

**UNIVERSIDAD EUROPEA DE MADRID**

**ESCUELA DE ARQUITECTURA, INGENIERÍA Y DISEÑO**

**GRADO EN INGENIERÍA MATEMÁTICA**

**APLICADA AL ANÁLISIS DE DATOS**

PROYECTO FIN DE GRADO

**AFFECTIVE COMPUTING: HERRAMIENTA DE ACOMPAÑAMIENTO BASADA EN EL ANÁLISIS DE EMOCIONES**

**MARTA ALMENDRO ÁLVAREZ**

**Dirigido por**

**ANA DEL VALLE CORRALES PAREDES**

**CURSO 2020-2021**

**TÍTULO**: AFFECTIVE COMPUTING: HERRAMIENTA DE ACOMPAÑAMIENTO BASADA EN EL ANÁLISIS DE EMOCIONES

**AUTOR**: MARTA ALMENDRO ÁLVAREZ

**TITULACIÓN**: GRADO EN INGENIERÍA MATÉMATICA APLICACA AL ANÁLISIS DE DATOS

**DIRECTOR/ES DEL PROYECTO**: ANA DEL VALLE CORRALES PAREDES

**FECHA**: JUNIO DE 2021

# RESUMEN

El resumen tiene entre 150-250 palabras. Resumir consiste en ofrecer información sobre cómo, dónde, cuándo y por qué se aplica el proyecto. Se realiza al finalizar el trabajo.

Debe incluir:

* Resumen del problema planteado y las aportaciones más importantes del proyecto al respecto
* Indica si procede si el proyecto se realizó en colaboración con una empresa, indicando nombre (siempre que haya autorización expresa por la empresa), y el sector industrial o empresarial
* Principales resultados del proyecto (diseño de un sistema para…, desarrollo e implementación de una solución para…, análisis de…, modelo de…)
* Resume las conclusiones más importantes del trabajo realizado

El resumen NO

* Da una información genérica
* Se refiere a datos aportados en el texto del proyecto.

**Palabras clave:** hasta un máximo de 6 conceptos (un concepto puede conllevar una o más palabras). Incluye tecnologías que hayas utilizado, conceptos relevantes del ámbito científico-técnico, y conceptos de la industria (algunos ejemplos: machine learning, Big Data, Smart City, visión artificial, social media, Twitter, computación afectiva, transformación digital, GDPR, DevOps, EEG, …)

# ABSTRACT

Resumen en inglés.

**Keywords:** Palabras clave en inglés

**AGRADECIMIENTOS**

En ocasiones se incluye este apartado para agradecer a aquellos que han ofrecido su ayuda en el desarrollo del trabajo, ya sea técnica o de otro tipo.

**Cita - frase célebre / Dedicatoria**

Esta página es del todo opcional, pero resulta una muy buena forma de presentar el trabajo académico más importante de todo el grado.

# TABLA RESUMEN

|  |  |
| --- | --- |
|  | DATOS |
| Nombre y apellidos: | Marta Almendro Álvarez |
| Título del proyecto: | Affective Computing: Herramienta de Acompañamiento Basada en el Análisis de Emociones |
| Directores del proyecto: | Ana del Valle Corrales Paredes |
| El proyecto ha implementado un producto:  (esta entrada se puede marcar junto a la siguiente) | SI |
| El proyecto ha consistido en el desarrollo de una investigación o innovación:  (esta entrada se puede marcar junto a la anterior) | SI |
| Objetivo general del proyecto: | Desarrollar una aplicación multi plataforma integrada con un sistema de reconocimiento de las emociones del usuario a través de su voz. |

**Índice**

[RESUMEN 6](#_Toc37845866)

[ABSTRACT 7](#_Toc37845867)

[TABLA RESUMEN 10](#_Toc37845868)

[Capítulo 1. RESUMEN DEL PROYECTO 15](#_Toc37845869)

[1.1 Contexto y justificación 15](#_Toc37845870)

[1.2 Planteamiento del problema 15](#_Toc37845871)

[1.3 Objetivos del proyecto 15](#_Toc37845872)

[1.4 Resultados obtenidos 15](#_Toc37845873)

[1.5 Estructura de la memoria 15](#_Toc37845874)

[Capítulo 2. ANTECEDENTES / ESTADO DEL ARTE 16](#_Toc37845875)

[2.1 Estado del arte 16](#_Toc37845876)

[2.2 Contexto y justificación 16](#_Toc37845877)

[2.3 Planteamiento del problema 16](#_Toc37845878)

[Capítulo 3. OBJETIVOS 17](#_Toc37845879)

[3.1 Objetivos generales 17](#_Toc37845880)

[3.2 Objetivos específicos 17](#_Toc37845881)

[3.3 Beneficios del proyecto 18](#_Toc37845882)

[Capítulo 4. DESARROLLO DEL PROYECTO 19](#_Toc37845883)

[4.1 Planificación del proyecto 19](#_Toc37845884)

[4.2 Descripción de la solución, metodologías y herramientas empleadas 19](#_Toc37845885)

[4.3 Recursos requeridos 19](#_Toc37845886)

[4.4 Presupuesto 19](#_Toc37845887)

[4.5 Viabilidad 20](#_Toc37845888)

[4.6 Resultados del proyecto 20](#_Toc37845889)

[Capítulo 5. DISCUSIÓN 21](#_Toc37845890)

[Capítulo 6. CONCLUSIONES 22](#_Toc37845891)

[6.1 Conclusiones del trabajo 22](#_Toc37845892)

[6.2 Conclusiones personales 22](#_Toc37845893)

[Capítulo 7. FUTURAS LÍNEAS DE TRABAJO 23](#_Toc37845894)

[Capítulo 8. REFERENCIAS 24](#_Toc37845895)

[Capítulo 9. ANEXOS 25](#_Toc37845896)

**Índice de Figuras**

Todas las figuras incluidas en la memoria deben estar referenciadas para incluir en este índice.

**Índice de Tablas**

Todas las tablas incluidas en la memoria deben estar referenciadas para incluir en este índice.

# RESUMEN DEL PROYECTO

## Contexto y justificación

El estudio de la relación entre las emociones y la salud en los humanos se remonta a la Antigüedad. En la antigua Grecia esta idea se sustentaba en el hecho de que el cuerpo y la mente debían estar en armonía. La filosofía “mens sana in corpore sano”, que se traduce a “mente sana en un cuerpo sano”, era uno de los pilares de la medicina hipocrática.

La comunicación hombre-máquina se ha vuelto cada vez más ‘habladora’: Alexa, Cortana, Siri [11], entre muchos otros sistemas inteligente de diálogo, han triunfado en el mercado de consumo de manera destacada, pero ¿y si estos realmente pudiesen notar nuestras emociones y reaccionar a ellas como un humano haría? Esto haría de la HMI una comunicación mucho más natural, efectiva y agradable. La disciplina de reconocimiento automático de las emociones humanas y estados afectivos a través de la voz es conocida como Speech Emotion Recognition (SER).

## Planteamiento del problema

El contacto físico hoy en día supone un peligro grave para muchos, por lo que para muchos el hogar es su día a día. Este día a día incluye el cuidado de la salud, de modo que, este proyecto quiere ayudar a crear una alternativa al contacto físico para proporcionar seguridad en momentos de mayor malestar emocional.

Este proyecto trata de conseguir una solución que pueda aliviar a los sistemas sanitarios, y al mismo tiempo incrementar el estado afectivo de los usuarios, que podrán mantener un estado físico y mental saludables desde la comodidad del hogar. Sin embargo, su uso puede ir más allá y puede ser de utilidad, por ejemplo, para trabajadores en oficinas.

## Objetivos del proyecto

Este proyecto propone una aplicación de reconocimiento de emociones a través de la voz, siguiendo la dinámica de la interacción hombre-maquina, y sumergiéndose en el campo del Affective Computing. Para ello se plantea desarrollar un algoritmo de reconocimiento, como estructura de red neuronal artificial, empleando diferentes estructuras y experimentando con diferentes parámetros. A su vez, se persigue construir una arquitectura front-end con una interfaz de usuario (UI) adecuada y un diseño que haga de la experiencia de usuario (UX) lo más dinámica y satisfactoria posible.

## Resultados obtenidos

Incluye el resumen de los resultados obtenidos al final de tu proyecto. En el capítulo correspondiente a resultados se detallan estos resultados.

## Estructura de la memoria

La estructura de la memoria corresponde a los siguientes apartados:

**Capítulo 1:** contiene un breve resumen del contexto del proyecto, el planteamiento del problema, los objetivos y los resultados,

**Capítulo 2:** pone en contexto el proyecto y se especifican aquellos aspectos que resguardan las técnicas estudiadas y aplicadas a través del estado del arte,

**Capítulo 3:** incluye una descripción detallada de los objetivos, general y específicos, y los beneficios del proyecto,

**Capítulo 4:** expone el desarrollo completo del proyecto, desde la planificación, las herramientas y recursos empleados, en este caso, mayoritariamente computacionales, hasta la viabilidad, el presupuesto estimado y los resultados obtenidos, además del componente de innovación,

**Capítulo 5:** plantea una discusión de los resultados obtenidos,

**Capítulo 6:** incluye las conclusiones obtenidas a raíz de los resultados,

**Capítulo 7:** propone las futuras líneas de trabajo consideradas tras la finalización del proyecto.

# ANTECEDENTES / ESTADO DEL ARTE

## Estado del arte

*Human-Machine Interaction*

Para poder emplear la tecnología en el campo de reconocimiento de emociones humanas, debe existir la comunicación entre un humano y una máquina. De aquí nace el concepto Interacción Hombre-Máquina (HMI), descrito como la interacción y comunicación entre un usuario humano y una máquina (sistema técnico dinámico), a través de una interfaz hombre-máquina [8]. Para esta última parte, hay que tener en cuenta tanto la User Interface (UI) y la User Experience (UX).

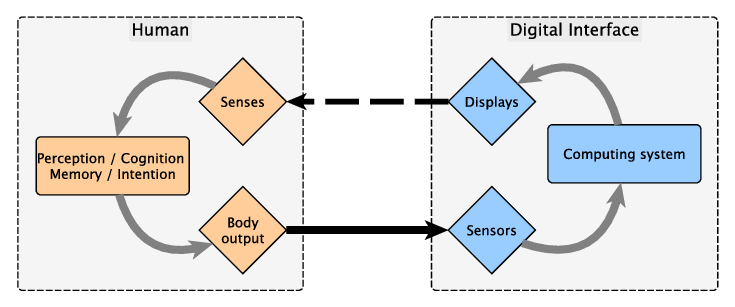


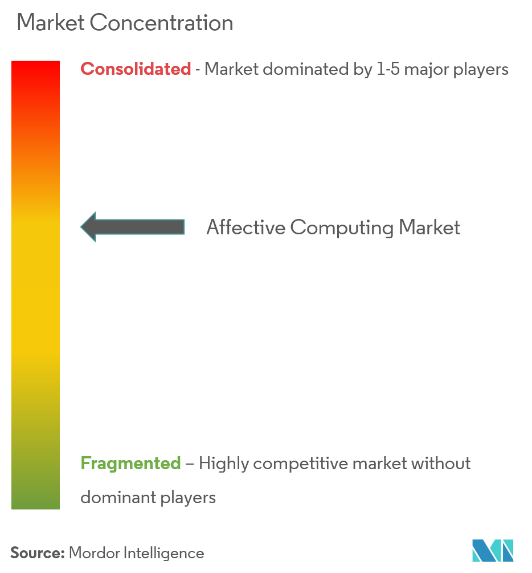
Figura 1: Sistema HMI: usuario humano, interfaz y máquina. Fuente: [8] [DIBUJOHMI].

*Affective Computing*

Una de las grandes fuentes de inspiración de este proyecto viene de la idea del Affective Computing (AC). Este término fue propuesto por Rosalind Picard en 1997 como la computación que se relaciona, emerge o influye en las emociones [6]. Esto lo convierte en una modalidad de computación que nos permite comunicarnos e interactuar con ordenadores, robots u otras tecnologías a través de nuestras emociones [9]. El AC persigue formar una base sólida en la investigación sobre la relación entre los estados afectivos, cognitivos y físicos del ser humano [10]. Un valor añadido de esta modalidad de computación es que se pueden acumular grandes cantidades de datos para el desarrollo de las ciencias cognitivas que servirán para el entrenamiento de tecnologías avanzadas que puedan, entre otras cosas, reforzar la capacidad de tomar decisiones. A través de tecnologías que monitorean y analizan las emociones, se pueden detectar amenazas a la salud tanto mental como física, incluso logrando reducir el problema de la detección de enfermedades tardías.

Según un informe de AllTheReasearch, los gigantes de la industria del AC son Google e IBM. Los productos se segregan entre hardware y software y entre sus más destacadas aplicaciones está la sanidad [11]. El segmento de la atención médica tiene algunas de las aplicaciones más avanzadas y comercializadas de AC. Empresas como DeepMind y Babylon Health están realizando importantes esfuerzos para aumentar la contribución al dominio [mercado]. La empresa Babylon lanzó en 2016 una aplicación de consultas médicas online; a través de un sistema de reconocimiento de voz y de inspección de historiales clínicos, ofrece una ruta de acción adecuada a la circunstancia médica en cuestión [14]. La contribución de Google en el sector es Google Health, una plataforma que promueve el descubrimiento de nuevas oportunidades en el campo de la AI con el fin de mejorar la eficacia de las tecnologías sanitarias a nivel global [15].

En cuanto al impacto que ha tenido la crisis del Covid-19 en este mercado, el uso generalizado de los dispositivos móviles y la penetración de internet en tantos rincones del mundo ha fomentado la inclinación progresiva hacia el uso de tecnologías digitales como el reconocimiento facial y de voz para mantener las conexiones de manera virtual. Además, aparecen las soluciones informáticas que permiten emplear estas tecnologías de reconocimiento para la detección de temperatura en el control de la propagación del virus. Muchas industrias están invirtiendo fuertemente en I+D para desarrollar software que ayude a controlar la propagación del Covid-19 [mercado2].



[mercado2]

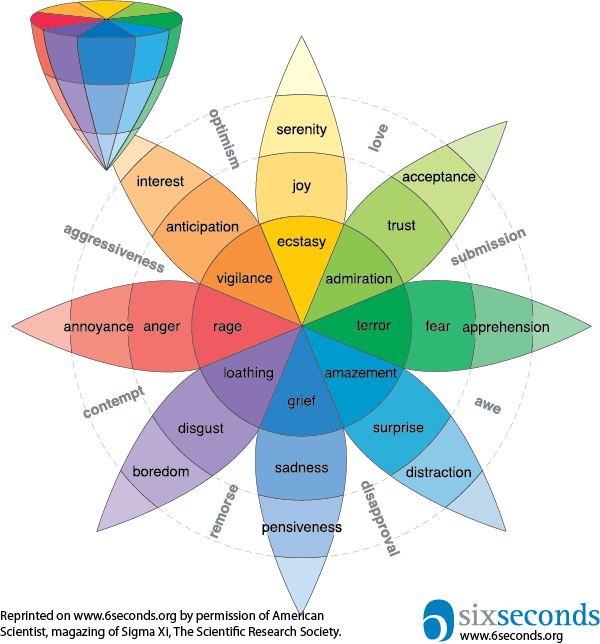
*Speech Emotion Recognition (SER)*

Emplear las tecnologías en el reconocimiento automático de las emociones humanas y estados afectivos a través de la voz, normalmente conocido como Speech Emotion Recognition o SER, lleva siendo objeto de investigación varios años ya [SER]. Ejemplo de ello son los experimentos realizados (France et al. 2000) donde se descubrieron las propiedades acústicas del habla como identificadores de depresión y riesgo de suicidio [expt]. Otros campos donde se han empleado sistemas SER son la educación, el entretenimiento, la industria automotriz, y en sistemas de síntesis de voz natural [sectors].

Los humanos poseemos una habilidad natural de emplear todos nuestros sentidos para obtener la máxima percepción de los estímulos y mensajes que recibimos. La detección de emociones es un proceso natural e innato en los humanos, pero es una tarea compleja para las máquinas, ya que carecen de estas cualidades humanas. Por lo tanto, si el problema se desglosa, la tarea inicial es definir cómo se miden las emociones humanas. El estado emocional no puede ser reconocido de manera directa, por lo tanto, se queda en manos de las expresiones emocionales. Estas últimas se rigen por el sistema motor a través de varias formas; las principales son la voz y las expresiones faciales [6]. Sin embargo, las señales de voz son de las formas más rápidas y naturales de comunicación humana [SER]. Por ello, se convierten en el candidato ideal para servir de input en una interacción hombre-máquina más veloz y eficiente.

Lo primero es lo primero: abordar el reconocimiento automático de la emoción requiere un modelo apropiado de representación de la emoción. Comenzando por representar la emoción de una manera adecuada, un primer modelo es a través de clases discretas, en el que las emociones son categorías. Por lo general, se admite un grupo básico e innato de categorías emocionales, o patrones de respuesta emocional, esenciales en la supervivencia y evolución de la especie. Darwin (1872/1998) sugirió que las emociones se han desarrollado para servir una función de comunicación en la especie, y por tanto se expresan y reconocen de manera parecida de manera intercultural [BasicEmotions].

El grupo de emociones primordiales son conocidas como las "seis grandes" de Ekman (1992): sorpresa, alegría, ira, miedo, asco y tristeza [BasicEmotions]. A su vez, tenemos el concepto de ‘tonos emocionales’, desarrollado por Manfred Clynes, que se refiere a los siete tonos/energías emocionales que transmitimos los seres humanos a través de una sutil modulación del sistema motor [7]. Plutchik (1980) sugirió ocho emociones (enfado, anticipación, felicidad, confianza, miedo, sorpresa y tristeza) a través de un modelo tridimensional que modela los niveles de intensidad y las relaciones entre las emociones [colors].



Estrella de emociones de Plutchik (Donaldson, 2017)

Uno de los mayores problemas a los que se enfrenta la disciplina del SER es que las emociones son conductas complejas, y cada persona siente emociones diferentes, y por lo tanto tiene formas distintas de mostrarlas [SERCNN)].  Las emociones se manifiestan y pueden medirse atendiendo a tres niveles o componentes diferentes [Ems]:

* **Cognitivo**: por ejemplo, a través de las expresiones verbales que manifiestan los sentimientos o experiencia subjetiva.
* **Conductual o motor**: por medio de cambios faciales o conductas de aproximación o retirada.
* **Fisiológico**: a través de los cambios viscerales u hormonales, y de los de tipo eléctrico y metabólico que se dan a nivel cerebral y periférico.

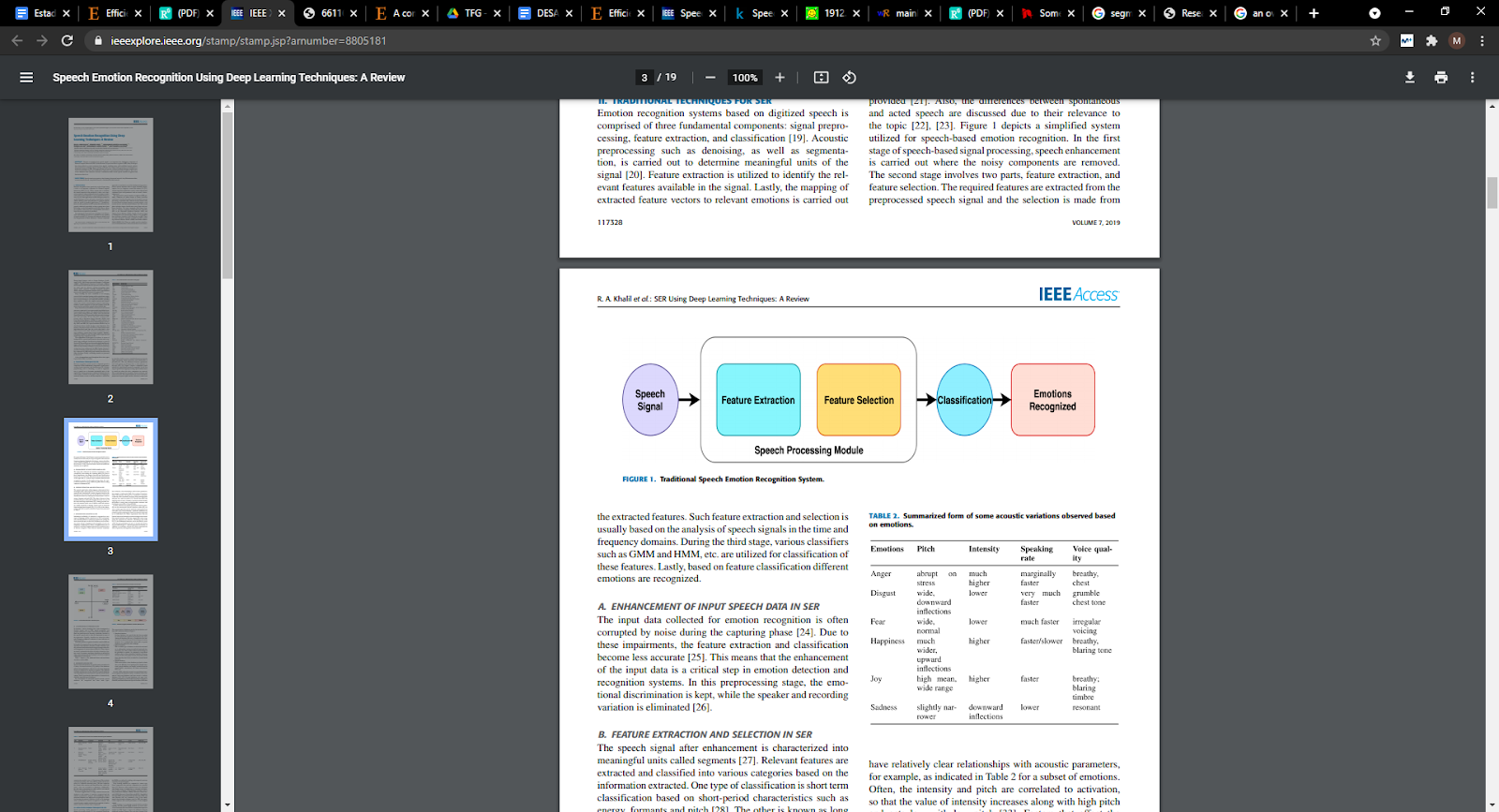
*SER: Feature Extraction*

El caso del tono de voz se cataloga como una respuesta fisiológica [6]. Realizar un reconocimiento de las emociones a través de la voz significa extraer estas emociones de las señales de voz de la persona en cuestión. Sin embargo, estas señales son complejas y almacenan información variada; el mensaje, quien habla, el género, el idioma y las emociones. SER es una tarea complicada ya que los tonos de voz son únicos para cada individuo [feature2]. Esta alta variabilidad del tono y la frecuencia puede resultar en que una misma frecuencia en cierto punto sugiera dos o más emociones. Por lo tanto, diferenciar entre las diferentes porciones es una tarea compleja [SER]. El feature extraction pretende extraer características particulares para cada una de estas porciones, denominadas speech utterances [feature3].

Las propiedades básicas de las señales de voz son las siguientes:

* spectral
* prosodic
* acoustic

El enfoque del SER está compuesto principalmente por las fases conocidas como la extracción de características (feature extraction), la selección de estas características (feature selection) y la clasificación de las características (emotion classification) [Feature1].



[DIBUJOS]

Las características del audio se categorizan generalmente en:

**Características temporales**

Estas características son fáciles de extraer, ya que se extraen directamente y por lo tanto requieren de menor poder de cómputo. Estas características proporcionan una manera más sencilla de analizar las señales de audio [SER2]. Estas incluyen:

* **Zero crossing rate**
* **short-term energy**
* **maximum amplitude**
* **minimum energy**
* **entropy energy**

**Características espectrales**

Relacionadas con el tracto vocal, suelen representar la distribución de la energía en la frecuencia del habla. Revelan patrones más profundos de las señales de audio, por lo que son las características que resultan más determinantes para determinar las emociones subyacentes [SER2].

(aquí meter tablas)

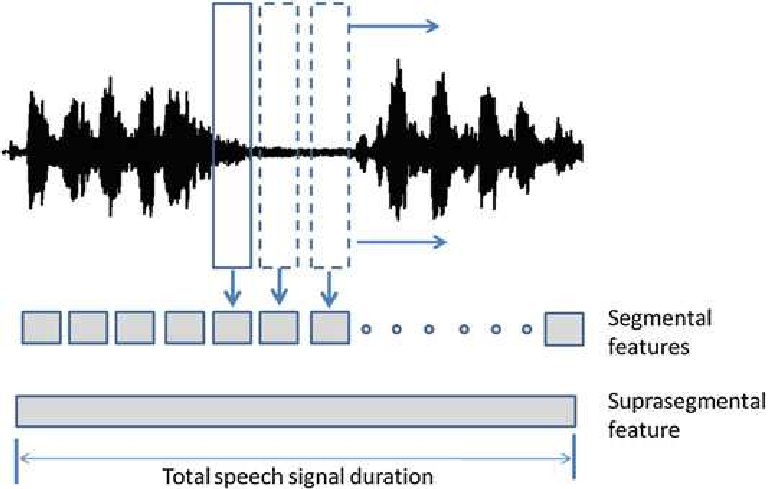
* **MFCC**
* **LPCC**
* **Formants**
* **DFT**
* **Spectral centroid**
* **Linear prediction**
* **Chroma coeficcients**

**Características prosódicas**

En muchos idiomas, son las responsables de transmitir información semántica al oyente, y forman las bases del comportamiento lingüístico. Por lo tanto, son útiles para detectar las emociones del locutor [SER3]. En varias investigaciones, estas características se han empleado para tareas de reconocimiento de emociones [SER4] [SER5].

* **noise**
* **stretch**
* **shift**
* **pitch**
* **higher\_speed**
* **lower\_speed**

Los vectores de características se dividen entre segmentales y suprasegmentales, acorde a su estructura temporal. Las características prosódicas se califican del segundo tipo, y estas no segmentan la señal, sino que se aplican sobre la extensión entera. Al contrario que las características segmentales (temporales y espectrales) a las que se les aplica técnicas de windowing y se calculan una vez cada cierto segmento de tiempo (unos 20-50 mseg) [feature4].



[feature4]

En estudios más tempranos, el foco se ponía principalmente en las características prosódicas, mientras que en investigación más reciente el uso de características espectrales ha crecido exponencialmente [feature4]. Escoger qué features son los más útiles para el entrenamiento puede ser una tarea muy confusa [SER]. Es importante que la selección de características no se vea afectada por la cultura, la región o el acento del locutor [feature2]. Dar con la combinación de características acústicas más robusta para el reconocimiento automático de la emoción de un locutor es uno de los pilares del SER [feature2].

(FEATURE SELECTION)

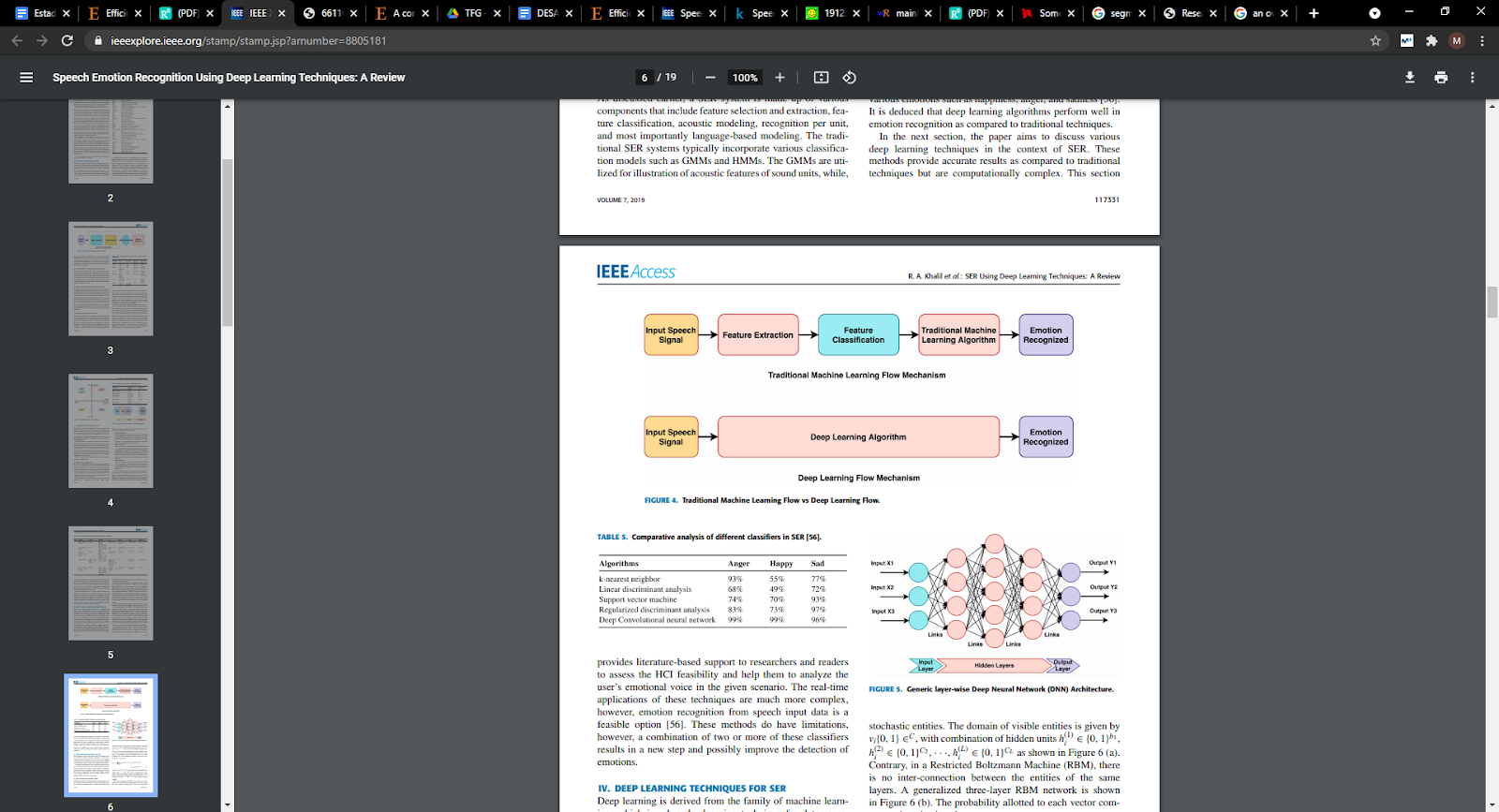
<https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e#:~:text=Feature%20Selection%20is%20the%20process,learn%20based%20on%20irrelevant%20features>.

*SER: Feature Classification*

Para la tarea de clasificación de características se emplean clasificadores lineales y no lineales. Los clasificadores lineales más usados incluyen las Bayesian Networks (BN) o el Maximum Likelihood Principle (MLP) y el Support Vector Machine (SVM). Sin embargo, la señal de voz se considera no estacionaria, por lo que un clasificador no lineal es óptimo para tareas SER [dibujos]. Existen muchos clasificadores no lineales, los más usados en la detección de emociones por medio de voz son el Gaussian Mixture Model (GMM), Hidden Markov Model (HMM) [16].

Las investigaciones más recientes se han desarrollado en el terreno de los sistemas de aprendizaje profundo. Estos sistemas infieren una representación jerárquica de los datos de entrada, lo que facilita su categorización [SERCNN2]. Estos sistemas forman parte de la rama de estudio conocida como el Deep Learning, cuya investigación y aplicación ha experimentado un fuerte auge durante los últimos años [DL]. Los modelos más utilizados en el campo del SER con los Convolutional Neural Networks (CNN) y las arquitecturas recurrentes como las Recurrent Neural Networks (RNN) y el Long Short-Term Memory (LSTM), que han resultado ser mucho más efectivos en las tareas de reconocimiento de voz como el Natural Language Processing (NLP) y el SER [SERCNN3].

Las técnicas de Deep Learning tienen ciertas ventajas sobre los métodos tradicionales. Una de las más destacadas es su habilidad de detectar estructuras complejas de las características (features) sin la necesidad de hacer una extracción manual [dibujos].

[DIBUJOS]

(falta hablar de arquitectura APP)

## Contexto y justificación

La medicina hipocrática defiende la necesidad de un estado de equilibrio y armonía entre el cuerpo y la mente. El padre y pionero de esta perspectiva de la medicina es Hipócrates (460-357 a.C.), una eminencia de la medicina de la Antigua Grecia [1]. Este mismo fue quien afirmó que para prevenir un ataque de asma, “el asmático debe protegerse de su propia ira” [asma]. Hoy en día, más de 2 mil años después, se persigue una filosofía similar: “La salud es un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades” [2], la definición que proporcionó la Organización Mundial de la Salud en 1948. De ahí la aparición de la medicina holística, que defiende que el cuerpo, la mente y el ambiente contribuyen de manera equitativa a la salud [3].

¿Puede la mente curar el cuerpo? La corriente de pensamiento budista ha fomentado sus creencias sobre las capacidades curativas de la mente desde hace más de dos mil años [buddist], a lo que se han ido sumando cada vez más los científicos occidentales. Prueba de ello son numerosos encuentros entre el Dalai Lama y prominentes psicólogos, médicos y profesores de meditación, que han arrojado nueva luz sobre esta conexión cuerpo-mente [Dalai lama].

Ya existen numerosos estudios que analizan el estado emocional y afectivo como factor determinante para el estado de salud de las personas. Cohen demostró que el estrés incrementa el riesgo de infecciones respiratorias agudas [4]. Con la aparición de la pandemia del Covid-19, es posible que un cuidado de la salud mental pueda fortalecer las defensas y ser un aliado a la hora de combatir esta enfermedad. Ejemplo de ello es un estudio que comprobó que mantener una actitud positiva durante momentos de elevado estrés promovía los niveles de S-Iga (inmunoglobulina-A salival), reconocida como una sustancia que protege contra las enfermedades respiratorias [5].

Si las emociones juegan un papel tan importante en la salud, fomentar un correcto estado de salud mental podría mejorar la salud de muchos ciudadanos. Especialmente ahora que vivimos en tiempos sin precedentes; la pandemia del Covid-19 ha impactado cada aspecto de nuestras vidas y nos ha llevado a una nueva realidad. Una realidad que está afectando a la salud tanto física como mental; es un buen momento para poder trabajar con ambas entrelazadas para asegurar el bienestar de aquellos más afectados. La monitorización masiva de la salud de la población es crucial, para así poder garantizar el bienestar de los mismos ciudadanos, como a su vez, ayudar a descongestionar los servicios sanitarios.

A su vez, vivimos en un mundo en el que se puede explotar mucha riqueza de las nuevas tecnologías. El auge de estas últimas amplía el horizonte de oportunidades y nos permite abordar problemas a través de soluciones menos tediosas, más rápidas y eficaces. Entonces, dado lo mucho que las máquinas se han integrado en nuestras vidas, ¿por qué no darles la oportunidad de que nos conozcan mejor? A muchos se les pueden venir a la cabeza historias terroríficas de ciencia ficción, resultado de un uso descontrolado. Sin embargo, sembrando las bases éticas, y con el control adecuado, el uso de las avanzadas tecnologías puede traer múltiples beneficios y mejoras a nuestro día a día. De la cuestión anterior, nace otra, ¿pueden las máquinas detectar nuestras emociones? La respuesta a estas dos cuestiones se plantea en el siguiente trabajo.

Watson de IBM es uno de los pioneros en computación cognitiva orientada a la salud, con una de sus mayores aplicaciones en el campo de la oncología. La empresa Babylon lanzó en 2016 una aplicación de consultas médicas online. Es a través de un sistema de reconocimiento de voz y de inspección de historiales clínicos que ofrece una ruta de acción adecuada a la circunstancia médica en cuestión [14]. En el mercado también está Google Health, una plataforma de Google que promueve el descubrimiento de nuevas oportunidades en el campo de la AI para mejorar la eficacia de las tecnologías sanitarias a nivel global [15].

## Planteamiento del problema

La crisis sanitaria producida por el Covid-19 ha sembrado sentimientos de miedo, preocupación y estrés en la población. Ante esta nueva y desafiante realidad es importante que cuidemos tanto nuestra salud física como mental. La vida cotidiana ha cambiado, el distanciamiento físico requerido para mantener un estilo de vida seguro ha desembocado en trabajar o estudiar desde casa, el desempleo temporal y la falta de contacto con seres queridos y amigos [OPS]. La salud tanto física como mental se está viendo afectada, por lo que es un buen momento para poder trabajar con ambas entrelazadas para asegurar el bienestar de aquellos más afectados.

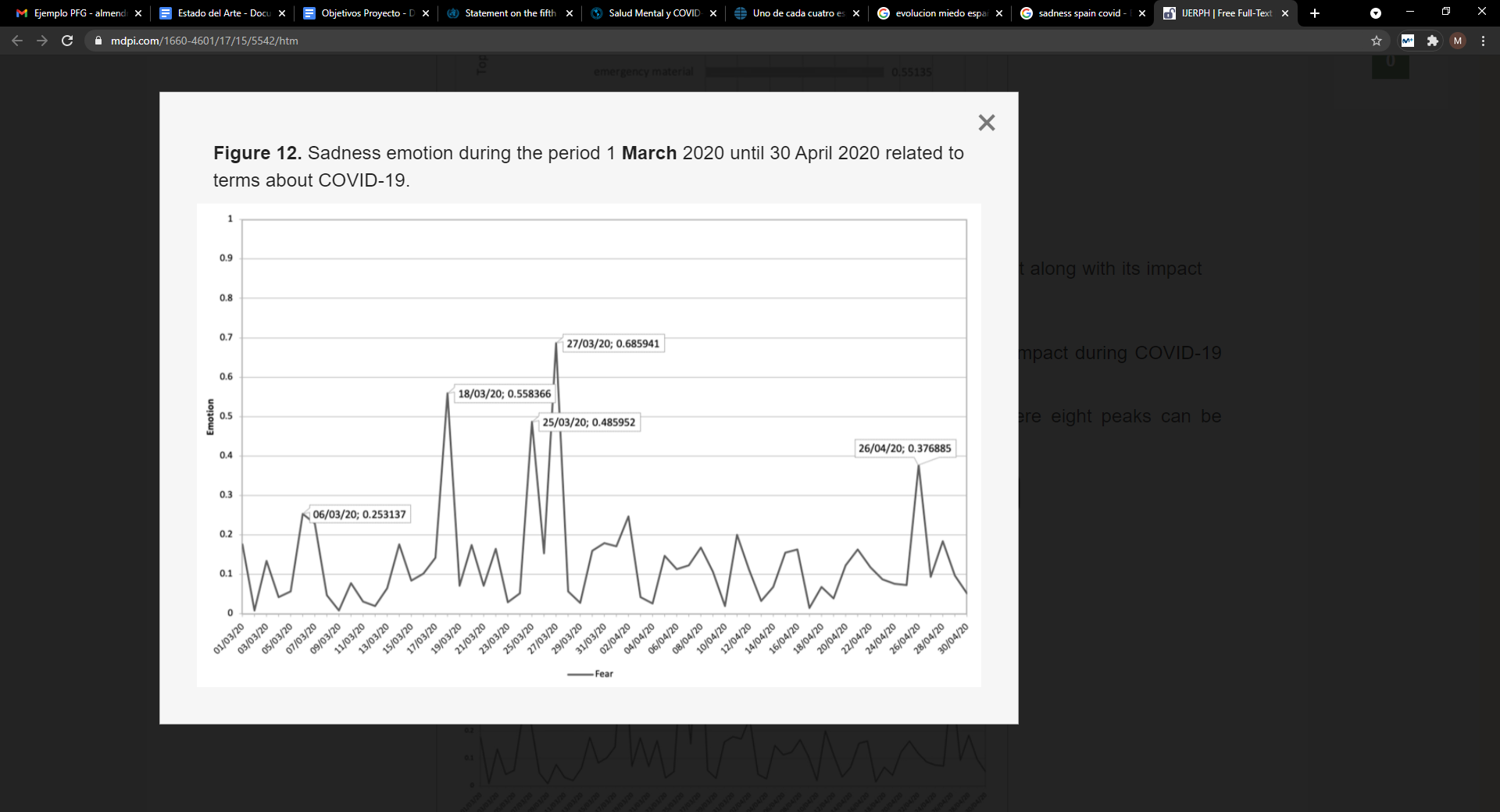
Un estudio sobre el análisis de sentimientos en España durante la pandemia del Covid-19 refleja el gran impacto que esta ha causado en el estado emocional de muchos españoles. Esta crisis es un reto para los gobiernos e instituciones sanitarias, a medida que los sentimientos de miedo, estrés y disgusto se propagan [16].

Figura 2: evolución del sentimiento del miedo en España durante el periodo del 1 del marzo 2020 hasta el 30 de abril 2020. Fuente: [16]

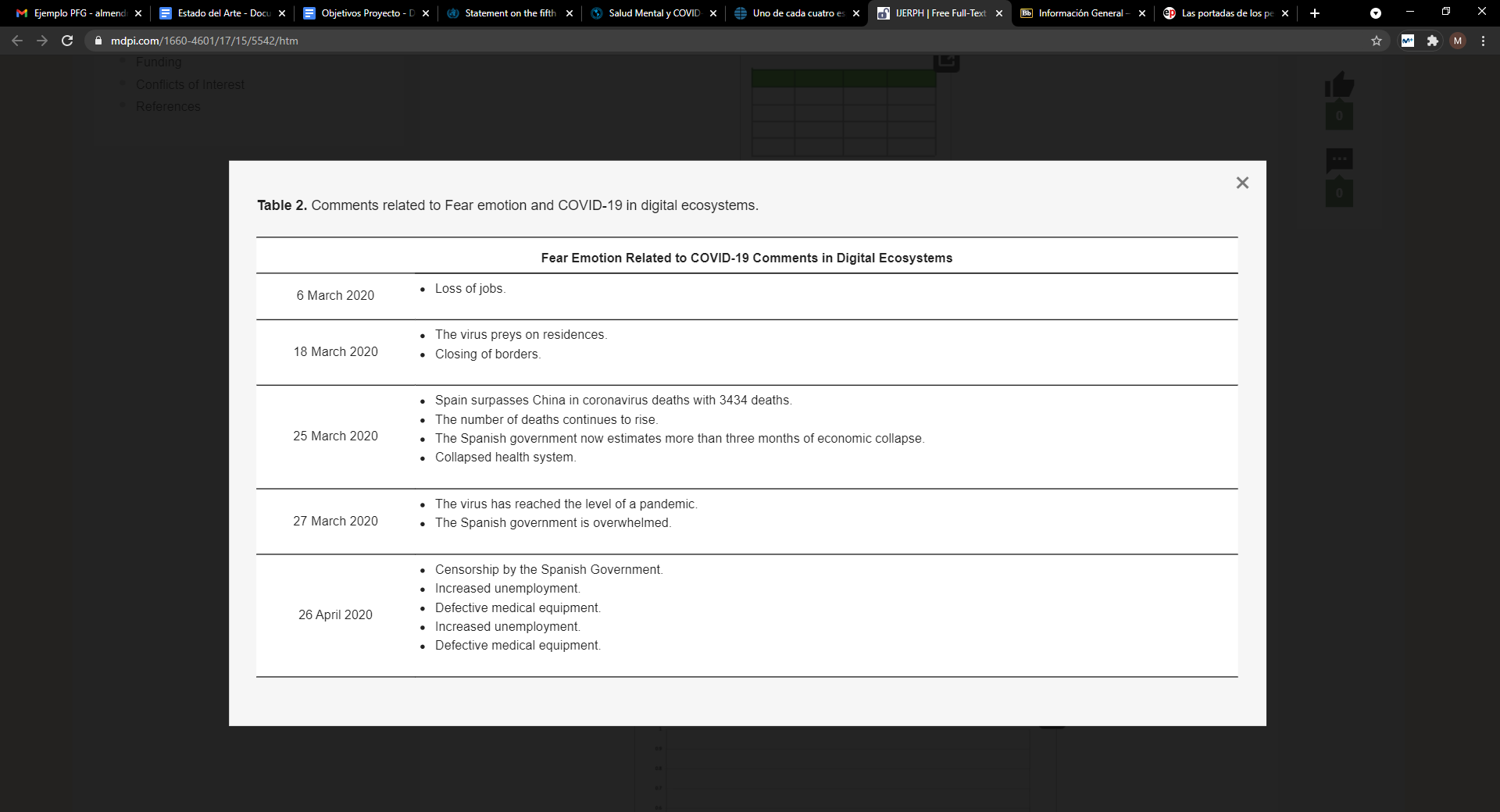
Se distinguen varios picos en los datos que coinciden con fechas específicas de ciertos titulares en los medios de comunicación. La tabla a continuación detalla la difusión de información en ciertos días del confinamiento que coinciden con los picos observados en la Figura 2.

Figura **x:** Noticias relacionadas con el miedo y el Covid-19 en los medios de comunicación.

La OMS hizo un llamamiento en la Declaración sobre la reunión realizada el 29 de octubre de 2020 a la necesidad de fortalecer los servicios de salud mental ante la crisis sanitaria del Covid-19 [OMSComite].

Mucha gente tiene que luchar sola contra estas situaciones emocionalmente complejas, lo cual puede suponer un empeoramiento de la salud mental y física. La población de avanzada edad es especialmente vulnerable; para muchos tener que asistir a la consulta de un médico de manera física supone un alto riesgo, muchas veces vital. Por lo tanto, se considera importante poder ofrecer alternativas al contacto físico que puedan proporcionar seguridad en momentos de mayor malestar emocional. Este proyecto trata de conseguir una solución que pueda aliviar a los sistemas sanitarios, y al mismo tiempo incrementar el estado afectivo de los usuarios, que podrán mantener un estado físico y mental saludables desde la comodidad del hogar.

El principal problema que afronta el AC según Picard es que los perfiles de cada persona son muy particulares, lo que convierte la tarea predecir los estados afectivos en una con una alta variabilidad [6]. lo tanto, este estudio plantea el uso de características extraídas de la voz para mejorar la comprensión del estado afectivo de los usuarios. La voz se considera un input adecuado ya que toma diferentes formas para distintas emociones [6].

Una tarea indispensable es que la aplicación aprenda de manera autónoma. En este caso, se busca la estrategia de aprendizaje óptima a través de un algoritmo de clasificación que de los mejores resultados en las fases de entrenamiento y predicción y empleando previamente las mejores prácticas de preparado de datos.

(no existe esta APP)

# OBJETIVOS

## Objetivos generales

El objetivo del presente trabajo es desarrollar una aplicación multi plataforma que sea capaz de detectar el estado emocional del usuario a través de un sistema de reconocimiento de voz.

## Objetivos específicos

La arquitectura propuesta estará formada por un componente back-end y un front-end.

La analítica de datos tomará un papel importante en este proyecto. Para ello, se diseña un proceso ETL sobre el dataset previo a la implementación del algoritmo. Este proceso se desglosa en las siguientes transformaciones:

1. Extract:

Encontrar y extraer los conjuntos de datos adecuados para el entrenamiento del algoritmo. Para la búsqueda de fuentes de datos adecuadas se va a utilizar la plataforma Kaggle [25], donde se encuentra una de las comunidades de científicos de datos más destacadas a nivel mundial.

1. Transform:
   * Solucionar los problemas de calidad de los datos en el caso de que los hubiese
   * Tratado de datos para su correcta ingesta por el algoritmo.
   * En el caso de tareas de clasificación del habla, se requiere una extracción de características (*feature extraction*) adecuada. Por lo tanto, requiere estudiar las opciones de extracción de características.
   * Seleccionar las características que optimizan el aprendizaje del algoritmo.
2. Load:

Cargar los datos finales transformados en un lugar centralizado. Una vez estén los datos preparados para su uso, podrán ser empleados para entrenar el algoritmo.

A continuación, se especifica la parte técnica de inteligencia artificial:

1. Estudiar las opciones de algoritmos de sistemas de reconocimiento de voz.
2. Diseñar el sistema de reconocimiento de voz: la idea principal consistirá en un algoritmo de clasificación, en concreto una red neuronal artificial.
3. Decidir cuáles serán las métricas más significativas para escoger el modelo óptimo.
4. Análisis de los resultados obtenidos para cada iteración; realizar los ajustes necesarios en los parámetros, con el fin de escoger el mejor modelo.
5. Escoger el algoritmo con mejores resultados, capaz de predecir la emoción en datos nuevos.

La importancia de que el modelo muestre destreza a la hora de completar e interpretar estas tareas correctamente va a ser determinante a la hora de proporcionar las conclusiones adecuadas. Una vez el algoritmo haya sido entrenado, se pasa a la fase de desarrollo de la aplicación móvil.

Una vez el dispositivo haya realizado la tarea de clasificación, debe ser capaz de responder al usuario.

El objetivo es conseguir que el usuario se sienta mejor, para ello, los principales sistemas integrados serán una serie de habilidades:

1. Captar los datos necesarios a través del micrófono
2. Preprocesamiento de los datos; repetir transformaciones especificadas anteriormente
3. Analizar y dar respuesta a los datos

Diseñar interfaz usuario-maquina

(APP)

## Beneficios del proyecto

Incluye una descripción del beneficio que aporta tu proyecto, en relación a los objetivos.

La herramienta fundamental para poder usar el producto será el smartphone, de uso generalizado.

Esta aplicación puede ser un beneficio muy grande para aquellos que se encuentren en una situación emocionalmente compleja, y deseen llevar una vida más autónoma, sin necesidad de desplazamientos. Con este sistema el usuario podrá mantener una monitorización del estado de sus emociones e impulsar una rutina hacia una vida más saludable.

(beneficios de la arquitectura cloud app)

# DESARROLLO DEL PROYECTO

## Planificación del proyecto

**Fase 1. Investigación previa**

1. Estudio del del contexto y el estado del arte de los conceptos y tecnologías del proyecto. Esto incluye la búsqueda de las alternativas y casos de uso actuales.
2. Análisis técnico de las tecnologías escogidas, tanto para el componente back-end y el front-end.

**Fase 2. Planificación y diseño**

1. Definir los objetivos del proyecto en base al planteamiento del problema detectado. Determinar el alcance de los objetivos y la distribución de tiempo correspondiente.
2. Definición de requisitos y recursos necesarios. Estos pueden ser requisitos tanto software como hardware.
3. Diseño de la arquitectura back- end del sistema. Esto consiste en canal de preparado de datos, extracción de características y en el algoritmo de reconocimiento de emociones y sus predicciones.
4. Diseño de la arquitectura front-end del sistema. Esta será una aplicación móvil capaz de captar la voz del usuario, enviarla al servidor donde se aloje el algoritmo, y recibir una predicción en tiempo real.
5. Diseño del proceso de validación y pruebas del sistema

**Fase 3. Preparación del entorno**

1. Reunir los requisitos necesarios para preparar el entorno de trabajo. Aplicaciones, programas y elementos como SDKs necesarios para el correcto funcionamiento en la maquina local.
2. Empezar a manejar aquellas tecnologías que son nuevas y no se hayan usado previamente al desarrollo del proyecto. Para ello se emplean videos y cursos online relacionados con estas competencias.

**Fase 4. Desarrollo e implantación de la arquitectura**

ETL:

1. Buscar los datasets adecuados para el entrenamiento. Los datos deben tener un formato parecido o igual y con un labelling adecuado.
2. Desarrollar un pipeline adecuado de ingesta y preparado/limpieza de datos. Incluye determinar las librerías necesarias para ello.
3. Llevar a cabo un análisis exploratorio previo de los datos. Se pretende encontrar patrones o indicios de fenómenos como en oversampling o el undersampling.

Desarrollo del clasificador SER:

1. Estudiar las opciones de algoritmos para la tarea de clasificación de emociones. Plantear una serie de pruebas con aquellos considerados más adecuados.
2. Determinar la estrategia de extracción de características de los datos y prepararlos para poder ser ingeridos por el algoritmo elegido. Para ello, se investigarán tareas de escalado de datos y separación entre entrenamiento, validación y test.
3. Iniciar la fase de entrenamiento. Probar con diferentes valores para los hiperparámetros de los modelos y fijar las métricas que se utilizarán para la evaluación.
4. Comparativa de modelos y escoger en base a la eficacia de predicción.

Interfaz de usuario:

1. Estudiar alternativas para el desarrollo de una aplicación web y móvil (multi plataforma). Con un sistema de registro e inicio de sesión personal incorporado y que sea capaz de reconocer las emociones en la voz de un usuario y devolver un resultado en tiempo real.
2. Diseño de la interfaz de usuario.
3. Implementar una arquitectura de cloud computing junto con un framework de programación de aplicaciones móviles para montar una aplicación con alojamiento web y móvil en la nube.
4. Conectar todos los componentes implementando una BBDD adecuada para centralizar los datos que se intercambiaran entre los componentes back-end y front-end.

**Fase 5. Pruebas, conclusiones y análisis de rendimiento**

1. Periodo de pruebas, refuerzo e identificación de disfunciones. Se deberán aplicar los parches y las medidas necesarias para arreglar posibles fallos de funcionamiento.
2. Desarrollar un modelo para la salida a producción de la aplicación. Para ello es necesario un análisis de como garantizar la seguridad y confidencialidad de los usuarios a través de protocolos de seguridad de datos (RGPD, LSSI).
3. Redactar la memoria a partir de todos los documentos empleados y redactados durante cada fase de desarrollo.
4. Defensa del TFG.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nombre de la tarea | Mes | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Enero | | | | Febrero | | | | | Marzo | | | | Abril | | | | | Mayo | | | | | Junio | | | | | N.º horas |
| **Investigación previa** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Contexto y estado del arte |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | 20 |
| Análisis técnico de las tecnologías empleadas |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| **Planificación y diseño** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Definir objetivos y el alcance |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | 30 |
| Definir requisitos y recursos |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Diseño de la arquitectura back-end |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Diseño de la arquitectura front-end |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Diseño del proceso de pruebas |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| **Preparación del entorno** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Instalar y configurar los requisitos computacionales necesarios |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | 90 |
| Aprendizaje de nuevas tecnologías |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Reunión de seguimiento con la tutora del TFG |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| **Desarrollo e implantación** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ETL | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Búsqueda de datasets para el entrenamiento |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | 50 |
| Desarrollo del pipeline de ingesta y preparado/limpieza de datos |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Análisis exploratorio de los datos limpios |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Reunión de seguimiento con la tutora del TFG |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Desarrollo del Clasificador SER | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Estudiar algoritmos de clasificación |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | 150 |
| Extracción de características |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Reunión de seguimiento con la tutora del TFG |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Entrenamiento del modelo y ajuste de hiperparémetros |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Reunión de seguimiento con la tutora del TFG |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Comparativa y elección del mejor modelo |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Interfaz de usuario | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Estudiar opciones para el desarrollo de una aplicación multi plataforma |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | 130 |
| Diseño de la interfaz de usuario |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Reunión de seguimiento con la tutora del TFG. Revisión de los algoritmos. |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Montar arquitectura de computación en la nube |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Conectar todos los componentes y administrar la BBDD |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Reunión de seguimiento con la tutora del TFG |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| **Pruebas, conclusiones y análisis de rendimiento** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Pruebas y ajustes de funcionamiento |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | 100 |
| Desarrollar modelo de puesta en producción y políticas de seguridad de datos |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Redacción de la memoria final |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Entrega primera versión completa de la memoria de TFG a la tutora |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |  |
| Realizar cambios sugeridos en el feedback de la tutora de TFG |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Entrega de la versión final de la memoria de TFG |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| **DEFENSA DEL TFG** |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |

## Descripción de la solución, metodologías y herramientas empleadas

Para un proyecto cuyo foco sea el ámbito científico y de investigación, aquí se describen las metodologías utilizadas para la realización del proyecto, incluye herramientas tecnológicas, programas de cálculo, dispositivos utilizados, modelos aplicados para el desarrollo del proyecto, descripción del análisis de datos cuyos resultados se presentarán en la sección de resultados, etc.

Para un proyecto cuyo foco sea el desarrollo de producto, aquí se describe la solución, modelos analizados, diseños realizados, algoritmos aplicados, desarrollos implementados, etc. incluye herramientas tecnológicas, programas de cálculo, dispositivos utilizados, modelos aplicados para el desarrollo del proyecto, descripción del análisis de datos cuyos resultados se presentarán en la sección de resultados, etc.

Puedes añadir una estructura en subsecciones si facilita la organización de esta sección.

El proyecto se divide en

### Bases de datos

El entrenamiento del algoritmo de clasificación de emociones es el núcleo del presente trabajo. Por lo tanto, encontrar un conjunto de datos adecuado para su entrenamiento era una tarea esencial. Se decide combinar dos datasets diferentes para poder tener un conjunto de entrenamiento grande y variado, pero evitando mezclar datos de diferentes formatos ya que puede dar lugar a incongruencias en el entrenamiento. A su vez, ambos datasets contienen voces categorizadas como masculinas como femeninas, lo cual permite desarrollar algoritmos especializados para cada género, aparte de uno generalizado. Los datasets elegidos son los siguientes:

* **RAVDESS** (Ryerson Audio-Visual Database of Emotional Speech and Song) (Livingston & Russo, 2018), datos publicados bajo una licencia de atribución de Creative Commons de uso no comercial [RAVDESS].

La RAVDESS contiene 7356 archivos, en los que participan 24 actores profesionales (12 mujeres, 12 hombres) con un acento norteamericano neutral. El habla incluye expresiones de calma, felicidad, tristeza, ira, miedo, sorpresa y disgusto, y se vocaliza una de dos frases. La canción contiene emociones tranquilas, felices, tristes, enfadadas y temerosas. Cada expresión tiene dos niveles de intensidad emocional (normal, fuerte), con una expresión neutra adicional.

Para este proyecto solo se van a usar archivos de audio y de tipo voz. El archivo de voz (Audio\_Speech\_Actors\_01-24.zip, 215 MB) contiene 1440 archivos: 60 ensayos por actor x 24 actores = 1440. La tabla x detalla las características de los archivos.

|  |  |  |
| --- | --- | --- |
| Nº | Característica | Opciones |
| 1 | Modalidad | 01 = AV completo, 02 = solo video, 03 = solo audio |
| 2 | Canal vocal | 01 = habla, 02 = canción |
| 3 | Emoción | 01 = neutral, 02 = calma, 03 = feliz, 04 = triste, 05 = enojado, 06 = temeroso, 07 = disgusto, 08 = sorprendido |
| 4 | Intensidad emocional | 01 = normal, 02 = fuerte (no hay intensidad fuerte para la emoción neutral) |
| 5 | Frase | 01 = “Kids are talking by the door”, 02 = “Dogs are sitting by the door” |
| 6 | Repetición | 01 = 1ª repetición, 02 = 2ª repetición |
| 7 | Actor | 01 a 24. Los actores impares son hombres, los actores pares son mujeres |

Tabla x Descripción de la identificación de archivos de audio de la base de datos RAVDESS

Ejemplo de nombre de archivo: 02-01-06-01-02-01-12.mp4

* Solo video (02)
* Discurso (01)
* Temeroso (06)
* Intensidad normal (01)
* Declaración "perros" (02)
* 1a repetición (01)
* 12 ° actor (12)
* Mujer, ya que el número de identificación del actor es par.
* **CREMA-D** (Crowd-sourced Emotional Multimodal Actors Dataset) es un conjunto de datos de 7,442 archivos de 91 actores. Las voces las componen 48 actores masculinos y 43 femeninos entre las edades de 20 y 74, provenientes de una variedad de razas y etnias (afroamericanos, asiáticos, caucásicos, hispanos y no especificados).

Los actores hablan una selección de 12 frases expresadas en una de seis emociones diferentes (ira, disgusto, miedo, feliz, neutral y triste) y cuatro niveles de emoción diferentes (bajo, medio, alto y no especificado).

|  |  |  |
| --- | --- | --- |
| Nº | Característica | Opciones |
| 1 | ID del actor | Número de 4 dígitos. Los IDs 1002-1013,1018,1020,1021,1024,1025,1028-1030,1037,1043,1046,1047,1049, 1052-1056,1058,1060,1061,1063,1072-1076,1078,1079,1082,1084,1089,1091 corresponden al sexo femenino, y el resto al masculino. |
| 2 | Frase | “It's eleven o'clock” (IEO).  “That is exactly what happened” (TIE).  “I'm on my way to the meeting” (IOM).  “I wonder what this is about” (IWW).  “The airplane is almost full” (TAI).  “Maybe tomorrow it will be cold” (MTI).  “I would like a new alarm clock” (IWL)  “I think I have a doctor's appointment” (ITH).  “Don't forget a jacket” (DFA).  “I think I've seen this before” (ITS).  “The surface is slick” (TSI).  “We'll stop in a couple of minutes” (WSI). |
| 3 | Emoción | Enfado (ANG)  Disgusto (DIS)  Miedo (FEA)  Feliz (HAP)  Neutro (NEU)  Triste (SAD) |
| 4 | Nivel de emoción | Bajo (LO)  Medio (MD)  Alto (HI)  Sin especificar (XX) |

Tabla x Descripción de la identificación de archivos de audio de la base de datos CREMA-D

Ejemplo de nombre de archivo: 1001\_DFA\_ANG\_XX.wav

* ID del actor (1001), actor masculino
* Frase “Don’t forget a jacket” (DFA)
* Enojado (XX)
* Nivel de emoción no especificado (XX)

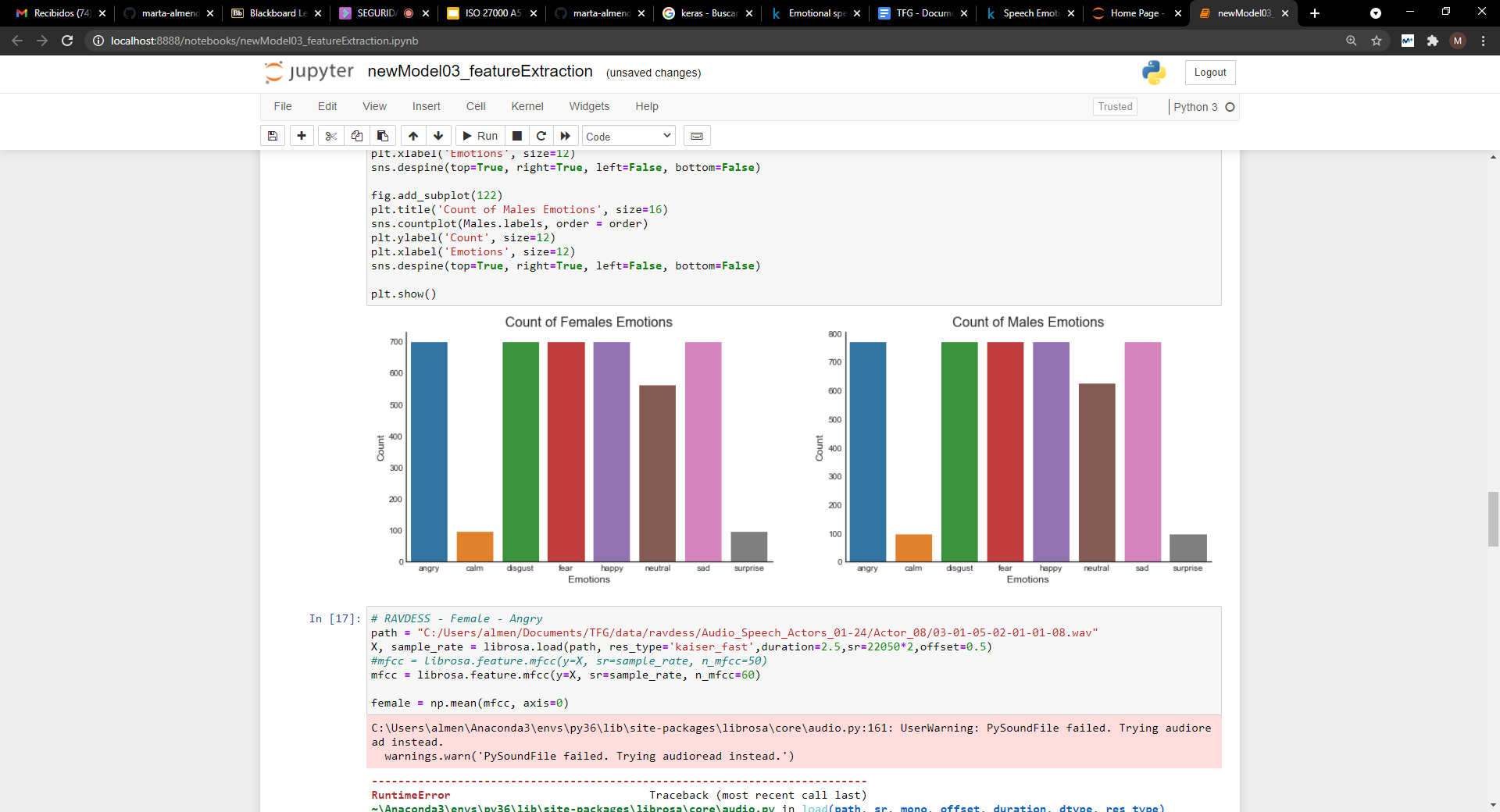
### Ingesta, limpieza y análisis exploratorio de datos

Para los procesos de tratado de datos y entrenamiento se implementa una arquitectura en el entorno de trabajo Jupyter Notebook. Una vez importados todos los datos, se organizan de tal manera que se tenga un valor y una etiqueta (emoción) para cada valor, para que el clasificador pueda aprender y diferenciar cada emoción.

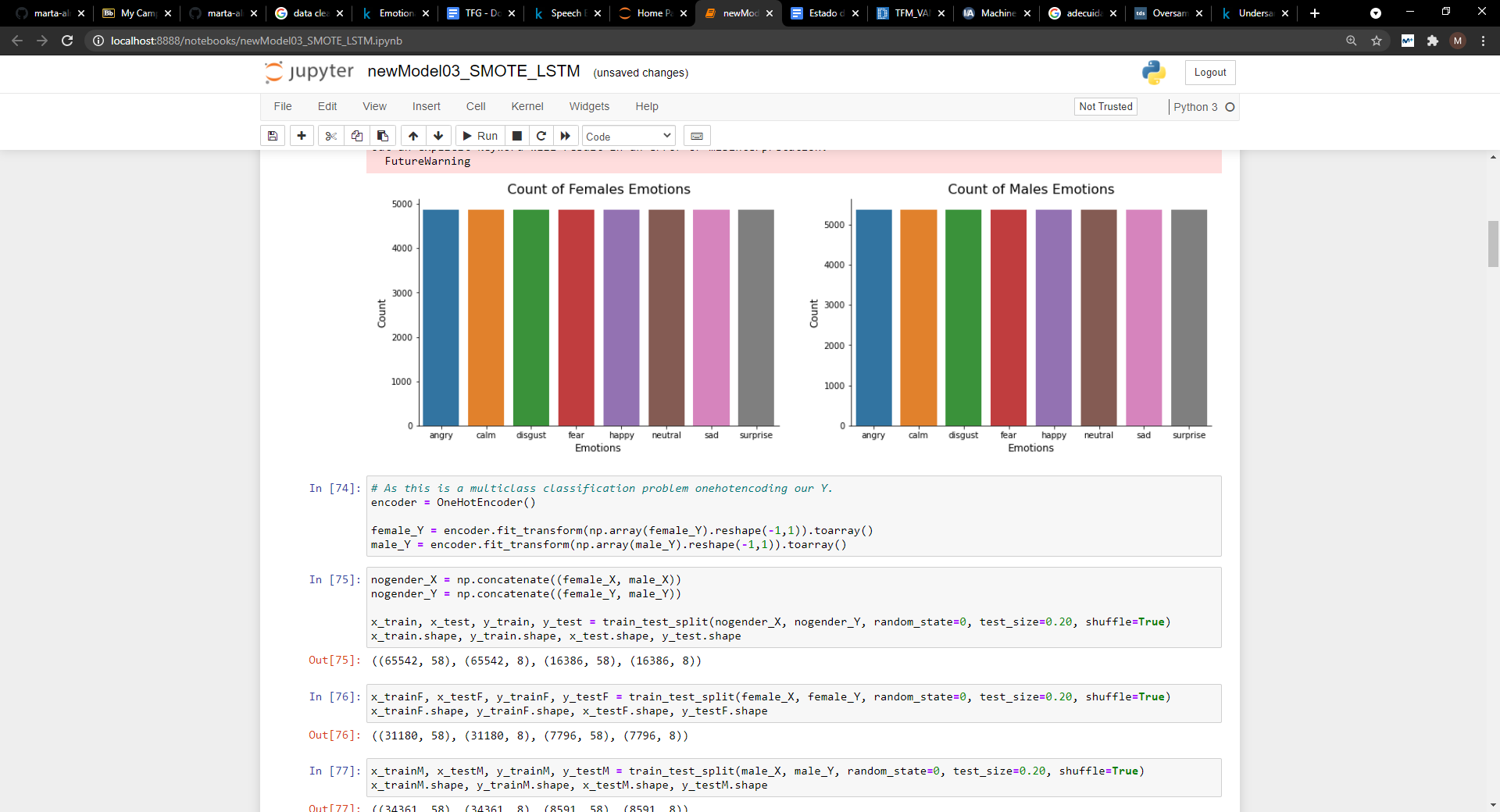
Una vez se haya realizado el labeling adecuado para cada tanda de datos, al estar trabajando con dos datasets diferentes, se crea un dataframe de *pandas*, donde se almacenan todos los datos con sus respetivas etiquetas. Este dataframe es el que se va a emplear para las próximas fases de preprocesamiento de datos y de extracción de características.

Un modelo predictivo de machine learning aprende de los datos que ingiere, y las tareas de automatización van a ser guiadas por la calidad de los datos y el nivel de representatividad del grupo de entrenamiento [cleandata2]. La limpieza de datos es un proceso fundamental en toda tarea de machine learning. Permite eliminar errores y datos redundantes para lograr un dataset más fiable [cleandata] lo cual va a mejorar la calidad de entrenamiento del clasificador. Sin embargo, los datasets escogidos ya estan en muy buen estado, al tratarse de fuentes de datos de alta difusión, por lo que no es necesario ningún tipo de transformación para eliminar, modificar o sustituir datos.

Parte del proceso de verificación de los datos para el entrenamiento es que haya equilibrio entre las clases. Para ello se hace un conteo de ellas (figura x) utilizando elementos visuales. Se hace un conteo tanto para las emociones de los hombres como de las mujeres, y se descubre que ambas se comportan igual; se identifica un destacado desbalance de clases que puede obstaculizar el proceso de entrenamiento.

(figura x)

Existe una carencia de emociones de ‘calm’ y de ‘surprise’. Es un desafío para las tareas de modelado predictivo trabajar con una distribución de clases severamente sesgada, causando desigualdad en los costos de la clasificación errónea [imbclass]. De tal manera, el algoritmo se obstruye prediciendo las clases mayoritarias y no aprenda sobre las más pequeñas. Una solución es eliminar estas categorías de datos, pero en este proyecto se consideran las ocho emociones igual de importantes, y un aspecto considerado importante es añadir variabilidad al modelo, dotándole de la capacidad de predecir un amplio abanico de emociones, el cual se pretende agrandar en futuras líneas de trabajo. Para problemas donde estas categorías más escasas son importantes, se aplican técnicas de *oversampling* o *undersampling* [imbclass2]. La primera consiste en generar más datos de las clases minoritarias y la segunda en reducir la cantidad de datos de las clases mayoritarias. En este caso, se aplica *oversampling*, ya que se considera importante trabajar con cantidades de datos elevadas, así el algoritmo tiene más información para el aprendizaje, razón por la que además de hacer un oversampling de las clases de ‘calm’ y ‘surprise’, también se hace del resto de clases.



|  |  |
| --- | --- |
| Librería | Descripción |
| os | Librería de Python que permite interactuar con el sistema operativo |
| pandas | Librería de Python especializada en el manejo y análisis de estructuras de datos |
| matplotlib | Librería de visualización de datos y trazado gráfico para Python |
| numpy | Librería de Python para trabajar con arrays y en el dominio del algebra lineal |
| seaborn | Librería de Python para visualización de datos basada en matplotlib. Especializada en gráficas y representaciones estadísticas |
| librosa | Librería de Python para el análisis de audio y música. Incluye capacidades de extracción de datos/características de los archivos de audio. |
| sklearn | Librería open source para aprendizaje automático en Python |
| Imblearn | Librería open source basada en sklearn que provee herramientas para tareas de clasificación con datos desbalanceados |
| tensorflow | Librería open source para aprendizaje automático en Python |
| keras | Librería open source de redes neuronales escrita en Python |

Tabla x Descripción de librerías

## Recursos requeridos

En este apartado debes enumerar los recursos que has utilizado para la ejecución del proyecto (recursos técnicos, dispositivos, material de laboratorio, asistencia de expertos, etc.). Si bien, en la sección donde se describe Metodología y Herramientas empleadas, se describen en detalle las mismas, en esta sección, únicamente debes enumerarlas.

## Presupuesto

El Presupuesto supone la evaluación económica total del proyecto.

No olvides que tu tiempo también vale dinero, no sólo hay que incluir el coste de los materiales empleados.

A continuación, se muestra un ejemplo de resumen de presupuesto. Puedes añadir entradas a esta tabla si lo necesitas, conforme a la naturaleza de tu trabajo. Todos los recursos requeridos indicados en el apartado anterior deberían estar recogidos en la tabla de costes, aunque hayan tenido coste cero.

Las herramientas de software utilizadas para el preprocesamiento de los datos de entrenamiento y desarrollo del algoritmo han sido de software libre, por lo tanto no ha sido necesario invertir en este aspecto.

|  |  |  |
| --- | --- | --- |
| **Tipo de coste** | **Valor** | **Comentarios** |
| Horas de trabajo en el proyecto | 470 | Añade entradas para otras personas que hayan participado en tu proyecto para contabilizar sus horas aproximadas. |
| Equipo técnico utilizado | € | Si el equipo no lo has tenido que adquirir, indica su valor aproximado en el mercado si lo tuvieras que adquirir nuevo. Utiliza una entrada en la tabla por cada equipo (desktop, servidor, etc.) |
| Software utilizado | € | Si el Software no lo has tenido que adquirir, indica su valor aproximado en el mercado si lo tuvieras que adquirir. Indica coste 0€ si es Software que no requiere pagar precio por licencia.  Ejemplos: IDE de desarrollo, herramienta de análisis de datos, programa de diseño gráfico, etc. |
| Estudios e informes | € | Si has tenido que adquirir algún informe de consultora o de revista de investigación, indica el coste. |
| Materiales empleados | € | P. ej.: material de laboratorio, sensores, etc. |

## Viabilidad

Este apartado es opcional, pero es interesante que incluyas un breve análisis de viabilidad económica del proyecto (relación coste / beneficio), y un análisis de sostenibilidad a futuro del resultado de tu proyecto.

## Resultados del proyecto

Conforme a los objetivos específicos, describe los resultados finales obtenidos.

Para proyectos con un enfoque de desarrollo de producto, debes incluir el resultado de tu plan de pruebas.

Puedes añadir una descripción de cambios durante el proyecto respecto a los objetivos iniciales, y comentarios que quieras incluir de cómo has desarrollado las distintas actividades.

# DISCUSIÓN

Esta sección es más habitual en trabajos de tipo "científico" o de "investigación", donde uno presenta brevemente los resultados principales y los discute, pero también puede utilizarse en otro tipo de trabajos.

Puedes incluir secciones específicas para discutir cuestiones como: limitaciones del estudio, limitaciones de la tecnología empleada, cambios respecto a objetivos planteados inicialmente.

Puedes incluir respuestas a preguntas del tipo: ¿la metodología inicialmente pensada ha sido útil?, ¿he tenido que adaptarme a cambios a lo largo del proyecto?, ¿qué cambios han sido y cómo he adaptado el proyecto para poder manejar esos cambios?, ¿qué impacto ha tenido el resultado de mi proyecto?

# CONCLUSIONES

## Conclusiones del trabajo

Breve descripción objetiva del resultado en relación al objetivo general de tu proyecto.

## Conclusiones personales

Describe tus impresiones y experiencia personal durante el desarrollo del proyecto, o destacar la importancia que tiene el tema para ti, lo que has aprendido, o la trascendencia que ha tenido para ti o para otros este proyecto.

# FUTURAS LÍNEAS DE TRABAJO

Es importante destacar en este apartado las líneas con las que se podría continuar tu trabajo. Indica todo lo que has anotado como futuro trabajos durante el desarrollo del proyecto, o aquellos aspectos que estaban fuera del alcance, pero que son interesantes para desarrollar a futuro el valor de los resultados de tu trabajo.

# REFERENCIAS

En este apartado figurará el conjunto de libros, revistas u otros textos que el autor considere de interés para justificar las soluciones adoptadas en el Proyecto. **Cita todas las fuentes** que has utilizado como consulta para elaborar el trabajo.

Sigue el estilo de cita que te indiquen las normas de estilo y respétalo a lo largo de todo el proyecto. Recuerda que has de citar todas las fuentes que hayas usado. Los estilos de cita más comunes son:

* ISO
* IEEE
* APPA
* Etc.

En ingeniería se suele usar el ISO o el IEEE.

Si puedes usar un gestor de citas bibliográficas te será más fácil. Si no tendrás que recurrir a las páginas web de las bibliotecas para saber cómo citar adecuadamente. Por ejemplo, lo encuentras en:

<http://biblioteca.uem.es/es/aprendizaje-y-formacion/citas-bibliograficas-documentos>

Referencias usadas en este manual de estilo:

**AENOR. 2010.** AEN/CTN 157 - PROYECTOS. *Normas y Publicaciones.* [En línea] 2010. [Citado el: 25 de abril de 2013.] http://www.aenor.es/aenor/normas/ctn/fichactn.asp?codigonorm=AEN/CTN%20157.

**Miró Julià, José. 2010.** Recursos para aprender a escribir. [En línea] 2010. http://bioinfo.uib.es/~joemiro/RecEscr/manual.pdf.

**UNE 157001. 2002.** Criterios generales para la elaboración de proyectos. *Escuela Universitaria de Ingeniería de Vitoria.* [En línea] 2002. [Citado el: 25 de abril de 2013.] http://www.coiib.es/coiib/documentos/DocumentosContenidos/Gu%C3%ADa%20de%20elaboraci%C3%B3n%20de%20proyectos/2-Electricidad/5\_PNE\_157701\_Criterios.pdf.

# ANEXOS

Sirven para incluir documentación complementaria (planos, circuitos, código, ficheros de configuración, especificaciones técnicas y hojas de características, fichas explicativas, resultado de encuestas, reglamentación y normativas requeridas, etc.).

[PÁGINA INTENCIONADAMENTE EN BLANCO]