectiva comportamientos más generales, graficandolos y observando desde un punto de vista más general los resultados, se observó como esto era cierto. Y es que los resultados plasmados en las figuras relacionadas con los experimentos más en esta línea dan sin duda una perspectiva muy buena sobre la capacidad de predicción que estos frameworks son capaces de alcanzar. Por otro lado, una de las mayores dificultades que se han puesto en manifiesto es la capacidad de los modelos para predecir un gran número de puntos hacia delante. Las series temporales parecen depender en gran medida de los últimos valores de entrada y las predicciones pierden mucha rigurosidad al intentar captar tendencias cuando las salidas son más largas. Este efecto es menos notable al realizar la predicción en horas. Por tanto podemos concluir que no es tanto la medida utilizada o el umbral predicho si no el tamaño de las salidas lo que puede ser más influyente en la red, probablemente requiriendo el caso de las salidas de 120 un entrenamiento con un número mucho mayor de ejemplos.

9.2. Trabajos Futuros

Este trabajo utiliza series temporales de un día completo de datos para realizar predicciones de dos horas hacia delante a partir de las entradas. A partir de ello, trabajos siguientes podrán estudiar la explicabilidad de los modelos de cara a entender cuales son las partes más relevantes en las series temporales de actividad física para poder realizar acotaciones de las mismas. Por otro lado, los resultados sobre los horizontes de predicción indican la posibilidad de realizar predicciones sobre umbrales de tiempo mayores ya que lo importante no es tanto la similitud entre los valores de entrada y de salida si no el número de puntos a predecir en el horizonte de predicción.

El multi-step forecasting es una posibilidad que se plantea prometedora a raíz de los resultados del horizonte en minutos. El modelo sufre en precisión al predecir una gran cantidad de puntos pero logra en media captar cómo van a ser los valores que debe predecir. Por tanto, un valor que prediga un número menor de puntos pero pueda ser retroalimentado para la obtención de horizontes mayores de predicción se presenta como una opción viable.

Bibliografía

- [1] I.N.E. Estimación del número de defunciones semanales. [Online]. Available: <https://www.ine.es/uc/W7RXPSZu>
- [2] ——. Encuesta de población activa (epa) segundo trimestre de 2021. [Online]. Available: <https://www.ine.es/daco/daco42/daco4211/epa0221.pdf>
- [3] P. Chujai, N. Kerdprasop, and K. Kerdprasop, “Time series analysis of household electric consumption with arima and arma models,” in *Proceedings of the international multiconference of engineers and computer scientists*, vol. 1. IAENG Hong Kong, 2013, pp. 295–300.
- [4] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, “Financial time series forecasting with deep learning: A systematic literature review: 2005–2019,” *Applied soft computing*, vol. 90, p. 106181, 2020.
- [5] E. Çiçek, S. Gören, and G. Memik, “Physical activity forecasting with time series data using android smartphone,” *Pervasive and Mobile Computing*, vol. 82, p. 101567, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1574119222000207>
- [6] A. Mamun, K. S. Leonard, M. P. Buman, and H. Ghasemzadeh, “Multi-modal time-series activity forecasting for adaptive lifestyle intervention design,” in *2022 IEEE-EMBS International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2022, pp. 1–4.
- [7] N. Arbel. How long short term memory networks solve the problem of vanishing gradients. [Online]. Available: <https://medium.datadriveninvestor.com/how-do-lstm-networks-solve-the-problem-of-vanishing-gradients-a6784971a577>
- [8] (2013) Desarrollo en cascada. [Online]. Available: https://es.wikipedia.org/wiki/Desarrollo_en_cascada
- [9] R. Vogl, “Deep learning methods for drum transcription and drum pattern generation,” Ph.D. dissertation, 11 2018.

- [10] Wikipedia. (2023) Pandemia de covid-19 en espana — wikipedia, la enciclopedia libre. [Online]. Available: https://es.wikipedia.org/wiki/Pandemia_de_COVID-19_en_Espa%C3%B1a
- [11] Postcovid-ai. [Online]. Available: <https://projects.ugr.es/postcovid-ai/es/>
- [12] O. M. de la Salud. (2022) Actividad física. [Online]. Available: <https://www.who.int/es/news-room/fact-sheets/detail/physical-activity>
- [13] M. A. Sanchez-Lastra, A. López-Valenciano, D. Suárez-Iglesias, and C. Ayán. Estimación del aumento global del tiempo de sedentarismo durante los confinamientos de la covid-19: una revisión sistemática y un metanálisis. [Online]. Available: https://www.sanidad.gob.es/biblioPublic/publicaciones/recursos_propios/resp/revista_cdrom/VOL96/REVISIONES/RS96C_202205042.pdf
- [14] G. U. Yule, “Vii. on a method of investigating periodicities disturbed series, with special reference to wolfer’s sunspot numbers,” *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 226, no. 636-646, pp. 267–298, 1927.
- [15] J. G. De Gooijer and R. J. Hyndman, “25 years of time series forecasting,” *International journal of forecasting*, vol. 22, no. 3, pp. 443–473, 2006.
- [16] G. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.
- [17] R. S. Tsay, “Time series and forecasting: Brief history and future research,” *Journal of the American Statistical Association*, vol. 95, no. 450, pp. 638–643, 2000. [Online]. Available: <http://www.jstor.org/stable/2669408>
- [18] B. Lim and S. Zohren, “Time-series forecasting with deep learning: a survey,” *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, p. 20200209, 2021.
- [19] G. Jujjuru. (2023) Learning time series analysis modern statistical models. [Online]. Available: <https://www.analyticsvidhya.com/blog/2023/01/learning-time-series-analysis-modern-statistical-models/>
- [20] R. Abdel-Aal and A. Z. Al-Garni, “Forecasting monthly electric energy consumption in eastern saudi arabia using univariate time-series analysis,” *Energy*, vol. 22, no. 11, pp. 1059–1069, 1997.

- [21] A. Tealab, “Time series forecasting using artificial neural networks methodologies: A systematic review,” *Future Computing and Informatics Journal*, vol. 3, pp. 334–340, 12 2018.
- [22] R. J. Hyndman, “A brief history of forecasting competitions,” *International Journal of Forecasting*, vol. 36, no. 1, pp. 7–14, 2020.
- [23] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “The m4 competition: Results, findings, conclusion and way forward,” *International Journal of Forecasting*, vol. 34, no. 4, pp. 802–808, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169207018300785>
- [24] P. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Harvard University, 1975. [Online]. Available: <https://books.google.es/books?id=z81XmgEACAAJ>
- [25] P. J. Werbos, “Generalization of backpropagation with application to a recurrent gas market model,” *Neural Networks*, vol. 1, no. 4, pp. 339–356, 1988. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/089360808890007X>
- [26] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks:: The state of the art,” *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169207097000447>
- [27] J. Lago, F. De Ridder, and B. De Schutter, “Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms,” *Applied Energy*, vol. 221, pp. 386–405, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S030626191830196X>
- [28] G. Ozogur, M. A. Erturk, and M. A. Aydin, “Prediction of physical activity times using deep learning method,” in *International Telecommunications Conference*, A. Boyaci, A. R. Ekti, M. A. Aydin, and S. Yarkan, Eds. Singapore: Springer Singapore, 2019, pp. 299–307.
- [29] J. Nabi. Recurrent neural networks (rnns). [Online]. Available: <https://towardsdatascience.com/recurrent-neural-networks-rnns-3f06d7653a85>
- [30] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 11 1997. [Online]. Available: <https://doi.org/10.1162/neco.1997.9.8.1735>
- [31] Metodología waterfall: Qué es y cuándo usarla. [Online]. Available: <https://www.iebschool.com/blog/metodologia-waterfall-que-es-y-cuando-usarla-management/>

- [32] Google. Activityrecognitionapi. [Online]. Available: <https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognitionApi#public-abstract-pendingresultstatus-requestactivityupdates-googleapiclient-client,-long-detectionintervalmillis,-pendingintent-callbackintent>
- [33] M. Terrera. (2011) Compendium of physical activities. [Online]. Available: <https://sites.google.com/site/compendiumofphysicalactivities/home?authuser=1>
- [34] S. Selvin, R. Vinayakumar, E. Gopalakrishnan, V. K. Menon, and K. Soman, “Stock price prediction using lstm, rnn and cnn-sliding window model,” in *2017 international conference on advances in computing, communications and informatics (icacci)*. IEEE, 2017, pp. 1643–1647.
- [35] Keras. Keras: Simple,flexible,powerful. [Online]. Available: <https://keras.io/>
- [36] TensorFow. Tensorflow. [Online]. Available: <https://www.tensorflow.org/>
- [37] (2022) Overfitting in machine learning: What it is and how to prevent it. [Online]. Available: <https://elitedatascience.com/overfitting-in-machine-learning#how-to-prevent>

Resultados completos

Tabla 1: **Valores medios a predecir horizonte en horas.**

split	POINT_TO_POINT	ALL METS
0	2294.15625	4588.3125
1	2318.01660	4636.0332
2	2342.07373	4684.1479
3	2446.64746	4893.2949
4	2190.43921	4380.8784
5	2293.51660	4587.0327
6	2224.74756	4449.4951
7	2310.45386	4620.9082
8	2405.63330	4811.2671
Mean	2313.96495	4627.9300
Std	79.73037	159.4608

Tabla 2: **Valores medios a predecir horizonte en minutos.**

split	POINT_TO_POINT	ALL METS
0	38.23594	4588.31299
1	38.63361	4636.03320
2	39.03456	4684.14795
3	40.77745	4893.29492
4	36.50732	4380.87842
5	38.22528	4587.03320
6	37.07913	4449.49561
7	38.50757	4620.90820
8	40.09389	4811.26709
Mean	38.56608	4627.93018
Std	1.32884	159.46074

Tabla 3: Resultados punto a punto para el horizonte en horas.

split	MAE Agregado	MSE Agregado		
0	108.4600	20878.4961		
1	112.0058	18370.3594		
2	93.3936	16894.1992		
3	100.5769	18140.5859		
4	106.3322	25480.0664		
5	99.2600	19694.8398		
6	91.8067	18794.6895		
7	105.5283	17588.0215		
8	112.1542	18843.7344		
Mean	103.2797	19409.4436		
Std	7.4912	2552.5405 height		

split	MAE Individual	MSE Individual	MAE Matricial	MSE Matricial
0	132.9869	36027.0877	140.1739	36622.0391
1	140.1510	34286.4599	115.0322	23038.3320
2	115.8194	26874.1964	425.8986	238980.5625
3	150.3391	38340.6820	136.1953	29440.2930
4	126.7445	35686.4645	205.1356	78188.8984
5	142.3545	39524.3899	129.3496	33927.8906
6	106.5355	23082.0469	133.8848	33209.5547
7	127.0323	30025.0547	191.9583	64909.3242
8	122.1456	29778.8517	174.1503	49517.6016
Mean	129.3454	32625.0259	183.5310	65314.9440
Std	13.7258	5511.1000	95.9041	67535.8190

Tabla 4: **Resultados punto a punto para el horizonte en minutos.**

split	MAE Agregado	MSE Agregado
0	2.42831	10.84345
1	2.22463	8.30890
2	2.25486	10.25682
3	2.48959	11.30082
4	2.58837	11.81295
5	2.70630	13.77714
6	2.06076	8.65379
7	2.44680	10.55992
8	2.34474	9.49727
Mean	2.39382	10.55678
Std	0.19683	1.67598 height

split	MAE Individual	MSE Individual	MAE Matricial	MSE Matricial
0	2.77177	16.87611	2.36596	11.59576
1	3.01890	18.89226	2.02765	7.83467
2	4.00287	28.28013	2.83355	16.36296
3	4.59232	34.40728	2.84321	14.55958
4	2.95847	19.02886	2.38177	13.16939
5	2.95314	18.48764	2.80137	16.70181
6	2.45937	14.55684	2.08656	9.02671
7	2.71243	15.18324	2.78753	15.07467
8	3.41268	21.84806	2.68957	14.47241
Mean	3.20911	20.84005	2.53524	13.19977
Std	0.68459	6.52150	0.32659	3.12303

Tabla 5: **Resultados totales para el horizonte en horas.**

split	MAE Agregado	MSE Agregado
0	198.18858	69329.25781
1	214.05779	65507.47656
2	167.62901	53612.00000
3	175.08315	53003.94141
4	192.95720	83660.04688
5	180.87407	63192.08203
6	167.66629	64692.02344
7	192.35919	57334.28516
8	209.46983	63932.05859
Mean	188.69835	63807.01910
Std	17.06245	9295.48002

split	MAE Individual	MSE Individual	MAE Matricial	MSE Matricial
0	2.77177	16.87611	2.36596	11.59576
1	3.01890	18.89226	2.02765	7.83467
2	4.00287	28.28013	2.83355	16.36296
3	4.59232	34.40728	2.84321	14.55958
4	2.95847	19.02886	2.38177	13.16939
5	2.95314	18.48764	2.80137	16.70181
6	2.45937	14.55684	2.08656	9.02671
7	2.71243	15.18324	2.78753	15.07467
8	3.41268	21.84806	2.68957	14.47241
Mean	248.55073	114370.18947	307.83005	164328.66146
Std	28.27581	20540.39036	74.19322	71708.41684

Tabla 6: **Resultados totales para el horizonte en minutos.**

split	MAE Agregado	MSE Agregado
0	187.9326	63554.3984
1	162.6531	42341.9922
2	179.4447	61564.3984
3	206.7332	68484.5234
4	229.3883	87892.1484
5	230.4709	94236.1875
6	145.4077	48998.6680
7	202.2085	70641.5547
8	171.9048	47541.9727
Mean	190.6826	65028.4271
Std	29.1411	17716.0615

split	MAE Individual	MSE Individual	MAE Matricial	MSE Matricial
0	287.2740	162446.6482	219.7171	86102.1875
1	345.3096	204977.0015	172.1610	52177.8320
2	473.6166	315291.0526	299.0818	155386.7188
3	532.9575	398736.8692	279.0779	116141.2734
4	337.3240	198179.6239	234.3429	113060.0078
5	309.0123	172476.6788	272.4054	145774.1250
6	261.0446	140888.4595	186.1871	62293.1016
7	281.1599	137073.3656	289.5085	140330.7813
8	397.9997	231076.6749	257.6194	120559.6328
Mean	358.4109	217905.1527	245.5668	110202.8511
Std	92.7840	86975.9928	45.4035	36414.9174

Tabla 7: **Comportamiento a largo plazo para el horizonte en horas.**

split	Agregado	Individual	Matricial	Real
0	106281.3594	103788.3984	100404.3672	105584.734375
1	102983.0859	104059.0781	102442.8828	106464.71875
2	106478.1719	107617.4922	111657.5859	107659.5546875
3	110947.3906	113532.5156	109555.9063	112477.5234375
4	99422.1328	100424.2969	91594.9297	100688.3671875
5	103930.6016	104400.1406	100471.3828	105515.6328125
6	101982.2109	102704.4453	96886.8203	102448.46875
7	105777.8047	106130.2500	98708.5469	106605.0703125
8	108054.2344	108999.6484	106346.1172	110424.2734375
Mean	105095.2214	105739.5851	102007.6155	106429.815972222
Std	3436.5910	3887.9082	6320.3196	3613.16860992526

Tabla 8: Comportamiento a largo plazo para el horizonte en minutos.

split	Agregado	Individual	Matricial	Real
0	106281.3594	103788.3984	100404.3672	105584.734375
1	102983.0859	104059.0781	102442.8828	106464.71875
2	106478.1719	107617.4922	111657.5859	107659.5546875
3	110947.3906	113532.5156	109555.9063	112477.5234375
4	99422.1328	100424.2969	91594.9297	100688.3671875
5	103930.6016	104400.1406	100471.3828	105515.6328125
6	101982.2109	102704.4453	96886.8203	102448.46875
7	105777.8047	106130.2500	98708.5469	106605.0703125
8	108054.2344	108999.6484	106346.1172	110424.2734375
Mean	105095.2214	105739.5851	102007.6155	106429.815972222
Std	3436.5910	3887.9082	6320.3196	3613.16860992526

Repositorio del proyecto



GitHub

Puedes encontrar el código fuente en mi repositorio GitHub: <https://github.com/sanntana21/TFG>.