



UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA

Trabajo Fin de Máster del  
Máster en Ingeniería y Ciencia de Datos

**Detección de emociones con  
pulsera inteligente Fitbit**

Ernesto P. Vallejo Luengo

Dirigido por: Olga C. Santos Martín

Raúl Cabestrero Alonso

Curso: 2021-2022: <1<sup>a</sup>> Convocatoria



# Agradecimientos

Quiero dedicar este trabajo a todos los amigos que me han apoyado durante el proceso: Asier, Aritz, Patri, Sara, bea, Rafa, Dani, Álvaro, Yara, David, Éder, María, Cris y Julia. Y especialmente quiero darle las gracias a mis padres y a mi hermana por todo lo que ha hecho por mí.



## Resumen

En el campo de la detección de emociones, las pulseras inteligentes o *smartbands* son un posible sustituto de dispositivos más especializados. Este proyecto estudia la capacidad de la pulsera comercial Fitbit Sense de detectar emociones. Se diseñó un experimento en el que se inducían distintas emociones a los participantes mientras la pulsera registraba su frecuencia cardíaca. Los participantes debían además evaluar su propio estado emocional. Se aplicó una serie de transformaciones a los datos obtenidos y se extrajeron varias características. Se eligió el modelo de regresión lineal múltiple para intentar predecir las emociones a partir de las medidas realizadas. Tras realizar varias pruebas los modelos finales no consiguieron una buena precisión diferenciando entre estados emocionales. No se ha podido determinar si la pulsera Fitbit Sense puede usarse para detectar emociones.

## **Abstract**

In the field of emotion detection, smartbands are a possible substitute for more specialized devices. This project studies the ability of the commercial smartband Fitbit Sense to detect emotions. An experiment was designed in which different emotions were induced to the participants while the smartband recorded their heart rate. Participants also had to assess their own emotional state. A series of transformations were applied to the data obtained and several characteristics were extracted. The multiple linear regression model was chosen to try to predict emotions from the measurements made. After performing several tests, the final models did not achieve good accuracy differentiating between emotional states. It has not been possible to determine whether the Fitbit Sense can be used to detect emotions.

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Tema . . . . .	1
1.2. Antecedentes . . . . .	2
1.3. Objetivos . . . . .	2
1.4. Estructura del documento . . . . .	3
<b>2. Estado del Arte</b>	<b>5</b>
2.1. Procesado de datos y extracción de características . . . . .	5
2.2. Modelos de aprendizaje automático . . . . .	6
<b>3. Recogida de datos</b>	<b>7</b>
3.1. Pulsera Fitbit Sense . . . . .	7
3.2. Descripción de las emociones: valencia y <i>arousal</i> . . . . .	8
3.3. Recogida de datos . . . . .	9
3.3.1. Sincronización de relojes . . . . .	10
3.3.2. Participantes . . . . .	10
3.3.3. Errores en los datos recogidos . . . . .	11
<b>4. Procesamiento de los datos</b>	<b>15</b>
4.1. Suavizado . . . . .	15
4.2. Estandarización . . . . .	16
4.3. Interpolación . . . . .	16
4.4. Filtrado de observaciones por número de medidas . . . . .	17
4.5. Obtención de características . . . . .	17
4.5.1. Parámetros que caracterizan una distribución . . . . .	17
4.5.2. Frecuencia fundamental de la serie temporal . . . . .	18
4.5.3. Autocorrelación . . . . .	18
4.5.4. Valor cuadrático medio . . . . .	18

---

<b>5. Regresión lineal</b>	<b>19</b>
5.1. Afinación de parámetros del procesado de datos . . . . .	19
5.2. Selección de variables . . . . .	22
5.3. Evaluación resultados . . . . .	22
5.3.1. Valencia . . . . .	22
5.3.2. <i>Arousal</i> . . . . .	24
5.3.3. Evaluación resultados . . . . .	29
5.4. Análisis por clip . . . . .	29
5.5. Análisis por persona . . . . .	30
<b>6. Conclusión</b>	<b>33</b>
<b>A. Instrucciones API web de Fitbit</b>	<b>39</b>

# Índice de figuras

3.1. Escala SAM . . . . .	9
3.2. Pulsaciones participante 14 . . . . .	12
3.3. Pulsaciones participante 2 . . . . .	13
4.1. Procesamiento de los datos . . . . .	15
5.1. P-valor frente a $med_{min}$ . . . . .	21
5.2. P-valor frente a $f_s$ . . . . .	21
5.3. Regresión y residuos valencia . . . . .	23
5.4. <i>Heatmap</i> correlaciones valencia . . . . .	25
5.5. Regresión y residuos <i>arousal</i> . . . . .	27
5.6. <i>Heatmap</i> correlaciones <i>arousal</i> . . . . .	28



# Índice de tablas

3.1. Lista de películas . . . . .	10
3.2. Porcentaje datos perdidos . . . . .	12
5.1. <i>Stepwise forward selection</i> valencia . . . . .	23
5.2. VIF modelo valencia . . . . .	25
5.3. <i>Stepwise forward selection arousal</i> . . . . .	26
5.4. VIF modelo <i>arousal</i> . . . . .	27
5.5. Promedio residuos por clip valencia . . . . .	30
5.6. Promedio residuos por clip <i>arousal</i> . . . . .	30
5.7. Promedio residuos por persona valencia . . . . .	31
5.8. Promedio residuos por persona <i>arousal</i> . . . . .	31



# Capítulo 1

## Introducción

### 1.1. Tema

La detección automática de emociones consiste en determinar el estado emocional de una persona a partir de parámetros fisiológicos. En este contexto, el término «emoción» se refiere a la experiencia de corta duración que siente una persona en respuesta a un estímulo. La detección de emociones se lleva a cabo utilizando aprendizaje automático. Gracias a este, se pueden entrenar modelos que toman los parámetros fisiológicos como *input* y devuelven la emoción como *output*. La detección de emociones es importante en áreas muy diversas, como la robótica, el entretenimiento o la educación.

Existen varios parámetros que se pueden utilizar para detectar emociones, como el electromiograma, el electrocardiograma o la actividad electrodérmica. Los instrumentos que realizan estas medidas son normalmente dispositivos de laboratorio con certificación médica. Esos dispositivos no solo son caros, sino que también requieren personal especializado para utilizarlos, lo que hace complicado usarlos fuera del laboratorio. Además, algunas de estas medidas requieren de procesos invasivos.

Un posible sustituto a estos instrumentos de laboratorio son las pulseras inteligentes, también conocidas como *smartbands*. Estas pulseras son dispositivos pequeños y ligeros que se llevan en la muñeca y que son capaces de monitorear distintos parámetros del cuerpo. Las pulseras inteligentes tienen la ventaja de que se pueden usar en mayor variedad de entornos y que no requieren de personal especializado para ser usadas, además de ser relativamente baratas. Una de estas pulseras es la Fitbit Sense, que se anuncia como una pulsera inteligente con varios sensores y componentes para medir diferentes variables fisiológicas. Entre ellas se encuentran la actividad electrodérmica y la variabilidad de la frecuencia cardíaca. Ambos parámetros han sido usados en la detección de emociones, lo que sugiere que la pulsera Fitbit podría usarse también con este fin.

## 1.2. Antecedentes

Este trabajo de fin de máster se desarrolla dentro de una investigación conjunta entre las facultades de Psicología e Informática en el proyecto ENFOQUE DE DESARROLLO INTELIGENTE E INTRASUJETO PARA MEJORAR ACCIONES EN SISTEMAS ADAPTATIVOS EDUCATIVOS QUE CONSIDERAN EL ESTADO AFECTIVO (INT2AFF) y en el proyecto de innovación docente del Marco Lógico Colaborativo (MLC). El MLC es una herramienta colaborativa basada en la web cuyo objetivo es crear entornos que permitan el aprendizaje a partir de la exploración, el debate y la interacción entre participantes [Santos and Boticario (2014)]. Puesto que el estado emocional juega un rol importante en los procesos de aprendizaje, el MLC utiliza herramientas inteligentes que permitan monitorear las emociones de los estudiantes de manera automática, liberando al equipo docente de esta tarea [Santos et al. (2013)]. Por estos motivos, se está evaluando si la pulsera Fitbit es capaz de detectar emociones, puesto que en caso afirmativo podrá ser utilizada en el MLC.

Ya se ha realizado otro trabajo de fin de máster [Herráiz (2022)] evaluando la pulsera Fitbit en el ámbito de la detección de emociones. En él se estudió la viabilidad de usar modelos de aprendizaje automático no supervisados para detectar cambios en el estado emocional a partir de los datos obtenidos por la fitbit. Para ello se desarrolló una aplicación móvil que procesaba los datos de la pulsera en tiempo real y se consiguió que detectara automáticamente los cambios afectivos. Este trabajo tiene además la particularidad de que usó las series temporales proporcionadas por la pulsera directamente sin realizarles ningún cambio de representación.

## 1.3. Objetivos

El objetivo del presente trabajo es comprobar si los datos fisiológicos proporcionados por la pulsera Fitbit Sense pueden usarse para detectar estados emocionales. Este estudio se diferenciará del realizado en Herráiz (2022) en los siguientes puntos:

- Se les aplicará un cambio de representación a las series temporales obtenidas por la pulsera. Concretamente, se obtendrán una serie de características de las series, que serán las que se usen en los modelos de aprendizaje automático.
- Se utilizarán modelos de aprendizaje supervisado. Por este motivo, será necesario un crear un conjunto de datos correctamente etiquetado.

Para llevar a cabo este estudio se proponen los siguientes tres subobjetivos:

- Diseñar una experiencia en la que se eliciten emociones a participantes voluntarios mientras llevan puesta la pulsera *fitbit*. Durante esta experiencia se pedirá a los participantes que evalúen su estado emocional. Estas autoevaluaciones de los participantes y la información recogida por la pulsera conformarán el conjunto de datos etiquetados que se usará en los siguientes pasos del trabajo.

- Procesar los datos obtenidos en el paso anterior y prepararlos para entrenar un modelo de aprendizaje supervisado. Puesto que los datos recogidos por la pulsera fitbit son series temporales habrá que estudiar qué transformaciones y cambios de representación son necesarios para que sean útiles.
- Entrenar uno o varios modelos de aprendizaje supervisado y estudiar si a partir de los datos de la pulsera se puede predecir el estado emocional y con qué precisión.

## 1.4. Estructura del documento

En el capítulo 2 se describen algunos trabajos similares que se han usado como punto de partida para este proyecto. El capítulo 3 describe cómo se ha utilizado la pulsera Fitbit Sense y qué datos han podido obtenerse con ella. También se explican las escalas con las que se han cuantificado las emociones y el experimento realizado para obtener los datos. En el capítulo 4 se explica el procesado que se le ha aplicado a los datos en crudo para poder utilizarlos en un modelo de aprendizaje supervisado. En el capítulo 5 se detallan todas las pruebas que se han realizado con el modelo de regresión lineal y qué resultados se han obtenido con ellas. Por último, en el capítulo 6 se encuentran las conclusiones del trabajo.



# Capítulo 2

## Estado del Arte

### 2.1. Procesado de datos y extracción de características

Normalmente, los parámetros fisiológicos medidos para realizar detección de emociones u otros procesos cognitivos tienen forma de serie temporal. Estas series pueden ser usadas directamente como input de los modelos de aprendizaje automático, pero también es común obtener un conjunto de características y usar estas para entrenar los modelos. En este trabajo se ha adoptado este segundo enfoque y a continuación se describen el preprocesado que se suele aplicar a las series temporales y las características más utilizadas.

Antes de extraer ninguna información de las series temporales, se pueden realizar varias transformaciones a las series. Para eliminar sesgos relacionados con las condiciones del experimento o de los participantes se aplica un estandarizado (restar la media y dividir entre la desviación típica) o un escalado MinMax a la serie (Rahman et al. (2018); Giannakos et al. (2019); Wijasena et al. (2021)). Para eliminar irregularidades abruptas de la serie temporal se aplica un suavizado o *smoothing* (Nogueira et al. (2013)).

Una vez realizado el preprocesamiento de la serie se pueden extraer las características. Para identificar las bandas de frecuencia más importantes se realiza una transformada de Fourier y se calculan sus máximos (Giannakos et al. (2019); Sitnikova et al. (2009); Guo et al. (2015)). Para estudios que intentaban clasificar tareas cognitivas se han utilizado coeficientes de autoregresión (Rahman et al. (2018)). Para describir la forma de la envolvente de la serie temporal se utilizan varios estadísticos como la media, el valor máximo y el mínimo, la desviación típica, la asimetría y la curtosis (Giannakos et al. (2019); Basu et al. (2015); Guo et al. (2015)). Por último, el valor cuadrático medio (*root mean square*) que se utiliza para medir el nivel de cansancio (Rui Xu (2018)).

En este área llama especialmente la atención el artículo "Fitbit for learning: Towards capturing the learning experience using wearable sensing" Giannakos et al. (2020). En este trabajo, se midieron 4 parámetros fisiológicos: la frecuencia cardíaca, la actividad electrodermal, la temperatura de la piel y la presión arterial. A estas medidas, se les aplicaron todas las transformaciones y se extrajeron todas

las características ya mencionadas. Finalmente, utilizaron estos datos para evaluar la experiencia de aprendizaje de los participantes con bastante éxito. Cabe destacar que a pesar de que mencionan la pulsera fitbit en el nombre del artículo, en realidad utilizaron otra marca distinta.

## 2.2. Modelos de aprendizaje automático

Se han utilizado con éxito distintos algoritmos de aprendizaje automático supervisado para reconocer emociones. A continuación se comentan algunos de los resultados más relevantes. En tareas de clasificación, el modelo de *support vector machine* (SVM) ha sido capaz de distinguir entre 4 emociones, tristeza, enfado, estrés y sorpresa con una exactitud del 61.8% (Kim et al. (2004)). SVM también ha servido para distinguir entre 13 emociones distintas con una exactitud del 81.45% (Verma and Tiwary (2014)) y entre distintos niveles de estrés con una exactitud de 79.3% (Katsis et al. (2008)). Otros modelos de clasificación a destacar son *k nearest neighbours* que distinguió entre 6 emociones con 72.3% de exactitud (Lisetti and Nasoz (2004)) y *linear discriminant analysis* que distinguió entre tres niveles de estrés con exactitud de 97% (Healey and Picard (2005)).

En cuanto a modelos de regresión, también hay varios con buenos resultados. Se han utilizado modelos de regresión lineal para predecir el *arousal* y la valencia, consiguiendo un  $R^2$ ajustado de 0.9 y 0.88 respectivamente (Nogueira et al. (2013)). El mismo estudio consiguió mejorar hasta un  $R^2$ ajustado de 0.95 utilizando una regresión polinómica. Utilizando *support vector regression* se ha conseguido una tasa de reconocimiento del 89.2% al distinguir entre tristeza, miedo y placer (Chang et al. (2013)), y utilizando *regression trees* se distinguió entre aburrimiento dolor y sorpresa con una precisión del 93% (Jang et al. (2012)).

# Capítulo 3

## Recogida de datos

En esta sección se empezará discutiendo qué variables fisiológicas registra la pulsera y cómo acceder a esos datos. A continuación se explica cómo se puede caracterizar el estado emocional de una persona y cómo describirlo cuantitativamente. Por último se describirá la experiencia diseñada para realizar la recogida de datos.

### 3.1. Pulsera Fitbit Sense

La pulsera está pensada para usarse como periférico de un *smartphone*. La descarga de una aplicación en el teléfono y la creación de una cuenta de usuario en la misma permite que ambos dispositivos se sincronicen mediante *Bluetooth*. Esta sincronización es necesaria para acceder a los datos recogidos por la pulsera, ya que es en la aplicación del móvil desde donde se suben los datos al servidor de Fitbit.

Una vez que los datos están en el servidor hay distintas maneras de acceder a ellos:

- Mediante el registro como usuario desde la página <https://www.fitbit.com/> se puede pedir una exportación del archivo de toda la vida útil de la cuenta. De todas las variables fisiológicas discutidas, este archivo solo incluye la frecuencia cardíaca con una resolución de 1 minuto.
- Otra manera posible es utilizar la API web que proporciona Fitbit <sup>1</sup>. Con ella se puede descargar un registro de la frecuencia cardíaca con resolución de 5 segundos.
- Por último, Fitbit provee un entorno web para el desarrollo de aplicaciones donde se puede crear una aplicación que se ejecute en la pulsera. Desarrollando una aplicación así, se podría recoger la frecuencia cardíaca con una resolución de hasta 1 segundo.

Ninguna de las 3 opciones permite descargar otra variable que no sea la frecuencia cardíaca, así que no es posible utilizar ni la actividad electrodermal ni la variabilidad de la frecuencia cardíaca.

---

<sup>1</sup><https://dev.fitbit.com/build/reference/web-api/>

Desarrollando una aplicación se conseguiría la mejor resolución posible, pero el mismo proceso de desarrollar la aplicación consumiría gran parte del tiempo del proyecto. Finalmente se ha optado por utilizar la API ya que es un método rápido y una resolución de 5 segundos es aceptable. Existe además una implementación de esta API en Python <sup>2</sup> que es la que se ha utilizado ya que el resto del proyecto también se ha desarrollado en Python.

Con la API se obtiene una tabla con dos columnas: la primera columna contiene la frecuencia cardíaca expresada en pulsaciones por minuto. La segunda columna contiene la fecha y la hora minuto y segundo en el que se ha medida esa frecuencia cardíaca. Aunque en teoría la resolución de la medida es de 5 segundos, se ha comprobado que a veces los intervalos entre mediciones son de 10 o hasta 15 segundos, por lo que hay mediciones que no se están realizando o que se están perdiendo. En la subsección 3.3.3 se detallan los errores encontrados en los datos recogidos y la cantidad de estos que se han perdido.

## 3.2. Descripción de las emociones: valencia y *arousal*

Aunque todos tenemos una idea intuitiva de qué son las emociones, clasificarlas o medirlas no es trivial. A la hora de tratar esta cuestión hay dos enfoques. Uno es clasificar las emociones como entidades discretas (enfado, miedo, alegría, tristeza etc). El otro consiste en evaluar las emociones en función de dos dimensiones distintas y ortogonales, la valencia y el *arousal*. La valencia, también llamada tono hedónico, es la medida en que una emoción resulta agradable o desagradable, mientras que el *arousal* es la medida en que una emoción es activadora o relajante (Dzedzickis et al. (2020)).

Para que los participantes voluntarios puedan autoevaluar sus emociones mediante la valencia y el *arousal*, se ha utilizado la escala *Self-Assessment Manikin* (SAM). El SAM consiste en una serie de imágenes para cada una de las dimensiones emocionales y una escala numérica asociada. Una vez que se les ha explicado las escalas a los participantes, estos pueden escoger un valor numérico para la valencia y el *arousal* de la emoción que han sentido (Bynion and Feldner (2017)). Como se puede ver en la figura 3.1, en nuestro proyecto las escalas de ambas dimensiones emocionales van del 1 al 9. En el caso de la valencia un 1 representa una emoción muy desagradable, un 5 una neutral y un 9 una muy agradable. Para el *arousal* un 1 representa una emoción relajante, un 5 una emoción algo activadora y un 9 una emoción muy activadora.

Se puede ver que con estas dos variables es relativamente intuitivo representar varias emociones. Por ejemplo, la euforia tendría una valencia alta por ser agradable y un *arousal* alto por ser excitante. En cambio el aburrimiento tendría una valencia algo bajo por ser desagradable y una activación baja por ser poco activador.

---

<sup>2</sup><https://github.com/orcasgit/python-fitbit>

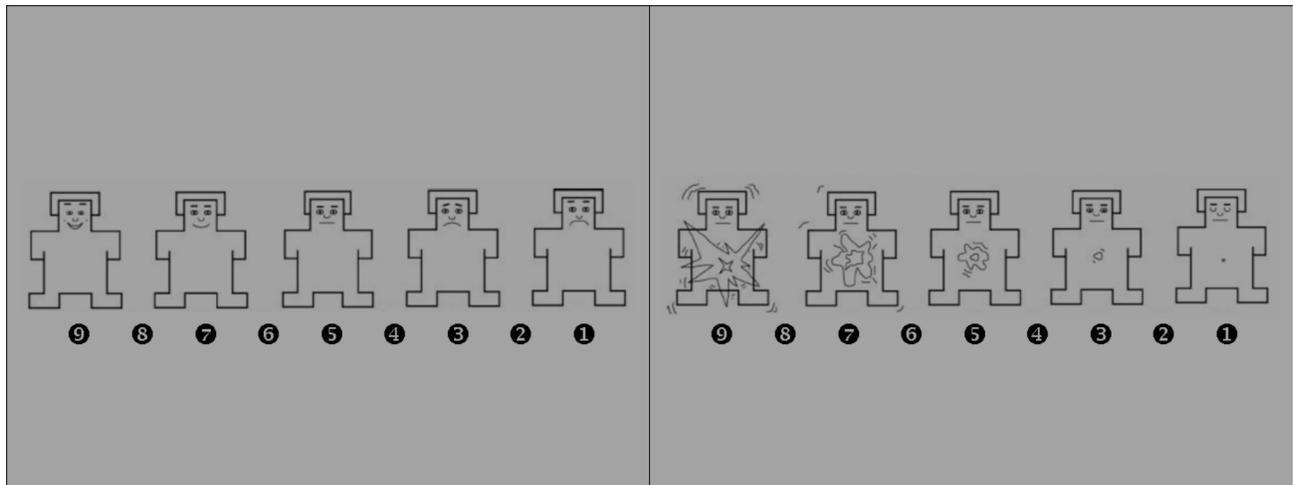


Figura 3.1: Izquierda: Escala SAM para la valencia. La escala va desde el 9 (muñeco muy sonriente) hasta el 1 (muñeco muy triste). Derecha: Escala SAM para el arousal. La escala va del 9 (muñeco muy activado) hasta el 1 (muñeco completamente relajado).

### 3.3. Recogida de datos

Para recoger datos era necesaria una experiencia en la que los participantes sintieran distintas emociones mientras llevaban puesta la pulsera Fitbit Sense. Uno de los métodos más usados para provocar emociones es el visionado de películas o fragmentos de ellas (Fernández et al. (2012)). Por lo tanto, la experiencia que se ha diseñado consiste en que participantes que llevan la pulsera vean fragmentos de distintas películas comerciales (clips) y evalúen ellos mismos cómo se han sentido mediante la valencia y el *arousal*. A continuación se describe en detalle el experimento.

Al empezar la experiencia se le explicaba al participante que iba a ver unos clips y que iba a tener que evaluar como se había sentido viéndolos. Se le explicaba también qué son la valencia y el *arousal* para que pudiera hacer la evaluación. Una vez explicadas las instrucciones el participante escuchaba música relajante durante 90 segundos. Después se reproduce un clip y se le pide al participante que evalúe cómo se ha sentido. Este primer vídeo es un ejemplo para que el participante pueda familiarizarse con la interfaz y los datos recogidos no son tenidos en cuenta. Tras el vídeo de práctica se reproducen consecutivamente otros 10 clips y tras cada uno de ellos se pide de nuevo que evalúe su estado emocional. Para terminar, el participante vuelve a escuchar 90 segundos de música relajante igual que al empezar. En general cada experimento duró alrededor de media hora.

Los periodos de 90 segundos de relajación sirven para medir la línea base. La línea base es la serie de respuestas fisiológicas que se dan en ausencia de estímulos emocionales. En este experimento pueden diferenciarse 2 líneas base, la previa al experimento y la posterior. De aquí en adelante cuando se hable de la línea base se estará hablando del conjunto de medidas de ambas, la previa y la posterior. El conocer la línea base nos permite estandarizar el resto de medidas del experimento de manera que se eliminen los efectos situacionales provocados por el experimento, como por ejemplo los nervios. La manera exacta en que se estandariza la frecuencia cardíaca se encuentra en la subsección 4.2.

Película	Duración clip (segundos)
El amante	44
Algo pasa con Mary	178
El piano	44
American History X	78
El campeón	153
Tres colores: Azul	24
City of Angels	277
El Exorcista	100
Un pez llamado Wanda	176
Scream	215

Tabla 3.1: Lista de películas con la duración del clip escogido en segundos

Las películas utilizadas y la duración del clip se encuentran en el cuadro 3.1. Estas han sido elegidas de una batería de películas comerciales, que han sido específicamente validadas para inducir distintas emociones (Fernández et al. (2011)).

La interfaz con la que interactúan los participantes se diseñó usando OpenSesame. OpenSesame es un programa pensado para llevar a cabo experimentos psicológicos que permite reproducir vídeo y audio y también recoger las respuestas de los participantes. Al acabar cada una de las experiencias OpenSesame generaba un fichero csv con los valores de *valencia* y *arousal* y con la hora, minuto y segundo a la que habían empezado tanto los clips como los audios de relajación. Registrar el momento exacto en el que empiezan y terminan los videos y los audios es necesario para saber con qué medida de la pulsera corresponden.

### 3.3.1. Sincronización de relojes

Para saber qué medidas tomadas por la pulsera corresponden con qué clips (y por lo tanto con qué valores de *arousal* y *valencia*) es necesario hacer corresponder los *timestamps* proporcionados por la pulsera con los proporcionados por el experimento en OpenSesame. Desgraciadamente, el reloj de la pulsera y el del ordenador dónde se ejecuta OpenSesame no estaban sincronizados perfectamente y se llevaban unos segundos de diferencia.

Para solucionar este problema, antes y después de cada experimento se medía y registraba la diferencia de hora entre ambos dispositivos. Esta medición se hacía manualmente, ya que como la pulsera solo toma datos cada 5 segundos no hace falta que la sincronización fuera muy precisa. Las diferencias medidas se corregían luego en los *timestamps* de comienzo de cada clip.

### 3.3.2. Participantes

Han realizado el experimento un total de 18 personas, siendo la mitad mujeres y la mitad hombres. Sus edades están comprendidas entre los 23 y los 66 años. De entre todos ellos, 7 son usuarios

habituales de pulsera inteligentes. Antes de comenzar la experiencia todos leyeron y firmaron un consentimiento informado siguiendo las indicaciones del comité de ética de la UNED y validado para el proyecto INT2AFF. En él, se les informaba de los contenidos de la prueba, sus objetivos y se les indicaba que podían retirarse del estudio en cualquier momento si así lo deseaban.

Los experimentos con los 3 primeros participantes sirvieron como prueba piloto. Gracias a ellos se identificaron algunas mejoras que hicieron la experiencia más fácil de entender:

- La valencia y el *arousal* se mencionaban nada más empezar la experiencia antes de haberlos explicado, lo que confundía a los participantes. Se cambió el texto para que supieran que se iban a explicar ambos conceptos poco después.
- Al escribir sus puntuaciones de valencia y *arousal* la interfaz no daba ningún *feedback* a los participantes, por lo que estos no estaban seguros de que su respuesta se hubiera registrado. Se modificó el experimento para que tras valorar sus emociones apareciera un breve mensaje con los números que habían pulsado.
- Un participante expresó incomodidad debido a que inmediatamente después de acabar la música relajante empezara un clip súbitamente. Se añadió una transición corta entre la música y el clip para que los participantes no se sorprendieran.

Aparte de en las pruebas piloto, también se preguntó su opinión sobre la experiencia al resto de participantes. En general, todos declararon que las instrucciones les habían parecido claras y concisas y que no habían tenido dudas mientras realizaban el experimento. Algunos comentaron haberse sentido incómodos al ver el clip de la película “American History X”.

### 3.3.3. Errores en los datos recogidos

Como ya se ha comentado en la subsección 3.1, la pulsera no siempre recoge la frecuencia cardíaca cada 5 segundos, sino que a veces se producen intervalos de hasta 10 o 15 segundos sin que se produzca ninguna medida. En el cuadro 3.2 se muestra la cantidad de datos que se han perdido en cada uno de los experimentos. Además, en 2 de los experimentos la cantidad de datos perdidos ha sido significativamente mayor y se explican a continuación.

Se ha observado que durante gran parte del experimento 14 la pulsera dejó de tomar medidas. Como se puede ver en la figura 3.2, solo se registraron las pulsaciones al comienzo y al final de la experiencia. No se ha podido determinar la causa del error.

En los datos del experimento número 2 se ha observado un aumento repentino e inusual del ritmo cardíaco. Como se puede ver en la figura 3.3, las pulsaciones por minuto casi se duplican en menos de 10 segundos, y permanecen muy altas durante varios minutos. Se informó de esto al participante, que declaró no haber sentido nada extraño durante la experiencia. Por este motivo se ha concluido que la causa de esta anomalía está en la pulsera, y se han eliminados estos datos extraños del conjunto final.

Experimento	Porcentaje datos perdidos
1	43.5 %
2	53.6 %
3	22.6 %
4	33.7 %
5	36.4 %
6	23.7 %
7	30.7 %
8	40.1 %
9	29.0 %
10	29.4 %
11	22.2 %
12	35.0 %
13	32.4 %
14	81.9 %
15	23.6 %
16	36.6 %
17	25.4 %
18	29.3 %
Total	34.95 %

Tabla 3.2: Porcentaje de datos perdidos en cada uno de los experimentos. En el experimento 2 también se cuentan como datos perdidos los datos anómalos.

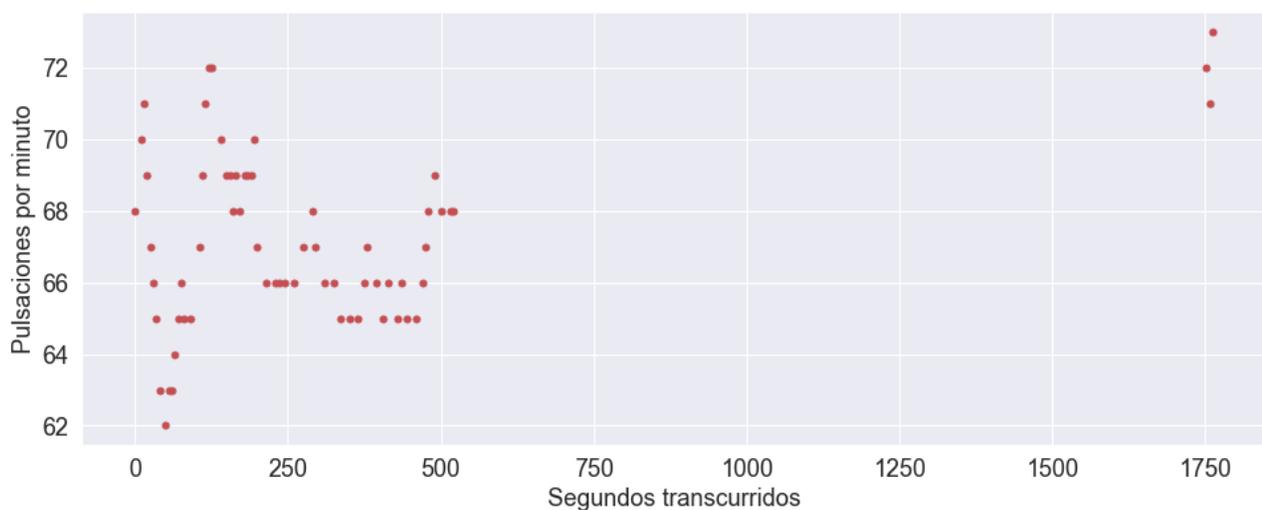


Figura 3.2: Frecuencia cardíaca medida al participante 14. Hay un intervalo de 20 minutos sin datos



Figura 3.3: Frecuencia cardíaca medida al participante 2. Se puede observar que las pulsaciones por minutos aumentan repentinamente de alrededor de 60 a 110.



# Capítulo 4

## Procesamiento de los datos

Las medidas de la frecuencia cardíaca proporcionadas por la pulsera conforman una serie temporal. Puesto que se realizó el experimento con 18 personas, se dispone de 18 series temporales. Como cada uno de los experimentos consta de 10 clips, cada una de las series temporales se puede dividir en 10 subseries. A cada una de estas subseries se les denominará observación. En principio, se dispone de 180 observaciones pero por los errores en la recogida de datos ya comentados, el número real de observaciones útiles es 171. A cada una de estas observaciones le corresponde un valor de *arousal* y uno de valencia. Por lo tanto, el planteamiento de este trabajo es usar las series temporales de cada uno de las observaciones para entrenar un modelo que prediga el *arousal* y otro que prediga la valencia.

Como se adelantó en la sección 2.1, es necesario realizar varias transformaciones a estas series temporales y obtener un conjunto de características de ellas. En la figura 4.1 se muestran en orden los pasos que se van a seguir. A continuación se detallan las transformaciones realizadas a las series temporales y cómo se han calculado las características

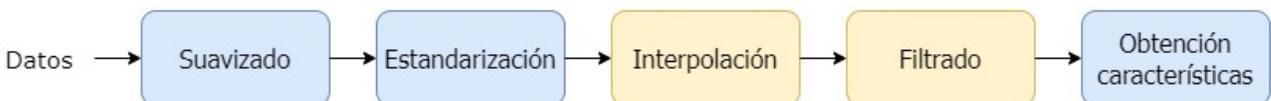


Figura 4.1: Pasos en orden del procesamiento de los datos. En azul las transformaciones cogidas de Giannakos et al. (2020). En amarillo las aportaciones propias.

### 4.1. Suavizado

El suavizado o *smoothing* es una transformación utilizada para eliminar las irregularidades abruptas de las series temporales, lo cual permite apreciar mejor los patrones subyacentes (Nogueira et al. (2013)). De los distintos métodos de suavizado, se ha elegido el de medias móviles (*moving averages*) por su simpleza y por la facilidad de implementación. Se ha decidido que el suavizado sea el primer paso en el procesado de datos puesto que así los posibles *outliers* introducirán menos ruido en los

siguientes pasos.

*Moving averages* consiste en sustituir cada valor de la serie temporal por su media aritmética con los valores adyacentes. Es decir, se sustituiría el valor de  $x_t$  por el de  $\frac{x_{t-n}+x_{t-n+1}+\dots+x_t+\dots+x_{t+k-1}+x_{t+k}}{n+k+1}$ . Cuanto mayores sean los valores de  $k$  y  $n$  más suave será la curva. Se ha decidido que el valor de  $n$  sea 0 para que la media no incluya valores anteriores de la serie temporal, pues esto provocaría que los valores de una observación influyesen en el siguiente. Este mismo problema no existe para  $k > 0$  ya que después de cada observación el participante existe una pequeña pausa durante la cual la emoción elicitada por el clip aún perdura.

Se ha denominado factor de suavizado a  $f_s = k + 1$  y se ha comprobado que su valor influye significativamente en los resultados obtenidos. En la subsección 5.1 se discute qué pruebas se han realizado con el factor de suavizado y sus resultados.

## 4.2. Estandarización

Después del suavizado se ha aplicado una estandarización a cada una de las series temporales. Para ello se ha tomado la media y la desviación típica de la línea base de cada experimento. A continuación se ha restado esa media de todos los valores de la serie temporal y luego se han dividido entre la desviación. Esto se ha hecho así para eliminar las posibles diferencias entre los participantes y eliminar los efectos situacionales provocados por el experimento (Rahman et al. (2018); Giannakos et al. (2019)).

## 4.3. Interpolación

Como se ha explicado en la subsección 3.3.3, la pulsera Fitbit no siempre toma una medida cada 5 segundos. Para sustituir estas medidas que no llegan a realizarse se ha aplicado una interpolación a las series temporales. Esta interpolación se ha hecho utilizando *splines*. La regresión mediante *splines*, al contrario que una regresión polinómica normal, no ajusta todos los datos a un polinomio para todos los valores de  $x$ , sino que ajusta un polinomio distinto para regiones distintas de  $x$ . La frontera entre dos de estas regiones se llama nudo o *knot*, y normalmente se añade la restricción de que en los nudos los valores de los polinomios y sus derivadas sean continuos.

Para implementar la interpolación con *splines* se ha utilizado la clase *interpolate* de *scipy*. Se ha introducido la serie temporal en el método *splrep* que calcula la regresión, y a continuación se ha usado el método *splev* para evaluar la regresión en los puntos dónde faltan las medidas.

## 4.4. Filtrado de observaciones por número de medidas

Como los clips tienen duraciones dispares cada una de las observaciones tiene un número distinto de medidas de frecuencias cardíacas. Por ejemplo, una observación con el clip de «Scream» va a tener  $\frac{215\text{segundos}}{5\frac{\text{segundos}}{\text{medida}}} = 43$  medidas mientras que el de «El exorcista» tendrá  $\frac{100\text{segundos}}{5\frac{\text{segundos}}{\text{medida}}} = 20$  medidas. Además, ya se ha comentado en la subsección 3.1 que de manera aleatoria los intervalos entre mediciones son de más de 5 segundos, por lo que observaciones con el mismo clip pueden tener distinto número de medidas.

Se ha utilizado un filtro que elimina las observaciones que no tengan un número mínimo de medidas  $med_{min}$ . Cuantas más medidas tenga una observación más información proporciona al modelo, pero a la vez si el modelo se entrena con pocos observaciones el resultado es menos significativo. Para comprobar como afecta el número mínimo de medidas se ha probado con varios valores. En la subsección 5.1 se discute en detalle.

## 4.5. Obtención de características

Una vez realizadas estas transformaciones el siguiente paso es extraer las características de las series temporales de cada una de las observaciones. En este apartado se explicará cómo se han calculado algunas de ellas. Las características elegidas han sido las mismas que en Giannakos et al. (2020). Esto se debe a los buenos resultados obtenidos en Giannakos et al. (2020) y a la similitud entre su trabajo y este (ambos parten de frecuencias cardíacas). Las características extraídas son:

- Parámetros que caracterizan una distribución: el máximo, el mínimo, la media, la mediana, la desviación típica, la asimetría y la curtosis.
- La frecuencia fundamental de la serie temporal.
- El valor cuadrático medio (*root mean square*) de la serie temporal.
- Los 5 primeros coeficientes de autocorrelación.

### 4.5.1. Parámetros que caracterizan una distribución

La mayoría de parámetros en esta categoría no requieren explicación: el máximo, el mínimo, la media, la mediana, la desviación típica son estadísticos sencillos por todos conocidos. Por el contrario, al ser la asimetría y la curtosis menos habituales se van a explicar brevemente.

La asimetría es una medida de lo asimétrica que es una distribución. Si la asimetría de una distribución es 0 es que es perfectamente simétrica. Una asimetría positiva indica que la distribución está «inclinada» hacia la izquierda, y una asimetría negativa que está inclinada hacia la derecha. la asimetría de una variable aleatoria se calcula como  $asimetría = \frac{3 * media - mediana}{desviación}$ .

La curtosis es una estadística que determina en qué medida los valores de una distribución están concentrados en la zona central. Cuanto más concentrados mayor será la curtosis. Para calcularla se ha utilizado el *scipy*, que calcula la curtosis de Fisher.

### 4.5.2. Frecuencia fundamental de la serie temporal

La transformada de Fourier se utiliza para transformar una señal del dominio temporal al dominio de frecuencias o viceversa. Por ejemplo, si se aplica la transformada a una onda de sonido se obtendría una función que nos indica la intensidad de cada una de las frecuencias de la onda. Formalmente, la transformada de Fourier se define como  $\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i \xi x} dx$ . Una vez realizada la transformada, la frecuencia fundamental se define como el máximo de la parte real de  $\hat{f}(\xi)$ . Puesto que las series temporales son un conjunto de valores discretos se les ha aplicado la transformada discreta de Fourier, concretamente la proporcionada por la librería *scipy*.

### 4.5.3. Autocorrelación

Intuitivamente se puede definir la correlación como el grado en que los valores de una serie temporal determinan los siguientes. De manera más formal, es la correlación de una serie temporal con una copia de si misma pero con atrasada  $n$  pasos. Se ha utilizado *statmodels* para calcular los 5 primeros coeficientes de autocorrelación, es decir, las autocorrelaciones para  $n \in [1, 5]$ . Esto implica que cada observación debe tener al menos una serie temporal de 5 puntos, por lo que no todas las observaciones son útiles.

### 4.5.4. Valor cuadrático medio

El valor cuadrático medio de un conjunto de  $N$  valores viene dado por la fórmula  $RMS = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_N^2}{N}}$  (Meza (1989)).

# Capítulo 5

## Regresión lineal

Como se ha visto en la sección 2.2, existen varios modelos de aprendizaje supervisado entre los que elegir. Para empezar se decidió utilizar un modelo de regresión, ya que las variables objetivo, la valencia y el *arousal*, son numéricas. De entre los modelos de regresión, se ha comprobado que uno tan simple como la regresión lineal tiene bastante éxito prediciendo ambas dimensiones emocionales (Nogueira et al. (2013)).

Además de por los buenos resultados obtenidos, se ha elegido utilizar la regresión lineal por mínimos cuadrados por las siguientes razones:

- Al ser un modelo muy simple se introduce menos error debido a la varianza de los datos. Nuestro *dataset* es pequeño comparado con otros con los que se suele trabajar, solo tiene 180 observaciones. Por lo tanto un modelo complejo como una red neuronal tendría mucho *overfitting*.
- Usando el *F-statistic* se puede comprobar si hay relación entre las variables explicativas y la variable respuesta. Aunque la regresión no se ajuste bien a los datos y nivel de error sea grande, gracias al *F-statistic* se puede saber si por lo menos hay alguna relación entre las medidas de la pulsera fitbit y el estado emocional.
- La regresión lineal requiere poco tiempo de cálculo. Esto permite realizar multitud de pruebas para afinar los parámetros del procesamiento de datos (apartado 5.1) o para seleccionar las variables más importantes (apartado 5.2) en relativamente poco tiempo.

A continuación se explican las distintas pruebas que se han llevado a cabo con la regresión lineal y sus resultados.

### 5.1. Afinación de parámetros del procesado de datos

En la sección 4 se ha hablado de 2 parámetros que afectan al procesado de datos, el factor de suavizado  $f_s$  y el mínimo de medidas por observación  $med_{min}$ . Puesto que a priori no se sabía cómo

escoger estos parámetros se hicieron pruebas con varios valores.

Para empezar, se descubrió que  $med_{min}$  debía ser al menos 5. El motivo es que si una serie tiene menos de 5 puntos, no se pueden calcular los 5 coeficientes de correlación. También se comprobó que apenas había series con más de 35 puntos, por lo que se estableció  $med_{min} = 35$  como cota superior. Una vez establecidos un máximo y un mínimo, se decidió probar todos los valores de  $med_{min}$  con incrementos de 5, es decir; 5, 10, 15, etc. En cuanto a  $f_s$ , su valor mínimo es 1, que equivale a no aplicar suavizado. En principio se podría experimentar con todos los valores de  $f_s$  mayores que 1, pero para mantener el tiempo de computación en un margen razonable se decidió probar todos los valores del 1 al 15.

Al haber 7 posibles valores de  $med_{min}$  y 15 de  $f_s$ , hay  $7 * 15 = 105$  posibles combinaciones. Se hizo el procesamiento de datos para cada una de estas combinaciones, por lo que se generaron 105 *datasets* de variables explicativas. Con cada uno de estos *datasets* se entrenaron dos modelos, uno con la valencia como variable objetivo y otro con el *arousal*.

Para comparar los distintos modelos entre sí, se ha usado el p-valor de la *F-statistic*. Se ha representado la mediana de los p-valores en función de  $med_{min}$  y de  $f_s$  tanto para los modelos que predicen el *arousal* como para los que predicen la valencia.

En la figura 5.1 se ha representado la mediana de los p-valores en función de  $med_{min}$ . A la izquierda están los p-valores de cuando la variable objetivo es la valencia, y a la derecha cuando la variable objetivo es el *arousal*. En ambas figuras se observa que los p-valores son menores para  $med_{min} = 25$ . Esto indica que las observaciones con menos de 25 medidas no son significativas a la hora de predecir emociones. Por otro lado, para valores de  $med_{min}$  mayores de 25 se tienen menos observaciones por lo que los resultados son menos significativos. Cabe destacar que solo 49 de las 180 observaciones originales tienen 25 o más medidas.

En la figura 5.2 se ha representado la mediana de los p-valores en función de  $f_s$ . De nuevo a la izquierda están los valores relativos a la valencia y a la derecha los relativos al *arousal*. En este caso las gráficas tienen varios mínimos relativos, y el valor óptimo de  $f_s$  no es el mismo en ambas. Para la valencia la mediana toma su valor mínimo en  $f_s = 13$ , mientras que para el *arousal* lo toma en  $f_s = 8$ . Esto puede indicar que los resultados no son realmente significativos sino fruto del azar.

Tras realizar el análisis, se decidió fijar el parámetro  $med_{min}$  en 25. Las gráficas de la figura 5.1 indican que este valor es un buen compromiso entre el número de observaciones y su relevancia. Este resultado también apunta a que en futuros trabajos se debe intentar trabajar con series temporales de al menos 25 puntos. Para  $f_s$ , se decidió fijar su valor en 13 en los modelos de valencia, y en 8 en los de *arousal*. En retrospectiva, esta decisión no fue acertada. Los p-valores parecen variar aleatoriamente al variar  $f_s$  y no se ha encontrado ninguna explicación teórica de por qué estos valores de  $f_s$  son mejores que otros. Probablemente  $f_s$  se ha sobreajustado para estos *datasets* en concreto.

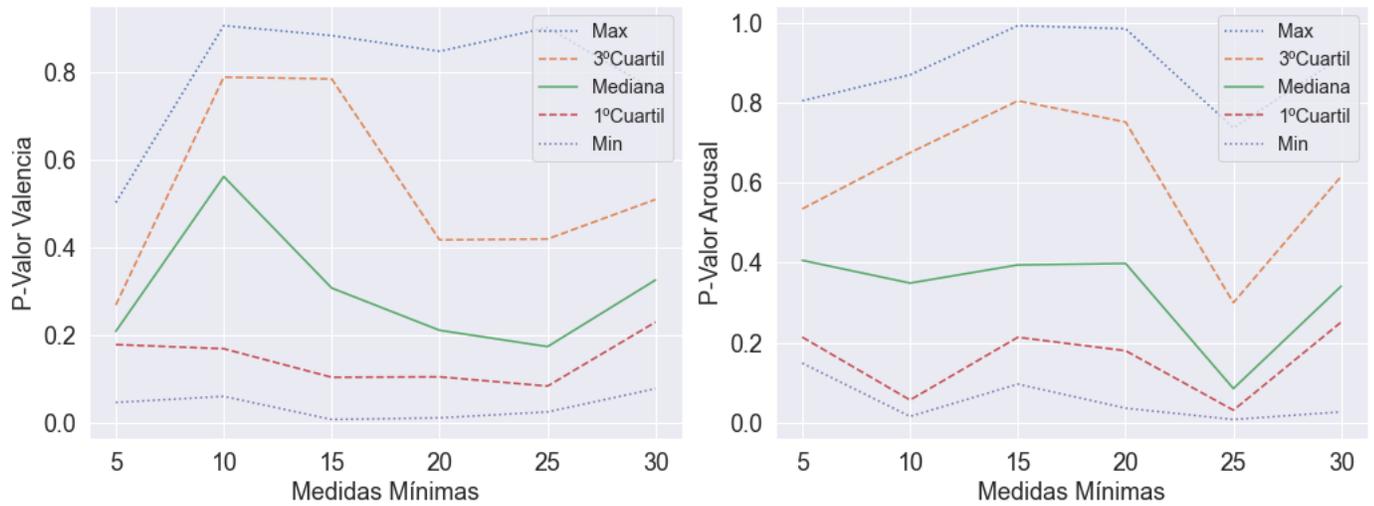


Figura 5.1: Izquierda: Variación del p-valor del modelo de la valencia en función de  $med_{min}$ . Al haber varios valores para cada  $med_{min}$  se han representado la mediana los cuantiles 0.25 y 0.75 los máximos y los mínimos. Derecha: Variación del p-valor del modelo de el *arousal* en función de  $med_{min}$ .

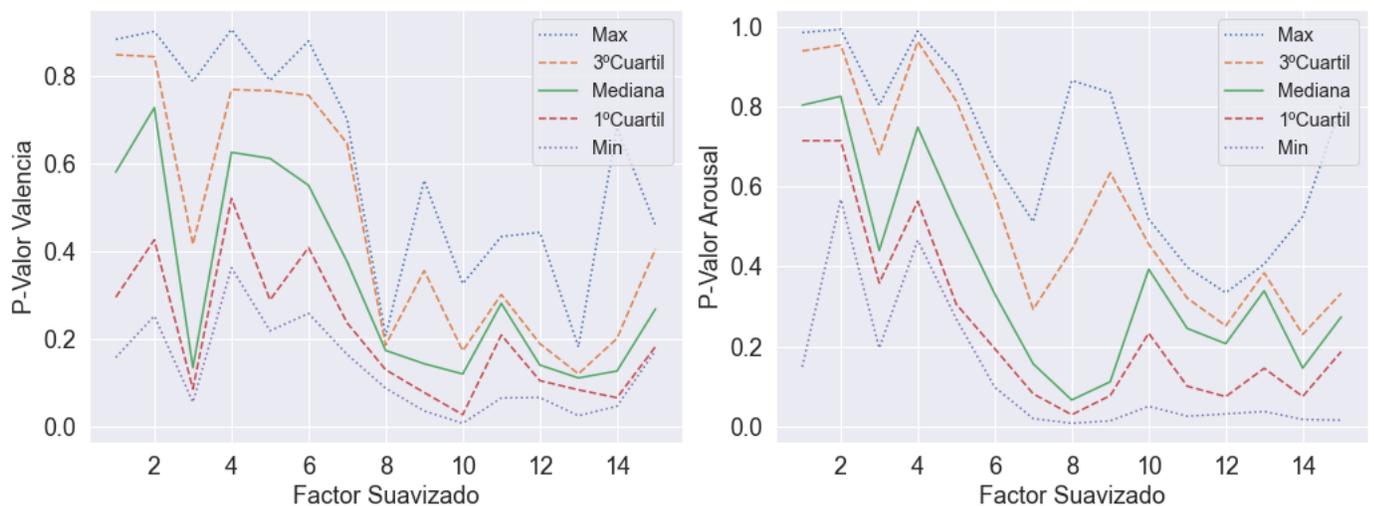


Figura 5.2: Izquierda: Variación del p-valor del modelo de la valencia en función de  $f_s$ . Al haber varios valores para cada  $f_s$  se han representado la mediana los cuantiles 0.25 y 0.75 los máximos y los mínimos. Derecha: Variación del p-valor del modelo de el *arousal* en función de  $f_s$ .

## 5.2. Selección de variables

Nuestro dataset contiene 14 variables explicativas. Para saber cuales de ellas tienen relación con el *arousal* y la valencia se van a probar modelos con distintas combinaciones de ellas. El número de posibles combinaciones es  $2^{14} = 16384$ . Cuanto mayor es el espacio de búsqueda mayor es la probabilidad de encontrar resultados que parecen significativos con nuestros datos, pero que no son generalizables. Por lo tanto se requiere de un método que no examine absolutamente todas las combinaciones de variables.

Se ha decidido utilizar *stepwise forward selection* por ser fácil de implementar y porque solo evalúa  $1 + \frac{p*(p+1)}{2}$  modelos, que para  $p = 14$  es 105. *Stepwise forward selection* consiste en empezar con un modelo sin variables explicativas. En cada paso se evalúa cómo mejora el modelo al añadir cada una de las variables y se selecciona la mejor. Este proceso se repite hasta que se han añadido todas las variables. Al final del proceso se dispone de un número de modelos igual al número de variables para comparar entre sí. La métrica elegida para evaluar los distintos modelos es  $R^2$  ajustado.

## 5.3. Evaluación resultados

### 5.3.1. Valencia

En el cuadro 5.1 se muestran los resultados de aplicar *stepwise forward selection* en el modelo de regresión para la valencia. El modelo con mejor resultado es el que contiene las variables *freq1*, *curtosis*, *autocorr3*, *skew*, *median*, *mean*, *stdev*, *autocorr5*, *autocorr4* y *autocorr2*. Este modelo tiene un  $R^2$  ajustado de 0.327 y el p-valor de su *F-Statistic* es 0.00345.

Atendiendo a los p-valores asociados a cada variable, solo hay tres que estén por debajo de 0.05: *freq1* ( $pvalue = 0,001$ ), *mediana* ( $pvalue = 0,019$ ) y la *media* ( $pvalue = 0,022$ ).

En la figura 5.3 se ha representado a la izquierda los valores reales de la valencia frente a los predicho por el modelo. La línea recta representa los puntos en los que el valor real es igual al predicho. Esta gráfica nos sirve para comprobar a simple vista si hay *outliers* o puntos con *high leverage*.

En cuanto a los *outliers* se observa que hay dos puntos que están significativamente más alejados de la recta que los demás. Eliminando esos dos *outliers* del modelo se obtendría una notable mejora en los resultados.  $R^2$  ajustado se incrementaría de 0.327 a 0.423 y el p-valor del *F-Statistic* disminuiría de 0.00345 a 0.000481. Sin embargo, no está justificado quitar estos puntos, ya que no hay nada que indique que son diferentes del resto.

Los puntos con *high leverage* son aquellos en los que la variable objetivo tiene un valor inusual. Pueden suponer un problema debido a que estos puntos tienen un impacto mucho mayor que el resto en el resultado de la regresión. Visualmente no se aprecian puntos con un *leverage* demasiado alto.

En la derecha de la figura 5.3 se ha representado los residuos frente a las predicciones del modelo.

Última variable añadida	$R^2_{ajustado}$
freq1	0.0985
curtosis	0.157
autocorr3	0.228
skew	0.269
median	0.276
mean	0.303
stdev	0.317
autocorr5	0.307
autocorr4	0.312
autocorr2	0.327
autocorr1	0.316
min	0.307
rms	0.291
max	0.272

Tabla 5.1: Resultado del *stepwise forward selection* en el modelo de la valencia.

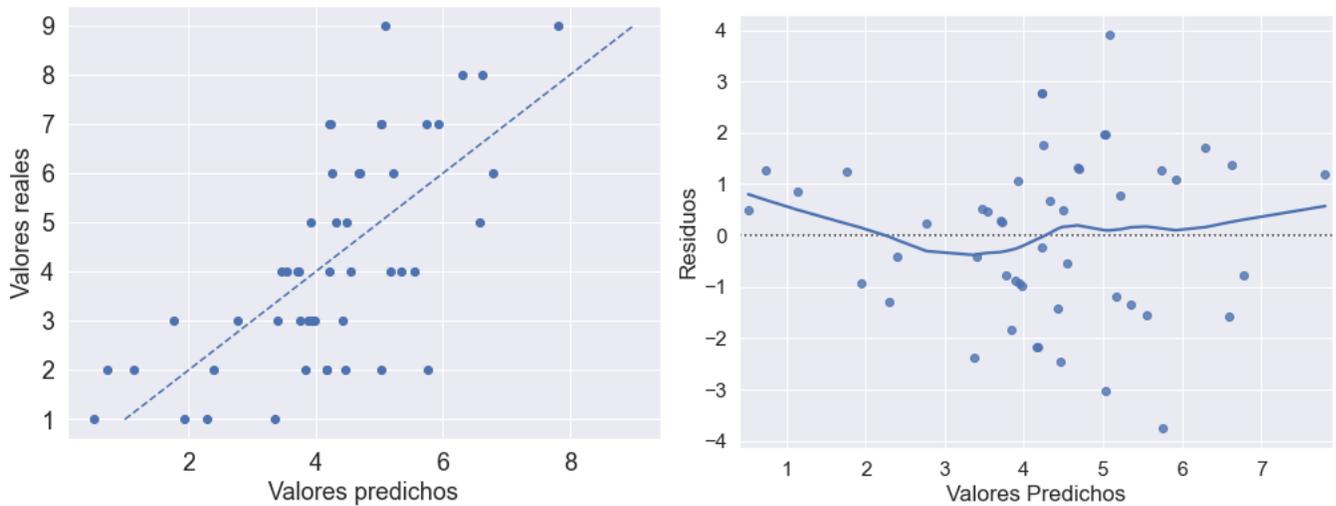


Figura 5.3: Izquierda: Valores reales de la valencia frente a los valores predichos por el modelo de regresión lineal. La línea discontinua marca dónde los valores reales son igual a los predichos. Derecha: Residuos de la regresión lineal frente a los valores predichos. La curva azul es un ajuste a los puntos que permiten ver si hay alguna tendencia.

Esta gráfica puede servir para identificar 3 problemas: la falta de linealidad en los datos, falta de homocedasticidad, y la correlación en los residuos.

La falta de linealidad en los datos se refiere al hecho de que la relación entre las variables explicativas y las objetivo no es lineal. Esto es un problema ya que el modelo de regresión lineal, como su nombre indica, asume una relación lineal. Cuando esto ocurre puede detectarse si en la gráfica de los residuos hay algún patrón o tendencia. En nuestra gráfica no hay ningún patrón evidente, por lo que no hay pruebas de falta de linealidad, aunque con un número tan reducido de puntos podría ser difícil ver un patrón aunque lo hubiera.

La homocedasticidad es una propiedad de una regresión que implica que la varianza del error es constante a lo largo de las observaciones. La falta de homocedasticidad provoca que las observaciones en el gráfico de residuos tengan forma de embudo, es decir que las observaciones están más dispersas cuanto más crece el valor predicho. En nuestro gráfico se observa que para valores pequeños del valor predicho hay menos varianza en el error. Aun así es difícil decir si esto se debe a la falta de homocedasticidad o a que hay muy pocas observaciones cuyo valor predicho sea menos que 3.

Queda comprobar si hay correlación en los residuos, ya que la regresión lineal parte de la suposición de que no la hay. En la gráfica no se observa ninguna correlación. Además, calculando los 5 primeros coeficientes de correlación, se obtiene que ninguno de ellos es mayor que 0.25 en valor absoluto. Se puede concluir por lo tanto que no hay correlación en los residuos.

Por último se va a estudiar si hay colinealidad. La colinealidad se da cuando dos o más variables están relacionadas entre sí, de manera que la información que aportan al modelo es redundante. La presencia de colinealidad es un problema porque enmascara la importancia real de las variables explicativas.

Para medir la colinealidad se puede usar el *variance inflation factor* (VIF). Una variable está contribuyendo a la colinealidad si su VIF está por encima de 10. En el cuadro 5.2 se puede ver que la mayoría de variables tienen un VIF desmesuradamente grande. En la figura 5.4 se ha representado un *heatmap* con las correlaciones entre las variables explicativas y se puede observar que muchas de ellas están altamente correlacionadas.

Para eliminar la colinealidad por completo, es necesario un modelo en el que las variables explicativas sean *freq1*, *stdev*, *curtosis*, *mean* y *autocorr3*. Calculando el VIF de las variables se ve que ninguno es mayor que 4. Sin embargo al hacer la regresión con estas variables los resultados son peores que antes.  $R^2$  ajustado se disminuye de 0.327 a 0.212 y el p-valor del F-Statistic aumenta de 0.00345 a 0.00858.

### 5.3.2. *Arousal*

A continuación se repetirá el mismo análisis que se ha hecho para el modelo de la valencia pero para el del *arousal*.

En el cuadro 5.3 se muestran los resultados de aplicar *stepwise forward selection* en el modelo de

VIF	Variable
10481.9	autocorr3
8200.1	autocorr4
2078.2	autocorr2
1894.4	autocorr5
171.3	median
161.5	mean
22.1	skew
4.7	curtosis
4.5	stdev
2.0	freq1

Tabla 5.2: VIF de las distintas variables explicativas para el modelo de la valencia

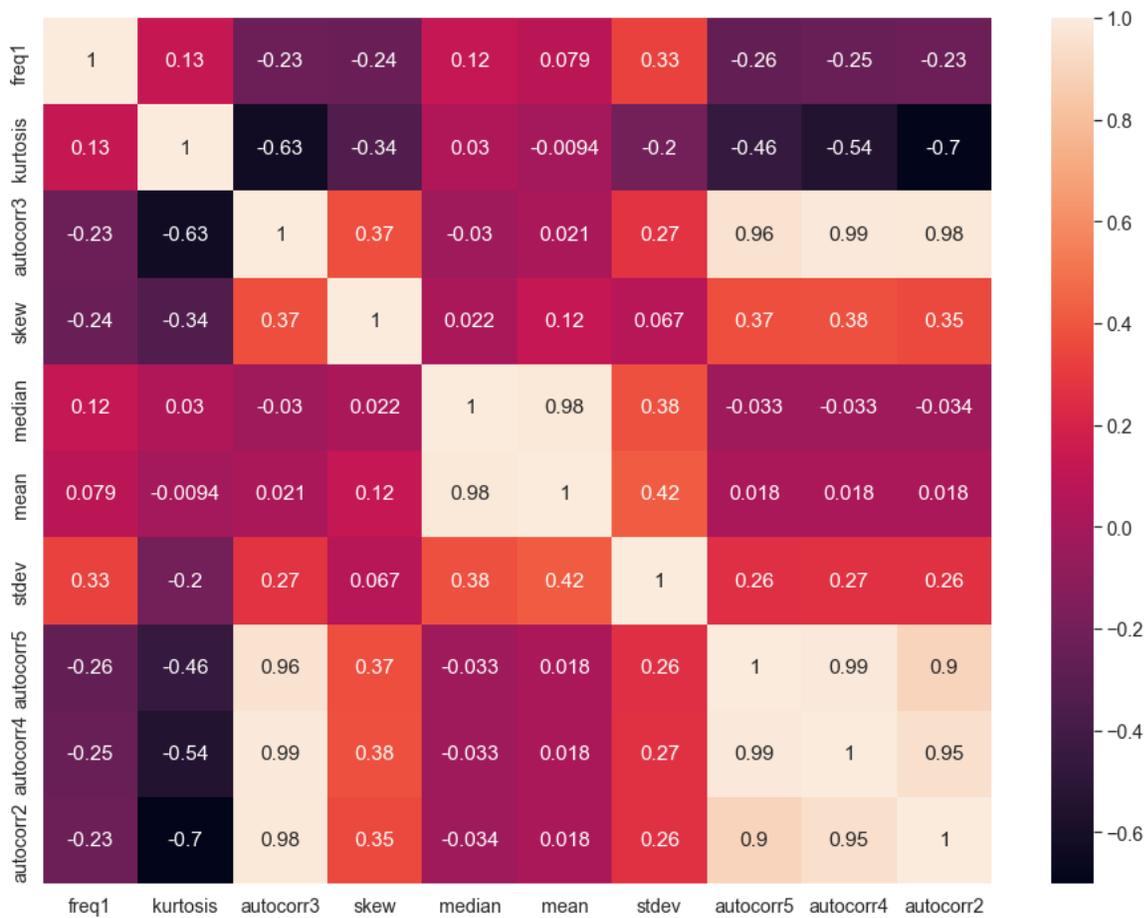


Figura 5.4: Heatmap con las correlaciones entre las variables explicativas para el modelo de la valencia

Última variable añadida	$R^2$ ajustado
autocorr1	0.225
autocorr2	0.300
autocorr5	0.362
freq1	0.387
autocorr3	0.398
skew	0.402
autocorr4	0.401
stdev	0.398
rms	0.412
curtosis	0.398
min	0.382
median	0.374
mean	0.359
max	0.342

Tabla 5.3: Resultado del *stepwise forward selection* en el modelo de el *arousal*

regresión para el *arousal*. El modelo con mejor resultado es el que contiene las variables autocorr1, autocorr2, autocorr5, freq1, autocorr3, skew, autocorr4, stdev y rms. Este modelo tiene un  $R^2$  ajustado de 0.412 y el p-valor de su *F-Statistic* es 0.000279. Atendiendo a los p-valores asociados a cada variable, ninguno de ellos está por debajo de 0.05.

En la figura 5.5 se ha representado a la izquierda los valores reales de la valencia frente a los predicho por el modelo. En este caso, no se observa que ningún punto sea claramente un *outlier* ni que tenga *high leverage*.

En la figura 5.5 a la derecha se han representado los residuos frente a los valores predichos por el modelo. No se observa ninguna tendencia que indique que la relación entre las variables explicativas y objetivo no sea lineal.

Sí que se observa que la varianza de los residuos parece disminuir con el valor del *arousal* predicho. Esto puede indicar la presencia de heterocedasticidad, aunque al haber tan pocos datos con valor predicho  $> 8$  es difícil decirlo.

Se han calculado los 5 primeros coeficientes de autoregresión y ninguno es mayor en valor absoluto que 0.25, por lo se puede concluir que los residuos no están correlacionados.

Por último, se han calculado los VIF de las variables para ver si hay colinealidad. En el cuadro 5.4 se pueden ver los valores, que de nuevo son extremadamente altos, por lo que si hay colinealidad. En la figura 5.6 se ha representado en un *heatmap* las distintas correlaciones entre variables.

Para eliminar la colinealidad es necesario que las únicas variables del modelo sean freq1, skew, rms. De esta manera ninguno de los VIF es mayor que 3. El problema es que este modelo obtiene resultados bastante pobres con un  $R^2$  ajustado de 0.012 y el p-valor de la *F-Statistic* de 0.322.

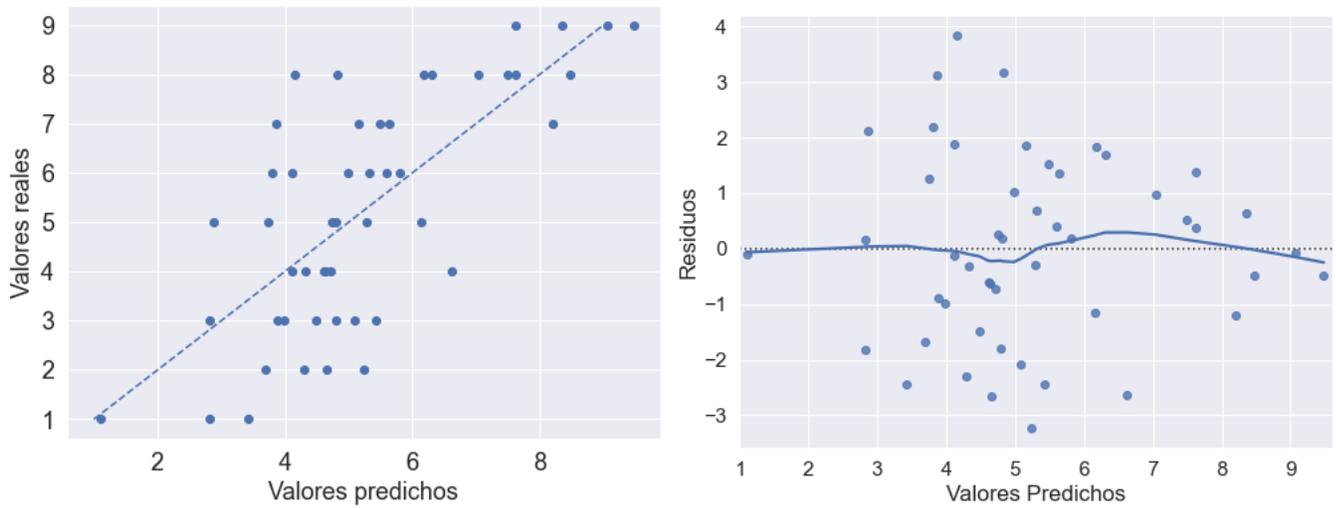


Figura 5.5: Izquierda: Valores reales de el *arousal* frente a los valores predichos por el modelo de regresión lineal. La línea discontinua marca dónde los valores reales son igual a los predichos. Derecha: Residuos de la regresión lineal frente a los valores predichos. La curva azul es un ajuste a los puntos que permiten ver si hay alguna tendencia.

Variables	VIF
autocorr1	2144.8
autocorr2	13565.2
autocorr5	1699.3
freq1	2.0
autocorr3	18321.2
skew	9.6
autocorr4	8944.2
stdev	8.8
rms	6.6

Tabla 5.4: VIF de las distintas variables explicativas para el modelo de el *arousal*

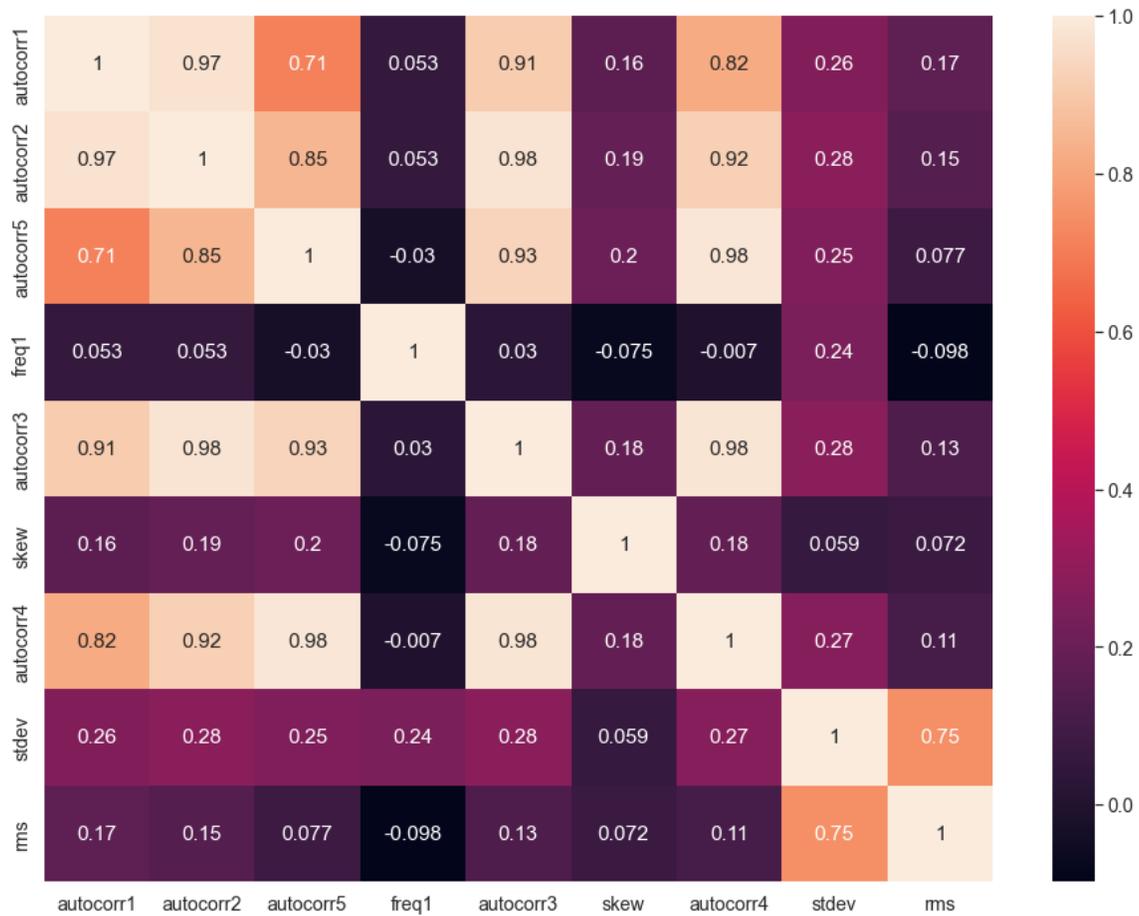


Figura 5.6: *Heatmap* con las correlaciones entre las variables explicativas para el modelo de el *arousal*

### 5.3.3. Evaluación resultados

El mejor modelo para la valencia ha obtenido un  $R^2$  ajustado de 0.327, y el mejor modelo para el *arousal* un  $R^2$  ajustado de 0.412. Estos resultados muestran que los modelos no son en absoluto precisos, sobretodo tras compararlos con los modelos de Nogueira et al. (2013) que alcanzaron  $R^2$  ajustado de 0.88 y 0.9 para el *arousal* y la valencia respectivamente.

Por otro lado, los p-valores de los modelos son ambos menores que 0.01: 0.00345 para la valencia y 0.000279 para el *arousal*. Esto podría indicar que aunque los modelos no predicen el estado emocional con precisión, por lo menos confirman una relación entre los datos de la pulsera y las emociones. Sin embargo, como se ha comentado en 5.1, es probable que se haya sobreajustado el parámetro  $f_s$  para este dataset, por lo que los resultados no son concluyentes.

En definitiva, no se ha conseguido predecir las emociones a partir de los datos de la pulsera, y no se ha demostrado que exista relación.

## 5.4. Análisis por clip

Una vez examinados los resultados del modelo de regresión lineal cabe preguntarse por qué algunas observaciones se ajustan mejor que otras a la recta de regresión. Un motivo podría ser que algunos clips son más eficaces a la hora de elicitar emociones que puedan ser captadas por la pulsera fitbit. También es posible que alguna de las emociones elicítadas produzcan una mayor respuesta que otras en el sistema cardiovascular. Si esto fuera cierto, examinar cuales son esos clips y qué características comunes tienen puede servir a la hora de diseñar futuros experimentos.

Para realizar este análisis se va a estudiar la relación entre el clip de la observación y el valor absoluto del residuo. Si la media de los residuos para un clip es significativamente mayor que para el resto se podría concluir que este clip no es muy útil para este tipo de observaciones. Ocurriría lo contrario si la media de los residuos es baja.

Puesto que se quiere estudiar la relación entre una variable nominal (el clip) y una respuesta continua (el valor absoluto del residuo) se va a utilizar el test de Kruskal-Wallis. Se ha elegido este test en vez de el de ANOVA de un factor porque este último requiere que la variable respuesta sea normal, condición que los valores absolutos de los residuos no cumplen.

La hipótesis nula del test de Kruskal-Wallis es que la medias de la variable respuesta es igual para todos los grupos, y la hipótesis alternativa es que al menos una de las medias es distinta.

Primero se examina el modelo de la valencia. En el cuadro 5.5 se muestra el promedio de los valores absolutos de los residuos para cada uno de los clips. El test de Kruskal-Wallis resulta en un estadístico con un p-valor de 0.150, que al ser mayor que 0.05 no permite rechazar la hipótesis nula. Esto indica que no hay diferencias significativas en los resultados de cada clip.

Para el modelo del *arousal* se muestra en el cuadro 5.6 el promedio de los valores absolutos de los residuos para cada uno de los clips. El test de Kruskal-Wallis resulta en un estadístico con un p-valor

Clip	Promedio valor absoluto residuos	# observaciones
Algo pasa con Mary	1.97	10
El campeón	2.47	1
City of angels	1.16	17
Scream	1.16	13
Un pez llamado Wanda	1.14	8

Tabla 5.5: Promedio de el valor absoluto de los residuos para cada clip para el modelo de la valencia. También se indica el número de observaciones con cada clip

Clip	Promedio valor absoluto residuos	# observaciones
Algo pasa con Mary	1.09	10
El campeón	0.711	1
City of angels	1.71	17
Scream	1.12	13
Un pez llamado Wanda	1.27	8

Tabla 5.6: Promedio de el valor absoluto de los residuos para cada clip para el modelo de el *arousal*. También se indica el número de observaciones con cada clip

de 0.287, que de nuevo no permite rechazar la hipótesis nula. En este caso tampoco hay evidencias suficientes de que el clip influya en el residuo.

## 5.5. Análisis por persona

En esta sección se van a examinar las diferencias en los resultados para los distintos participantes. Los resultados entre participantes pueden variar por distintos motivos: la reacción fisiológica a las emociones no es igual en todo el mundo, algunas personas son mejores autoevaluando sus emociones, etc. Aunque es imposible saber a priori qué participantes van a dar buenos resultados en un experimento como este, si que se pueden descartar datos a posteriori si alguno de los participantes ha producido resultados especialmente malos.

De nuevo se va a utilizar el test de Kruskal-Wallis para ver si la media de los valores absolutos de los residuos varía significativamente entre los participantes. En el cuadro 5.7 se muestran los valores para el modelo de la valencia. El test de Kruskal-Wallis resulta en un estadístico con un p-valor de 0.250, por lo que no se puede rechazar la hipótesis nula y se concluye que no hay diferencias por participante.

En el cuadro 5.8 se muestran los resultados para el modelo del *arousal*. El test de Kruskal-Wallis resulta en un estadístico con un p-valor de 0.564, por lo que no se rechaza la hipótesis nula y se concluye que no hay diferencias por participante.

Id del participante	Promedio valor absoluto residuos	# observaciones
1	2.08	2
2	1.07	1
3	0.581	4
4	1.22	2
5	0.862	2
6	1.19	4
7	0.982	3
8	0.816	2
9	2.17	2
10	0.410	2
11	2.27	4
12	1.50	3
13	1.35	3
15	1.72	5
16	1.38	3
17	1.74	3
18	1.09	4

Tabla 5.7: Promedio de el valor absoluto de los residuos para cada participante para el modelo de la *valencia*. También se indica el número de observaciones de cada participante

Id del participante	Promedio valor absoluto residuos	# observaciones
1	1.33	2
2	2.43	1
3	1.28	4
4	0.541	2
5	2.23	2
6	1.41	4
7	2.42	3
8	1.30	2
9	1.81	2
10	0.945	2
11	0.913	4
12	1.82	3
13	0.702	3
15	0.819	5
16	1.78	3
17	0.782	3
18	1.38	4

Tabla 5.8: Promedio de el valor absoluto de los residuos para cada participante para el modelo de el *arousal*. También se indica el número de observaciones de cada participante



# Capítulo 6

## Conclusión

En esta sección se presentan de forma resumida los puntos más importantes del trabajo y las conclusiones a las que se han llegado tras realizarlo.

El principal objetivo de este proyecto era investigar si la pulsera inteligente Fitbit Sense puede usarse para detectar emociones. Se comenzó diseñando una experiencia en la que se elicitaban emociones a participantes voluntarios mientras llevaban la pulsera y autoevaluaban su estado emocional. Esto ha permitido la creación de un dataset consistente en series temporales de la frecuencia cardíaca etiquetadas en base a 2 dimensiones emocionales, la valencia y el *arousal*.

Se aplicó una serie de transformaciones a las series temporales y se extrajeron varias características de ellas imitando los métodos utilizados en Giannakos et al. (2020). Estas características fueron utilizadas como *input* en varios modelos de regresión lineal múltiple. Con estos modelos de regresión se afinaron los parámetros  $f_s$  y  $med_{min}$  de el procesado de datos y se eligieron los modelos con las mejores variables usando *stepwise forward selection*. También se estudió si había diferencias en lo bien que ajustaba el modelo para distintos clips y para distintas personas.

Tras analizar los resultados, se ha concluido que no se ha conseguido hacer un modelo que prediga las emociones en función de los datos proporcionados por la pulsera Fitbit Sense. Los mejores resultados obtenidos han sido un  $R^2$  ajustado de 0.327 al predecir la valencia, y 0.412 para predecir el *arousal*, que aunque son mejor que una predicción aleatoria, dejan mucho que desear en términos de precisión.

El haber conseguido un  $R^2$  ajustado mayor que 0 y que los p-valores de ambos modelos sean menores que 0.01 podría indicar que por lo menos existe una relación entre las medidas y el *arousal* y la valencia. Sin embargo, al haber afinado  $f_s$  para que los modelos tuvieran un p-valor bajo, hay una gran probabilidad de que este parámetro se haya sobreajustado y de que en realidad los resultados no sean significativos. Por lo tanto, no se ha llegado a demostrar que exista relación entre las medidas tomadas por la pulsera y el estado emocional.

Por otro lado, si se ha comprobado que se requiere un número mínimo de puntos en cada serie temporal para que esta sea útil, concretamente 25. Dadas las características de la Fitbit Sense, que

mide las pulsaciones cada 5 segundos y pierde el 35 % de las medidas realizadas, esto se traduce en que los clips deben durar un mínimo de 170 segundos para ser útiles. Más allá de la duración, no se ha podido encontrar si hay otras características que hagan los clips mejores o peores para este tipo de trabajos.

Para futuros trabajos, para comprobar la utilidad de la pulsera Fitbit Sense se podría intentar algo más sencillo, como distinguir entre 3 o 4 emociones discretas en vez de ubicar una emoción concreta a lo largo de dos dimensiones emocionales. Una vez que se haya comprobado si la pulsera es capaz de diferenciar entre unas pocas emociones, se podrá intentar ser más preciso y distinguir entre un mayor número de emociones.

# Bibliografía

- Basu, S., Bag, A., M, M., Mukherjee, J., and Guha, R. (2015). Affect detection in normal groups with the help of biological markers. In *2015 Annual IEEE India Conference (INDICON)*, pages 1–6.
- Bynion, T.-M. and Feldner, M. T. (2017). *Self-Assessment Manikin*, pages 1–3. Springer International Publishing, Cham.
- Chang, C.-Y., Chang, C.-W., Zheng, J.-Y., and Chung, P.-C. (2013). Physiological emotion analysis using support vector regression. *Neurocomputing*, 122:79–87. Advances in cognitive and ubiquitous computing.
- Dzedzickis, A., Kaklauskas, A., and Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. *Sensors*, 20(3).
- Fernández, C., Pascual, J. C., Soler, J., Elices, M., Portella, M. J., and Fernández-Abascal, E. (2012). Physiological responses induced by emotion-eliciting films. *Applied Psychophysiology and Biofeedback*, 37(2):73–79.
- Fernández, C., Pascual, J. C., Soler, J., and García, E. (2011). Validación española de una batería de películas para inducir emociones. *Psicothema*, 23:778–785.
- Giannakos, M. N., Sharma, K., Papavlasopoulou, S., Pappas, I. O., and Kostakos, V. (2020). Fitbit for learning: Towards capturing the learning experience using wearable sensing. *International Journal of Human-Computer Studies*, 136:102384.
- Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., and Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, 48:108–119.
- Guo, H. W., Huang, Y. S., Chien, J. C., and Shieh, J. S. (2015). Short-term analysis of heart rate variability for emotion recognition via a wearable ecg device. In *2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, pages 262–265.

- Healey, J. and Picard, R. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166.
- Herráiz, C. I. (2022). Análisis de pulsera inteligente para la detección de estados afectivos mediante aprendizaje no supervisado sobre series temporales.
- Jang, E., Park, B., Kim, S., Chung, M., and Sohn, J. (2012). Classification of three emotions by machine learning algorithms using psychophysiological signals. *International Journal of Psychophysiology*, 85(3):402–403. PROCEEDINGS OF THE 16TH WORLD CONGRESS OF PSYCHOPHYSIOLOGY of the International Organization of Psychophysiology (IOP) Pisa, Italy September 13-17, 2012.
- Katsis, C. D., Katertsidis, N., Ganiatsas, G., and Fotiadis, D. I. (2008). Toward emotion recognition in car-racing drivers: A biosignal processing approach. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 38(3):502–512.
- Kim, K. H., Bang, S. W., and Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing*, 42(3):419–427.
- Lisetti, C. L. and Nasoz, F. (2004). Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP Journal on Advances in Signal Processing*, 2004(11):929414.
- Meza, M. M. (1989). Tópicos de estadística descriptiva y probabilidad. page 86.
- Nogueira, P. A., Rodrigues, R., Oliveira, E., and Nacke, L. E. (2013). A hybrid approach at emotional state detection: Merging theoretical models of emotion with data-driven statistical classifiers. In *Web Intelligence and Intelligent Agent Technology, IEEE/WIC/ACM International Conference on*, volume 2, pages 253–260, Los Alamitos, CA, USA. IEEE Computer Society.
- Rahman, M. M., Chowdhury, M. A., and Fattah, S. A. (2018). An efficient scheme for mental task classification utilizing reflection coefficients obtained from autocorrelation function of eeg signal. *Brain informatics*, 5:1–12.
- Rui Xu, C. Z. (2018). How physical activities affect mental fatigue based on eeg energy, connectivity, and complexity. *Front. Neurol.*
- Santos, O. C. and Boticario, J. G. (2014). Involving users to improve the collaborative framework. *The Scientific World Journal*.
- Santos, O. C., Boticario, J. G., Cabestrero, R., and Quirós, P. (2013). Extending collaborative learning modeling with emotional information.

- Sitnikova, E., Hramov, A. E., Koronovsky, A. A., and van Luijtelaar, G. (2009). Sleep spindles and spike-wave discharges in eeg: Their generic features, similarities and distinctions disclosed with fourier transform and continuous wavelet analysis. *Journal of neuroscience methods*, 180:304–16.
- Verma, G. K. and Tiwary, U. S. (2014). Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals. *NeuroImage*, 102:162–172. Multimodal Data Fusion.
- Wijasena, H. Z., Ferdiana, R., and Wibirama, S. (2021). A survey of emotion recognition using physiological signal in wearable devices. In *2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, pages 1–6.



# Anexo A

## Instrucciones API web de Fitbit

A continuación se muestra un documento elaborado durante la realización del trabajo, que indica cómo utilizar la API web de Fitbit Sense para descargar los datos obtenidos por la pulsera. También se muestra el código al que hacen referencia las instrucciones.

### Instrucciones

Para poder descargar datos recogidos por la pulsera fitbit, esta debe estar conectada con un dispositivo (smartphone, tableta u ordenador) que tenga la aplicación 'Fitbit'. Para conectar la pulsera a un dispositivo usa la siguiente guía

[https://help.fitbit.com/articles/en\\_US/Help\\_article/1873.htm?Highlight=add%20device](https://help.fitbit.com/articles/en_US/Help_article/1873.htm?Highlight=add%20device)

**IMPORTANTE!** Antes de hacer cualquier descarga de datos hay que asegurarse de que la pulsera se ha sincronizado con la aplicación del móvil en algún momento desde que se realizó la recogida de datos. Si no se ha hecho, los datos estarán en la pulsera pero no en la cuenta y no se podrán descargar.

Para realizar la sincronización basta con abrir la aplicación 'Fitbit' y arrastrar la pantalla hacia abajo (en dispositivos táctiles). Alternativamente, desde la aplicación se puede entrar en cuenta->sense->sincronizar ahora.

#### INSTRUCCIONES PARA DESCARGAR LOS DATOS DE FITBIT USANDO LA WEB API

- Primero hay que seguir los pasos del 1 al 4 de esta guía <https://towardsdatascience.com/using-the-fitbit-web-api-with-python-f29f119621ea>

- En el paso 2 te explican dónde están la OAuth 2.0 Client ID y el Client Secret, que necesitarás más adelante.

- Una vez completado el paso 3 ya no hace falta que sigas la guía. Ahora abre el archivo `descarga_datos.py`

- En él tendrás que escribir el OAuth 2.0 Client ID y el Client Secret como se te indica

- También tienes que escribir la fecha del día cuyos datos quieres. Si no hay datos de ese día el

dataframe resultante estará vacío.

- Ahora ejecuta el archivo `descarga_datos.py`
- Cuando ejecutes el archivo se debe abrir una pestaña en el navegador pidiéndote que te loguees con tu cuenta de fitbit. Una vez logueada ya puedes cerrar la pestaña.
- Si te sale el error "No module named 'gather\_keys\_oauth2'" prueba a copiar el archivo 'gather\_keys\_oauth2.py' (que está en la api fitbit que te has bajado 'python-fitbit-master') a la carpeta `/Anaconda3/Lib/site-packages`. Si estás usando un entorno, cópialo en la carpeta `/Lib/site-packages` del entorno.
- Si te sale el error 'No module named cherrypy' intenta instalar cherrypy con `pip install CherryPy`.
- Esos son los errores que me han salido a mí, si tienes otros la guía tienen al final una sección con errores comunes y posibles soluciones.
- Finalmente comprueba que se ha creado un fichero llamado `datos_piloto.csv` en la misma carpeta donde esté `descarga_datos.py`. Si se ha creado este fichero y tiene datos dentro da igual que salgan otros errores o warnings.

## Código `descarga_datos.py`

```
import gather_keys_oauth2 as OAuth2
import fitbit import pandas as pd
import datetime
# Escribe aquí tu OAuth 2.0 Client ID
CLIENT_ID = ''
# Y aquí el Client Secret
CLIENT_SECRET = ''

server = OAuth2.OAuth2Server(CLIENT_ID, CLIENT_SECRET)
server.browser_authorize()
ACCESS_TOKEN=str(server.fitbit.client.session.token['access_token'])

REFRESH_TOKEN=str(server.fitbit.client.session.token['refresh_token'])

auth2_client=fitbit.Fitbit(CLIENT_ID, CLIENT_SECRET, oauth2=True, /
access_token=ACCESS_TOKEN, refresh_token=REFRESH_TOKEN)

# Escribe aquí el año, mes y día del día que quieres obtener datos
year =
month =
```

```
day =
oneDate = datetime.datetime(year = year, month = month, day = day)
oneDayData = auth2_client.intraday_time_series('activities/heart', oneDate, /
detail_level='1sec')
df = pd.DataFrame(oneDayData['activities-heart-intraday']['dataset'])

df.to_csv('datos_piloto.csv')
```