T

E

S

I

S

Alumno: Santillán Cooper, Martín

Director: Armentano, Marcelo

Co-Directora: Tomassel, Antonela

# 

# 

# Introducción

La definición de comportamiento sedentario ha ido evolucionando a lo largo de los años, al mismo tiempo que lo hizo la forma de medirlo. La Red de investigación del comportamiento sedentario[[1]](#footnote-0) define el comportamiento sedentario como cualquier comportamiento de vigilia caracterizado por un gasto de energía ≤1.5 MET (equivalente metabólico de tareas) mientras se está sentado o reclinado (Tremblay 2012). MET mide la intensidad de una actividad en múltiplos del gasto energético en reposo. Ejemplos de actividades sedentarias son mirar televisión (1.0 MET), comer sentado (1.5 MET), jugar videojuegos (1.0 MET) y conducir (1.3 MET).

La investigación realizada en el área demuestra asociaciones sólidas y consistentes entre el tiempo sedentario y la diabetes, las enfermedades cardiovasculares y la mortalidad por todas las causas (Wilmot et al. 2012) (Carter et al. 2017). Sin embargo, las asociaciones informadas fueron en gran medida independientes de la actividad física. Por lo tanto, es importante tener en cuenta que el comportamiento sedentario no representa lo contrario de la actividad física y que es posible que un individuo tenga niveles altos de actividad física moderada a vigorosa (MVPA, por sus siglas en inglés) y comportamiento sedentario. En general, se ha demostrado que el tiempo sedentario está asociado de forma perjudicial con la salud y con marcadores de riesgo metabólico en diversos grupos de población. Además, se ha destacado la importancia de no solo estimular la MVPA sino también de reducir el tiempo de sedentarismo, ya que la conducta sedentaria es un factor de riesgo para la mortalidad independiente de la MVPA (Koster et al. 2012).

Otra línea de investigación se centró en las asociaciones entre los descansos breves en el tiempo sedentario con los resultados metabólicos (Paing et al. 2018) y con la optimización de las operaciones cognitivas (Felez-Nobrega et al. 2018) (Magnon, Vallet y Auxiette 2018) (Falck, Davis, y Liu-Ambrose 2017). Benatti y Ried-Larsen (Benatti y Ried-Larsen 2015) afirmaron que existe evidencia suficiente para demostrar los efectos positivos de romper el tiempo prolongado de estar sentado en los resultados metabólicos.

Los métodos utilizados para medir la conducta sedentaria se pueden clasificar en subjetivos (cuestionarios y diarios de autoevaluación) y objetivos (sensores comunes de dispositivos ubicuos). Los métodos subjetivos están siendo superados por las nuevas tecnologías que pueden proporcionar, para todos los grupos de la población, información de segundo a segundo sobre la postura, el movimiento (o la falta de movimiento) y los patrones dentro y entre días. (Atkin et. Al. 2012).

Si bien los dispositivos móviles pueden considerarse como una de las causas del comportamiento sedentario (He y Agu 2016b), también ofrecen nuevas oportunidades para prevenirlo. Hoy en día, los dispositivos móviles portátiles, como teléfonos inteligentes, relojes inteligentes y rastreadores de ejercicios están equipados con una amplia variedad de sensores que se pueden usar para la actividad humana y el análisis de comportamiento. El uso de métodos objetivos para evaluar el comportamiento sedentario está creciendo en popularidad a medida que los costos de los dispositivos móviles portátiles disminuyen y son más fáciles de usar. En 2018, el 91% de las personas entre 18 y 49 años en los EE. UU. Posee un teléfono inteligente[[2]](#footnote-1). En este contexto, estos dispositivos pueden verse como una oportunidad para desarrollar métodos objetivos complejos para medir el comportamiento sedentario.

Se han implementado muchas aplicaciones para teléfonos inteligentes con el objetivo de alertar al usuario cuando se reconoce un comportamiento sedentario (He y Agu 2014; Grundgeiger et al. 2017; Fahim et al. 2017). Predecir conductas sedentarias futuras puede ayudar a habilitar intervenciones preventivas, como recordatorios y sugerencias para diferentes actividades basadas en la Teoría de la conducta planificada (TPB) (Ajzen, 1991). TPB postula que es más probable que un sujeto participe en las intervenciones recomendadas para reducir las conductas sedentarias si tales actividades están incluidas en sus planes. Siguiendo esta idea, nuestra hipótesis es que si pudiéramos predecir en el momento t que un sujeto será sedentario en el tiempo t + 1, podríamos recomendar actividades para cambiar su rutina sedentaria a largo plazo. Este tipo de intervenciones puede dar lugar a mejores oportunidades para cambiar el comportamiento de los sujetos a resultados más saludables.

Definimos la predicción del comportamiento sedentario futuro (FSBP) como la predicción de si la actividad física de un usuario superará o no, en promedio, 1,5 MET en un futuro próximo. El problema de predecir el comportamiento sedentario futuro se ha abordado previamente analizando solo el tiempo inactivo / estacionario de un sujeto y se ha medido y comparado el rendimiento de varios modelos. Sin embargo, aunque MET es una métrica estándar en el área de la salud para medir la intensidad de una actividad en términos del gasto de energía, el uso de esta métrica para predecir el comportamiento sedentario mediante el uso de dispositivos portátiles y dispositivos móviles permanece inexplorada. En este artículo, presentamos un enfoque novedoso para predecir el comportamiento sedentario futuro de un sujeto en términos de su nivel de MET (Tremblay 2012) a partir de diferentes algoritmos de Deep Learning.

En resumen, la pregunta de investigación que abordamos en este artículo es si es posible predecir el futuro comportamiento sedentario de un sujeto, basado en la observación de valores obtenidos de múltiples sensores de dispositivos portátiles y móviles.

# Trabajos relacionados

Monitoreo y modelado del comportamiento humano utilizando smartphones

A partir de que los dispositivos móviles y “wearables”, como smartphones, comenzaron a volverse populares, varios estudios fueron llevados a cabo para encontrar los patrones de uso que permiten relacionar, inferir y predecir diferentes tipos de comportamientos humanos en el contexto de la salud y el bienestar.

Harari et al. [(Harari et al. 2017)](https://paperpile.com/c/FQGDmo/VGeQ), en su investigación, parte del hecho de que los recientes avances en tecnología de sensado móvil ha revolucionado la evaluación del comportamiento al permitir el seguimiento no obstrusivo y continuo del comportamiento a través de dispositivos móviles (como acelerómetros y micrófonos) incluidos en smartphones.

Para examinar la viabilidad de usar métodos móviles de sensado para obtener estilos de vida diarios de comportamiento de estudiantes, los investigadores presentaron un estudio utilizando un dataset generado por una aplicación de sensado en smartphones llamada StudentLife para medir, objetivamente, la actividad natural de estudiantes y comportamientos sociales a lo largo del curso de un ciclo académico de 10 semanas de duración. Al hacerlo, abordaron una brecha en la literatura de mHealth existente al proporcionar una descripción de los patrones finos de estabilidad y cambio que caracterizan las conductas relacionadas con la salud de los estudiantes durante un período académico.

Los resultados de este estudio indican que, a medida que el ciclo académico progresa, los estudiantes tendían a participar en más comportamientos sedentarios y relacionarse menos con otros estudiantes, lo que puede estar relacionado a la necesidad de concentrándose en estudiar y preparar exámenes.

En conclusión, Harari et. al. demostraron la viabilidad de utilizar métodos objetivos de sensado basados en smartphones para realizar el seguimientos de comportamientos relacionados a la salud en el contexto de la vida cotidiana de estudiantes. Los identificadores identificaron varias correlaciones significativas entre los datos objetivos de los sensores de los dispositivos móviles y el bienestar mental y el rendimiento académico.

Gong et al. [(Gong et al. 2019)](https://paperpile.com/c/FQGDmo/TxFr), en una colaboración entre ingenieros y psicólogos, utilizaron los datos del acelerómetro y del GPS de smartphones en los momentos antes, durante y después de realizar llamadas y enviar mensajes de texto. Los investigadores hipotetizaron que los movimientos sutiles (medidos con acelerómetros) covarían con el nivel de ansiedad social y el contexto social. Por ejemplo, las manos pueden temblar más cuando el estudiante está en un ambiente público, debido a que la chance de evaluación negativa por otros es alta. Los autores denominan a estos movimientos como comportamientos dinámicos del usuario.

Para realizar el estudio, 52 estudiantes (20.5±3.8) fueron reclutados del grupo de participantes del departamento de psicología de la Universidad de Virginia. De cada participante se tomaron los datos del acelerómetro, el GPS y datos relacionados a las llamadas y a los mensajes de texto. Luego, modelaron los datos como un sistema lineal dinámico. Para ello, primero aplicaron el algoritmo Foursquare para obtener los puntos de interés a partir de los datos crudos del GPS de cada participante, y luego, los dividieron en sus locaciones semánticas (hogar, transporte, vida personal, lugar de comida). Respecto a los datos sociales (llamadas y mensajes de texto), los autores analizan por separado períodos de observación minutos antes y después de una llamada, para detectar por ejemplo, como varían los indicadores de ansiedad de un sujeto cuando debe realizar una llamada y no recibe una respuesta a la llamada realizada.

Los resultados obtenidos en el estudio mostraron una relación estadística significativa entre los datos de los sensores de smartphones y la ansiedad social de los estudiantes. A partir de los experimentos realizados, los autores concluyeron que hay una diferencia significativa en los comportamientos de movimiento ligados al nivel de ansiedad social y que estos comportamientos varían en diferentes ubicaciones semánticas. Estos resultados sugieren que es posible integrar datos de acelerómetros y de GPS para identificar marcadores de comportamiento asociados con la ansiedad social en varias locaciones. Por lo anterior, este trabajo abre nuevas posibilidades para monitorear pasivamente marcas de comportamiento de ansiedad social a través de la integración los datos del acelerómetro y los datos de GPS.

Kanjo et al. [(Kanjo et. al. 2019)](https://paperpile.com/c/FQGDmo/zeAi) adoptó un enfoque de Deep Learning para la clasificación de emociones a partir de datos de diferentes tipos de sensores portables. La clasificación de emociones a menudo requiere del modelado de tipos de datos de entrada de diferentes modalidades, como señales fisiológicas, datos ambientales, video y datos de movimiento y ubicación. Este problema ya había sido abordado desde la utilización de algoritmos del área de Machine Learning, pero aunque los procesos de ingeniería de características a menudo integrados en estos algoritmos son beneficiosos para la modelización de emociones, heredan algunas limitaciones críticas que pueden dificultar el desarrollo de modelos fiables y precisos. En este trabajo, los investigadores adoptaron un enfoque de aprendizaje profundo para la clasificación de emociones a través de un proceso iterativo mediante la adición y eliminación de un gran número de señales de sensores de diferentes modalidades. El conjunto de datos fue recogido en un estudio del mundo real desde teléfonos inteligentes y dispositivos portátiles. Combina la interacción local de tres modalidades de sensores: en el cuerpo, el medio ambiente y la ubicación en un modelo global que representa la dinámica de la señal junto con las relaciones temporales de cada modalidad. El enfoque propuesto emplea una serie de algoritmos de aprendizaje que incluyen un enfoque híbrido usando la Red Neuronal Convolucional y la Red Neural Recurrente de Memoria Corta a Largo Plazo (CNN-LSTM) en los datos crudos del sensor, eliminando las necesidades de extracción manual de características e ingeniería. Los resultados muestran que la adopción de enfoques de aprendizaje profundo es efectiva en la clasificación de la emoción humana cuando se utiliza un gran número de sensores de entrada (precisión media 95% y F-Measure=%95) y los modelos híbridos superan a la red neuronal profunda tradicional totalmente conectada (precisión promedio 73% y F-Measure=73%). Además, los modelos híbridos superan los algoritmos Ensemble previamente desarrollados que utilizan la ingeniería de características para entrenar la precisión media del modelo 83% y F-Measure=82%)

Wu et al. [(Wu et al. 2018)](https://paperpile.com/c/FQGDmo/H0Qg) utilizaron redes de encuentro basadas en los registros de encuentros Bluetooth para monitorear el nivel de stress de los usuarios. Bluetooth es un sensor móvil ampliamente disponible que puede detectar el entorno social de un individuo basado en la proximidad física. Los autores señalan que en la bibliografía existente falta la ingeniería sistemática de características y el examen del valor predictivo utilizando datos de encuentros Bluetooth. En este trabajo se extrajeron características de los datos de encuentro de Bluetooth de los teléfonos inteligentes de 49 estudiantes a lo largo de 66 días, provenientes del dataset StudentLife. El análisis de correlación sugiere que los características generadas pueden explicar la varianza en el stress significativamente mejor que los modelos nulos.

Por otro lado, en este trabajo los autores evaluaron la predictabilidad del nivel de stress que podía conseguirse utilizando las redes de encuentro de Bluetooth, más otros datos tales como la calidad del sueño y otros datos de sensores como GPS, acelerómetro, uso del teléfono, etc. Sus predictores pueden ser clasificados de acuerdo diferentes características: predecir el nivel de stress o el cambio del mismo, utilizar la técnica de evaluación Leave-One-Subject-Out (LOSO) o Leave-one-observation-out (LOOO), predecir el presente o el futuro. Los modelos utilizaron la técnica Random Forest en todos los casos. Los resultados indican que las características extraídas de las redes de encuentro representan una mejora en términos de predictabilidad para la predicción del nivel de stress y del cambio de este en el presente, pero no en el futuro. Estos resultados proveen evidencia del valor potencial de mejora al incorporar las redes de encuentro basadas en datos de redes Bluetooth a aplicaciones de monitoreo de stress.

Zia et al. [(Zia et al. 2016)](https://paperpile.com/c/FQGDmo/SRc7) destacan que utilizar los datos de dispositivos portables puede ayudar a que pacientes con Parkinson vivan mas segura e independientemente. Partiendo de esta idea, el objetivo de este trabajo consistió en determinar la viabilidad de utilizar redes neuronales para predecir la congelacion de la marcha, que es un síntoma del Parkinson. Particularmente, los investigadores usaron una clase de red neuronal conocidad como Redes Recurrentes en capas (LRN). El dataset utilizado para entrenar las redes fue uno público donado por el Repositorio de Aprendizaje Profundo de la Universidad de California y comprende un registro de los datos de movimiento (3 acelerómetros) de 3 sujetos, donde los momentos en los cuales se predice el Congelamiento de la Marcha están anotados como un *timestamp.* El trabajo no especifica las características obtenidas a partir de los datos del acelerómetro. Los parámetros que fueron variando en las diferentes redes neuronales testeadas fueron la cantidad de capas, las cantidad de neuronas por capas, el tamaño de batch y el factor de *down sampling*- esto es, la cantidad de ejemplos de entrenamiento que se quitaron del dataset-.

Los resultados mostraron que las que las redes que mejor precisión y recall tuvieron son aquellas de una sola capa, baja cantidad de neuronas y alto *down sampling*. Los resultados para cada uno de los 3 sujetos evaluados fueron 47, 30 y 27 de recall y 42, 89 y 57 de precisión. Los autores propusieron, como trabajo futuro, utilizar otros tipos de datos complementándolos con los datos de movimiento.

Predicción de actividades

La predicción de actividades es un área de investigación que consiste en hipotizar información sobre las actividades en las que un sujeto va a estar involucrado. En este contexto, la investigación se basó en la predicción del comportamiento sedentario. A nuestro entender, Q. He and E. Agu son los únicos dos autores que han propuesto, analizado y comparado modelos para la predicción de comportamiento sedentario futuro. A diferencia del enfoque propuesto en el presente trabajo, Q. He and E. Agu determinan el nivel de sedentarismo como el porcentaje de registros de actividad estacionarios que se tomaron de los datos de actividad física de cada usuario en cada hora, con respecto a los demás tipos de registros de actividad (corriendo, caminando, etc.).

Primero, propusieron un algoritmo de dominio de frecuencia para detectar patrones sedentarios recurrentes a partir de los datos de los usuarios. Este trabajo buscó ajustar funciones periódicas (seno, coseno) al porcentaje de registros de actividad estacionaria de cada hora y de cada usuario, tratando así de identificar los patrones subyacentes en su actividad física.

En un segundo estudio, los mismos autores exploraron la idea de si el contexto de un usuario puede ser usado para predecir su comportamiento sedentario futuro. Usando regresión logística, fueron capaces de clasificar las variables de contexto de cada usuario para predecir si iba a ser ¨muy sedentario¨ en la próxima hora, con una precisión de 73.1% y un recall de 87.7% (He and Agu 2016b). Paralelamente, propusieron un enfoque para descubrir automáticamente los patrones de actividad sedentaria de los usuario usando modelos auto-regresivos (He and Agu 2016c).

Finalmente, se enfocaron en detectar los ritmos predominantes de comportamiento sedentario modelando los ritmos cíclicos y ritmos lineales expuestos por la filosofía de Lefebvre (He and Agu 2017).

Es importante remarcar que en ninguno de los trabajos anteriormente expuestos se toma en cuenta el MET de cada actividad, que es el estándar en el área de la salud para decidir si una actividad es sedentaria o no. Consideramos que es muy importante tener en cuenta el concepto de MET al momento de diseñar modelos para la predicción de comportamiento sedentario, ya que este problema es de carácter interdisciplinario, por lo que para que dichos modelos tengan su mayor performance, es necesario tener en cuenta, además, los avances de las ciencias médicas. Adicionalmente, es importante resaltar que hasta el momento no se han utilizado modelos que usen algoritmos de Deep Learning para tratar el problema de la predicción del comportamiento sedentario.

Redes Neuronales y time-series

Los nuevos sensores remotos poseen gran resolución, tanto espectral como espacial, lo que ha beneficiado a las Series de Tiempo de Imágenes Satelitales (SITS, por sus siglas en inglés). Las series de imágenes son un componente clave para la sistemas de clasificación que apuntan a obtener mapas actualizados y precisos de la cobertura terrestre de las superficies de la Tierra. Pelletier et al. [(Pelletier, Webb, and Petitjean 2019)](https://paperpile.com/c/FQGDmo/qKiY) propone el uso de redes convolucionales temporales, es decir, aquellas que aplican la convolución en la dimensión temporal, para la

Dividir trabajos relacionados en: aquellos de activity prediction, aquellos de prediccion con redes neuronales (se podrian hacer 3 subsecciones de aceurto a rnn tcn y cnn) y prediccion a partir de datos de smartphones

[https://www.google.com/url?q=https://arxiv.org/pdf/1811.10166.pdf&sa=D&ust=1557167443720000&usg=AFQjCNFHEi-ii6qbZWtSebRrGoY61l0Kqw](https://arxiv.org/pdf/1811.10166.pdf)

# Descripción del problema

El problema que nos proponemos resolver está enmarcado dentro de la predicción del comportamiento sedentario futuro (FSBP, por sus siglas en inglés) y es el de predecir el gasto de energía (METs) que tendrá una persona en el futuro en base a datos recolectados de sensores de Smartphone. El gasto energético es medido en MET (Metabolic Equivalent of Tasks), que es la medida estándar en la comunidad científica que estudia la salud en relación con la actividad física. En el ámbito de la salud, se ha llegado a un acuerdo entre los investigadores en determinar cómo actividad física a toda aquella actividad cuyo MET asociado sea menor o igual a 1.5.

El problema de predicción a resolver es de regresión, ya que, como queremos predecir el gasto de energía, y éste es un valor continuo, debemos usar algoritmos de regresión. Más específicamente, de todos los tipos de redes neuronales vistas durante la cursada, utilizaremos Redes Neuronales Recurrentes (RNN) y Redes Neuronales Convolucionales, que son especialmente aptas para el problema que trataremos porque los datos que usamos están organizados en series de tiempo, por lo que tenemos un problema donde la secuencialita juega un papel importante. Para generar las features que serán utilizadas como entrada para diferentes Redes Neuronales, pre-procesaremos un dataset llamado StudentLife[[3]](#footnote-2).

Es importante notar que el trabajo desarrollado plantea un nuevo enfoque al problema de FSBP ya que laforma de estimar el nivel de sedentarismo utilizada hasta ahora en el estado del arte no está de acuerdo a la propuesto por la comunidad científica ya que no toma en cuenta el valor de MET de cada tipo de actividad física llevada a cabo, a diferencia de nuestro trabajo, en el que si fue llevado a cabo el cálculo de MET.

En este trabajo, trataremos predecir el valor de MET para un usuario únicamente en la hora siguiente. Es decir, dados los datos de las horas *t-n, ..., t-2, t-1, t* utilizaremos NNs para hallar el gasto energético en el tiempo *t+1.* Nuestro objetivo principal es evaluar si es factible utilizar redes neuronales que puedan tratar con problemas de predicción que deban aprender de una secuencia de datos para el problema de FSBP. Además, nos interesa comparar la performance de las RNN y las CNN (mas específicamente, las TCN), con el objetivo de comprobar los resultados empíricos conseguidos en “An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling”, por S. Bai et. al. Para evaluar nuestras hipótesis diseñaremos arquitecturas con RNN, CNN normales, TCN y simple NN.

# Descripción del dataset

Analizamos y pre-procesamos el dataset StudentLife[[4]](#footnote-3) [(Wang et al. 2014)](https://paperpile.com/c/FQGDmo/eCuKI) para probar la validez de nuestros modelos. Este dataset contiene datos de los sensores de los Smartphone de 48 estudiantes a lo largo de 10 semanas en la primavera de 2013.

El dataset se recogió de 30 estudiantes de pregrado y 18 graduados. De todo el grupo de estudiantes, 38 eran varones y 10 mujeres. Del grupo de estudiantes de pregrado, dos eran de primer año, 14 de segundo año, 6 de tercer año y ocho de cuarto año. También hubo 13 estudiantes de maestría de primer año y 1 de segundo año, y 3 estudiantes de doctorado. Los participantes fueron racialmente diversos, con 23 caucásicos, 23 asiáticos y 2 afroamericanos. Los datos disponibles incluyen: datos de actividad física, datos de conversación, datos de ubicación, entre otros.

Diferentes tipos de datos fueron registrados con diferentes frecuencias, dependiendo del tipo de sensor y la carga de trabajo global del Smartphone. Por ejemplo, la actividad física se registró cada 2–3 segundos en 1 de cada 4 minutos, el uso de aplicaciones fue se registró cada 20 minutos, y los registros de ubicación cada 10 minutos.

# Pre-procesamiento del dataset

Todos los datos del dataset StudentLife poseen un *timestamp* asociado. Decidimos discretizar las series de tiempo en *buckets* de una hora ya que esta es la granularidad utilizada en la mayor parte del trabajo relacionado para la predicción de sedentarismo. Por lo tanto, todas las *features* que se generaron corresponden a una combinación de usuario / hora en particular y son algún cómputo que resume algún aspecto de los datos de algún usuario / hora particular. Por ejemplo, un *bucket* específico puede corresponder al usuario 10 y la hora 2013-04-24 19: 00–20: 00. Los *buckets* para los cuales no teníamos información sobre la actividad física del usuario fueron eliminados. Además, se eliminó la información de un usuario en particular que presentaba inconsistencias en sus registros de actividad física. En total, obtuvimos 60,819 *buckets* para los 48 usuarios.

Los registros de actividad de cada usuario se están clasificados en el dataset como *estacionario*, *caminando* o *corriendo.* A partir de estos registros se calculó el valor del gasto energético para cada *bucket.* Tal como fue explicado anteriormente, el gasto energético es comúnmente medido de términos de Metabolic Equivalent of Tasks (MET). A cada tipo de actividad se le asignó un valor estático de MET de acuerdo al Compendio de Actividades Físicas[[5]](#footnote-4). Finalmente se calculó el promedio de MET para cada *bucket.*

Dado que discretizamos todos los datos en *buckets*  de una hora, tuvimos que tomar en cuenta que información iba a ser útil para predecir el valor del MET en las horas siguientes. Las *features* usadas fueron elegidas basándonos en previas investigaciones [(He and Agu 2016)](https://paperpile.com/c/FQGDmo/GSfPw)[(Cook and Krishnan 2015)](https://paperpile.com/c/FQGDmo/DZqSa), así como también otras fueron agregadas.

A continuación definimos las 20 *features* generadas pre-procesando el dataset StudentLife:

**GPS features**

* distanceTraveled: la distancia *haversine* recorrida calculada a partir de los registros de GPS;
* locationVariance: utilizado para medir la variabilidad en la ubicación GPS de un participante en cada *bucket*, computado como se describe en [(Saeb et al. 2015)](https://paperpile.com/c/FQGDmo/TbLIy).

**Time features**

* hourSine: transformación de seno de la hora;
* hourCosine: transformación de coseno de la hora;
* dayOfWeek: día de la semana;
* pastMinutes: el número de minutos transcurridos desde el comienzo del día; remainingMinutes: el número de minutos que quedan para terminar el día;

**Physical activity features**

* Stationary level: el porcentaje de instancias de actividad física clasificadas como estacionario/a’ en cada *bucket;*
* Walking level: el porcentaje de instancias de actividad física clasificadas como ‘caminando’ en cada *bucket;*
* Running level: el porcentaje de instancias de actividad física clasificadas como de ‘corriendo’ en cada *bucket;*
* activityMajor: el tipo de actividad física con más instancias en cada *bucket*;

**Audio features**

* SilenceLevel: el porcentaje de instancias de audio clasificadas como ‘silencio’ en cada bucket;
* voiceLevel: el porcentaje de instancias de audio clasificadas como ‘voz’ en cada bucket;
* noiseLevel: el porcentaje de instancias de audio clasificadas como ‘ruido’ en cada bucket;
* numberOfConversations: el número de conversaciones que ese estudiante tuvo en cada bucket;

**Other features**

* isCharging: si el Smartphone se estaba cargando;
* isLocked: si el Smartphone estaba bloqueado;
* isInDark: si el Smartphone estaba en la oscuridad;
* hasCalendarEvent: si tenía un evento programado en el calendario;
* wifiChanges: el número de cambio de conexiones wifi en cada bucket;
* sLevel: el valor de MET para la hora actual.

Finalmente, dado que las features *dayOfWeek* y *activityMajor* son categóricas, se realizó dummy-encoding sobre ellas. Con esta transformación, se obtuvieron 7 dummy features representando las categorías para la *feature* *dayOfWeek* y 3 para la feature *activityMajor*. Como resultado de esta última transformación, se finalizó el proceso con 28 *features* smartphone que serán luego usadas por los modelos predictivos para predecir el comportamiento sedentario de los usuarios.

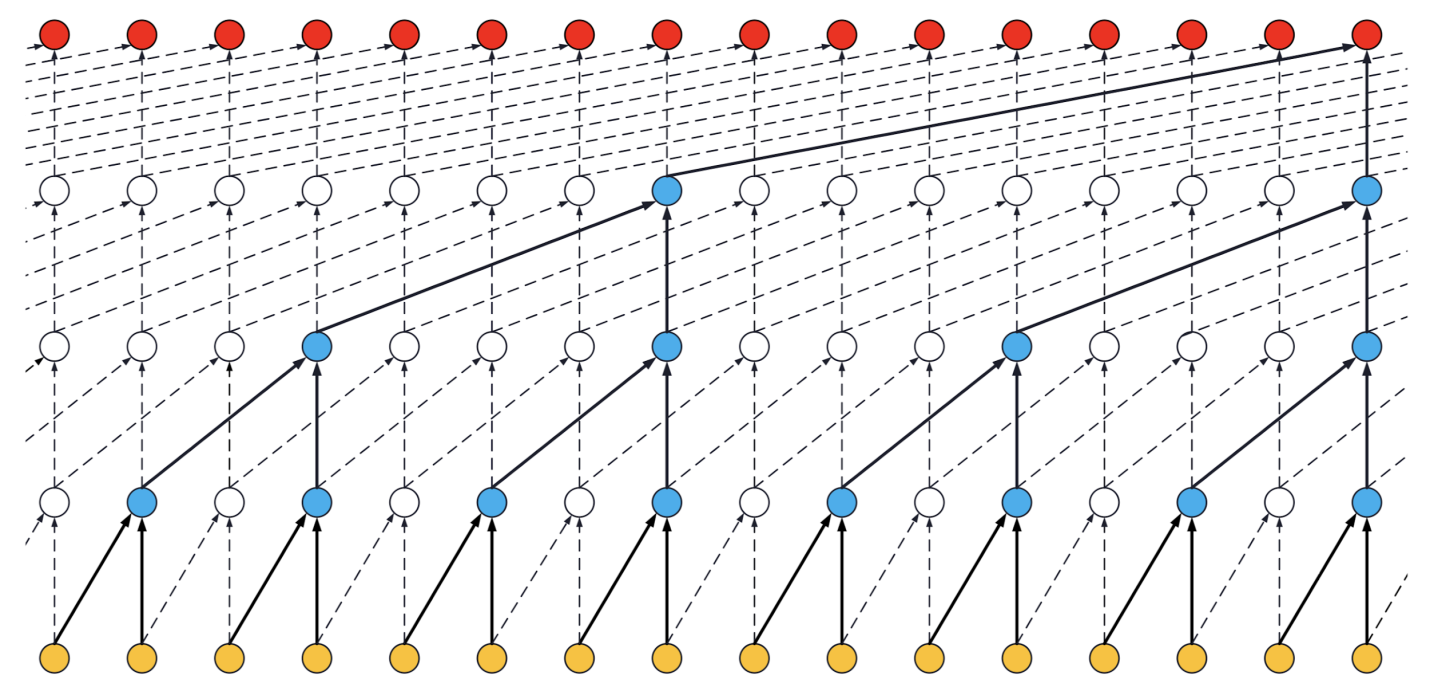
El siguiente paso del pre-procesamiento consistió en generar los time-lags. La cantidad de time-lags depende de cada arquitectura propuesta, siendo en algunos casos cero la cantidad de time-lags. Para ello, a una instancia del dataset que corresponde a un tiempo *t*  se le agregan las features de tiempos anteriores, dependiendo de la cantidad de time-lags que utilice la arquitectura. Este proceso aumenta la cantidad de features tantas veces como time-lags utilice la arquitectura, aumentando así el tamaño del dataset. Luego, se separa el dataset entre *X* e *y,* siendo *X* todas las *features* diferentes al *metLevel* del tiempo t (tiempo presente)*.* Por último, se dividen *X* e *y* entre casos de entrenamiento y de testeo. Para esta última división se usó un proporción de 2 a 1 en todos los casos.

El último paso del pre-procesamiento consiste en normalizar los datos. Para este paso se utilizaron normalizadores proveídos por la librería *Scikit-learn*. Varios normalizadores fueron tenidos en cuenta y, finalmente, se utilizaron diferentes según las características de cada usuario. Para evitar la fuga de datos, o también llamado *data lickeage,* para preparar los datos de entrenamiento de cada modelo primero realizamos una normalización de los datos de entrenamiento y, con la distribución aprendida por el normalizador, normalizamos los datos de entrenamiento. El *data lickeage* debe ser siempre tenido en cuenta ya que si se normalizan todos los datos al mismo tiempo se estaría perdiendo la posibilidad de evaluar la capacidad generalización de los predictores, ya que les estaríamos dando ya información de los datos que serán luego utilizados para testeo.

# Arquitecturas propuestas

Las arquitecturas propuestas fueron aquellas aptas para tratar con series de tiempo. Ellas son: Redes Neuronales Recurrentes (RNN) y Redes Neuronales Convolucionales (CNN). Clásicamente, las RNN fueron la elección para cualquier problema en el que la secuencia de los datos fuera importante (por ejemplo: traducción de texto). Recientemente, ciertos tipos de CNN han comenzado a utilizarse en los problemas para los que las RNN conseguían el mejor desempeño. Más específicamente, las CNN mencionadas utilizan operaciones convolucionales de una única dimensión. Entre las CNN propuestas, se encuentran aquellas implementadas con Keras por nosotros y las Temporal Convolutional Networks (TCN)[[6]](#footnote-5).

Las TCN son CNN que aplican la convolución de una dimensión, son causales, utilizan conexiones residuales y dilataciones para aumentar el campo receptivo. Las TCN demostraron superar ampliamente a las clásicas RNN basadas en GRU o LSTM en los problemas típicos como el problema de sumar o de copiar. Además, las TCN han demostrado tener más memoria que redes recurrentes de la misma capacidad. Por otro lado y no menos importante, las TCN y las CNN en general, poseen la característica de ser altamente paralelizables, en contraste con las RNN, por los que el proceso de la búsqueda óptima del valor de los hiper-parametros -proceso también llamado “tuneo”- se realiza de una forma más rápida. La implementación utilizada para las TCN puede hallarse en el siguiente [link.](https://github.com/philipperemy/keras-tcn#why-temporal-convolutional-network) En Fig. 1 se puede observar un ejemplo de una sencilla TCN. En d icha figura se ve claramente el carácter causal de la red, ya que nunca se toma información del futuro. La red del ejemplo posee un único bloque residual y dilaciones de 1, 2, 4 y 8.

Fig. 1

Uno de los hiper-parámetros que varía en las distintas arquitecturas propuestas es la cantidad de time-lags utilizados. Los time-lags son la cantidad de información que se le da a la arquitectura sobre el pasado. Es decir, si se quiere predecir el gasto de energía de un usuario en un tiempo t, si la cantidad de time-lags es 3, el input de la red neuronal estará formado por las features del tiempo anterior (t-1), así como también las features de los tiempos t-2 y t-3.

En general, los hiper-parámetros a tunear son: tipo de capas utilizadas, cantidad de capas, cantidad de unidades por capa, técnica de regularización a utilizar, time-lags a utilizar, algoritmo de optimización y la función de pérdida. Además las CNN poseen ciertos parámetros que no poseen las demás arquitecturas, como la cantidad de filtros por capa o el tamaño del kernel. Más aun, las TCN poseen hiper-parámetros que no poseen las CNN: como la cantidad de bloques residuales o la lista de dilataciones. Estos últimos hiperparámetros se tunean de una forma menos aleatoria que los demás, ya que lo que se intenta calcular en este caso es que el campo receptivo sea igual o mayor a la cantidad de time-lags.

En resumen, las arquitecturas 1,2 y 3 son RNN, las arquitecturas 4 y 5 son CNN y la arquitectura número 6 es una NN simple. La arquitectura número 6 es una NN simple que utiliza únicamente información de la hora anterior para predecir el valor del MET de la hora siguiente y fue agregada a modo de comparación y para evaluar la utilidad de utilizar datos del pasado para FSBP. Esto quiere decir que si la arquitectura 6 supera a las RNN y CNN que utilizan información del pasado para predecir el valor de MET puede significar que, para este problema, la información del pasado no ayuda a mejorar la calidad de los predictores.

A continuación se muestra un cuadro que describe las arquitecturas propuestas. Las características más detalladas de cada arquitectura pueden ser analizadas en la función *get\_model* del script *tp\_dl\_train.py.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Arq. 1 | Arq. 2 | Arq. 3 | Arq. 4 | Arq. 5 | Arq. 6 |
| Tipo de capas | LSTM - FC x 3 | LSTM x 2 - FC x 1 | LSTM x 3 - FC x 4 | Ver *model.summary()* | Conv1D x 1 - Flatten x 1 - FC x 1 | Dense x 3 |
| Neuronas por capa | 32-16-8-1 | 64-32-1 | 128-64-32-32-16-8-1 | Ver *model.summary()* | 32 - No aplica - 1 | 64 - 32 - 1 |
| Técnica de regularización | Dropout | Dropout | Dropout | Dropout | Dropout y Batch Normalization | Dropout |
| Número de time-lags | 8 | 12 | 12 | 8 | 4 | 1 |
| Algoritmo de optimización | Adam | Adam | Adam | Adam | Adam | Adam |
| Función de pérdida | MSE | MSE | MSE | MSE | MSE | MSE |
| Cantidad de épocas | 256 | 128 | 128 | 64 | 128 | 256 |
| Tamaño de batch | 64 | 64 | 64 | 64 | 64 | 128 |
| Relación Train/Test | ⅔-⅓ | ⅔-⅓ | ⅔-⅓ | ⅔-⅓` | ⅔-⅓ | ⅔-⅓ |
| Número de bloques residuales | No aplica | No aplica | No aplica | 1 | No aplica | No aplica |
| Lista de dilataciones | No aplica | No aplica | No aplica | 1-2-4 | No aplica | No aplica |
| Tamaño del kernel | No aplica | No aplica | No aplica | 2 | 2 | No aplica |
| Número de filtros | No aplica | No aplica | No aplica | 8 | 32 | No aplica |

FC = Fully Connected

# Métrica de evaluación seleccionada

La métrica utilizada para evaluar el desempeño de las redes diseñadas es el Error Cuadrático Medio o MSE (Mean Squared Error). El MSE es una medida de desempeño cuantitativa utilizada comúnmente para medir el error que hay entre el valor real y el valor estimado. En este contexto MSE es igual a la sumatoria de los errores cuadráticos. En comparación con el Error Medio Absoluto o MAE, MSE amplifica y penaliza con mayor fuerza aquellos errores de mayor magnitud. La fórmula de cálculo del MSE se muestra a continuación:

Donde:

* es el resultado en el tiempo t.
* es el pronóstico de valor en el tiempo t.

diferencias MAE y MSE

-MAE es robusto frente outliers

-MSE penaliza a los errores mas grandes

-el MAE es linealmente dependiente de la diferencia entre el valor predicto y el valor real

-en MSE el error crece cuadráticamente.

-MAE expresa el error en una unidad del tipo de la variable a predicta

-MSE no lo hace, RMSE si.

-MSE y RMSE le da mayor peso a los errores mas grandes.

MSE o RMSE deben ser usados cuando los errores grandes son particularmente indeseables.

Creo que en nuestro problema fallar en los outliers no lastima tanto, pq aca lo importante es saber cuando es mayor que 1.5 el MET, no cambia tanto si es 2 o 8.

En otras palabras, compara un valor predicho y un valor observado o conocido.

# Resultados

Es interesante notar que el dataset que generamos está formado por una gran cantidad de usuarios, donde cada usuario posee su propia rutina y características personales, que no tienen porque parecerse a la de los demás usuarios. Además, algunos usuarios poseen menos datos que otros. Debido a esto, es difícil hallar una arquitectura que funcione correctamente para todos los usuarios por las idiosincrasias de cada uno de ellos. Por este motivo, decidimos realizar un proceso de búsqueda de los hiper-parámetros que conseguían los mejores resultados para ciertos usuarios. Tres usuarios fueron seleccionados para medir la performance de las arquitecturas propuestas, ya que, realizar un análisis tan extenso (48 usuarios) no era el objetivo de la cátedra. Los usuarios seleccionados fueron elegidos específicamente debido a que tenían patrones de gasto energético diferentes. Los 3 usuarios son: 50 (bajo MET), 31 (MET medio), 4 (MET alto). Además los usuarios seleccionados poseen un diferente grado rutinario en su actividad física. Lo anterior puede corroborarse con el cálculo de la correlación entre los gastos energéticos de las mismas horas de los diferentes días de la semana. En Fig. 2, Fig. 3 y Fig. 4 se muestra el porcentaje del valor de MET a lo largo de los días de la semana y las horas del día de los 3 usuarios, respectivamente.

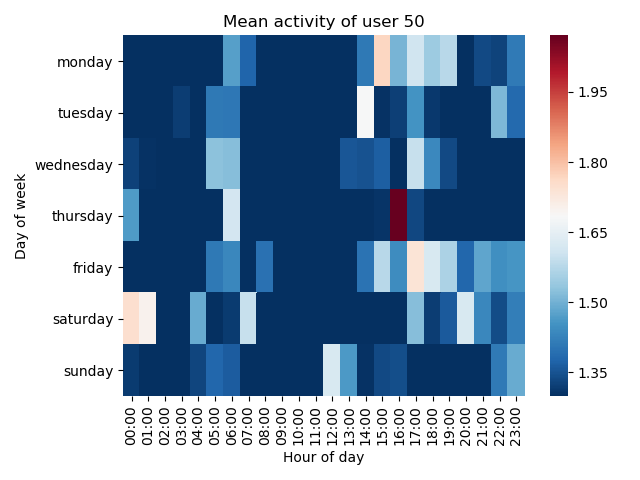


Fig. 2

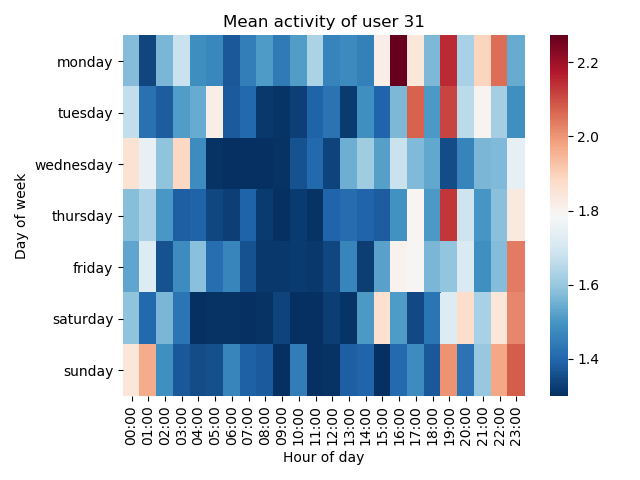


Fig. 3

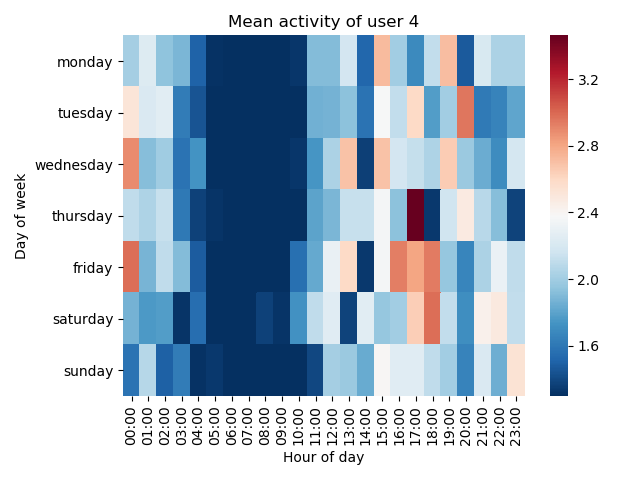


Fig. 4

A continuación, se presenta una tabla que muestra los valores de *mse* conseguidos para cada predictor. Se presenta el *mse* para entrenamiento y para el testeo. A pesar de que para evaluar la capacidad de generalización que posee un predictor debe proveerse su error prediciendo a partir de un conjunto de datos que no fue utilizado para entrenarlo, mostramos en la tabla el *mse* conseguido en el entrenamiento para poder analizar si los predictores produjeron *overfitting*.

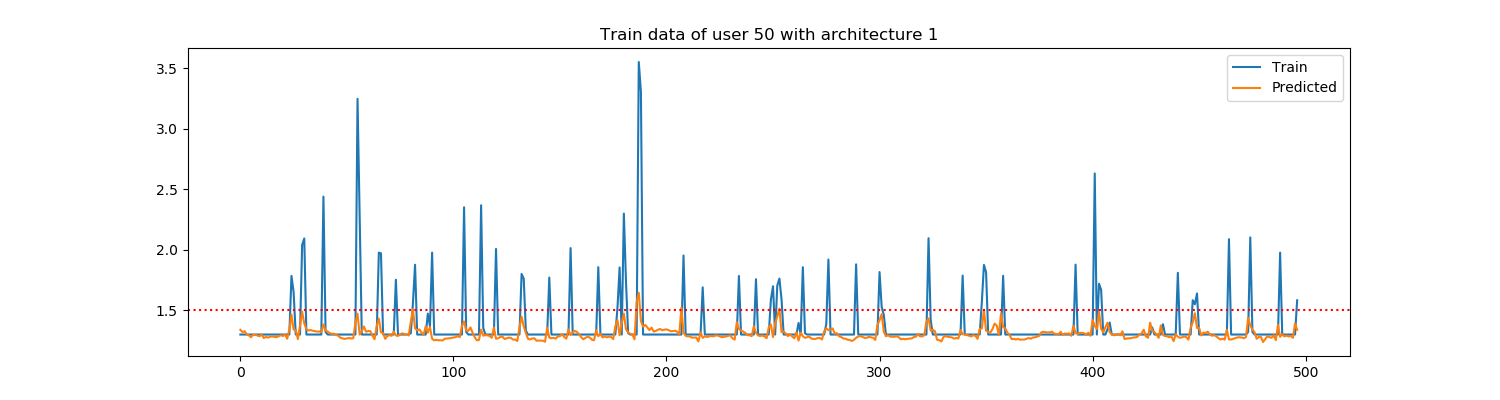
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| MSE Train/ MSE Test | Arquitectura 1 | Arquitectura 2 | Arquitectura 3 | Arquitectura 4 | Arquitectura 5 | Arquitectura 6 |
| Usuario 50 | 0.058/0.028 | 0.039/.0.04 | 0.043/0.035 | 0.05/0.028 | 0.042/0.029 | 0.034/0.047 |
| Usuario 31 | 0.143/0.309 | 0.047/0.0355 | 0.052/0.382 | 0.116/0.321 | 0.108/0.307 | 0.069/0.342 |
| Usuario 4 | 0.444/0.608 | 0.197/1.002 | 0.168/0.73 | 0.364/0.619 | 0.308/0.59 | 0.281/0.651 |

Si el entrenamiento se realiza nuevamente, los resultados de cada predictor pueden variar ya que no se consiguió que los resultados sean replicables fijando semillas mediante los generadores de números pseudo-aleatorios de TensorFlow y Numpy. Esto puede deberse a que las librarías que utiliza CudaNN para que los modelos sean entrenados aprovechando la GPU introduzcan cierto grado de aleatoriedad.

# Discusión

A continuación pasarán a discutirse los resultados obtenidos. Es interesante notar qué regresor tuvo un mejor desempeño para cada uno de los usuarios y analizar el porqué de dicho resultado.

El usuario más desafiante para los predictores resultó ser el número 50, ya que fue el menos rutinario y con menos casos de comportamiento no sedentario. Por lo tanto, resultó difícil que los NN aprendieran, a partir de las features del pasado, cuándo iba a tener un comportamiento no sedentario. Las arquitecturas 1 y 4 fueron las que mejor desempeño presentaron. Las Fig. 5 y 6 muestran las predicciones de entrenamiento y testeo para esta arquitectura.

Fig. 5

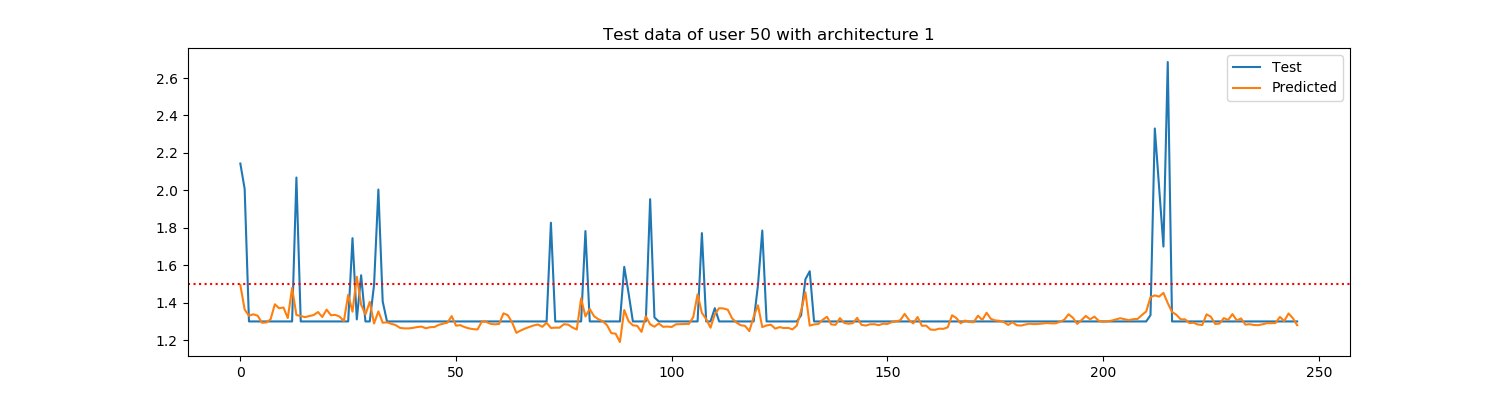


Fig. 6

En el caso del usuario 31, se puede observar un comportamiento más rutinario. Teniendo en cuenta que en el pre-procesamiento se le dio especial atención a las *features* de tiempo, es entendible que los regresores se desempeñen mejor en estos casos. La arquitectura 5 (CNN) fue la que mejor performance tuvo en este caso, con un *mse* de testeo de 0.307. Las Fig. 7 y 8 muestran las predicciones de entrenamiento y testeo para esta arquitectura.

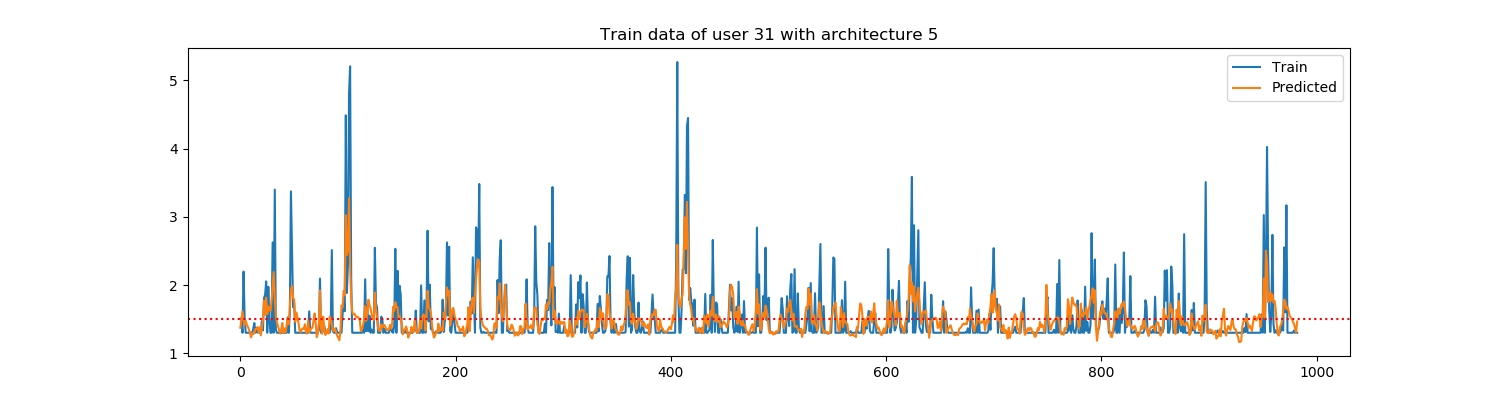


Fig. 7

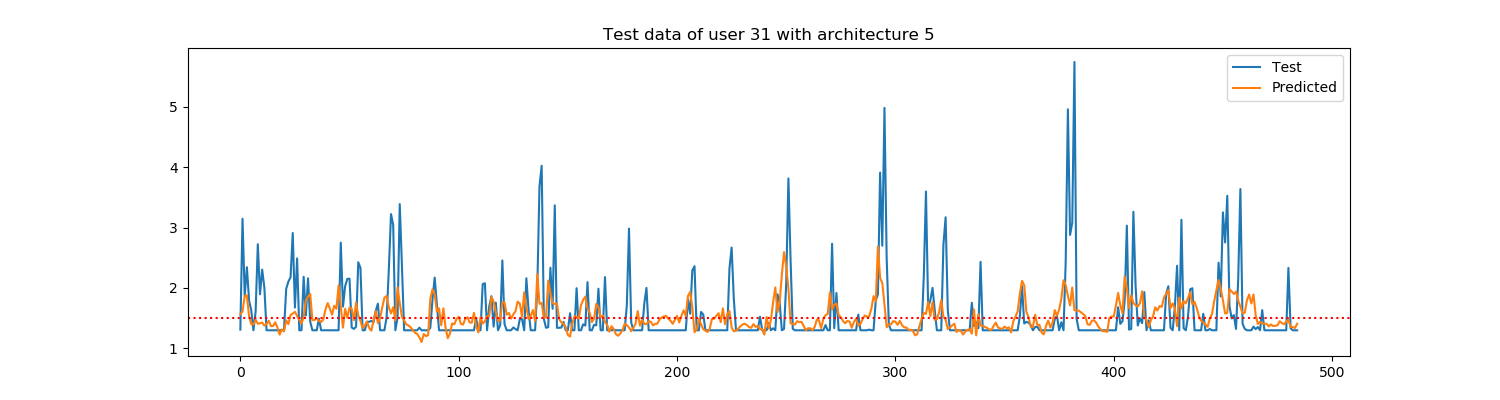


Fig. 8

Por último, los modelos sobre el usuario 4 fueron los que mejor desempeño tuvieron, a pesar de tener un mayor mse que los modelos sobre los otros usuarios. Esto se debe a que este usuario posee muchos mas casos de comportamiento no sedentario, por lo que para que el regresor se desempeñe bien debe aprender los patrones de comportamiento del usuario a partir de las features provistas. En el caso del usuario 50, el predictor podría solo aprender a predecir un valor cercano a 1.3 METs para obtener un bajo *mse.* La arquitectura que mejor se desempeñó en este caso fue la número 4 (TCN). Esto, y el hecho de que para todos los modelos las CNN tuvieron, en general, una mejor performance que las RNN, confirma nuestra hipótesis sobre que las CNN, y en particular las TCN, son más adecuadas para problemas de predicción que deban aprender de una secuencia de datos. Las Fig. 9 y 10 muestran las predicciones de entrenamiento y testeo para esta arquitectura.

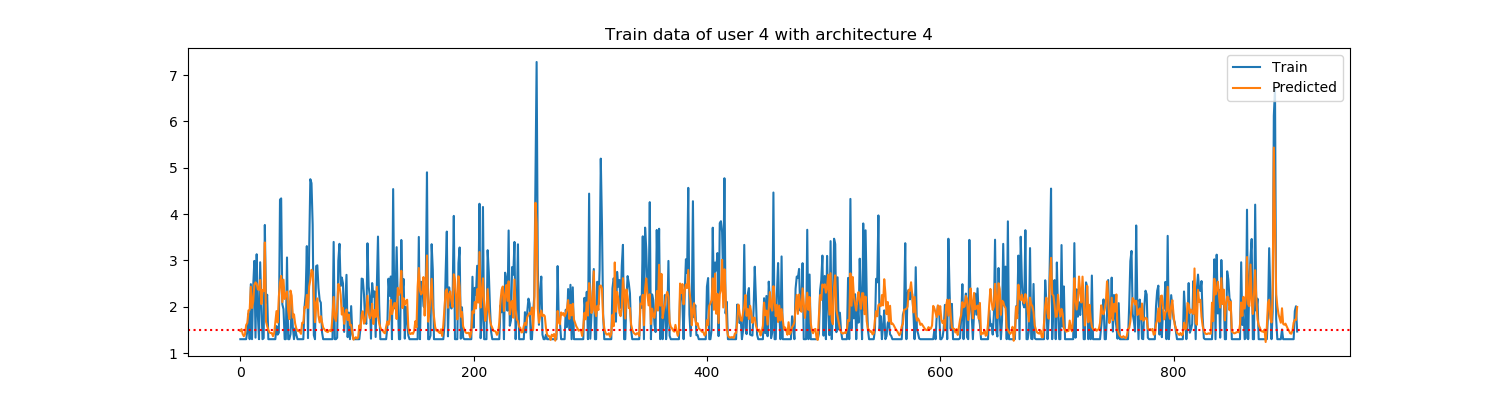


Fig. 9

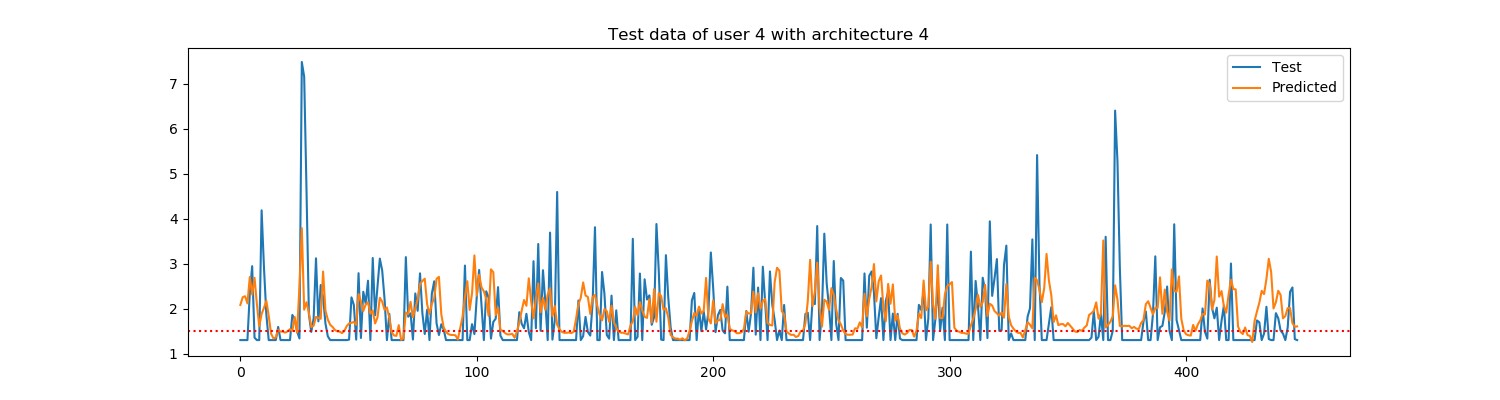


Fig. 10

# Conclusiones

En este trabajo nos propusimos comparar diferentes tipos de redes neuronales para evaluar su performance en el problema de la predicción del comportamiento sedentario. Este trabajo es el primero en evaluar el rendimiento de redes neuronales que intenten sacar provecho al carácter secuencial del problema tratado. Se presentaron usuarios con diferentes niveles de gasto energético, lo que hace que muy difícil diseñar una arquitectura que tenga una buena performance para todos ellos.

A pesar de haber utilizado un alto grado de regularización, las NN más extensas son propensas a producir overfitting. En consecuencia, tendrán un mal rendimiento en la etapa de *testing*. Este es un *trade-off* que fue considerado fue el nivel de la complejidad de las redes. Al momento de diseñar cada red, se buscó que tuvieran, cada una, una performance aceptable para cada uno de los usuarios seleccionados. Por lo tanto, es importante que la complejidad de cada red sea la menor posible para que posean la capacidad necesaria de generalización.

Por otra parte, como solo se toman los datos de un usuario en particular para entrenar las redes, la cantidad de datos es muy baja y no permite llegar a entrenar apropiadamente a redes más complejas y con mayor cantidad de parámetros. Este fue un *trade-off* al momento de diseñar las arquitecturas propuestasy podría llegar a ser resuelto aumentando la granularidad en el pre-procesamiento y tomar, por ejemplo, *buckets* de tiempo de 30 minutos o 15 minutos, resultando así en más casos de entrenamiento.

Otro de los trade-off que tuvimos que enfrentar fue la cantidad de time-lags a tomar. Es lógico pensar que cuanta más información del pasado se le dé a la arquitectura, más información tendrá y así podrá hacer mejores predicciones. Esto no es cierto por dos razones. La primera de ellas es que el MET que tendrá un usuario en el tiempo *t+1* puede no estar relacionado con las features que poseemos del mismo usuario en el tiempo *t-n,* siento n un entero arbitrario. Esto, al tener pocos casos de entrenamiento, puede generar ruido en el modelo. La segunda razón es que, como en muchos casos no se poseen datos sobre todas las horas que duró el experimento, cuantos más time-lags se tengan, mayor es la probabilidad de que un caso de entrenamiento sea descartado. Por ejemplo, si la cantidad de time-lags es *n* y si se tienen los datos del usuario *x* desde el tiempo *t* al *t-n+1* pero no del tiempo *t-n,* el caso de entrenamiento se desecha.

Un punto importante a destacar es que, como los estudiantes mantuvieron, por lo general, un comportamiento totalmente sedentario, los clasificadores que intentaban siempre predecir valores cercanos a cero MET fueron descartados, ya que conseguían un MSE bajo pero no eran útiles. Este problema fue tratado en el trabajo realizado por M. Santillán Cooper y M. Armentano con las técnicas de *over-sampling* y asignación de peso a las clases.

Con respecto a las hipótesis planteadas al comienzo del informe, todas ellas pudieron ser evaluadas y comprobadas. Se lograron diseñar arquitecturas con un mejor desempeño que aquellas arquitecturas que no utilizan información del pasado. Además, se comprobó que las CNN pueden llegar a desempeñarse de una mejor manera que las RNN.

# Bibliografía

[Cook, Diane J., and Narayanan C. Krishnan. 2015. *Activity Learning: Discovering, Recognizing, and Predicting Human Behavior from Sensor Data*. John Wiley & Sons.](http://paperpile.com/b/FQGDmo/DZqSa)

[Gong, Jiaqi, Yu Huang, Philip I. Chow, Karl Fua, Matthew S. Gerber, Bethany A. Teachman, and Laura E. Barnes. 2019. “Understanding Behavioral Dynamics of Social Anxiety among College Students through Smartphone Sensors.” *Information Fusion*. https://doi.org/](http://paperpile.com/b/FQGDmo/TxFr)[10.1016/j.inffus.2018.09.002](http://dx.doi.org/10.1016/j.inffus.2018.09.002)[.](http://paperpile.com/b/FQGDmo/TxFr)

[Harari, Gabriella M., Samuel D. Gosling, Rui Wang, Fanglin Chen, Zhenyu Chen, and Andrew T. Campbell. 2017. “Patterns of Behavior Change in Students over an Academic Term: A Preliminary Study of Activity and Sociability Behaviors Using Smartphone Sensing Methods.” *Computers in Human Behavior*. https://doi.org/](http://paperpile.com/b/FQGDmo/VGeQ)[10.1016/j.chb.2016.10.027](http://dx.doi.org/10.1016/j.chb.2016.10.027)[.](http://paperpile.com/b/FQGDmo/VGeQ)

[He, Qian, and Emmanuel O. Agu. 2016. “Smartphone Usage Contexts and Sensable Patterns as Predictors of Future Sedentary Behaviors.” In *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*. https://doi.org/](http://paperpile.com/b/FQGDmo/GSfPw)[10.1109/hic.2016.7797695](http://dx.doi.org/10.1109/hic.2016.7797695)[.](http://paperpile.com/b/FQGDmo/GSfPw)

[Kanjo, Eiman, Eman M. G. Younis, and Chee Siang Ang. 2019. “Deep Learning Analysis of Mobile Physiological, Environmental and Location Sensor Data for Emotion Detection.” *Information Fusion*. https://doi.org/](http://paperpile.com/b/FQGDmo/zeAi)[10.1016/j.inffus.2018.09.001](http://dx.doi.org/10.1016/j.inffus.2018.09.001)[.](http://paperpile.com/b/FQGDmo/zeAi)

[Pelletier, Charlotte, Geoffrey Webb, and François Petitjean. 2019. “Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series.” *Remote Sensing*. https://doi.org/](http://paperpile.com/b/FQGDmo/qKiY)[10.3390/rs11050523](http://dx.doi.org/10.3390/rs11050523)[.](http://paperpile.com/b/FQGDmo/qKiY)

[Saeb, Sohrab, Mi Zhang, Christopher J. Karr, Stephen M. Schueller, Marya E. Corden, Konrad P. Kording, and David C. Mohr. 2015. “Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study.” *Journal of Medical Internet Research* 17 (7): e175.](http://paperpile.com/b/FQGDmo/TbLIy)

[Wang, Rui, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. 2014. “StudentLife.” In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp ’14 Adjunct*. https://doi.org/](http://paperpile.com/b/FQGDmo/eCuKI)[10.1145/2632048.2632054](http://dx.doi.org/10.1145/2632048.2632054)[.](http://paperpile.com/b/FQGDmo/eCuKI)

[Wu, Congyu, Mehdi Boukhechba, Lihua Cai, Laura E. Barnes, and Matthew S. Gerber. 2018. “Improving Momentary Stress Measurement and Prediction with Bluetooth Encounter Networks.” *Smart Health*. https://doi.org/](http://paperpile.com/b/FQGDmo/H0Qg)[10.1016/j.smhl.2018.07.017](http://dx.doi.org/10.1016/j.smhl.2018.07.017)[.](http://paperpile.com/b/FQGDmo/H0Qg)

[Zia, Jonathan, Arash Tadayon, Troy McDaniel, and Sethuraman Panchanathan. 2016. “Utilizing Neural Networks to Predict Freezing of Gait in Parkinson’s Patients.” *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS ’16*. https://doi.org/](http://paperpile.com/b/FQGDmo/SRc7)[10.1145/2982142.2982194](http://dx.doi.org/10.1145/2982142.2982194)[.](http://paperpile.com/b/FQGDmo/SRc7)

1. https://www.sedentarybehaviour.org/ [↑](#footnote-ref-0)
2. http://www.pewresearch.org/fact-tank/2018/09/28/internet-social-media-use-and-device-ownershipin-

   u-s-have-plateaued-after-years-of-growth/ [↑](#footnote-ref-1)
3. http://studentlife.cs.dartmouth.edu/ [↑](#footnote-ref-2)
4. Link al dataset: http://studentlife.cs.dartmouth.edu/ [↑](#footnote-ref-3)
5. https://sites.google.com/site/compendiumofphysicalactivities/ [↑](#footnote-ref-4)
6. https://arxiv.org/abs/1803.01271 [↑](#footnote-ref-5)