



Universidad Nacional  
de Educación a Distancia



Universidad  
Complutense de Madrid

Escuela Técnica Superior de  
Ingeniería Informática

Facultad de Informática

# GENERACIÓN DE PLAN ESTRATÉGICO DE INVERSIÓN MEDIANTE RECOPILOCIÓN DE DATOS Y ANÁLISIS DE SENTIMIENTO DE LAS REDES SOCIALES CON REDES NEURONALES

Joan Serra Ginestar

Directores: Matilde Santos Peñas

Raúl Fernández Fernández

Trabajo de Fin de Máster

Máster Universitario  
en Ingeniería de Sistemas y de Control

Curso 2023-24



# GENERACIÓN DE PLAN ESTRATÉGICO DE INVERSIÓN MEDIANTE RECOPILOCIÓN DE DATOS Y ANÁLISIS DE SENTIMIENTO DE LAS REDES SOCIALES CON REDES NEURONALES

Joan Serra Ginestar

Directores: Matilde Santos Peñas

Raúl Fernández Fernández

Trabajo de Fin de Máster

Máster Universitario  
en Ingeniería de Sistemas y de Control

Curso 2023-24





## Autorización

Autorizamos a la Universidad Complutense y a la UNED a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a sus autores, tanto la memoria de este Trabajo Fin de Máster, como el código, la documentación y/o el prototipo desarrollado.

Firmado:

Firma del alumno

## **Agradecimientos**

La realización de este trabajo no hubiera sido posible sin haber conocido antes las asignaturas impartidas por mis directores, Matilde y Raúl, los cuales me han apoyado en mi afán de seguir aprendiendo sobre la Inteligencia Artificial y la Ciencia de Datos. Gracias a ambos.

Me gustaría también agradecer a mi familia y amigos, especialmente a Andrea, por el apoyo durante estos meses y entender mi baja disponibilidad ante sus planes.

Finalmente, agradecer inmensamente el apoyo incondicional de mi pareja, Ana Paula, cuya paciencia y ayuda en mi vida diaria ha sido clave para poder finalizar mis estudios de Máster. Te estaré siempre agradecido.



## Resumen

Las redes sociales cuentan con cientos de miles de datos, los cuales suponen una gran fuente de información. En plataformas como Twitter se publican cerca de 500 millones de tuits al día. Esta información, compartida directamente por usuarios de estas plataformas, supone una valiosa percepción de cómo se comporta la sociedad y, sobre todo, de lo que piensa la sociedad respecto a diferentes temas.

Las redes sociales se han convertido en un escaparate para muchos locales de restauración. Es muy común hoy en día entre los consumidores compartir con los demás el local en el que se encuentran, la comida que les han servido, y su opinión acerca de la comida o el local en cuestión. Todo este conjunto de información se puede utilizar para conocer el sentimiento de la gente hacia ese tipo de gastronomía en concreto. Por ello en este trabajo se ha recopilado este tipo de datos para analizar el mercado gastronómico, entendiendo la percepción de los usuarios hacia diferentes tipos de comida, y encontrando correlaciones entre las distintas fuentes de información. Esto ha permitido generar una propuesta de inversión que indica qué tipo de restaurante es más probable que sea bien acogido según los gustos actuales de los consumidores.

Se ha hecho uso de las plataformas Twitter y Google Maps, mediante las cuales se han recopilado las opiniones de los distintos usuarios para crear una fuente de datos con técnicas de raspado web. Posteriormente se ha entrenado un modelo de aprendizaje automático basado en redes neuronales el cual es capaz de generar una predicción sobre el sentimiento de la gente hacia ciertos tipos de comida, con un elevado porcentaje de acierto, permitiendo así elaborar una estrategia de inversión en base a sus resultados. La recolección de datos sobre restaurantes así como la aplicación de los resultados se ha centrado en la ciudad de Valencia. Este trabajo pretende poder adaptarse en futuras investigaciones a otras fuentes de datos disponibles para aumentar su base de conocimiento y ofrecer un resultado más fiable. Además, se puede extrapolar su aplicación a otras ciudades o incluso a otros sectores totalmente distintos.

**Palabras clave:** redes sociales, comida, redes neuronales, análisis del sentimiento, análisis de datos.



## Abstract

Social media platforms contain hundreds of thousands of data points, making them a significant source of information. On platforms like Twitter, nearly 500 million tweets are published daily. This information, shared directly by users of these platforms, provides valuable insight into societal behavior and, most importantly, what people think about various topics.

Social media has become a showcase for many restaurants. It is now very common for consumers to share their experiences, the restaurant they are visiting, the food they have been served, and their opinions about the meal or the establishment itself. All this data can be used to understand the sentiment people have toward a specific type of cuisine. For this reason, this study has gathered such data to analyze the gastronomic market, understanding user perceptions of different types of food and finding correlations between the various sources of information. This has allowed us to generate an investment proposal that indicates which type of restaurant is more likely to be well-received based on current consumer preferences.

To achieve this, we utilized social media platforms like Twitter and Google Maps, through which user opinions were collected using web scraping techniques to create a data source. Subsequently, a machine learning model based on neural networks was trained, which can generate a prediction about people's sentiment towards certain types of food with a high degree of accuracy. This allows us to develop a sound investment strategy based on the results. The data collection focused on restaurants, and the application of the results was centered on the city of Valencia. This work aims to adapt to other available data sources in future research to increase its scope of knowledge and offer even more reliable results. Furthermore, its application can be extrapolated to other cities or even to entirely different sectors.

**Keywords:** food, social networks, neural networks, sentiment analysis, data analysis.



# Índice general

<b>Índice general</b> .....	<b>9</b>
<b>Índice de figuras</b> .....	<b>10</b>
<b>Índice de tablas</b> .....	<b>12</b>
<b>Capítulo 1. Introducción</b> .....	<b>1</b>
1.1. Motivación.....	2
1.2. Propuesta y objetivos.....	3
1.3. Estructura del documento.....	4
<b>Capítulo 2. Fundamentos y estado del arte</b> .....	<b>6</b>
2.1 Software utilizado .....	10
<b>Capítulo 3. Desarrollo y generación de la base de datos</b> .....	<b>11</b>
3.1 Recopilación de datos mediante web scraping .....	11
3.1.1 Base de datos de Twitter .....	12
3.1.2 Desarrollo del programa de web scraping para recopilación de datos de Twitter .....	17
3.1.3 Base de datos de Google Maps.....	24
3.1.4 Desarrollo del programa de web scraping para recopilación de datos de Google Maps .....	27
3.2 Preprocesamiento de los datos recopilados .....	29
3.3 Resultados tras el preprocesamiento de los datos.....	33
3.4 Análisis de los datos recopilados y preprocesados.....	39
<b>Capítulo 4. Aplicación de la red neuronal y propuesta de inversión</b> .....	<b>46</b>
4.1 Base de datos de entrenamiento de la red neuronal.....	46
4.2 Programación de la red neuronal para predicciones del sentimiento .....	48
4.3 Resultados de las predicciones con la CNN.....	54
4.4 Propuesta de sistema de inversión .....	56
<b>Capítulo 5. Conclusiones y trabajos futuros</b> .....	<b>62</b>
<b>Bibliografía</b> .....	<b>64</b>
<b>Apéndice A</b> .....	<b>66</b>
A.1 Comprobación manual de la precisión de la red neuronal para la muestra tomada ...	66
A.2 Muestra de resultados completos de la recopilación de datos de Google Maps .....	73

# Índice de figuras

Figura 1: Búsqueda de palabra clave / Herramienta de desarrollador.....	13
Figura 2: Ubicación de la palabra clave dentro del código web.....	14
Figura 3: Respuesta que contiene la información a procesar. ....	14
Figura 4: Estructura de objetos dentro de la respuesta procesada. ....	15
Figura 5: Texto y metadatos del tuit procesado.....	16
Figura 6: Clave de autenticación de nuestro usuario en Twitter. ....	20
Figura 7: Cursor utilizado para la paginación. ....	21
Figura 8: Contenido del archivo "searches.txt".....	23
Figura 9: Diagrama de flujo del programa de recopilación de datos de Twitter. ....	24
Figura 10: Respuestas de Google Maps al ejecutar la herramienta para desarrolladores.....	25
Figura 11: Ubicación de metadatos para la hamburguesería "Hundred" tras ejecutar búsqueda en Google Maps.....	26
Figura 12: Diagrama de flujo del programa de recopilación de datos de Google Maps. ....	29
Figura 13: Diagrama de flujo del programa de procesamiento y limpieza de tuits.....	31
Figura 14: Diagrama de flujo del procesamiento y limpieza de resultados de Google Maps. ....	33
Figura 15: Datos recopilados de Twitter antes de ser procesados.....	34
Figura 16: Ejemplo de tuit recopilado por error debido al nombre del usuario. ....	34
Figura 17: Ejemplo de tuit con expresiones típicas de las redes sociales y difícilmente comprensibles por un programa.....	35
Figura 18: Representación gráfica del número de tuits recopilados por tipo de comida.....	36
Figura 19: Datos recopilados de Google Maps antes de ser procesados. ....	37
Figura 20: Datos de Google Maps tras ser procesados y almacenados en un archivo de texto.....	38
Figura 21: Representación gráfica del número de restaurantes recopilados por tipo de comida.....	39
Figura 22: Representación gráfica del número de reseñas y tuits totales según categoría. ....	41
Figura 23: Representación gráfica del número de restaurantes y tuits totales según categoría.....	41
Figura 24: Representación gráfica de la puntuación media y tuits totales según categoría.....	42
Figura 25: Representación gráfica del precio medio y tuits totales según categoría.....	43
Figura 26: Archivo original de la base de datos utilizada para el entrenamiento de la red neuronal. ....	47
Figura 27: Representación gráfica de la arquitectura de la red neuronal. ....	52
Figura 28: Evolución del rendimiento del modelo a lo largo de las épocas. ....	54
Figura 29: Representación gráfica de los porcentajes de distribución del sentimiento según categoría. ....	56
Figura 30: Distribución de los porcentajes de sentimiento según categoría tras haber contabilizado los favoritos y los retuits. ....	57
Figura 31: Representación gráfica del número de tuits positivos y de reseñas según categoría. ....	58
Figura 32: Representación gráfica del número de tuits positivos, tuits negativos y restaurantes según	

categoría. ....59

Figura 33: Representación gráfica del número de tuits positivos y de la puntuación media por categoría...59

Figura 34: Representación gráfica del número de tuits positivos y el precio medio por categoría. ....60

# Índice de tablas

Tabla 1: Categorías según tipo de comida.....	11
Tabla 2: Resultados de la recopilación de Twitter por tipo de comida. ....	35
Tabla 3: Número de restaurantes recopilados por tipo de comida. ....	38
Tabla 4: Estadísticas de Google Maps según categoría.....	45
Tabla 5: Estadísticas de Twitter según categoría. ....	45
Tabla 6: Resumen del modelo de red neuronal configurado.....	52
Tabla 7: Resultados del análisis de sentimiento según categoría. ....	55
Tabla 8: Comprobación manual de las predicciones de la red neuronal. ....	66
Tabla 9: Muestra de datos completos extraídos de Google Maps. ....	73



# Capítulo 1. Introducción

En las últimas décadas las redes sociales han transformado radicalmente la forma en que las personas se comunican, comparten información y expresan sus opiniones. Entre estas plataformas, Twitter se destaca por su capacidad de generar y difundir rápidamente una gran cantidad de contenido, con cerca de 500 millones de tuits publicados diariamente (Sayce, 2022). La fuente de datos creada por los propios usuarios ofrece una ventana valiosa en tiempo real para la observación y análisis del comportamiento social y la opinión pública.

En el contexto de la gastronomía, las redes sociales han adquirido un papel fundamental como escaparate de los restaurantes y demás establecimientos de comida (Nikhil Yadav, 2024). Los consumidores de la actualidad se han acostumbrado a compartir sus experiencias en internet, publicando los locales que visitan, los platos que prueban, y sus opiniones sobre el servicio y la calidad de la comida. Estas publicaciones no solo hacen referencia a elecciones personales, sino que pueden dar cuenta de tendencias más amplias en los gustos gastronómicos de la sociedad.

La ciudad de Valencia, con una multitud de opciones gastronómicas en constante evolución, representa un terreno idóneo para investigar la relación entre la actividad en redes sociales y la popularidad de los restaurantes. En este trabajo se propone utilizar datos de las plataformas Twitter y Google Maps para realizar un análisis de sentimiento que permita identificar patrones y correlaciones entre los comentarios de los usuarios y el éxito de diferentes tipos de establecimientos gastronómicos.

Para abordar este desafío será necesario emplear técnicas avanzadas de recopilación de datos web y aprendizaje automático, específicamente modelos basados en redes neuronales, que han demostrado ser herramientas poderosas para el análisis y la predicción en varios campos (Sneha Sukheja, 2024). La base de datos con la cual alimentaremos este modelo estará compuesta por comentarios extraídos de Twitter, donde se puede encontrar una base de datos de comentarios de usuarios, mientras que los datos extraídos de Google Maps aportarán información sobre la popularidad de los restaurantes de Valencia, en forma de reseñas, precios, localización del establecimiento y nombre, entre otros.

El objetivo final de este estudio es desarrollar una propuesta de inversión a partir de la

predicción de sentimiento que identifique nuestro modelo, pudiendo predecir tendencias en la popularidad de los distintos tipos de comida en función de los datos de interacción en redes sociales. Este modelo podría ser implementado para, por ejemplo, dar recomendaciones en cuanto al lugar para abrir un nuevo establecimiento de comida, el tipo de comida que muy probablemente será popular, y el rango de precios más adecuado. A largo plazo, el enfoque propuesto puede ser adaptado para más ciudades y mercados, proporcionando a empresarios en la gastronomía una herramienta valiosa para el análisis del mercado.

La metodología propuesta incluye la recopilación de datos, el preprocesamiento y adaptación de estos, la implementación de modelos de redes neuronales, y la comparación de los distintos datos junto a la predicción de sentimiento del modelo. Los resultados serán analizados, evaluando distintas métricas, para identificar las tendencias y los factores cruciales en el éxito del establecimiento.

## **1.1. Motivación**

La idea del trabajo surge a raíz del aumento constante en la demanda de nuevos locales de restauración y la fama que han ganado en las redes sociales las cuentas dedicadas a exhibir diferentes tipos de gastronomía, así como el actual sentimiento de la sociedad de compartir y opinar en base a sus experiencias. Este trabajo busca contribuir al campo de la investigación sobre el uso de redes sociales como fuente de datos para el análisis de mercado. A pesar de la gran cantidad de información disponible en estas plataformas, su potencial para generar visiones prácticas y actuaciones está aún en gran medida sin explotar en algunos sectores. Al combinar técnicas de recopilación de datos con modelos de aprendizaje automático se espera demostrar cómo los datos pueden ser utilizados de manera efectiva para entender las preferencias de los consumidores.

Mediante la elaboración de este trabajo de investigación se busca obtener un rendimiento de la constante información compartida por los usuarios en plataformas como Twitter o Google Maps, consiguiendo ventaja en el mercado laboral frente a otros inversionistas. El uso de datos de Twitter en la investigación del comportamiento del consumidor relacionado con la alimentación se ha explorado previamente en estudios donde se demuestran cómo las redes sociales pueden proporcionar información espontánea y en tiempo real sobre las preferencias de los consumidores (Larissa S. Drescher, 2023). Este estudio destaca que las metodologías tradicionales, basadas en preguntas directas, pueden resultar en respuestas sesgadas debido a la sobre-racionalización de los consumidores, y propone el análisis de redes sociales como una alternativa viable para obtener evaluaciones más naturales y espontáneas en contextos reales de consumo.

Nuestro trabajo se centra en la ciudad de Valencia y hace uso de otras plataformas como

es Google Maps, pudiendo llegar a ser utilizado con otros fines, como es el análisis estadístico del consumo gastronómico de una población o un país. Se podrían así analizar estos resultados en búsqueda de campañas para promover, por ejemplo, por parte del gobierno, otro tipo de alimentos más saludables o identificar si la gastronomía local de un territorio se va perdiendo debido a la apertura de nuevos locales con gastronomía extranjera. También se podría utilizar como fuente de ideas para campañas de marketing, conociendo los gustos actuales de la gente.

Sin embargo, a la hora de acceder a este tipo de información de manera masiva y automatizada, encontramos que el acceso está limitado a través de una API (Interfaz de Programación de Aplicaciones) de pago. Recientemente, estas funciones que permiten integrar sistemas permitiendo que sus funcionalidades se puedan utilizar dentro de otros softwares, se han convertido en servicios de pago, como es el caso de las APIs de Twitter y Google Maps, lo que limita su accesibilidad para proyectos académicos sin presupuesto asignado. Ante esta situación, se propone utilizar técnicas de “web scraping” para la recopilación de datos, aprovechando herramientas de código abierto y desarrollando programas específicos para extraer información relevante de estas plataformas.

De este modo, además de diseñar un modelo de aprendizaje automático capaz de arrojar una estrategia de inversión lo suficientemente fiable, se va a llevar a cabo un estudio sobre el modo de funcionamiento de estas plataformas y cómo sus interfaces responden a las interacciones del usuario. Entendiendo el comportamiento de estas plataformas ante las distintas acciones se pueden capturar estas respuestas que contienen la información que buscamos para generar nuestra base de datos.

## **1.2. Propuesta y objetivos**

El objetivo principal de este trabajo es investigar la relación entre la interacción de los usuarios de Twitter y Google Maps con la popularidad de los restaurantes en Valencia, utilizando técnicas de análisis de sentimiento y aprendizaje automático. Mientras algunos trabajos similares se centran en conocer las preferencias de los consumidores mediante la búsqueda genérica de “desayuno, almuerzo, merienda y cena” (Vida, 2015), en nuestro trabajo vamos a centrarnos en la búsqueda en concreto de diferentes tipos de gastronomía. Además, se pretende generar un programa capaz de ser adaptado a otros sectores, es decir, que recopile otro tipo de información completamente distinta, y que se pueda utilizar con fines de análisis estadístico de cualquier tema. Como objetivo complementario, se pretende entender la manera en la que estas plataformas almacenan los datos, sin necesidad directa de consultar a su API de pago.

Para conseguir nuestros objetivos, en primer lugar se implementarán técnicas de “web scraping” para obtener los datos de Twitter y Google Maps. Para ello se creará una cuenta de

usuario expresamente para el trabajo de investigación. En el caso de Twitter se ha creado la cuenta @TFMSGJ, mientras que para el caso de Google Maps se ha creado la cuenta de Google “[tfmsgj@gmail.com](mailto:tfmsgj@gmail.com)”. A través de estas cuentas se va a interactuar con los elementos de cada plataforma para ver cómo se comportan las peticiones de la web, observando estas mediante las herramientas de desarrollador que ofrece el navegador, en este caso Mozilla Firefox. Con esto se pretende diseñar un programa en Python capaz de automatizar la búsqueda de tuits y el almacenamiento de estos en un archivo del tipo “JSON”, de fácil manejo, para poder trabajar posteriormente los datos.

Según el límite de peticiones que permitan las plataformas se montará una base de datos con una cantidad de información lo suficientemente grande como para poder entrenar el modelo. Antes se preprocesarán estos, de manera que se eliminen signos de puntuación, caracteres especiales, imágenes, emojis, u otros símbolos que puedan entorpecer el aprendizaje de nuestro modelo. Una vez generada la base de datos, se desarrollará el código de la red neuronal mediante Python, el cual hará uso de una serie de datos previamente clasificados sacados de internet los cuales le ayudarán a aprender cuándo un comentario es positivo o negativo, para así poder posteriormente entender la información recopilada en la base de datos y dar los resultados esperados.

### **1.3. Estructura del documento**

La estructura de este trabajo fin de máster se organiza en varias secciones que cubren todos los aspectos de la investigación, desde la introducción hasta las conclusiones y recomendaciones. Cada sección se desarrolla en detalle para proporcionar una comprensión completa del estudio y sus resultados.

En la Introducción se plantea el contexto del problema, así como la relevancia del estudio, describiendo brevemente el impacto de las redes sociales en nuestra sociedad moderna, centrándonos en Twitter y Google Maps, así como la importancia de estas plataformas como fuentes de información para el análisis de preferencias de los consumidores. Se define el problema que se aborda en el estudio: la relación entre la interacción en redes sociales y la popularidad de los restaurantes en Valencia. Además de este objetivo principal, se definen los objetivos específicos para alcanzarlo:

- Implementar técnicas de “web scraping” para recopilar datos.
- Aplicar análisis de sentimiento a los datos obtenidos mediante un modelo predictivo basado en aprendizaje automático.
- Analizar los resultados de la recopilación de datos y análisis de sentimiento,

comparando distintas métricas entre sí.

- Evaluar la correlación entre las diferentes métricas observadas y estudiar la viabilidad de diseñar una estrategia de inversión basada en éstas.

En el Estado del Arte llevaremos a cabo un análisis detallado de estudios previos relacionados con el uso de redes sociales para la investigación del comportamiento del consumidor y de las técnicas de análisis de sentimiento y aprendizaje automático utilizadas. Además se identificarán áreas que no han sido suficientemente exploradas en investigaciones previas.

Para desarrollar el modelo, primeramente, se recopilarán los datos necesarios de las redes sociales Twitter y Google Maps para posteriormente preprocesarlos. Se aplicarán técnicas de limpieza de datos, procesos de “tokenización”, eliminación de enlaces, emojis u otros, y se normalizarán los datos. Se definirá la arquitectura de la red neuronal, el proceso de entrenamiento y validación del modelo, haciendo uso de una base de datos previamente clasificada.

En los resultados de las predicciones del modelo se compararán las métricas de rendimiento. Basado en estas métricas, y tras analizar la correlación entre los distintos datos recopilados, se generarán recomendaciones sobre la viabilidad de las estrategias de inversión. Se discutirán las tendencias observadas y su relevancia, comentando posibles factores que podrían haber afectado los resultados obtenidos y las restricciones inherentes, así como el impacto potencial de los resultados en el sector gastronómico y las contribuciones del estudio al campo del análisis de datos de redes sociales.

## Capítulo 2. Fundamentos y estado del arte

Debido a la gran influencia que ejercen hoy en día las redes sociales sobre la sociedad, existen múltiples estudios que tratan la recopilación de datos de éstas para investigar el comportamiento o los gustos de los usuarios de tales plataformas. Muchos de estos trabajos hacen uso de técnicas de Aprendizaje Automático, también conocido en inglés como “Machine Learning” (ML) (Hayat, 2019). El Aprendizaje Automático es una disciplina dentro del campo de la Inteligencia Artificial. Esta, mediante algoritmos que procesan una gran cantidad de datos y se encargan de analizarlos, puede aprender cómo estos se relacionan entre ellos y, por lo tanto, realizar predicciones sobre el comportamiento de los mismos. Dentro del ML podemos encontrar tres categorías de algoritmos (Sandoval, 2018):

- Aprendizaje supervisado: El usuario o programador del algoritmo alimenta éste con una fuente de datos previamente etiquetada, es decir, le indica desde un inicio cómo debe interpretar cada uno de los valores. El algoritmo deberá encontrar cuál es la causa o relación entre estos datos para haber sido clasificados así, aprendiendo de las observaciones y siendo capaz de realizar predicciones lo suficientemente precisas.
- Aprendizaje no supervisado: En este tipo de algoritmos, a diferencia del anterior, no se cuenta con una fuente de datos etiquetada. El algoritmo deberá encontrar patrones para organizar los datos e intentar describir su estructura.
- Aprendizaje por refuerzo: El algoritmo aprende en base a su propia experiencia, es decir, imitará un proceso de aprendizaje motivado por recompensas que el usuario del algoritmo define previamente. Estas recompensas definen el objetivo de la tarea de forma no explícita. Se trata, por tanto, de un tipo de algoritmo que se nutre tanto de la filosofía del aprendizaje supervisado como del no supervisado.

En nuestro caso de estudio vamos a emplear algoritmos basados en aprendizaje supervisado usando redes neuronales artificiales.

Las redes neuronales artificiales son un tipo de tecnología dentro del campo del aprendizaje automático e inteligencia artificial inspiradas en la estructura y el funcionamiento del cerebro humano. Estas se basan en una serie de capas de nodos que, al igual que las neuronas, están interconectados entre sí, procesando información. Estas son capaces de reconocer patrones complejos en datos, pudiendo aprender de estos y tomar decisiones acerca de predicciones futuras. Por ello, en el contexto de nuestro problema, estas capacidades son

especialmente útiles para el análisis de sentimiento de los tuits.

Para este estudio vamos a hacer uso de una red neuronal del tipo convolucional, o CNN por sus siglas en inglés (“Convolutional Neural Network”). En el caso de la información obtenida de Twitter, contamos con multitud de fragmentos de texto generalmente cortos, los cuales presentan diferentes desafíos: cuentan con un lenguaje informal, utilizan abreviaturas, menciones a otros usuarios, o expresiones propias de la plataforma. Por ello escogemos una red neuronal convolucional, pues éstas han demostrado ser eficaces tanto en tareas de procesamiento de lenguaje natural como el análisis de sentimientos en textos cortos (Wang, 2018).

La base de este tipo de redes neuronales es la capa de convolución. Esta capa aplica un filtro sobre la entrada para producir un mapa de características. En este contexto, la entrada de la red neuronal es una secuencia de palabras o “tokens”, sobre la cual se realiza la convolución. Un filtro es un pequeño conjunto de pesos (de las neuronas) que se aplica de forma deslizante sobre la secuencia de texto. En cada posición, el filtro realiza una multiplicación punto a punto entre los pesos del filtro y los valores de entrada, generando un valor que se almacena en el mapa de características. A medida que el filtro se va “deslizándose” por la secuencia de texto genera una serie de valores que formarán el mapa de características definitivo. Este mapa resaltará patrones locales que se hayan detectado en la secuencia, como la presencia de combinaciones de palabras que puedan ser indicativas de un sentimiento específico: “me encanta”, “odio” “excelente servicio”, etc. Esto es una tarea muy complicada en nuestro trabajo pues el lenguaje utilizado en Twitter es muy variado, siendo diferente la manera en la que las personas expresan su sentimiento positivo o negativo sobre un tema.

Después de la fase de convolución se suele aplicar una capa de “pooling”, la cual reduce la dimensionalidad del mapa de características. Esto se hace tomando el valor máximo dentro de una ventana de características, reteniendo aquellas más importantes mientras se reduce la complejidad de computación. Una vez extraídos y reducidos los patrones locales, las redes neuronales convoluciones suelen hacer uso de capas densas, las cuales toman los mapas de características procesados y los utilizan para realizar la clasificación final. En nuestro caso, esta clasificación será del tipo binaria (positivo o negativo).

En comparación con otras arquitecturas como las redes neuronales recurrentes (RNN) del tipo memoria larga a corto plazo (LSTM), que suelen ser más pesadas computacionalmente, las CNN son más eficientes. Los tuits son inherentemente cortos y a menudo carecen de la complejidad estructural que justificaría el uso de modelos más complejos como LSTM. Las CNN pueden extraer las características necesarias para clasificar el sentimiento sin necesidad de modelar relaciones a largo plazo en el texto.

El uso de este tipo de algoritmos en el análisis de sentimientos ha demostrado ser altamente eficaz, particularmente cuando se trata de texto no estructurado como pueden ser los comentarios en redes sociales. En diferentes investigaciones se ha demostrado que las redes neuronales pueden lograr resultados muy buenos en tareas de clasificación de texto (Soni, 2022). En concreto, los modelos del tipo CNN han mostrado ser considerablemente mejores en comparación a los modelos estadísticos tradicionales en la predicción de sentimientos a partir de tuits (Liao, 2017).

Algunos estudios previos hacen uso de técnicas de “Machine Learning” para analizar tendencias de marketing en redes sociales como Twitter debido a su enorme potencial. Twitter se trata de una plataforma de tipo “microblogging” que permite comunicarse e interactuar con otros usuarios a través de unidades de contenido textual breve, principalmente pensados para compartir opiniones y comentarios sobre una variedad diversa de temas.

Tal es la capacidad de estos algoritmos que en campañas de marketing como “Black Friday” se llevó a cabo un estudio empleando este tipo de técnicas para analizar el sentimiento de los consumidores a través de los comentarios que realizaban en Twitter (Saura, 2018). Este estudio se centró en aquellas discusiones referentes a las distintas plataformas de venta online y las ofertas que publicaban durante el periodo que cubrió la campaña. Los resultados arrojaron información valiosa sobre cómo los usuarios percibían grandes empresas como Amazon, Ebay, MediaMarkt, o el Corte Inglés, permitiendo a estas compañías entender cómo estaban resultando sus estrategias de marketing durante la campaña y si cabía margen de mejora. Otro estudio también basado en el uso de la API de Twitter se empleó para analizar los mensajes en español de Twitter que hablaban de la empresa de comida rápida “KFC” (Morales, 2022). Para ello los tuits fueron capturados durante todo un trimestre, descubriendo cuáles eran los sentimientos relacionados con cada mensaje y entendiendo así la polaridad de estos entre positivo y negativo. Los estudios sirvieron para comprobar cómo esta empresa obtuvo una percepción mayormente positiva en la plataforma durante el primer trimestre del 2022, comprobando así si sus estrategias de marketing estaban siendo efectivas.

Si bien existen múltiples estudios que analizan el sentimiento de los usuarios de Twitter respecto a una compañía o un tema en concreto, resultan más escasos los estudios dedicados a analizar la percepción sobre un amplio rango de productos alimenticios. Encontramos el uso de Twitter como fuente de datos para analizar la percepción sobre alimentos ultraprocesados (Amarante, 2019), pero éste se basa en un diccionario de palabras sin llegar a hacer uso de técnicas de inteligencia artificial para obtener unos resultados más precisos y que se adapte al lenguaje típico de una red social como Twitter.

En otro estudio, elaborado en 2016 (Park, 2016), diferentes estudiantes recolectaron alrededor de 86.000 tuits en relación a cuatro tipos de restaurantes asiáticos: chino, japonés,

coreano y tailandés. A través del análisis de estos tuits, los resultados mostraron que los restaurantes chinos recibieron una percepción significativamente más negativa en comparación con los otros tres tipos de restaurantes asiáticos. Este estudio sirvió como marco metodológico para entender los puntos negativos y positivos de estos restaurantes, pudiendo observar que los tuits positivos hacían referencia a la calidad de la comida, mientras los negativos señalaban problemas relacionados con la calidad del servicio.

Sin embargo, nuestro estudio se diferencia de éste y otros estudios previos en varios aspectos clave. En primer lugar, nos centramos no sólo en un tipo de restaurante o cultura gastronómica, sino que abarcamos un rango más amplio recopilando diferentes tipos de comida. Esto resultará en una investigación mucho más nutritiva a la hora de hacer un análisis de mercado, aportando una visión más global sobre cómo diferentes cocinas son percibidas en un mercado local específico como es la ciudad de Valencia. Este enfoque permitirá desarrollar estrategias de inversión más variadas y versátiles basadas en el análisis de los resultados de nuestra recopilación de datos.

Por otro lado, los datos ofrecidos por Google Maps son también cada vez más utilizados en la investigación de mercados, especialmente en estudios que buscan optimizar la localización de negocios o analizar su popularidad. En uno de ellos se utilizaron datos de Google Maps para estudiar la ubicación de diferentes restaurantes y su correlación con la demografía y el poder adquisitivo de los barrios bajo estudio (Yang, 2016). En nuestro trabajo utilizaremos los datos ofrecidos por esta red social para complementar nuestro análisis de sentimientos extraído de Twitter, proporcionando un contexto que permitirá correlacionar estos resultados con el éxito de los restaurantes en Valencia. De este modo el enfoque aquí presentado se diferencia también en la integración de diferentes fuentes de datos, algo que no suele ser común en otros estudios.

A diferencia de la gran mayoría de estudios basados en Twitter, los cuales hacen uso de la API (interfaz de programación de aplicaciones), en nuestro caso no contamos con acceso directo a este tipo de consultas. Esto es debido a que, recientemente, el acceso a esta API ha sido privatizada y se requiere de un pago o en su lugar de una aceptación previa por parte de la administración de Twitter. Por ello en nuestro estudio se ha diseñado y elaborado una herramienta propia de recopilación automática de datos, basada en técnicas conocidas como “web scrapping”.

El “web scraping” o raspado web es una técnica utilizada para extraer información de sitios web de manera automatizada (Lofti, 2020). Esta técnica implica la creación de programas que navegan por páginas web, descargan su contenido y extraen datos específicos, como textos, imágenes, enlaces, u otros elementos de interés. A diferencia de la extracción manual de datos, el “web scraping” permite obtener grandes volúmenes de información de

manera eficiente y rápida, lo que lo convierte en una herramienta poderosa para diversas aplicaciones, desde la investigación académica hasta el análisis de mercados y la monitorización de datos.

En este estudio utilizaremos el “web scraping” para recopilar datos tanto de Twitter como de Google Maps, específicamente comentarios y menciones relacionados con diferentes tipos de comida en la ciudad de Valencia, así como los restaurantes que entre sus servicios ofrecen estas gastronomías. Dado que se ha denegado el acceso a las “APIs” oficiales de estas redes sociales, el “web scraping” se presenta como una alternativa viable para obtener los datos necesarios.

## 2.1 Software utilizado

Para el desarrollo de nuestro trabajo, se ha seleccionado Python como lenguaje de programación, debido a su versatilidad y amplia variedad de librerías disponibles para el análisis de datos y machine learning. Python es especialmente utilizado en la comunidad científica debido a su capacidad de manejar grandes volúmenes de datos.

A continuación, se mencionan las principales bibliotecas utilizadas en cada una de las etapas del trabajo, según su finalidad:

- Web scraping y recolección de datos: Para las fases de recopilación y procesamiento de datos de Twitter y Google Maps se utilizarán las librerías de “Requests”, “Browsercookie”, “Selenium”, “PyAutoGUI”, “JSON” y “Re”.
- Gestión del tiempo y control de flujo: Para gestionar las fechas y tiempos de ejecución se utilizarán las librerías “Time” y “DateTime”.
- Red neuronal y análisis de sentimientos: Utilizaremos las librerías de “TensorFlow” y “Keras” (Google) la creación y entrenamiento de la red neuronal, mientras que “NumPy” y “Matplotlib” nos permitirán realizar operaciones matemáticas y visualizar los resultados.

# Capítulo 3. Desarrollo y generación de la base de datos

## 3.1 Recopilación de datos mediante web scraping

Para recopilar datos que sirvan para nuestro análisis aplicaremos técnicas de “web scraping”, para así extraer información en la que se mencionen palabras clave relevantes, como "paella", "tapas españolas", y "hamburguesa", entre otras que consideramos puedan ser de las más nombradas en Twitter. Además, estas palabras servirán para generar diferentes búsquedas en Google Maps y recopilar la mayor cantidad de restaurantes posibles. Esta selección de palabras clave se ha hecho en base a las categorías de gastronomía más demandadas en la ciudad de Valencia, agrupando estas conforme se puede apreciar en la Tabla 1.

*Tabla 1: Categorías según tipo de comida.*

Categoría	Alimento
Comida valenciana	Paella, arroz a banda, arroz al horno, fideua, esmorzar, esgarraet
Comida española	Tapas españolas, patatas bravas, tortilla española, calamares a la romana, croquetas, ensaladilla
Comida italiana	Comida italiana, pasta, pizza, lasaña, tiramisú, carbonara
Pescado	Marisco, mariscada, pescado, gambas, langosta, pulpo a la brasa
Carne	Carne, carne a la brasa, hamburguesa, asado, barbacoa, chuletón
Comida internacional	Comida china, comida japonesa, kebab, comida india, comida mexicana, comida latina
Cafetería/pastelería	Horchata, helado, café, churros, fartons, pastel

Para las categorías de comida valenciana, comida española y comida italiana, se han seleccionado platos tradicionales considerados pilares de estas gastronomías, como puede ser

la “paella” para la cocina valenciana, las “tapas” para la comida típica española y la “pizza” para la comida italiana. Estos alimentos son de reconocimiento mundial y es muy fácil encontrar tuits así como restaurantes centrando las búsquedas en estas palabras, por lo que resultará más eficiente la recopilación de datos.

Para las categorías de carne y pescado se han escogido alimentos muy comunes en Valencia. Por un lado, nos hemos centrado en comida típica de ciudades con mar, como puede ser la “mariscada” o las “gambas”; mientras que para los restaurantes que ofrecen carne, hemos buscado palabras claramente identificativas de este tipo de locales, como puede ser “asado” o “chuletón”, así como “hamburguesa”, pues es bien sabido que Valencia es una de las ciudades con más hamburgueserías de España.

En cuanto a la comida internacional, hemos decidido incluir de manera genérica aquellos tipos de comida de otros países que consideramos más frecuente de encontrar en Valencia tras haber observado en webs como TripAdvisor que la lista de búsquedas más recurrentes en la ciudad contienen platos típicos de la gastronomía mexicana, india, oriental, árabe y latina. Además, cabe destacar la gran multiculturalidad de Valencia, la cual cuenta con un gran barrio chino en el centro de la ciudad.

Por último, para recopilar datos de los locales valencianos que ofrecen cafés o postres, nos hemos centrado en clásicos de la ciudad como la “horchata” o “fartons”, pero abarcando también palabras más genéricas como “café” o “helado” que ayudarán a recopilar más resultados.

La recopilación de estos datos será fundamental para formar nuestra propia base de datos con tuits a partir de los cuales analizar su sentimiento y, por tanto, el sentimiento hacia este tipo de comida. Además, se pretende recopilar la gran mayoría (si no todos) de los restaurantes ubicados en Valencia, formando así una segunda base de datos con información valiosa sobre el rendimiento de estos restaurantes y otros aspectos clave que servirán a la hora de buscar relaciones con los datos obtenidos de Twitter.

### **3.1.1 Base de datos de Twitter**

Para realizar la extracción de datos de Twitter hemos creado una cuenta de usuario dedicada exclusivamente a esta tarea. Con el fin de comprender mejor la estructura interna y las interacciones dentro de Twitter, hemos iniciado sesión en esta cuenta y hemos realizado un análisis detallado del código web de la plataforma. Este análisis permite desentrañar las peticiones que se generan al interactuar con los tuits, las respuestas del servidor y, sobre todo, la localización de los datos relevantes, como son las interacciones y métricas asociadas a los tuits. Para llevar a cabo este análisis técnico hemos utilizado el navegador Mozilla Firefox,

que cuenta con herramientas de desarrollo para inspeccionar el código de la página, el cual está escrito en lenguaje “JavaScript”. La herramienta de desarrollador de Firefox se trata de un set de desarrollador web dentro de Firefox el cual se puede utilizar para examinar, editar y “debuguear” código en HTML, CSS y JavaScript (Firefox, 2024).

Para estudiar cómo funciona internamente Twitter vamos a ejecutar una búsqueda controlada utilizando la palabra clave “hamburguesa”. Esta búsqueda permitirá rastrear y monitorizar cómo se generan y cómo se presentan las respuestas de Twitter a nivel de código. Se han anonimizado todos los nombres de usuario así como imágenes para proteger la privacidad de las personas.

Lanzando la aplicación incluida en nuestro navegador “Herramienta para desarrolladores”, procedemos a inspeccionar las respuestas de la web (Figura 1).

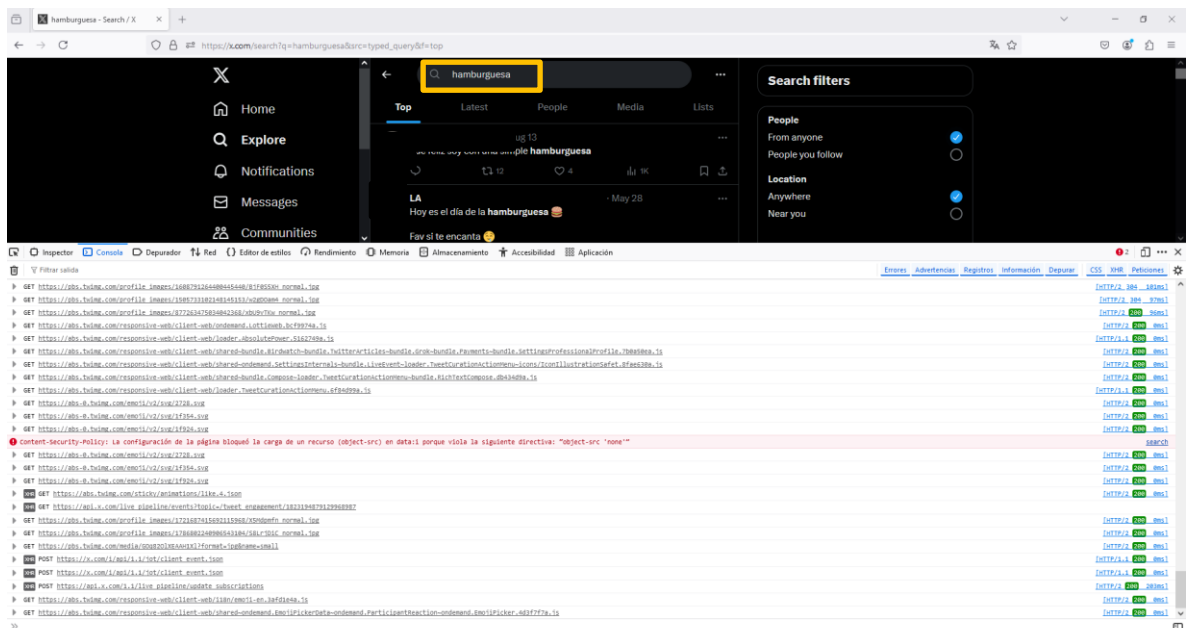


Figura 1: Búsqueda de palabra clave / Herramienta de desarrollador.

Al iniciar la herramienta de desarrolladores (Figura 1), podemos apreciar como en la consola aparecen multitud de respuestas de la web ante las diferentes solicitudes internas de la plataforma, y cómo éstas van aumentando constantemente, pues se van actualizando en tiempo real. Esto se debe a que Twitter, una plataforma con un flujo masivo de datos, implementa diversas tecnologías avanzadas para gestionar y optimizar la entrega de información a sus usuarios.

Una de estas tecnologías clave que utiliza la plataforma es GraphQL, un lenguaje de consulta flexible y eficiente, desarrollado originalmente por Facebook en 2012, y liberado como código abierto en 2015. Dado que Twitter maneja grandes cantidades de datos, como

tuits, retuits, respuestas, menciones, y más, GraphQL ofrece una solución eficiente para gestionar estas consultas. Por ejemplo, un cliente de Twitter puede utilizar GraphQL para solicitar sólo los tuits o respuestas de un usuario específico, o para obtener datos de múltiples fuentes (como tuits y sus correspondientes metadatos) en una sola solicitud.

Para optimizar nuestro proceso de “web scraping” hemos decidido explorar las consultas realizadas a través de GraphQL en la consola de desarrolladores, buscando la palabra “GraphQL” dentro de la propia consola. Al realizar búsquedas específicas dentro de esta interfaz podremos identificar de manera más precisa los tuits y sus metadatos correspondientes, lo que permite extraer exactamente la información que necesitamos para nuestro análisis.

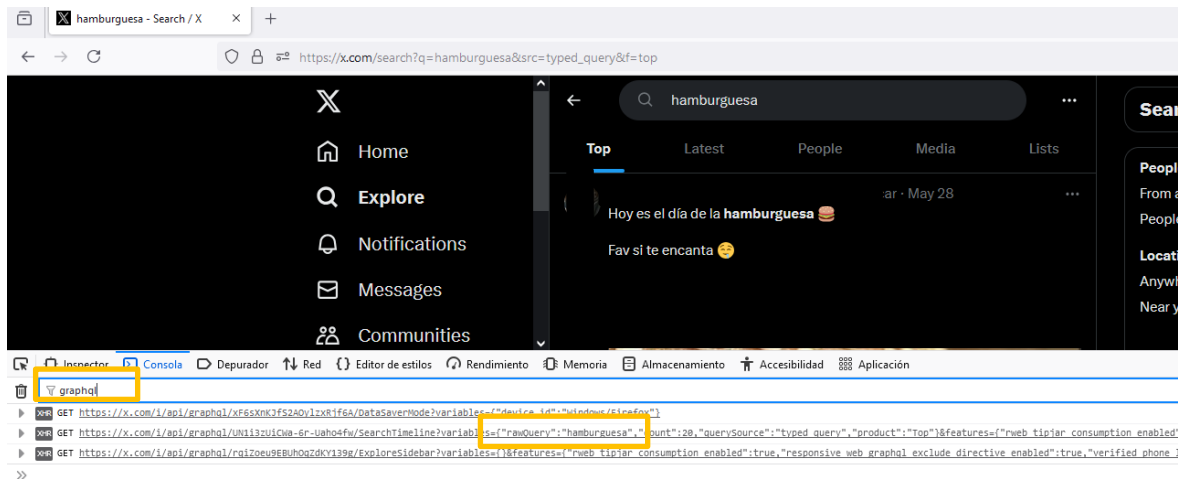


Figura 2: Ubicación de la palabra clave dentro del código web.

Al inspeccionar el código y las respuestas de la consola de desarrollador observamos que la palabra clave "hamburguesa" utilizada en la búsqueda en Twitter aparece vinculada a una variable llamada “rawQuery” dentro de las peticiones realizadas por el navegador (Figura 2). Esta observación es crucial ya que “rawQuery” actúa como un identificador de la consulta específica.

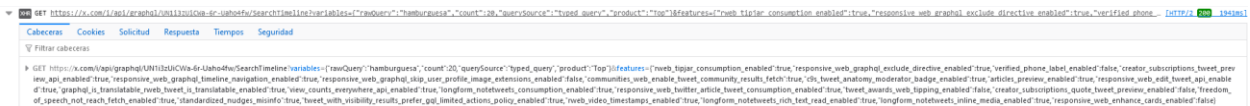


Figura 3: Respuesta que contiene la información a procesar.

Procedemos entonces a desglosar la respuesta proporcionada por la consulta, puesto que en el lenguaje “JavaScript” este tipo de datos se representan de manera jerárquica. Este formato de datos recibe el nombre de “JSON” (JavaScript Object Notation), el cual se trata de un formato de texto ampliamente utilizado para el intercambio de datos entre un servidor y una aplicación web (ECMA, 2024). Dentro de la estructura JSON devuelta (Figura 3),

identificamos el elemento principal denominado “data”, que agrupa la información solicitada. Así mismo, dentro de “data” encontramos un subelemento denominado “entries” el cual contiene una lista de objetos correspondientes a los tuits recuperados para la búsqueda de "hamburguesa" (Figura 4). Cada objeto dentro de “entries” representa un tuit individual asociado a nuestra búsqueda. Observamos que, de manera habitual, el número de “entries” que aparece en cada búsqueda es 20, aunque en algunas ocasiones este número cambia en una unidad más o menos. Esto significa que los tuits visibles en una primera instancia serán 20 o menos.

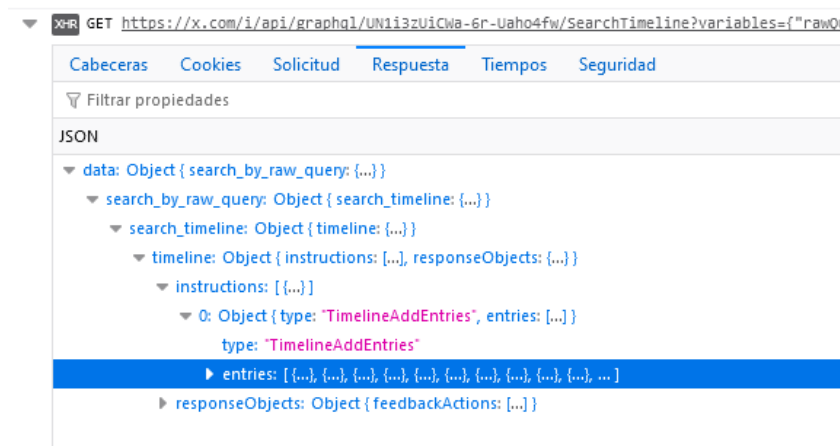


Figura 4: Estructura de objetos dentro de la respuesta procesada.

Al explorar estos objetos “entries” de manera jerárquica, siguiendo la estructura “Object > content > itemContent > tuit\_results > result > legacy”, finalmente llegamos al contenido textual de cada tuit, almacenado bajo la variable “full\_text”. Además del texto del tuit, esta ruta lleva a otros metadatos relevantes que Twitter asocia con cada publicación, como el número de favoritos (“favorite\_count”), el número de retuits (“retuit\_count”), el identificador único del tuit (“id\_str”), el identificador del usuario que publicó el tuit (“user\_id”), y el número de respuestas (“reply\_count”). Cada uno de estos elementos se encuentra claramente delimitado dentro de la estructura JSON y puede ser identificado fácilmente por su nombre de variable (Figura 5).

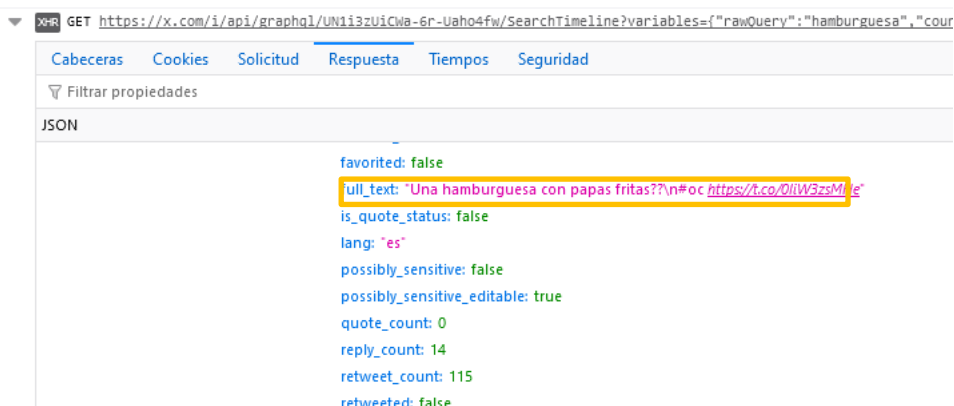


Figura 5: Texto y metadatos del tuit procesado.

Habiendo identificado la estructura de datos y el modo en que Twitter organiza la información dentro de sus respuestas del tipo JSON, el siguiente paso es la elaboración de un programa automatizado para la extracción de tuits. Este programa se encargará de realizar consultas dirigidas y de procesar los resultados obtenidos para generar un archivo en formato JSON. Este archivo replicará la estructura jerárquica empleada por Twitter, de modo que en este paso ahorramos la necesidad de transformaciones adicionales reduciendo los errores durante la recolección. Además, facilitará futuras manipulaciones y análisis de los datos obtenidos, ayudando a realizar las inspecciones visuales en búsqueda de errores durante las primeras recolecciones de datos.

Para nuestro caso de estudio, los elementos clave que se almacenarán en este archivo incluyen:

- **Identificador del tuit:** Un código único (“id\_str”) que permite identificar de manera inequívoca cada tuit.
- **Identificador del usuario:** El “user\_id”, que vincula el tuit al autor correspondiente.
- **Contenido del tuit:** El texto completo del tuit (“full\_text”), que es el objeto principal de nuestro análisis.
- **Número de favoritos:** El contador (“favorite\_count”) que indica cuántos usuarios han marcado el tuit como favorito.
- **Número de retuits:** El contador (“retuit\_count”) que muestra cuántas veces ha sido retuiteado el tuit.
- **Citas:** La cantidad de citas o comentarios adicionales que el tuit ha generado, almacenado en la variable “quote\_count”.

- **Guardados:** Un indicador del número de veces que el tuit ha sido guardado por otros usuarios, almacenado en la variable “bookmark\_count”.
- **Respuestas:** El número de respuestas directas (“reply\_count”) que ha recibido el tuit.

La inclusión de elementos como el identificador del tuit y del usuario servirá para descartar tuits duplicados en el futuro, puesto que la recolección se va a ejecutar en distintas tandas y cabe la posibilidad de recolectar tuits ya procesados. Así mismo, datos como el número de favoritos o de retuits ayudará a añadir importancia al sentimiento del tuit en cuestión, pues estos serán indicadores de que más gente comparte esa opinión. Si un tuit contiene favoritos o retuits, por cada uno de estos se contabilizará como una persona adicional que comparte la opinión de dicho tuit.

Aunque no todos estos datos serán utilizados directamente en el análisis de este trabajo (identificador del usuario, guardados, o citas), se ha decidido incluirlos en el archivo generado. Esta información podría ser de utilidad en futuras investigaciones o en la implementación de mejoras adicionales al sistema que estamos diseñando. La preservación de estos metadatos también ofrece una mayor flexibilidad para realizar análisis complementarios o explorar nuevas preguntas de investigación que puedan surgir en el transcurso del proyecto.

### 3.1.2 Desarrollo del programa de web scraping para recopilación de datos de Twitter

Para llevar a cabo el desarrollo del programa para la recolección masiva de datos vamos a utilizar el lenguaje de programación Python, debido a su versatilidad y amplia disponibilidad de bibliotecas especializadas en la manipulación de datos web, razón por la cual es conocido como un estándar en la comunidad de ciencia de datos y análisis web, siendo a fecha de hoy el lenguaje de programación más usado en el mundo (Baldur, 2023; Tiobe, 2024). El programa desarrollado en este trabajo hace uso de varios módulos de Python, los cuales se describen a continuación junto a la razón de su elección:

- **requests:** Este módulo es fundamental para realizar solicitudes HTTP, lo que permite interactuar con la API de Twitter. Es conocido por su simplicidad y su capacidad para gestionar cookies y sesiones de manera integrada, lo cual lo convierte en una herramienta perfecta para tareas que requieren autenticación continua y manejo de cookies, como es el caso de nuestra implementación.
- **browsercookie:** Este módulo se utiliza para extraer cookies de un navegador web, lo que permite que nuestro programa se autentique automáticamente utilizando las

cookies ya existentes de una sesión iniciada en el navegador. Esto es crucial para acceder a plataformas que requieren autenticación persistente, como Twitter, sin necesidad de solicitar credenciales repetidamente. En nuestro caso hemos empleado este módulo para asegurar que las solicitudes realizadas por el programa sean tratadas como si fueran originadas desde un navegador autenticado.

- **json:** Este módulo es indispensable para trabajar con datos en formato JSON. Dado que los datos obtenidos de Twitter incluyen información estructurada en este formato, es una herramienta clave para procesar y almacenar estos datos de manera organizada.
- **sys y time:** Estos módulos proporcionan funciones auxiliares. EL módulo “sys” se utiliza para manejar argumentos de línea de comandos y control del flujo de ejecución, mientras que “time” permite la implementación de pausas en la ejecución del programa, lo cual es útil para evitar superar los límites de tasa de las APIs y para sincronizar las solicitudes realizadas a la API de Twitter. También ayuda a evitar ser detectados como un “bot”, lo cual conllevaría a la suspensión de nuestra cuenta en estas plataformas.
- **datetime:** Este módulo se emplea para gestionar fechas y horas, lo cual es de gran importancia pues la fecha actual será usada para nombrar los archivos generados.

Cabe indicar que Twitter muestra distintos resultados para una misma palabra en función de si el usuario escoge leer entre los tuits más relevantes (con más interacciones) o los tuits más recientes. Para nuestro estudio hemos escogido el segundo caso ya que al escoger los tuits más relevantes se dificulta mucho la recolección de nuevos datos; esto es debido a que la aparición de tuits en la sección de relevantes no se actualiza de manera inmediata, a diferencia de los tuits recientes que aparecen en tiempo real según vayan siendo escritos por los usuarios.

En nuestro programa se define la clase “TweetData” como primera pieza clave para el funcionamiento. Esta clase encapsulará todos los datos relevantes de cada tuit en un objeto, facilitando así su manipulación y almacenamiento. Esta define varios atributos que corresponden a la información esencial de un tuit, como su identificador (“tweet\_id”), el identificador del usuario (“user\_id”), el texto del tuit (“text”) y métricas de interacción como favoritos (“favorites”), retuits (“retweets”), citas (“quotes”), guardados (“bookmarks”) y respuestas (“replies”).

Dentro de esta clase se crea el método “\_\_repr\_\_” para definir la representación en cadena del objeto en cuestión (tuit), cuya información contiene los atributos mencionados anteriormente. Esto permitirá obtener una descripción legible del objeto que facilitará la

interpretación de los datos que contiene ese tuit específico.

Crearemos un método estático `“from_tweet_entry”`, el cual permitirá crear instancias de `“TweetData”` a partir de una estructura de datos JSON, que es la forma en la que los tuits son devueltos. En éste verificaremos si los datos necesarios para extraer la información del tuit están presentes en la entrada JSON proporcionada. Al lanzar las consultas a Twitter hemos observado como, en alguna ocasión, la página devuelve los datos en diferentes estructuras que, si bien son muy parecidos entre sí, presentan pequeñas diferencias. Por esta razón el método `“from_tweet_entry”` se encargará de manejar estas variaciones, verificando primero si el contenido del tuit (`“itemContent”`) y el resultado de la consulta (`“tweet_results”`) están presentes dentro de la respuesta JSON. Si cualquiera de estos elementos falta, la función retornará `“None”`, indicando que no es posible crear un objeto `“TweetData”` válido.

Una vez confirmada la presencia de los datos anteriores, se extraerán los datos del tuit en su formato `“legacy”` analizado con anterioridad a través de la herramienta para desarrolladores. Este formato se refiere, en el contexto del código de Twitter, a una versión de datos más sencilla que agrupa toda la información relevante de un tuit en un único bloque de datos, facilitando la extracción de la información deseada.

Finalmente utilizamos la información extraída para crear una instancia de `“TweetData”` asignando así todos los datos a los atributos correspondientes del objeto. Una vez asignados todos los valores se devuelve la instancia de `“TweetData”` completada, lista para ser utilizada.

Para poder establecer la sesión que se utilizará para realizar solicitudes HTTP a la API de Twitter se crea una nueva función llamada `“setup(auth_token)”`, la cual garantiza que las solicitudes se realicen de manera segura y autenticada, recuperando las `“cookies”` de nuestro navegador. Para esto es necesario haber iniciado sesión previamente, lo cual hemos cumplido al haber estado analizando la página a través de la herramienta de desarrolladores y al haber aceptado las `“cookies”` previamente. Esta función recibirá como argumento la clave o `“token”` de autenticación necesario para acceder a los servicios de Twitter. Este se puede obtener muy fácilmente iniciando sesión en nuestra cuenta y lanzando de nuevo la consulta de ejemplo `“hamburguesa”`. En la misma cabecera de la petición encontraremos nuestro token de autenticación.

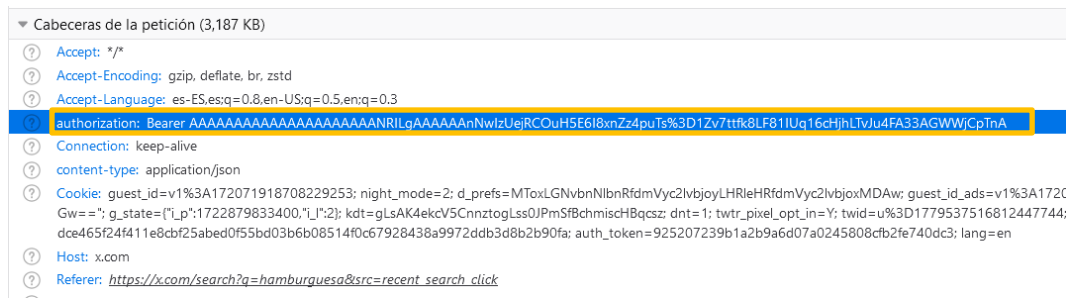


Figura 6: Clave de autenticación de nuestro usuario en Twitter.

Una vez hayamos creado una nueva sesión HTTP, utilizaremos una función basada en la librería “request” para almacenar las cookies y las cabeceras a través de las solicitudes, lo que es esencial para mantener la autenticación activa mientras se realizan múltiples solicitudes a la API. Utilizaremos el módulo “browsercookie” para extraer las cookies del navegador y añadirlas a la sesión creada. Finalmente, configuramos los encabezados HTTP que se incluirán en las solicitudes realizadas a través de esta sesión, incluyendo el “token” de autenticación junto a otro “token” del tipo “CSRF” (Cross-Site Request Forgery) (Kirsten, 2024), que utiliza Twitter como medida de seguridad y suele ser el mismo para todos los usuarios. La función devolverá el objeto “session” correctamente configurado para realizar las solicitudes.

Para obtener un conjunto de resultados de tuits iniciales desarrollaremos una función llamada “get\_tweet\_search\_data\_without\_cursor”. Esta función tiene el nombre “sin cursor” debido a que no utiliza un cursor para paginación, pues durante las pruebas de desarrollo se observó que tras los primeros resultados de búsqueda no se encontraban nuevos tuits. Esto es debido a que Twitter por defecto carga alrededor de 20 tuits en la primera página de resultados, los cuales no aumentarán en número a no ser que el usuario se desplace o haga “scroll” hacia abajo de la interfaz. Es en este momento donde, analizando las solicitudes y respuestas de la página, vemos que aparece una nueva respuesta la cual contiene información del cursor que utiliza Twitter para saber en qué parte de la interfaz se ubica el usuario y, por tanto, cargar más resultados. Basándonos en esta nueva estructura, posteriormente podremos crear una nueva función en la cual se recoja el cursor en cuestión para simular que estamos desplazándonos a través de la interfaz de Twitter y así poder recopilar una mayor cantidad de información en una misma búsqueda.



Figura 7: Cursor utilizado para la paginación.

Construiremos la URL de GraphQL de Twitter, la cual se utiliza para realizar la consulta de búsqueda. En la URL insertamos la variable “line”, que corresponde a la consulta en cuestión (esta variable será sustituida por el tipo de comida buscado), así como el número de resultados “count”, entre otros parámetros que especifican cómo debe comportarse la búsqueda. A continuación se realiza una solicitud a la API utilizando esta URL y la sesión autenticada, cuya respuesta contendrá los datos de los tuits que coinciden con la búsqueda especificada. La variable “line” viene especificada en un archivo de texto llamado “searches.txt”, que contiene todas las palabras clave que queremos buscar a través de nuestro programa.

Como en la mayoría de las páginas web existe una cantidad límite de solicitudes que se puedan realizar en un período de tiempo determinado. Si se excede este límite, la API devuelve un código de estado 429, el cual viene a indicar que se ha excedido el número máximo de solicitudes. En ese caso la función calculará el tiempo restante hasta poder volver a ejecutar las búsquedas. Si la respuesta de la API contiene un código de estado diferente al 429 (límite de peticiones excedido) o 200 (funcionamiento correcto) (Klaviyo, 2024), se imprime un mensaje de error por pantalla y la función retorna un “None”. Mientras la lectura de las búsquedas sea correcta, la función extraerá los datos de la respuesta en formato JSON.

Como indicamos anteriormente, los datos de los tuits se encuentran en el subelemento “entries”. Para llegar a éste desde el objeto principal “data” se debe seguir la jerarquía “data > search\_by\_raw\_query > search\_timeline > timeline > instructions > 0”. Así pues, nuestra función buscará la información dentro de esta estructura hasta encontrar el bloque de instrucciones. Si “entries” no se encuentra dentro del bloque “instructions[0]”, la función devolverá “None”, lo que indica que no se obtuvieron resultados.

Tras extraer el conjunto de “entries” de la respuesta, la función intentará encontrar y extraer el cursor que permitirá acceder a la siguiente página de resultados, buscando en la misma estructura de “entries” donde hemos podido identificar la aparición de este cursor a través de la herramienta de desarrolladores. Para manejar posibles errores no conocidos durante la extracción de datos usaremos un bloque “try-except”, que captura el error y lo imprime por pantalla para ayudar al usuario. Finalmente, la función convertirá cada entrada

en el campo “entries” en un objeto “TweetData”, devolviendo una lista de objetos “TweetData” así como el cursor para poder cargar automáticamente la siguiente página de resultados.

Para obtener los resultados adicionales que muestra Twitter al desplazarse o hacer “scroll” a través de su interfaz, crearemos una nueva función llamada “get\_tweet\_search\_data\_with\_cursor”. Esta será básicamente igual a la creada anteriormente, pero con la adición del manejo de un cursor para paginación, permitiendo continuar obteniendo más resultados de la misma consulta moviéndose a través de las páginas de resultados sucesivas. Al igual que anteriormente, construiremos la URL de GraphQL de manera similar pero esta vez incluyendo como variable el cursor para que sea un parámetro más de nuestra consulta. Esto permitirá a la API saber desde qué punto continuar obteniendo resultados adicionales de la misma búsqueda (segunda página, tercera página...). El resto del código de esta función se mantendrá igual, a diferencia de la obtención del nuevo cursor, el cual en este caso estará ubicado dentro de la tercera instrucción (“instructions”). Si no se encuentra el nuevo cursor, la función informará de ello y continuará con la siguiente palabra clave.

Para enlazar las dos funciones principales explicadas anteriormente será necesario crear una función que sirva para gestionar la paginación y acumular los resultados en una colección única. Esta función, llamada “get\_tweet\_search\_data”, inicializará una lista en la cual se almacenarán los resultados. Llamaremos a la función “get\_tweet\_search\_data\_without\_cursor” para obtener la primera página de resultados, agregando estos a la lista creada. A continuación, utilizamos un bucle “while” para continuar obteniendo más resultados mediante el uso del cursor hasta alcanzar el límite especificado o hasta no obtener más resultados, es decir, en caso de que el cursor sea “None”. Finalmente, la función devolverá la lista que contendrá el conjunto de datos a utilizar en nuestra investigación.

Al ejecutar el programa se le pedirá al usuario introducir su clave de autenticación de Twitter. Si no se proporciona la clave, el programa imprimirá un mensaje de uso y terminará, previniendo errores al asegurarse que no se ejecuta sin los parámetros esenciales. Una vez proporcionado el token se configurará la sesión HTTP con la API de Twitter, incluyendo las cookies del navegador y los encabezados de autenticación necesarios a través de la función “setup”. Seguidamente, el programa leerá las líneas del archivo “searches.txt” en el cual se almacenan las palabras clave que se utilizarán durante las búsquedas (Figura 8).



Figura 8: Contenido del archivo "searches.txt".

Con el uso de un diccionario se almacenarán todos los tuits recolectados, usando como identificador de estos el parámetro “tweet\_id”, el cual es único para cada tuit y asegura que no se almacenen dos tuits iguales. Finalmente, el bucle se detendrá al alcanzar el objetivo de búsquedas, generando un archivo único cuyo nombre se basa en la fecha y hora actual de ejecución del programa. El formato del archivo será JSON, que servirá para estructurar los datos de manera similar a cómo se presentan en la red social y así comprobar si la recolección ha sido satisfactoria. En la Figura 9, se puede apreciar el funcionamiento de este programa.

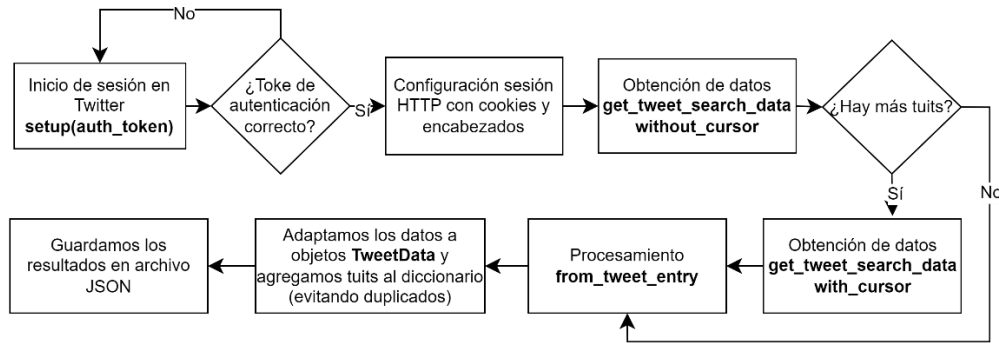


Figura 9: Diagrama de flujo del programa de recopilación de datos de Twitter.

### 3.1.3 Base de datos de Google Maps

Google Maps es una de las plataformas de geolocalización y reseñas más utilizada a nivel mundial. Proporciona una vasta cantidad de datos, desde ubicaciones hasta reseñas de usuarios, que resultan de gran utilidad en diversos contextos, incluyendo análisis de mercado y estudios de popularidad. Sin embargo, realizar “web scraping” en Google Maps presenta desafíos significativos debido a la complejidad y dinamismo de su estructura web, además de las restricciones impuestas por Google para proteger la privacidad de los datos y prevenir el abuso de su plataforma (Rayobite, 2024).

Uno de los mayores desafíos que se han encontrado al analizar la estructura web de Google Maps es que esta no es estática. Google cambia con frecuencia la disposición y los nombres de las clases del código, las cuales contienen la información que buscamos en nuestro estudio. Además, Google utiliza técnicas avanzadas de carga dinámica de contenido, como la renderización asincrónica mediante “Java”, que sólo muestra ciertos elementos cuando el usuario interactúa con la página. Un ejemplo de esto son las reseñas, las puntuaciones de la gente, o cierta información del restaurante, la cual sólo es visible haciendo “click” en la opción correspondiente dentro de la interfaz. Al intentar analizar las respuestas de la página es imposible encontrar el patrón de estructuración que utiliza Google para poder consultar a la API de manera totalmente automatizada, como ha sido el caso de Twitter.

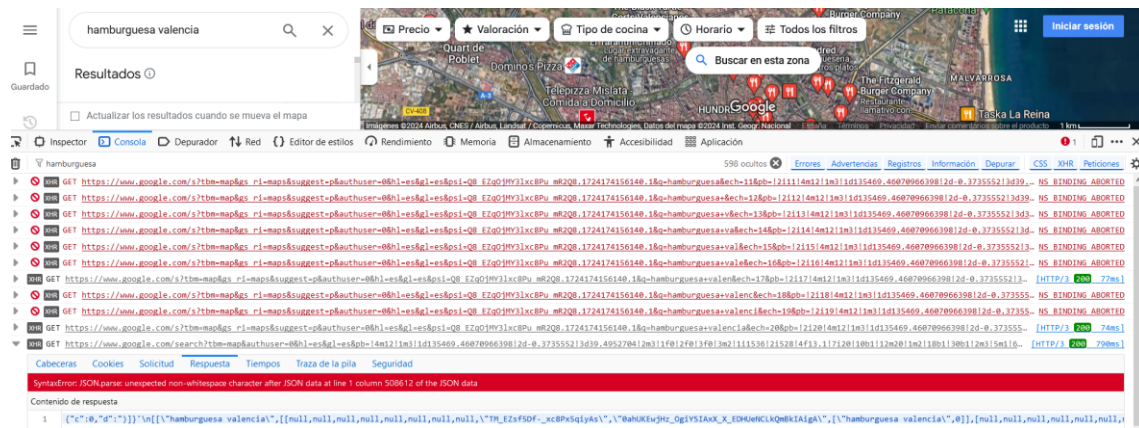


Figura 10: Respuestas de Google Maps al ejecutar la herramienta para desarrolladores.

Esta situación hace que sea muy complejo desarrollar un programa para lanzar consultas automatizadas. En su lugar, hemos realizado una simulación de navegación de un usuario real (Pandey, 2023), cargando completamente la página y realizando acciones como “scroll” o “clicks” para desencadenar la aparición de los datos para los distintos restaurantes. En este contexto, hemos elegido Selenium como la herramienta principal para llevar a cabo la recolección de datos. Selenium es una biblioteca de Python que permite automatizar un navegador web real (como Chrome o Firefox), simulando acciones típicas de un usuario. Esto es esencial para interactuar con la estructura dinámica de Google Maps, donde los elementos de interés pueden estar ocultos hasta que se desplaza por la interfaz o se realizan otras interacciones. Una vez simulada la interacción del usuario se obtendrá la información necesaria de los elementos dentro del código.

Se aprecia que, dentro de la respuesta, al seleccionar cada uno de los restaurantes aparecidos en las búsquedas los elementos de interés sí que parecen estáticos hoy en día y durante la elaboración de este estudio; sin embargo, estos pueden ser cambiantes a corto plazo. Para encontrar los elementos necesarios a incluir en nuestro código se ejecutan búsquedas con palabras clave que puedan ser fácilmente reconocibles dentro del código y sus elementos. A modo de ejemplo, tras haber seleccionado el restaurante de hamburguesas “Hundred” (ya que es la primera aparición al buscar “hamburguesa Valencia”, se busca el nombre en cuestión a través de las respuestas, identificando éste dentro de la cabecera “h1”. Esta cabecera contiene una clase que para todas las búsquedas se mantiene con el mismo nombre (“DUwDvf lfPIob”), la cual contiene el nombre del restaurante. Como se puede ver en la Figura 11, esta clase no es fácilmente identificable y se trata de una tarea laboriosa.

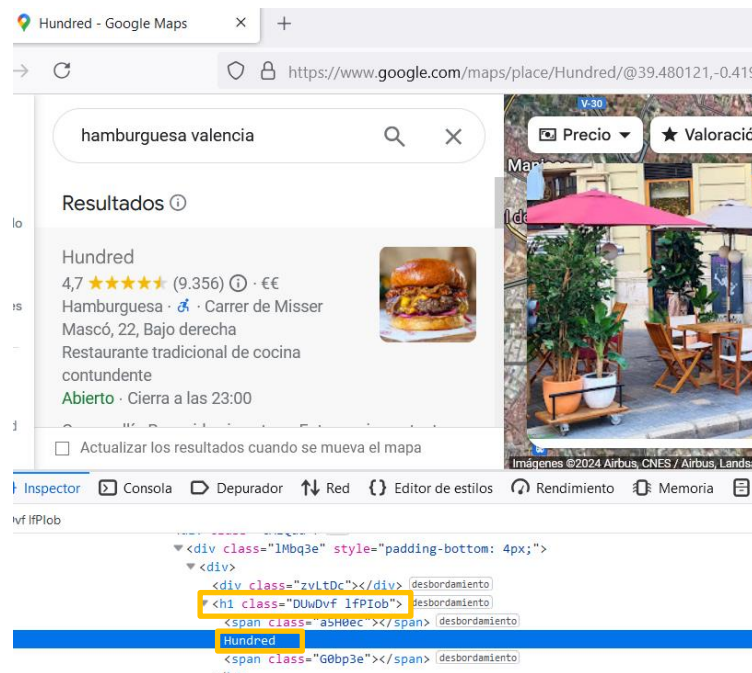


Figura 11: Ubicación de metadatos para la hamburguesería "Hundred" tras ejecutar búsqueda en Google Maps.

La información relevante de los restaurantes que se va a recopilar durante nuestro estudio será la siguiente:

- **Palabra clave:** El programa contará con una frase clave que ejecutará en el buscador de Google Maps, la cual está compuesta por las palabras “restaurante”+”palabra clave”+”Valencia”, de tal modo que se busquen restaurantes que ofrezcan ese tipo de comida en Valencia.
- **Nombre del restaurante:** Servirá para descartar resultados repetidos.
- **Coordenadas geográficas y dirección:** La ubicación geográfica del local ayudará a entender mejor las preferencias gastronómicas del barrio en cuestión.
- **Puntuación media:** Valoración media de los usuarios en Google Maps.
- **Precio:** Google Maps cuenta con una clasificación del coste medio del servicio en los restaurantes, la cual consiste en tres rangos de precios: asequible (0-10€), moderado (10-20€) y caro (+20€).
- **Número de reseñas:** Cantidad de reseñas que los usuarios han escrito en Google Maps acerca de su experiencia en el restaurante en cuestión.

Cada búsqueda se realiza sobre una URL base de Google Maps, construyendo la URL

final para cada palabra clave y lanzando la búsqueda en el navegador Firefox automatizado. A medida que se navega por cada restaurante los datos son capturados y almacenados en una lista de diccionario. Finalmente, todos estos datos son guardados en un archivo JSON, lo que facilita su análisis posterior. Con esta información podremos realizar un análisis de la popularidad de los restaurantes, comparando este análisis con los resultados del análisis de sentimiento aplicado a los comentarios de Twitter.

### 3.1.4 Desarrollo del programa de web scraping para recopilación de datos de Google Maps

Para desarrollar el programa para la recolección de datos de Google Maps usaremos principalmente las siguientes librerías de Python:

- **selenium**: Selenium es la herramienta principal utilizada para automatizar la navegación web en este programa. Permite controlar un navegador web a través de un tercer programa, lo que incluye realizar acciones como hacer clic, rellenar formularios y desplazarse por páginas. Dentro de nuestro contexto, Selenium permitirá navegar a través de los resultados de búsqueda de restaurantes y extraer la información necesaria de cada uno de ellos.
- **PyAutoGUI**: Es utilizada para realizar movimientos y clics del ratón, lo que es crucial para simular el desplazamiento en la página de resultados de Google Maps. Esta simulación es necesaria porque Google Maps carga más resultados a medida que el usuario se desplaza hacia abajo en la interfaz.
- **time**: Se utiliza para manejar pausas o retrasos en la ejecución del programa. Estos retrasos son fundamentales para simular la interacción humana (por ejemplo, tiempos de espera entre clics o desplazamientos) y para evitar ser bloqueado por realizar solicitudes en intervalos demasiado rápidos de tiempo.
- **datetime**: Este módulo se emplea para gestionar fechas y horas, lo que es crucial para la generación de nombres de archivos basados en la fecha actual.
- **JSON**: Permite la serialización y deserialización de los datos extraídos.
- **re**: La librería *re* es utilizada para realizar búsquedas y manipulaciones de cadenas de texto utilizando expresiones regulares. En nuestro programa servirá para extraer información de las coordenadas directamente de la URL del restaurante.

En el desarrollo del proyecto primeramente configuraremos las opciones del navegador a utilizar. Vamos a mantener el uso del navegador Firefox para la navegación automatizada.

Entre las opciones estableceremos un “user-agent” para evitar ser identificados como un “bot”, indicando a Google Maps que se está accediendo desde un dispositivo con sistema operativo Windows y a través del navegador Firefox. De este modo se intenta ocultar que la interacción la está realizando directamente un programa de “web scraping”. Desactivaremos la carga de imágenes para acelerar el tiempo de carga de las páginas, dejando eso sí habilitada la visualización de la navegación para comprobar si se está ejecutando correctamente nuestro programa.

Para ejecutar las búsquedas se definen los mismos tipos de comida utilizados durante la búsqueda en Twitter. Estas palabras serán utilizadas en el bucle principal, precedidas de la palabra “restaurante” y seguidas de la palabra “Valencia”. De este modo centraremos las búsquedas en los restaurantes que ofrezcan esos distintos tipos de comida en la ciudad de Valencia. Utilizaremos en nuestro caso la versión “.com” de Google Maps, definiendo así la URL base como “https://www.google.com/maps/search/”. A continuación, crearemos una lista llamada “restaurants\_data”, la cual usaremos para almacenar la información recopilada sobre los restaurantes. Esta información se almacenará en un archivo JSON, que tendrá como nombre la fecha y hora actual de manera que aseguramos tener un archivo único e identificado.

En el siguiente bloque de código, tras configurar nuestro navegador y los parámetros iniciales, se definen dos funciones que servirán para ejecutar acciones propias de un usuario. La primera de ellas, “accept\_cookies”, es la de aceptación de las cookies al lanzar la página inicial de Google Maps. Al ejecutar la URL base de Google Maps se solicitaba la aceptación de las cookies de Google en cada ejecución. Mediante el uso de “WebDriverWait”, de Selenium, solicitaremos al programa que espere hasta que cargue la página y aparezca el botón interactivo con el texto “Aceptar todo”. En ese momento hará clic en el botón en cuestión para eliminar la ventana emergente.

La segunda función, “scroll\_the\_page”, simulará el desplazamiento del usuario por la página, de modo que se vayan cargando más resultados en cada una de las búsquedas. Para ello indicaremos que se sitúe el ratón en la zona izquierda de la pantalla y, a continuación, realice un desplazamiento hacia abajo. Daremos un valor suficientemente negativo al movimiento del ratón como para que el desplazamiento sea largo y se carguen suficientes resultados. Entre desplazamientos esperaremos un tiempo para simular que la interacción la está realizando un ser humano y permitir que los restaurantes se carguen satisfactoriamente entre las apariciones de búsqueda.

Durante las pruebas se observa que las URL de los restaurantes en Google Maps son del tipo “https://www.google.com/maps/place/Hundred/@39.4753996,-0.381047,3411m/data...”. La función “extract\_lat\_long\_from\_url” servirá para extraer las

coordenadas de latitud y longitud de la URL pues, como se aprecia, las coordenadas están codificadas dentro de ésta, por lo que utilizaremos “re.search()” para buscar la expresión regular dentro del enlace.

En el bucle principal del programa iteraremos sobre cada palabra clave definida anteriormente para construir una búsqueda específica. Al ejecutar la primera búsqueda el programa esperará un tiempo aleatorio, entre tres y cinco segundos, para evitar ser detectados como un robot, para posteriormente aceptar las cookies que aparezcan y ejecutar los desplazamientos hacia abajo que hemos programado. Estos desplazamientos se harán en treinta ocasiones, pues hemos comprobado que es un número suficiente como para cargar todos los resultados posibles para la mayoría de las palabras clave. En algunas de ellas Google Maps directamente relaciona la búsqueda con locales a las afueras de Valencia, por lo que para optimizar el funcionamiento del programa y evitar resultados que no interesen limitamos la cantidad de desplazamientos a un número razonable. Una vez cargados los resultados el programa accede a la URL de cada restaurante, extrayendo la información deseada. Toda esta información será almacenada en un diccionario y luego se agregará a la lista “restaurants\_data”, guardando ésta en un archivo estructurado y cerrando correctamente el navegador para liberar recursos del sistema y evitar errores. En la Figura 12, se representa mediante un diagrama de flujo el funcionamiento de este programa.

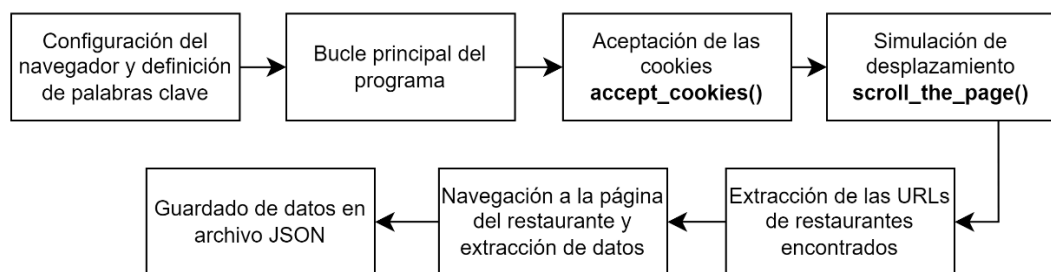


Figura 12: Diagrama de flujo del programa de recopilación de datos de Google Maps.

## 3.2 Preprocesamiento de los datos recopilados

El preprocesamiento de datos es un paso crucial en cualquier proyecto de análisis de datos, especialmente con datos textuales como los tuits. En este proceso se aplicarán diferentes técnicas para preparar y normalizar la información que hemos recopilado anteriormente. Esto facilitará el procesamiento de estos datos a la hora de analizarlos, evitando errores típicos en su interpretación y uso.

En el caso específico de Twitter, los datos recogidos incluyen un gran volumen de tuits, alrededor de casi 6.000. Algunos de estos pueden no contener información valiosa para nuestro

caso de estudio, pues el propio algoritmo de búsqueda de Twitter relaciona la palabra que estamos buscando con el nombre de los usuarios. Además, se va a descartar cierta información que, por motivos de tiempo y recursos, no va a poder ser utilizada en nuestra investigación. En este caso optaremos por conservar los datos de identificador del tuit, texto del tuit, número de favoritos y número de retuits. Se formatearán los tuits recopilados, contenidos en la estructura del JSON generado, para adaptarlos a un formato lineal y archivar el conjunto de todos los tuits en un mismo archivo de texto.

Los archivos de texto plano facilitarán la lectura y manipulación de los datos, siendo más sencillos de cargar y procesar en comparación con archivos JSON. Al extraer sólo los datos relevantes se va a reducir drásticamente el tamaño del archivo, lo que agilizará todavía más el procesamiento. Además, de este modo, adaptamos la información de manera similar a otros estudios previos, los cuales servirán como fuente de datos a la hora de investigar el sentimiento de los tuits extraídos, pues estos funcionarán como referencia de aprendizaje para la red neuronal que vamos a configurar.

En el caso de Twitter, el preprocesamiento implica varios pasos. Uno de estos es la detección de las palabras clave utilizadas durante las búsquedas, de modo que, si el texto del tuit en cuestión no contiene una de estas palabras, se elimina del archivo preprocesado. Para ello en nuestro programa de procesamiento definiremos esta lista de palabras clave, la cual es equivalente a la que podemos encontrar en el archivo “searches.txt” usado anteriormente durante la fase de recopilación de datos. Sin embargo, ya que Twitter ofrece resultados para diferentes equivalencias de la palabra buscada, vamos a considerar estas palabras tanto en sus versiones con acento como sin acento, de modo que se abarquen todas las maneras de escribirlas en Twitter. Mediante la función “contains\_keywords()” verificaremos si el texto de un tuit contiene alguna de estas palabras, convirtiendo todo el texto a minúsculas para asegurar que la búsqueda sea insensible a mayúsculas y minúsculas. Además, se aplicarán distintas operaciones de limpieza del texto, entre ellas la eliminación de enlaces, la eliminación de emojis y símbolos (manteniendo los “@” pues hacen mención a los usuarios), la eliminación de saltos de línea para mantener el contenido en una misma línea, y la normalización de los espacios reemplazando los espacios consecutivos por uno solo.

A continuación se traducirá el texto del tuit al español si no está ya en ese idioma. Como algunas de las palabras utilizadas son internacionales y no propias del lenguaje español (véase “kebab” o “pizza”), algunos de los tuits recopilados posiblemente estén en otros idiomas. Por ello en nuestro proceso de estandarización de los datos haremos uso de la librería “Deep\_translator”, que se utiliza para detectar automáticamente el idioma del texto y traducirlo al español, permitiendo un análisis uniforme de sentimiento y contenido. Para evitar traducir palabras claves y generar ruido en nuestro análisis, vamos a omitir la traducción del listado de palabras clave. Para ello, antes de la traducción, se extraen del texto de los tuits y se sustituyen con marcadores temporales.

Se procesará cada tuit del archivo JSON original y utilizaremos un “set” para ir almacenando los identificadores de los tuits ya procesados y así evitar duplicados. Inicializaremos una lista vacía “filtered\_tweet\_lines” la cual almacenará los tuits que pasen el filtro de palabras clave y hayan sido procesados. Si un tuit contiene alguna de las palabras clave se procesa: se traduce al español si es necesario, se limpia el texto, y luego se escribe dentro del archivo de texto. Esta línea se agrega a la lista “filtered\_tweets\_lines” y el identificador del tuit se marca como procesado. Cada uno de los tuits filtrados y procesados se guardarán en un archivo de texto único en la ubicación especificada, cuyo nombre corresponderá a su identificador del tuit, seguido del número de favoritos y de retuits con el que cuenta. Estos dos últimos datos servirán para dar peso al sentimiento del tuit. En la siguiente Figura 13, se representa el proceso de limpieza de los tuits.

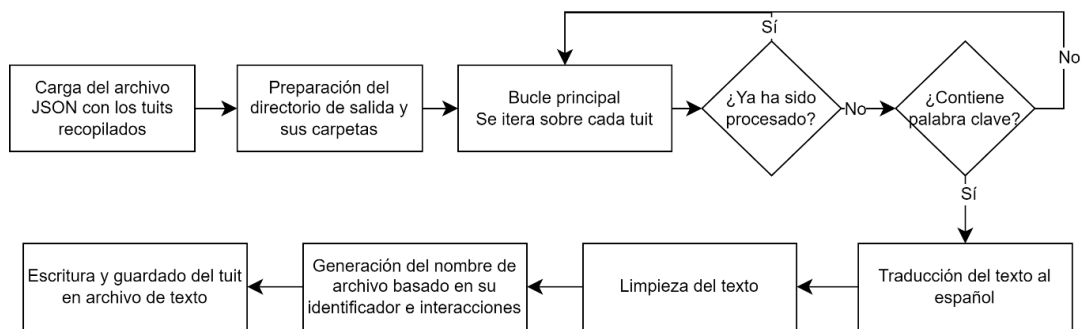


Figura 13: Diagrama de flujo del programa de procesamiento y limpieza de tuits.

Para el caso de Google Maps vamos también a seguir una lógica similar al preprocesamiento de los datos de Twitter. En este caso vamos a conservar la palabra clave del restaurante, es decir, el tipo de comida o servicio ofrecido por este; el nombre del restaurante, identificando así el restaurante y evitando duplicados; el barrio donde se ubica, de modo que podamos determinar el número de restaurantes por zonas de la ciudad; la puntuación media y el número de reseñas, para conocer la opinión de los usuarios; y por último, el precio, de modo que se pueda correlacionar la popularidad con la accesibilidad económica. En el archivo original el número de restaurantes asciende a 2038.

Durante el análisis manual de los datos recopilados hemos detectado que en la mayoría de los casos en la dirección del restaurante aparece el barrio donde se ubica, pero en otros no. En estos casos se trata principalmente de restaurantes ubicados en la playa o cercanos a ella. Por ello en la información de su dirección aparecen palabras como “playa”, “paseo marítimo” o “Malvarrosa” (el nombre de la playa más conocida de Valencia). Vamos a utilizar estas palabras para identificar los restaurantes ubicados en el barrio de “Poblats Marítims”, es decir, el barrio de la playa de Valencia. En muchos de los restaurantes el barrio aparece en su versión en valenciano, pero en otros aparece en castellano, por lo que se va a utilizar un diccionario para relacionar ambas traducciones.

En primer lugar, cargaremos el archivo JSON generado anteriormente mediante el programa de recopilación de información. Utilizaremos de nuevo un conjunto para almacenar el conjunto de palabra clave y nombre de restaurante que ya haya sido procesado; de este modo se evitarán duplicados para resultados de un mismo restaurante y tipo de comida. Sin embargo, conservaremos aquellos resultados de restaurantes que compartan nombre pero ofrezcan distintos tipos de comida, ya que si fueran eliminados no estaríamos reflejando en realidad la relevancia que tiene cada tipo de comida analizada.

A continuación definiremos una lista de los posibles barrios en Valencia y un diccionario con sus diferentes traducciones, de modo que se clasifiquen todos con una misma traducción. Además, definiremos la lista de palabras relacionadas con la playa y que indican que el restaurante está ubicado en la zona de “Poblats Marítims”. Esto será útil para clasificar los restaurantes en casos donde el barrio no está explícitamente mencionado. Se realizará de este modo la identificación de los barrios, puesto que identificar la zona del restaurante por la dirección completa sería muy complicado, teniendo que relacionar la calle con el barrio correspondiente.

Finalmente en el bucle principal del programa se procesará cada elemento en la lista de datos “data” (Figura 14). Se creará una tupla con la “keyword” y el nombre del restaurante (“keyword\_name\_combo”), añadiendo esta combinación al conjunto “seen\_keyword\_names” y evitando procesarlo de nuevo. Se analizará la dirección del restaurante para extraer el barrio correspondiente, verificando si es necesario traducirlo y, en caso de que no se encuentre el barrio, se buscará si la dirección contiene palabras relacionadas con la playa; por último, se creará una línea dentro del archivo de texto con un formato que separa la información del restaurante con un guión bajo (“keyword\_nombre\_barrio\_puntuación\_precio\_reseñas”). Los datos se guardarán en este archivo de texto con un nombre único utilizando la fecha y hora de ejecución del programa.

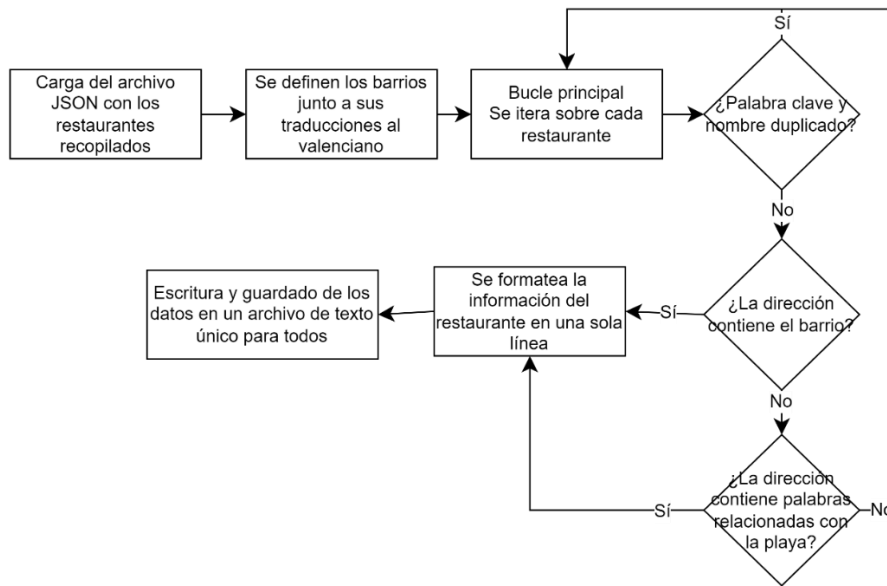


Figura 14: Diagrama de flujo del procesamiento y limpieza de resultados de Google Maps.

De este modo, conseguimos limpiar los datos de ambas redes sociales y tenerlos en un formato de texto similar para poder ser usados, tanto para el análisis de sentimiento de Twitter como para analizar la popularidad de los restaurantes de Google Maps.

### 3.3 Resultados tras el preprocesamiento de los datos

En este apartado se expondrán los datos recolectados tanto de la red social Twitter como de Google Maps mediante las técnicas de “web scraping” descritas en los apartados anteriores. Gracias al uso de esta técnica hemos podido recopilar un gran volumen de datos que dan una idea del volumen de interacciones que podemos encontrar en estas redes sociales acerca de estos tipos de comida del sector gastronómico, así como los datos más representativos de los diferentes restaurantes que podemos encontrar en la ciudad de Valencia.

Para cada uno de los diferentes tipos de comida definidos en el listado de búsqueda se han obtenido múltiples resultados procedentes de Twitter. El número de tuits recopilados inicialmente y antes de ser procesados ascendían a un total de casi 6.000 unidades. Sin embargo, como podemos apreciar a continuación, en muchas ocasiones se recopilaban datos irrelevantes para nuestro estudio. Esto es debido a que el algoritmo de búsqueda de Twitter muestra al usuario tuits que no están relacionados directamente con la búsqueda, pero que aparecen en pantalla debido a que, por ejemplo, el usuario que lo escribe cuenta con esa palabra en su nombre. Como ejemplo, en los datos iniciales dentro del archivo del tipo JSON, podemos apreciar la siguiente ocurrencia (Figura 15).

```

{
  "tweet_id": "1820119172875194557",
  "user_id": "1278438948872781833",
  "text": "¿Esta es la que supuestamente va a \"salvar\" Venezuela?\n\n☺",
  "favorites": 3,
  "retweets": 2,
  "quotes": 0,
  "bookmarks": 0,
  "replies": 0
},
{
  "tweet_id": "1820114326675390602",
  "user_id": "1278438948872781833",
  "text": "me convierto en mexicano y les mando a la chingada",
  "favorites": 0,
  "retweets": 0,
  "quotes": 0,
  "bookmarks": 0,
  "replies": 0
},

```

Figura 15: Datos recopilados de Twitter antes de ser procesados.

Si bien en los tuits anteriores no se aprecia directamente el nombre de los usuarios, es fácilmente comprobable que estos han sido recopilados erróneamente. Copiando el número asociado al “tweet\_id” del segundo tuit que aparece en la imagen anterior y buscándolo en Twitter, comprobamos como el nombre del usuario en cuestión contiene la palabra “pizza” (Figura 16).

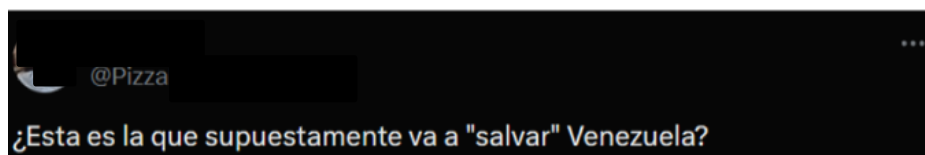


Figura 16: Ejemplo de tuit recopilado por error debido al nombre del usuario.

Dejando estos tuits de lado y contabilizando tan solo aquellos ya procesados (es decir, habiendo eliminado aquellos que no contuvieran una de las palabras clave o su identificador estuviera repetido), estos disminuyen a un total de 3.232 tuits únicos. Esta cifra se considera suficiente para las categorías de comida definidas en nuestro estudio (7) y para entender la relevancia de éstas entre los usuarios de Twitter. Aun existiendo una gran interacción en Twitter en la que se discutía sobre estos alimentos, hemos detectado que era necesario esperar varias horas para poder recopilar una toma de tuits totalmente nueva sin consultar los mismos que en la anterior búsqueda, por lo que la tarea de recopilación ha sido más complicada de lo esperado.

Si bien la gran mayoría de estos tuits contaba con información valiosa para nuestro análisis, parte de estos utilizaban expresiones típicas de las redes sociales, tales como emoticonos representados por caracteres (véase la expresión de risa “XD”). El uso de este tipo de expresiones o emoticonos complica en cierto modo la interpretación del sentimiento de los tuits. Esto es debido a que nuestra fuente de datos utilizada para el entrenamiento de la red

neuronal cuenta con expresiones de este tipo, pero al existir una variedad tan grande de las mismas así como diferentes contextos y maneras de usarlas, seguramente le resulte complicado aprender patrones para llegar a interpretarlas correctamente en cada situación. Un ejemplo de esto es el siguiente tuit (Figura 17), en el que podemos ver como el usuario en cuestión utiliza el emoticono “XD” para expresar incredulidad o darle un cierto tono de “risa irónica” a su mensaje, expresando que el precio de las hamburguesas de hoy en día le parece excesivamente caro.

1819684901110665246 @kriegeryllite en mi pueblo que esta en madrid sur al lado de mi casa ya han abierto un sitio donde te soplan 15 euros por una hamburguesa v 7.50 por cuatro tequeños XD 0 0

Figura 17: Ejemplo de tuit con expresiones típicas de las redes sociales y difícilmente comprensibles por un programa.

El mensaje anterior sería fácilmente interpretable por un ser humano, pudiéndolo clasificar como un mensaje negativo. Sin embargo, para nuestro modelo de red neuronal, este mensaje puede llevar a confusiones, lo que es algo para tener en cuenta.

En resumen, los resultados de recopilación de tuits para cada tipo de comida muestran las siguientes estadísticas (Figura 18 y Tabla 2):

Tabla 2: Resultados de la recopilación de Twitter por tipo de comida.

Tipo de comida	Número de tuits
Paella	131
Arroz a banda	32
Arroz al horno	77
Fideua	85
Esmorzar	99
Esgarraet	32
Tapas españolas	0
Patatas bravas	26
Tortilla española	14
Calamares a la romana	73
Croquetas	120
Ensaladilla	125
Comida italiana	20
Pasta	138
Pizza	125
Lasaña	1
Tiramisú	52
Carbonara	84
Marisco	138
Mariscada	96

Pescado	104
Gambas	91
Langosta	78
Pulpo a la brasa	64
Carne	139
Carne a la brasa	0
Hamburguesa	109
Asado	127
Barbacoa	114
Chuletón	118
Comida china	10
Comida japonesa	53
Kebab	93
Comida india	5
Comida mexicana	59
Comida latina	5
Horchata	120
Helado	110
Café	102
Churros	124
Fartons	56
Pastel	96

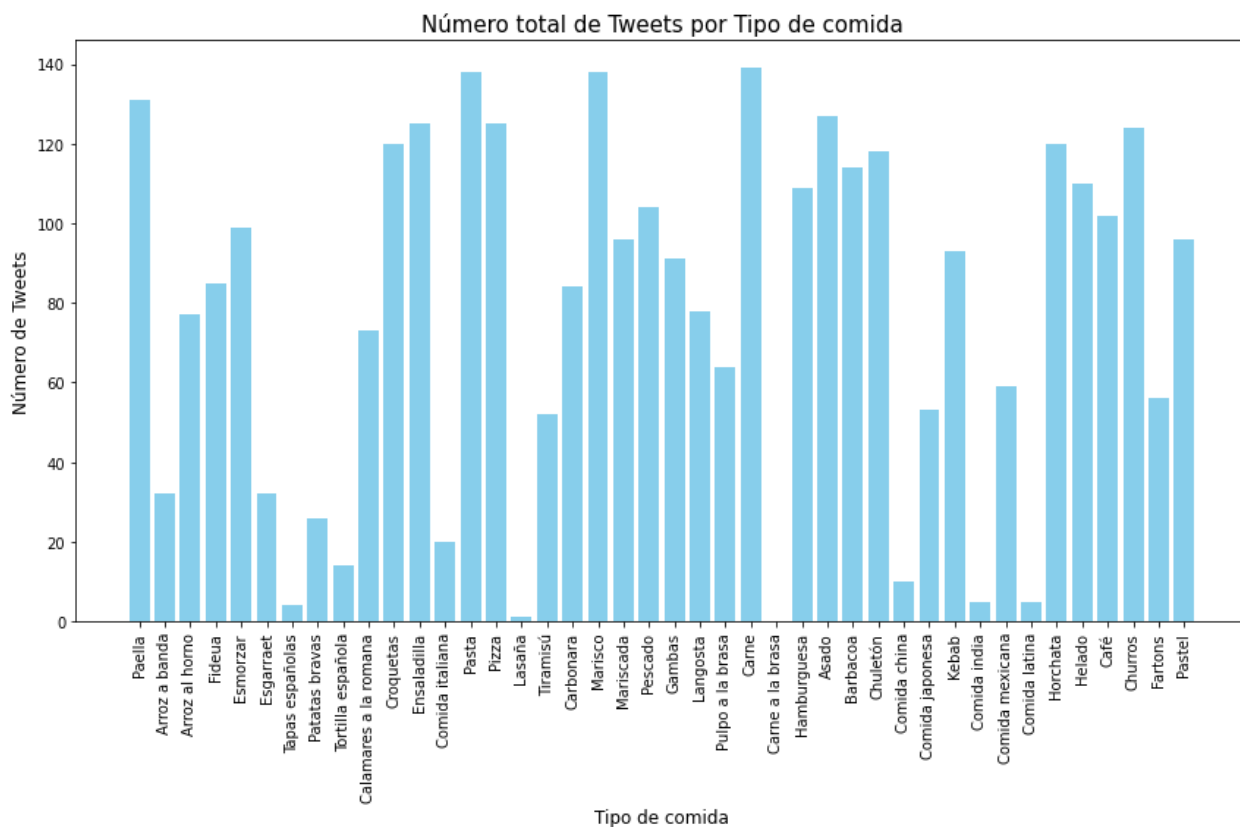


Figura 18: Representación gráfica del número de tuits recopilados por tipo de comida.

Para el caso de Google Maps, si bien la recopilación mediante “web scraping” no ha podido ser tan directa como en el caso de Twitter, ha sido más simple el recolectar la información de los múltiples restaurantes ubicados en Valencia que ofrecen los tipos de comida utilizados en nuestra investigación. Esto es debido a que todos los restaurantes publicados en Google Maps cuentan con una misma ficha técnica en la que podemos encontrar la misma información para todos: nombre, ubicación, puntuación, reseñas, etc. (Figura 19). Los resultados iniciales proporcionan información de unos 2.264 restaurantes dentro del archivo JSON generado.

```

{
  "keyword": "paella",
  "name": "Destino Puerto",
  "latitude": "39.4614626",
  "longitude": "-0.3230574",
  "address": "Marina Real Juan Carlos I, Carrer del Moll de la Duana,
s/n, Poblados Marítimos, 46024 Valencia",
  "rate": "4,0",
  "price": "Moderado",
  "reviews": "2824"
},
{
  "keyword": "paella",
  "name": "El Racó de la Paella",
  "latitude": "39.4831215",
  "longitude": "-0.3995919",
  "address": "Carrer de Mossèn Rausell, 17, Campanar, 46015 València,
Valencia",
  "rate": "4,4",
  "price": "Moderado",
  "reviews": "1685"
},

```

*Figura 19: Datos recopilados de Google Maps antes de ser procesados.*

Esta información, tras ser procesada, disminuye a 2.235 restaurantes. Esto es debido a que durante la investigación el programa de recopilación de datos se ha ejecutado en varias ocasiones, por lo que se han eliminado aquellos restaurantes que contaban con la misma palabra clave de búsqueda y el mismo nombre de restaurante, o aquellos en los que no ha sido posible identificar el barrio donde se encuentran. Sin embargo, se han almacenado aquellos restaurantes repetidos pero que contaban con una palabra clave distinta. Esto es debido a que han aparecido en diferentes tipos de búsqueda, puesto que en su servicio ofrecen distintos tipos de comida. Como en nuestro análisis es relevante tener en cuenta todos los restaurantes que ofrezcan el listado de comidas a investigar se conservan los resultados de estos, aunque el restaurante sea el mismo.

Para el capítulo de análisis de los resultados sí que agruparemos la representación de restaurantes según la categoría a la que pertenezcan, de modo que disminuirá aún más el número total de restaurantes, pero para esta primera fase de muestreo de los datos vamos a conservarlo así.

La información procesada y almacenada en el archivo de texto en cuestión resulta pues del siguiente modo (Figura 20):

```
paella_Destino Puerto_Poblats Marítims_4,0_Moderado_2824
paella_El Racó de la Paella_Campanar_4,4_Moderado_1685
paella_Restaurante Levante_Valencia_Campanar_4,5_Moderado_1356
paella_Restaurante Navarro_Ciutat Vella_4,6_Moderado_2760
paella_Restaurante Casa Carmela_Poblats Marítims_4,4_Caro_7327
paella_Arrocería Ricepaella_Delivery Take Away_Ciutat Vella_4,8_Asequible_760
```

Figura 20: Datos de Google Maps tras ser procesados y almacenados en un archivo de texto.

En resumen, los resultados recopilados de Google Maps arrojan las siguientes estadísticas según el tipo de comida utilizado en la búsqueda (Figura 21 y Tabla 3):

Tabla 3: Número de restaurantes recopilados por tipo de comida.

Tipo de comida	Restaurantes
Paella	94
Arroz a banda	82
Arroz al horno	91
Fideua	99
Esmorzar	88
Esgarraet	46
Tapas españolas	38
Patatas bravas	99
Tortilla española	94
Calamares a la romana	43
Croquetas	82
Ensaladilla	97
Comida italiana	75
Pasta	49
Pizza	54
Lasaña	81
Tiramisú	36
Carbonara	85
Marisco	21
Mariscada	25
Pescado	28
Gambas	34
Langosta	22
Pulpo a la brasa	39
Carne	68
Carne a la brasa	25
Hamburguesa	37
Asado	25
Barbacoa	38
Chuletón	70
Comida china	64

Comida japonesa	74
Kebab	30
Comida india	19
Comida mexicana	28
Comida latina	26
Horchata	21
Helado	60
Café	72
Churros	32
Fartons	32
Pastel	10

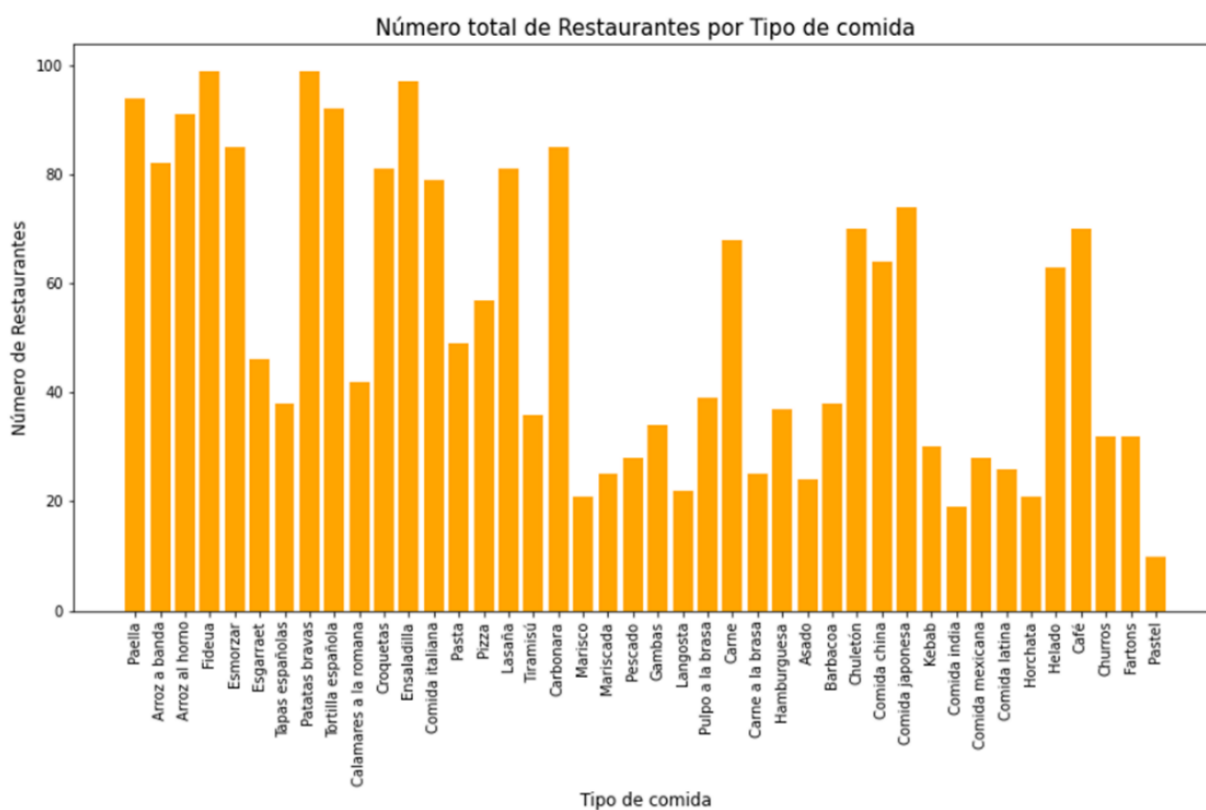


Figura 21: Representación gráfica del número de restaurantes recopilados por tipo de comida.

### 3.4 Análisis de los datos recopilados y preprocesados

Tras haber recopilado y procesado los datos necesarios provenientes de ambas redes sociales, procederemos a analizar la información resultante de forma que se intente encontrar algún tipo de correlación entre las interacciones de Twitter y la popularidad de los restaurantes publicados en Google Maps. Para ello vamos a realizar dos tipos de análisis:

- **Análisis de datos de los restaurantes:** Se analizarán los datos de los restaurantes,

los cuales incluyen el tipo y categoría de comida, el nombre del restaurante, el barrio en el que se encuentra, la puntuación media, el rango de precios y el número de reseñas. Se eliminan aquellos restaurantes que estén duplicados en una misma categoría, contabilizándolos tan solo en una ocasión. De este modo el resultado reflejado será más real, mostrando la información de 1.420 restaurantes únicos en cada categoría.

- **Análisis de datos de Twitter:** Se analizará el contenido de los tuits relacionados con los distintos tipos de comida considerados y sus interacciones.

Ambos análisis se integrarán en el mismo código y se representarán mediante diferentes métricas y gráficos para tener una percepción más visual de las relaciones entre las dos redes sociales. Se generará un archivo de hoja de cálculo Excel el cual contendrá toda la información de ambas fuentes con el objetivo de poder ser estudiado más en profundidad por el usuario y representar las métricas que considere más interesantes.

Para cargar y analizar los datos se han relacionado las palabras clave de estos con la categoría a la que pertenecen. Se han definido siete categorías distintas, como se ha indicado en capítulos anteriores. Estas categorías comprenden la comida valenciana, la comida española, la comida italiana, carnes, pescados, comida internacional y cafeterías. Con esta agrupación se pretende interpretar los datos y tener una visión más global de la opinión de la gente, puesto que un análisis individual para cada tipo de comida en concreto sería muy complicado, ya que es muy común que un restaurante cocine diferentes tipos de comida. Por esto se han intentado agrupar las diferentes palabras clave de la manera más lógica posible.

Para alcanzar el objetivo principal de nuestro estudio se puede intentar encontrar correlaciones analizando los datos desde diferentes ángulos. En nuestro caso vamos a ceñirnos a tres:

- **Número de reseñas y número de tuits por categoría:** Una correlación entre ambos datos puede sugerir que las categorías más discutidas en Twitter son también las que generan más reseñas y, por tanto, aunque la puntuación no sea la mejor, la que más afluencia de gente tienen (Figura 22).

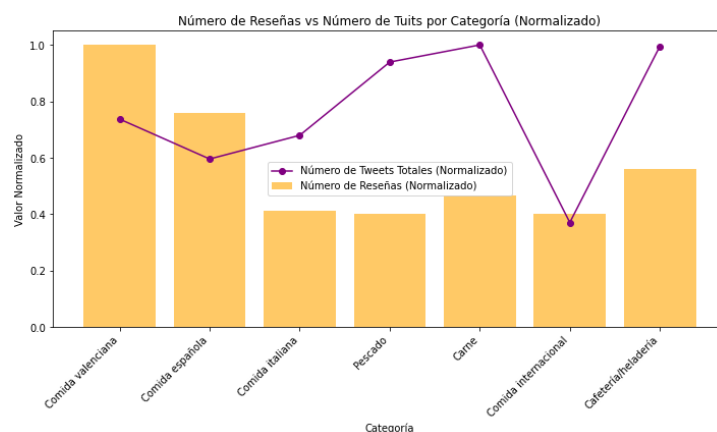


Figura 22: Representación gráfica del número de reseñas y tuits totales según categoría.

En este gráfico (Figura 22), en el cual previamente se han normalizado los datos, podemos detectar cómo en algunos casos existe una cierta relación entre el número de reseñas y las conversaciones que se generan en Twitter acerca de estas categorías. Aun así vemos como las categorías de “comida italiana”, “pescado”, “carne” y “cafetería/heladería” se desvían de esta tendencia, por lo que sugiere que el volumen de discusión en Twitter no siempre se traduce directamente en una mayor afluencia en estos restaurantes.

- **Número de restaurantes y número de tuits por categoría:** El resultado de este análisis podría indicar que los restaurantes responden a la demanda o a las gastronomías con más influencia en las redes sociales como Twitter.

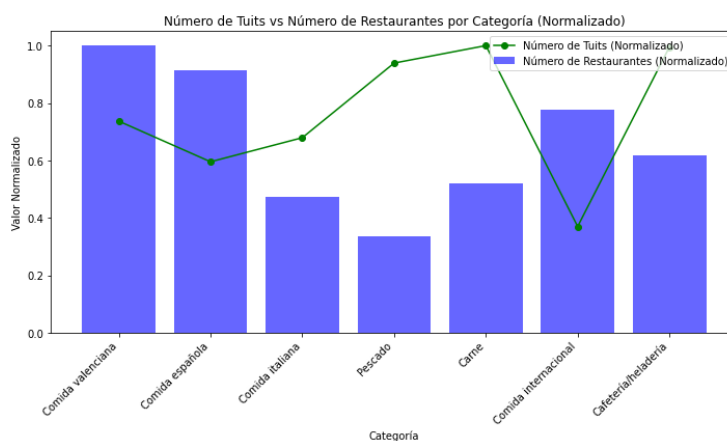


Figura 23: Representación gráfica del número de restaurantes y tuits totales según categoría.

Como se puede observar en la (Figura 23), también con los datos normalizados, observamos cómo existe una correlación pero menor entre el número de interacciones en Twitter y el número de restaurantes. Se puede ver como el número de tuits totales es en cierto modo proporcional al número de restaurantes para tres de las siete categorías estudiadas. Observamos una diferencia más notable en las categorías de pescado y de carne, seguidas por

la categoría de cafeterías/heladerías. Esto puede servir como indicador para seguir estudiando este camino en próximos apartados, pues parece que la oferta no llega a cubrir la demanda para esos casos, por lo que cabría la posibilidad de ser un indicador de idea de negocio.

- **Puntuación media de restaurantes y número de tuits por categoría:** Una correlación positiva aquí podría indicar que las categorías que reciben mejor puntuación en las reseñas también tienden a generar más comentarios en Twitter.

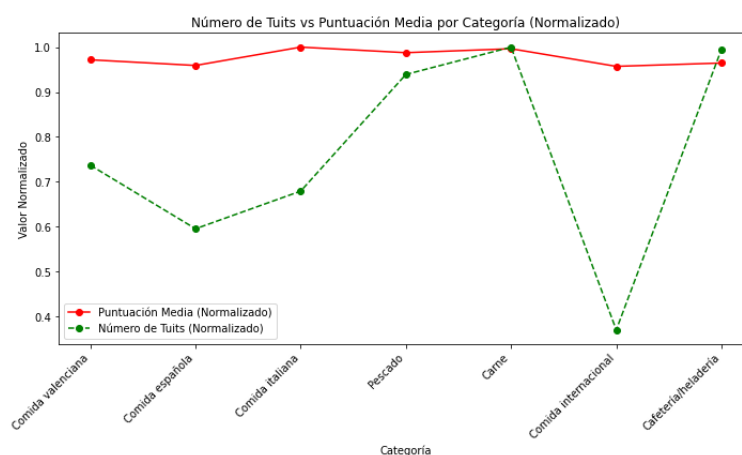


Figura 24: Representación gráfica de la puntuación media y tuits totales según categoría.

En la Figura 24, a diferencia de los anteriores, no vemos una clara tendencia entre el número de tuits y la puntuación media de estas categorías. Esto puede indicar que, aunque los restaurantes de mayor calidad generan buena reputación, no siempre son los que logran captar la atención en Twitter.

- **Precio medio y número de tuits por categoría:** Encontrar una tendencia entre el precio medio de los restaurantes y el número de tuits podría indicar de qué manera se podría ajustar nuestra propuesta de inversión para estar dentro de precio en el mercado actual.

La Figura 25 ofrece información clave para ajustar nuestra propuesta de inversión. Se aprecia cómo las categorías con un precio medio más alto, como “pescado” y “carne” tienden a tener también un mayor número de tuits. En general parece que las conversaciones en Twitter están alineadas con el precio medio por categoría. Vemos una ligera diferencia en las categorías de “carne” y “cafetería/heladería”, la cual va a servir más adelante para ajustar el precio de carta de nuestra propuesta.

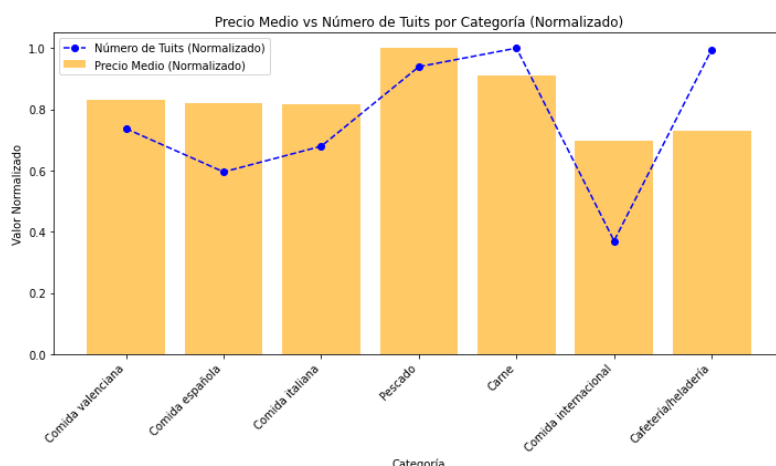


Figura 25: Representación gráfica del precio medio y tuits totales según categoría.

Para realizar este análisis hemos desarrollado un pequeño programa para calcular las estadísticas en base a los datos recopilados hasta el momento, y por otro lado para graficar las comparaciones planteadas anteriormente. Para ello hemos utilizado las siguientes librerías:

- **Os:** Proporciona funciones para interactuar con el sistema, en nuestro caso, navegar por los directorios.
- **Pandas:** Librería de manipulación y análisis de datos.
- **Matplotlib.pyplot:** Librería de gráficos que permite crear visualizaciones estáticas. En nuestro programa lo usaremos para generar los gráficos comparativos.
- **Openpyxl:** Permite leer y escribir en archivos Excel.
- **Datetime:** Permite trabajar con fechas y horas. Será utilizado para generar un nombre único a nuestros archivos generados.

En el programa de análisis primeramente hemos definido los directorios donde se encuentran los datos de los restaurantes y los tuits preprocesados, así como el directorio de salida para almacenar el archivo Excel que generaremos. Definimos también las categorías de comida agrupando los diferentes tipos de alimento según se ha detallado anteriormente, y hemos creado las listas y diccionarios para almacenar estadísticas de ambas fuentes de datos.

Creamos dos funciones para identificar, por un lado, si el texto en cuestión contiene alguna de las palabras clave mediante la función “contains\_keyword()”; y por otro lado, para identificar la categoría del restaurante o tuit según la palabra clave detectada mediante la función “identify\_category()”.

Al ejecutar el análisis de los datos obtenidos de Google Maps, procesamos línea por línea el archivo de texto que contiene la información y nos basamos en el separador que se utilizó anteriormente en el preprocesamiento “\_” para identificar cada una de las partes de la información. Se reemplazan los separadores decimales de los números con coma “,” por un punto “.”, para poder trabajar con ellos en Python. A continuación se almacenan todos los campos del restaurante en cuestión en la lista anteriormente creada llamada “restaurants\_stats”.

Para cada tuit se busca si este contiene alguna de las palabras clave utilizadas durante el análisis. Si contiene alguna palabra clave asociada a las categorías de comida previamente definidas se actualiza el contador en la estructura “tweet\_stats” para esa categoría, incrementando el número de tuits asociados a ésta. A continuación reorganizamos los datos de los tuits para poder ser almacenados y visualizados fácilmente en Excel, y creamos una nueva lista llamada “twitter\_stats\_data” que contiene un diccionario para categoría, donde se resumen las estadísticas calculadas.

Para poder guardar la información de manera eficiente en la hoja de cálculo se utiliza la librería “pandas” para crear dos “Dataframes” que organizan los datos de restaurantes y las estadísticas de Twitter. Estos “Dataframes” permiten convertir la información en una estructura de datos tabulares fácilmente exportable a Excel.

Realizamos la extracción de las métricas clave desde los “DataFrames” creados anteriormente, obteniendo el número de restaurantes, reseñas, puntuación media y tuits totales, todo ello organizado por categoría y sumando los valores para representarlas correctamente. Para poder graficar las diferentes estadísticas calculadas, normalizamos los datos. De este modo conseguimos tener una apreciación más real a la hora de comparar ambas fuentes, pues no es igual de relevante la información contenida en un simple tuit de un usuario que la ficha de Google Maps de un restaurante.

Finalmente, se generan los diferentes gráficos que utilizaremos para comparar las métricas normalizadas, principalmente el número de restaurantes, la valoración, las reseñas y el número de tuits por categoría. Estos gráficos se guardan como una imagen. Para la interpretación de los datos de precio medio de los restaurantes se ha utilizado una equivalencia numérica del siguiente modo: “asequible = 1”, “moderado = 2”, “caro = 3”. Así es más sencillo ejecutar cálculos con este parámetro y representarlos.

En las Tablas 4 y 5 se resumen todos los parámetros por categoría.

A modo de resumen, los métodos planteados durante el estudio son suficientes como para permitir la recolección, procesamiento y análisis de una amplia gama de datos provenientes

tanto de Google Maps como de Twitter. Con estos definidos, procedemos ahora a preparar y construir los datos necesarios para entrenar nuestro modelo de red neuronal. Si las predicciones de la red son satisfactorias, contaremos con una nueva fuente de datos a incorporar en la búsqueda de correlaciones y obtener así una visión más en detalle sobre la relación entre la presencia digital y el éxito de los establecimientos gastronómicos, permitiendo así una identificación más rápida de posibles vías de negocio en el sector de la restauración en base a las interacciones y al sentimiento de los usuarios de Twitter.

*Tabla 4: Estadísticas de Google Maps según categoría.*

Tipo de comida	Restaurantes	Barrio más común	Puntuación media	Reviews medias	Precio más común	Precio medio (numérico)
Comida valenciana	306	Ciutat Vella	4,26	1771	Moderado	1,63
Comida española	280	Ciutat Vella	4,21	1471	Moderado	1,61
Comida italiana	145	Ciutat Vella	4,38	1536	Moderado	1,60
Pescado	103	Ciutat Vella	4,30	2103	Moderado	1,96
Carne	159	L'Eixample	4,34	1584	Moderado	1,79
Comida internacional	238	L'Eixample	4,20	910	Asequible	1,37
Cafeterías/heladerías	189	Ciutat Vella	4,25	1601	Asequible	1,43

*Tabla 5: Estadísticas de Twitter según categoría.*

Tipo de comida	Tuits	Favoritos	Retuits
Comida valenciana	448	9683	1524
Comida española	362	1711	130
Comida italiana	413	13059	815
Pescado	571	2244	266
Carne	608	17537	1848
Comida internacional	225	5459	731
Cafeterías/heladerías	605	20321	1335

# Capítulo 4. Aplicación de la red neuronal y propuesta de inversión

## 4.1 Base de datos de entrenamiento de la red neuronal

Para entrenar el modelo neuronal utilizaremos un conjunto de datos ya etiquetados, que sirva como base para aprender a realizar la tarea de clasificación del sentimiento. Este conjunto de datos será utilizado tanto para el entrenamiento como para la validación y testeo. Durante el entrenamiento, el conjunto de datos se dividirá en ejemplos de tuits positivos y negativos. La red neuronal convolucional (CNN) procesará estos ejemplos, minimizando la diferencia entre sus predicciones y las etiquetas reales. Un subconjunto de estos datos se usará para validar el rendimiento del modelo mientras se ajustan sus hiperparámetros, y otro subconjunto se utilizará para probar la generalización del modelo una vez que haya sido entrenado. Esto garantizará que éste no solo funcione bien con los datos de entrenamiento, sino que también sea capaz de manejar datos nuevos nunca vistos.

La fuente de datos ya etiquetada que hemos utilizado se ha obtenido de uno de los recursos más reconocidos para el análisis de sentimientos en español, el TASS (Taller de Análisis de Sentimientos) (TASS, 2024). Este taller es un evento académico que se celebra en el marco de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN), el cual se lleva celebrando desde el año 2012, y se centra en la investigación del análisis de sentimientos en español (SEPLN, 2024).

A través de su página web oficial hemos podido cumplimentar un formulario para solicitar la base de datos de tuits clasificados con la que cuentan. Tras ello, hemos recibido un enlace de descarga en el cual contamos con las bases de datos utilizadas en todos los eventos celebrados desde el 2012 hasta el 2020.

Estos datos han sido clasificados en tres categorías: positivo (P), negativo (N) o neutro (NEU). En nuestro caso, en el que vamos a recopilar opiniones sobre los distintos tipos de comida, vamos a eliminar los ejemplos de tuits clasificados como neutros. Tras la descarga de estos datos contamos con una base de datos de tuits ya clasificados en formato JSON (Figura 26).

```

    "content": {
      "text": "@bichoclarke @AndreaVasRo El otro día me quejé que x un capuccino
obraban más de lo que vale 1 caja de leche y me mandaron a la mierda"
    },
    "date": {
      "text": "Sun Sep 11 21:58:54 +0000 2016"
    },
    "lang": {
      "text": "es"
    },
    "sentiment": {
      "polarity": {
        "value": {
          "text": "N"
        }
      }
    }
  },
  {
    "tweetid": {
      "text": "768874738089275393"
    },
    "user": {
      "text": "627838882"
    },
    "content": {
      "text": "El sábado se casa Adriel y eso me hace sentir super vieja mae cuando
bamos a pueblo viejo x olafos saliamos fuck off y yo era la niñera"
    },
    "date": {
      "text": "Thu Aug 25 18:16:37 +0000 2016"
    },
    "lang": {
      "text": "es"
    },
    "sentiment": {
      "polarity": {
        "value": {
          "text": "NEU"
        }
      }
    }
  }
}

```

Figura 26: Archivo original de la base de datos utilizada para el entrenamiento de la red neuronal.

En base a esta clasificación hemos generado un pequeño programa para automatizar el proceso de repartición de los archivos en dos carpetas distintas: “P”, para los positivos; y “N”, para los negativos. Además se han tenido que convertir del formato JSON a archivo de texto plano para su manejo posterior.

Para ello hemos definido dónde se ubica el archivo original, y dónde queremos almacenar los tuits ya separados (un tuit por archivo de texto) y clasificados según el sentimiento. Dentro del bucle principal del programa iteraremos sobre cada tuit para obtener de éste su identificador, el contenido del tuit y el sentimiento del mismo, el cual viene anotado dentro del elemento “sentiment” siguiendo la jerarquía de estructura “sentiment > polarity > value > text”. El tuit procesado se almacenará en la carpeta correspondiente según su polaridad y el nombre del archivo de texto en cuestión será el identificador del tuit procesado.

Como se puede ver, este programa servirá para generar una fuente de datos considerable con la que poder entrenar la red neuronal convolucional. Tras el procesamiento de los datos y la repartición contamos con un total de 2986 tuits positivos y 2785 tuits negativos ya clasificados.

Tras entrenar nuestro modelo de predicción de sentimiento con la base de datos anterior,

utilizaremos los datos recopilados con anterioridad a través de las técnicas de “web scraping” para predecir su sentimiento y comparar estos resultados con las estadísticas analizadas anteriormente.

## 4.2 Programación de la red neuronal para predicciones del sentimiento

En la programación de nuestra red neuronal se van a utilizar principalmente las siguientes librerías:

- **TensorFlow y Keras:** Biblioteca de aprendizaje automático de código abierto desarrollada por Google. Esta biblioteca proporciona herramientas para construir y entrenar la red neuronal, manejando operaciones matemáticas complejas y optimizando el rendimiento del modelo. Keras es una API ejecutada sobre TensorFlow, utilizada para crear las distintas capas del modelo.
- **Numpy:** Numpy se trata también de una biblioteca de código abierto que habilita la computación numérica en Python. Esta librería contiene métodos para la realización de operaciones estadísticas y matemáticas entre otras funciones.
- **Matplotlib:** Se trata de una biblioteca para la creación de gráficos en Python. Será utilizada para visualizar el rendimiento del modelo durante el entrenamiento de este facilitando la interpretación de su evolución.

El primer paso para implementar la CNN será asegurar que el entorno esté correctamente configurado y que los datos estén organizados de manera que faciliten el entrenamiento y la evaluación del modelo. Con este fin establecemos las rutas necesarias para acceder a los datos y realizamos una separación clara entre los conjuntos de entrenamiento y prueba.

Comenzamos creando las rutas a las carpetas donde se almacenarán los datos de entrenamiento y prueba. Esto es importante sobre todo a la hora de trabajar con grandes volúmenes de datos, como es el caso. Primero, definimos el directorio principal donde se guardarán todos los datos relacionados con el proyecto. A partir de este directorio base crearemos dos subdirectorios: “train” y “test”, para almacenar los datos de entrenamiento y de prueba, respectivamente. Si estos subdirectorios ya existen en el sistema de archivos no se ejecutará de nuevo la acción de creación.

Siguiendo con el proceso de copia y eliminación de archivos para la preparación del conjunto de datos, procederemos a dividir los datos en los dos directorios creados. Antes de nada, indicaremos al programa las subcarpetas que contienen los archivos a dividir, las

carpetas llamadas “P” y “N” generadas anteriormente en la preparación del conjunto de datos. Una vez definidas y detectadas estas subcarpetas, procederemos a iterar sobre cada una de ellas. Mezclaremos los archivos de cada una de éstas para evitar que se agrupen de manera que se puedan introducir sesgos en el modelo durante su entrenamiento, es decir, que siempre se tomen los mismos datos para el aprendizaje.

En este punto procederemos con la repartición de archivos. Esta división es esencial para poder evaluar la capacidad del modelo a la hora de generalizar los datos nunca vistos. En nuestro caso escogeremos el 20% de los datos (unos 1.019 tuits) para el conjunto de prueba, asignándolos a la lista “test\_files”, mientras que los datos de entrenamiento se asignarán a la lista “train\_files”.

Con los archivos ya divididos en dos conjuntos distintos procederemos a copiarlos a sus respectivas carpetas de destino. Realizaremos de nuevo una serie de comprobaciones para verificar que dentro de cada directorio la subcarpeta de destino exista, y si no es así, la crearemos, copiando cada archivo desde su ubicación original a esta nueva ubicación. Una vez terminada la copia de archivos se eliminarán las subcarpetas originales, manteniendo así el directorio base limpio de información innecesaria y evitando duplicaciones en los datos.

Siguiendo con la creación y preparación de los conjuntos de datos, procederemos a aplicar ciertas técnicas de estandarización y vectorización para transformar el texto de entrada en una forma que la red neuronal pueda interpretar. Para ello comenzaremos definiendo el número de ejemplos de datos que se procesarán juntos en una sola iteración del modelo, también conocido como el “batch size”. Cuanto más pequeño sea el tamaño de este lote, el entrenamiento será más lento, pero es probable que generalice mejor la información. En este caso escogemos un tamaño de lote de 8. Esto significa que el modelo tomará 8 ejemplos de datos, es decir, 8 tuits y los pasará por la red neuronal antes de realizar una actualización de sus pesos. Por otro lado, se establecerá una variable conocida en el ámbito de las redes neuronales como semilla o en inglés “seed”, la cual se encargará de iniciar un proceso aleatorio, pero de manera controlada. Esta variable se ajustará a un valor de 42 en nuestro trabajo, de manera que cada vez que se ejecute el código y entre en juego algún proceso aleatorio, los resultados sean predecibles y repetibles.

Una vez definidos los parámetros anteriores, procederemos a identificar los datos de texto en tres conjuntos: entrenamiento, validación y prueba. El conjunto de entrenamiento será el que se divida de manera que también se utilicen parte de esos datos para la validación, de modo que se pueda evaluar el rendimiento del modelo. Una vez asignados los conjuntos de datos, identificaremos las clases con las que estamos trabajando. En este caso estamos clasificando los tuits en dos categorías: “0” para negativos y “1” para positivos. De este modo proporcionaremos al modelo una referencia binaria sobre las etiquetas que debe aprender a

predecir.

Ya que el texto de la fuente de datos utilizada no es directamente aplicable a una red neuronal vamos a tener que preprocesarlo de manera similar al texto obtenido de Twitter, de modo que estandaricemos y eliminemos cualquier ruido que pueda interferir en el aprendizaje. Para ello convertiremos todo el texto a minúsculas, eliminando el uso de “hashtags”, menciones a otros usuarios, números, y signos de puntuación. Además, eliminaremos las etiquetas HTML y los enlaces, los cuales no aportan valor al análisis de sentimientos.

A continuación se definirá la fase de vectorización, la cual consiste en la conversión del texto a una serie de números enteros que representarán las palabras. Esto facilitará al modelo aprender a través de secuencias de números y así interpretar el texto. Cada palabra convertida a número se conoce como “token”. Por ejemplo, en la frase “Me gusta la pizza”, los tokens serían “Me”, “gusta”, “la”, “pizza”. Conociendo la diversidad de lenguaje utilizado en la red social, estableceremos un número máximo de “tokens” a almacenar en el vocabulario del modelo. Este límite lo estableceremos en las 10.000 palabras más frecuentes, descartando las que menos se repitan. Para asegurar que todas las entradas tienen la misma longitud de palabras, fijaremos una longitud máxima de procesamiento de 250 “tokens”, pues la mayoría de los tuits son cortos y contienen la información relevante en las primeras palabras.

Para preparar los datos de texto de manera que la red neuronal pueda procesarlos eficazmente, comenzaremos extrayendo sólo el texto de los ejemplos de entrenamiento y adaptaremos la capa de vectorización para que aprenda y represente el vocabulario en forma numérica. Esta vectorización se aplicará luego a todos los conjuntos de datos (entrenamiento, validación y prueba). Para asegurar un proceso de entrenamiento más eficiente, implementaremos las técnicas de “cache” y “prefetch”, que almacenarán los datos procesados en memoria, evitando recalcularlos en cada ciclo y preparando los siguientes lotes mientras el modelo está trabajando en el actual. Esto reducirá los tiempos de espera y optimizará el rendimiento general del entrenamiento.

Uno de los conceptos clave en problemas del procesamiento del lenguaje es el “embedding”. Este hace referencia a la representación vectorial de las palabras, donde cada una de las palabras se convierte en un vector de números. Si no se aplicara esta capa de “embedding” y se asignara un simple número a cada palabra, la red solo vería estos números como valores sin relación entre ellos, sin llegar a comprender que palabras como por ejemplo “bueno” y “excelente” son similares. Al convertir estos números en vectores, en este caso de 100 dimensiones, representamos las palabras como un vector del tipo [0.25, 0.10, 0.30...]. De este modo, los vectores para palabras como “bueno” y “excelente” serán similares y por tanto la red será capaz de relacionarlas.

A continuación, procederemos a construir el modelo secuencial de la CNN, compuesto por cinco capas distintas conectadas entre sí:

- **Capa de embedding:** Es la primera capa de la red. Esta se encargará de convertir las secuencias de enteros en vectores de longitud fija, ayudando a la red neuronal a captar las relaciones semánticas entre palabras.
- **Capas convolucionales:** En nuestro modelo, utilizaremos dos capas convolucionales. Estas son útiles para extraer características locales de las secuencias de texto, como frases o palabras clave. La primera capa tiene 128 filtros y la segunda 64, reduciendo la dimensionalidad de las secuencias y resaltando los patrones más relevantes.
- **MaxPooling y GlobalMaxPooling:** Después de cada capa convolucional, aplicamos una capa de operación que seleccionará los valores máximos, reduciendo aún más la dimensionalidad y ayudando a retener solo la información más significativa. La capa “GlobalMaxPooling1D” tomará los valores máximos a lo largo de toda la secuencia.
- **Capas densas:** Las capas densas conforman la parte final del modelo. Estas, permiten a la red combinar la información extraída en las anteriores capas, realizando la clasificación final. La activación “relu” se utiliza para introducir no linealidades en el modelo. En la capa de salida se utiliza una función de activación “softmax”, la cual convierte las salidas del modelo en probabilidades sobre las dos clases posibles (positivo o negativo).
- **Dropout:** Esta se trata de una técnica de regularización con motivo de evitar el sobreajuste. Esta situación se da cuando el modelo se ajusta demasiado bien a los datos de entrenamiento y por tanto no llega a generalizar correctamente. En este caso utilizamos la capa de “Dropout” para desactivar el 70% de las unidades de las capas densas de forma aleatoria, forzando al modelo a ser más robusto.

En la Tabla 6 y la Figura 27 se puede ver un resumen del modelo con los detalles de la arquitectura y el número de parámetros que debe aprender durante cada capa de entrenamiento.

Una vez definida la estructura de la CNN procederemos a implementarla. Utilizaremos como función de pérdida “SparseCategoricalCrossentropy”, la cual medirá la diferencia entre las predicciones del modelo y las etiquetas reales, intentando minimizar esta diferencia; y utilizaremos como métrica principal la precisión, “accuracy”, la cual mide el porcentaje de

predicciones correctas realizadas. Se utilizará además el optimizador del tipo Adam (del inglés, “Adaptive Moment Estimation”), el cual se encargará de ajustar automáticamente la tasa de aprendizaje, que inicialmente se fija a un valor bajo de 0.0001 para que el modelo converja de manera más estable.

Tabla 6: Resumen del modelo de red neuronal configurado.

Capas (tipo)	Forma Salida	Parámetros	Descripción
Embedding	(None, None, 100)	1000100	Representación de palabras en 100 dimensiones
Conv1D	(None, None, 128)	64128	Convolución 1D para captar patrones en secuencias
MaxPooling1D	(None, None, 128)	0	Reduce la dimensionalidad, conservando valores clave
Conv1D	(None, None, 64)	41024	Segunda convolución 1D con 64 filtros
GlobalMaxPooling1D	(None, 64)	0	Pooling global que selecciona el valor máximo de cada filtro
Dropout	(None, 32)	0	Desactiva aleatoriamente el 70% de las unidades para evitar sobreajuste
Dense	(None, 16)	528	Capa densa con 16 neuronas, activación ReLU (MLP)
Dropout	(None, 16)	0	Dropout para regularización
Dense	(None, 2)	34	Capa final MLP con activación softmax para clasificación binaria

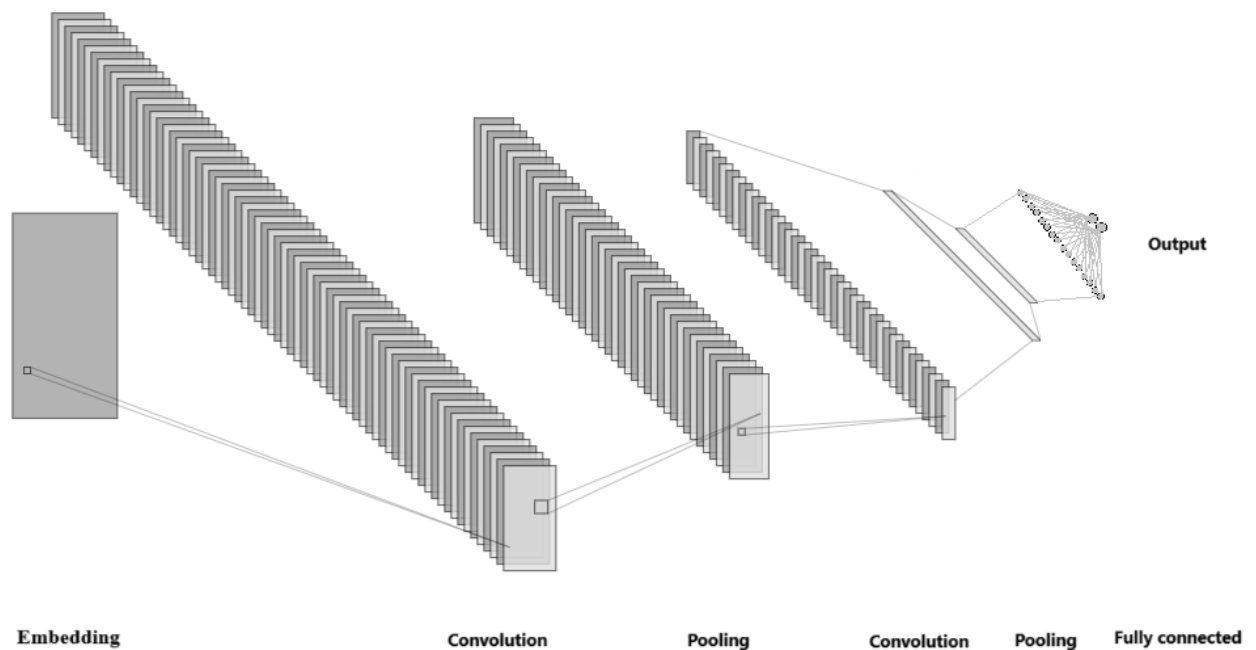


Figura 27: Representación gráfica de la arquitectura de la red neuronal.

Antes de iniciar el entrenamiento de la red definiremos diferentes funciones conocidas como “callbacks”, las cuales se ejecutarán automáticamente durante el entrenamiento en ciertos momentos predefinidos, permitiendo controlar y ajustar el proceso si vemos que el modelo evoluciona por un camino que consideramos que no es el óptimo. En nuestro caso haremos uso de tres tipos distintos de “callbacks”:

- **Early Stopping:** Monitorea el valor “val\_loss”, que es la función de pérdida calculada sobre el conjunto de datos de validación al final de cada época, de modo que si no se observa una mejora en esta métrica durante 2 épocas consecutivas el entrenamiento se detiene, evitando sobreajustes.
- **Model Checkpoint:** Se encarga de guardar el mejor modelo encontrado durante el entrenamiento, basado en la pérdida de validación. En caso de que el entrenamiento se detenga, el modelo puede ser recuperado para ser utilizado.
- **Reduce Learning Rate:** Este “callback” se encarga de reducir la tasa de aprendizaje en un factor de 0.1 si la pérdida de validación no mejora durante 2 épocas consecutivas.

Finalmente el modelo neuronal se entrenará utilizando los datos estandarizados y aplicando los “callbacks” definidos anteriormente para asegurar que el entrenamiento sea eficiente y el modelo generalice correctamente. Estableceremos las épocas, es decir, el número de veces que se ejecutará el entrenamiento, en 16, aunque siempre se podrá detener antes si las métricas lo requieren.

Tras el entrenamiento del modelo se visualiza en pantalla los resultados, graficando la evolución de la pérdida y de la precisión a lo largo de las épocas del entrenamiento. Este será exportado para aplicarlo al conjunto de tuits obtenido en la primera fase del proyecto y así realizar las predicciones de sentimiento. Durante la exportación combinaremos en el modelo la capa de vectorización de modo que, a pesar de haber ya aplicado un preprocesamiento a los tuits obtenidos, aseguraremos que están en el mismo formato que el conjunto de datos de entrenamiento. Una vez exportado se compilará de nuevo para asegurar que está listo para hacer las predicciones, utilizando la misma configuración de pérdida, optimizador, y métricas que en el modelo original.

Para utilizar el modelo entrenado en los nuevos datos de tuits, prepararemos el entorno especificando las rutas del archivo que contiene la información sin clasificar y la carpeta donde se almacenarán los resultados. Nuestra intención es que los tuits clasificados se almacenen en dos carpetas, según el sentimiento (positivo o negativo). Una vez cargado el archivo a procesar y creadas las subcarpetas, cada tuit pasará a través del modelo exportado

para determinar si es positivo o negativo. Para que el modelo pueda realizar las predicciones será necesario expandir la dimensión de los datos de entrada, ya que el modelo durante el entrenamiento recibía tensores con al menos dos dimensiones (tamaño del lote y la secuencia de tokens), por lo que ahora debemos preparar los datos de predicción para que tengan el formato adecuado que el modelo espera. Dependiendo de la predicción, el tuit será movido a la carpeta correspondiente. Al ser una gran cantidad de datos se creará un registro del progreso para saber cuántos tuits han sido procesados y clasificados durante la ejecución del modelo.

### 4.3 Resultados de las predicciones con la CNN

En esta sección del trabajo se muestran los resultados de la predicción de los sentimientos expresados en los tuits relacionados con las categorías de comida consideradas. El modelo de red neuronal construido ha permitido categorizar cada tuit como positivo o negativo. A través de este proceso hemos podido observar cómo se perciben los diferentes tipos de comida y cuáles de estos son los más discutidos en Twitter.

El modelo empleado es la red CNN, compuesta por diferentes capas diseñadas para capturar las características más relevantes en los textos y predecir la polaridad de estos. En nuestro caso se ha entrenado durante 16 épocas, con un punto de parada temprana debido a la evolución de la pérdida en el conjunto de validación. A continuación se muestra la gráfica de evolución del modelo (Figura 28).

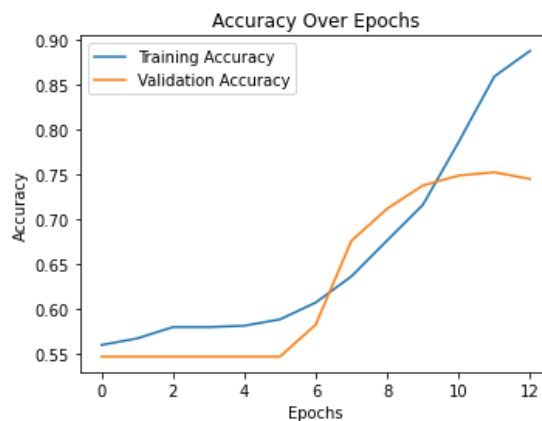


Figura 28: Evolución del rendimiento del modelo a lo largo de las épocas.

Al principio del entrenamiento del modelo se observó que la pérdida disminuía durante el entrenamiento, y también en la validación. Este comportamiento indica que el modelo estaba comenzando a aprender a distinguir entre los tuits positivos y negativos. Durante las últimas épocas del entrenamiento el modelo continúa mejorando, alcanzando una precisión del 88% en el conjunto de entrenamiento. No obstante, la pérdida en el conjunto de validación

comenzaba ligeramente a aumentar a partir de la época 10, lo que sugiere un posible sobreajuste.

La red neuronal artificial entrenada con la base de datos externa se usa posteriormente para clasificar los tuits obtenidos mediante “web scraping” según el sentimiento de estos. Con esto se busca obtener estadísticas acerca de cuántos tuits positivos o negativos encontramos referidos a cada categoría y, por tanto, poder complementar nuestro análisis de datos anterior.

Tras finalizar el entrenamiento del modelo vemos cómo se han creado satisfactoriamente la carpeta de resultados junto a sus subcarpetas para almacenar los tuits positivos y negativos. Del total de 3.232 tuits recopilados, vemos cómo 1.873 han sido clasificados como positivos, y 1.359 han sido clasificados como negativos. Si bien la plataforma es utilizada para expresar opiniones de todos los tipos, el hecho de contar con un usuario anónimo facilita la publicación de opiniones negativas. Sin embargo las estadísticas indican categorías de comida con una cierta apreciación positiva por parte de los usuarios, a pesar de predominar los tuits negativos.

Complementando nuestro programa de análisis de datos expuesto anteriormente se identificará el sentimiento de los archivos de Twitter según la carpeta donde estén almacenados, pues habrán sido clasificados con anterioridad a través de la red neuronal. Para cada tuit se buscará si contiene alguna de las palabras clave utilizadas durante nuestro análisis. Si contiene alguna palabra clave asociada a las categorías de comida previamente definidas, se actualizará el contador en la estructura “tweet\_stats” para esa categoría, incrementando el número de tuits positivos o negativos asociados a esa categoría. A continuación reorganizaremos los datos de los tuits para poder ser almacenados y visualizados fácilmente en Excel, y crearemos una nueva lista llamada “twitter\_stats\_data” que contendrá un diccionario para categoría, donde se resumirán las estadísticas calculadas.

*Tabla 7: Resultados del análisis de sentimiento según categoría.*

Tipo de comida	Número de tuits	Tuits positivos	Tuits negativos
Comida valenciana	448	320	128
Comida española	362	220	142
Comida italiana	413	210	203
Pescado	571	333	238
Carne	608	364	244
Comida internacional	225	105	120
Cafetería/heladería	605	321	284

Como se puede apreciar en la Tabla 7, la mayoría de los diferentes tipos de comida presentan una distribución de sentimientos relativamente equilibrada. En la Figura 29 aparece representado en un gráfico de barras el porcentaje de percepción de cada categoría en base a

estos resultados.

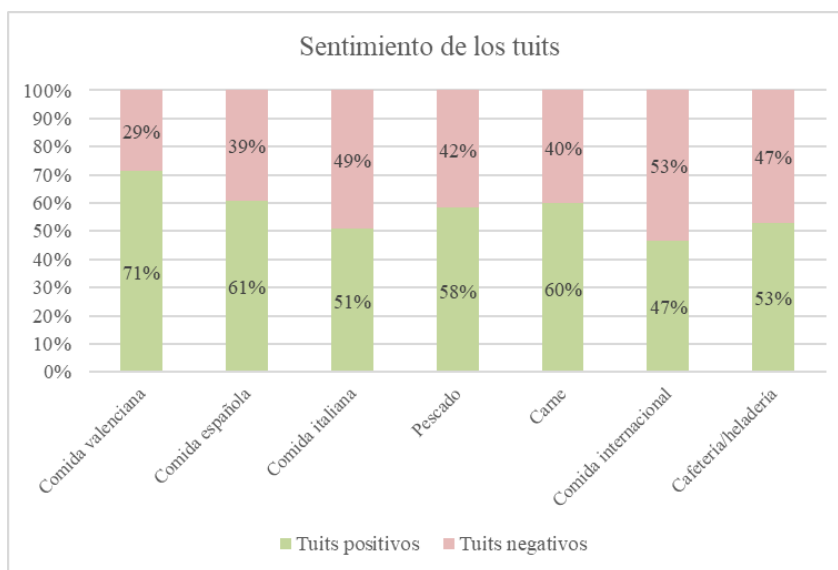


Figura 29: Representación gráfica de los porcentajes de distribución del sentimiento según categoría.

En general, según los resultados de la clasificación, podemos apreciar como el modelo entrenado ha estado bastante acertado, identificando correctamente el sentimiento de cada tuit. Para comprobar la precisión del modelo se ha tomado una muestra representativa de 120 tuits (60 para cada sentimiento), los cuales se han revisado manualmente.

Los resultados en esta muestra arrojan un acierto del 80% para los tuits positivos, mientras que para los tuits negativos este porcentaje de acierto baja hasta el 77%. Esto puede ser debido a que muchos de estos tipos de comida se usan como forma de insulto o como crítica hacia algunos aspectos. Por ejemplo, “pastel” se puede usar en la frase “menudo pastel tienen montado...”. Este uso de las palabras puede llevar a confusiones. Se adjunta este análisis en el Apéndice de la memoria.

## 4.4 Propuesta de sistema de inversión

En este apartado el objetivo será identificar posibles correlaciones entre los datos procesados de ambos conjuntos de modo que se pueda llevar a cabo una propuesta de sistema de inversión basada en el sentimiento de los tuits. Por ello, en base al análisis anterior de los datos recopilados, se realiza de nuevo algunas comparaciones que hemos observado que tienen más relevancia, añadiendo ahora sí el factor de percepción del sentimiento hacia estas categorías.

Al representar el número de tuits positivos para cada categoría vamos a tener en cuenta el

número de favoritos y de retuits de cada tuit en cuestión. Esta información es relevante puesto que contar con alguna de estas interacciones significa que algún usuario más ha compartido la misma opinión que el tuit original. Cada retuit o favorito con el que cuente un tuit se contabilizará como una unidad más de tuit positivo o negativo, según el sentimiento del tuit en cuestión. De este modo el nuevo resultado teniendo en cuenta la interacción sobre cada tuit, viene representado por la Figura 30.

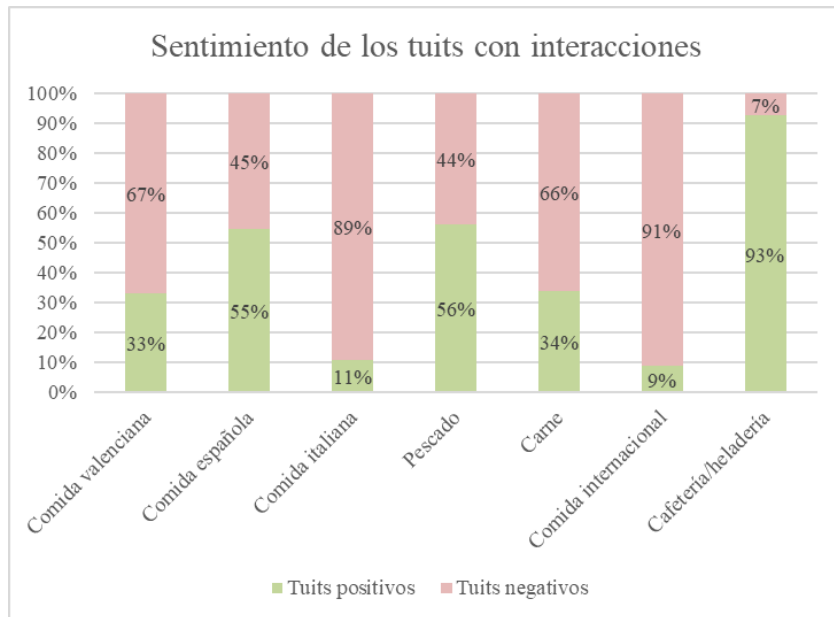


Figura 30: Distribución de los porcentajes de sentimiento según categoría tras haber contabilizado los favoritos y los retuits.

Observamos cómo se incrementa aún más la crítica negativa hacia muchas de las categorías, a diferencia de la categoría “cafeterías/heladerías”, la cual resulta beneficiada.

Tras normalizar los datos recopilados proponemos a analizar las siguientes métricas de los datos recopilados:

- Número de reseñas y tuits positivos por categoría.
- Número de restaurantes y tuits positivos y negativos por categoría.
- Puntuación media y tuits positivos por categoría.
- Precio y tuits positivos por categoría.

En el primer caso vemos como hay una percepción mayormente positiva en Twitter para aquellos restaurantes especializados en carne y para la categoría de cafeterías/heladerías. A éstas les precede la comida valenciana, como se puede observar en la Figura 31.

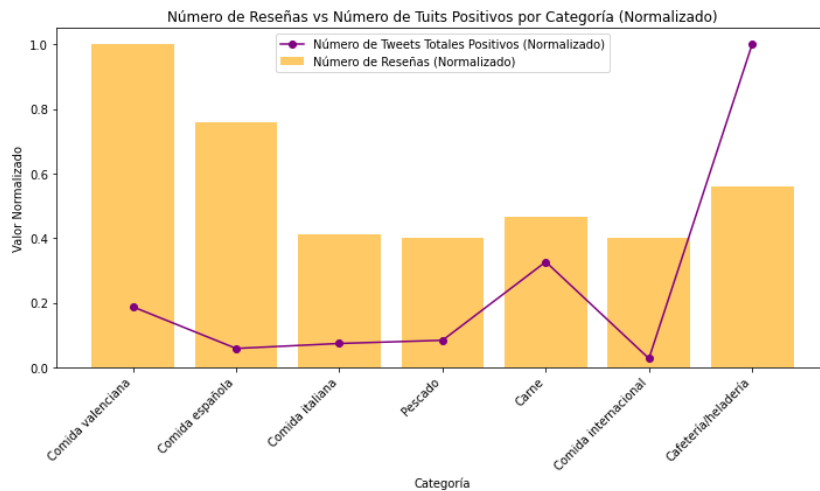


Figura 31: Representación gráfica del número de tuits positivos y de reseñas según categoría.

Parece existir una correlación entre los tuits positivos y el número de reseñas, pues la línea de tendencia de tuits positivos sigue la misma curva que la tendencia de reseñas. Es en el último caso, en la categoría de cafeterías, donde apreciamos una diferencia muy notable entre el número de reseñas y los tuits positivos, a pesar de que la línea de tuits positivos sigue de nuevo la tendencia alcista de las reseñas para esta categoría. Esta información indica que para esta categoría parece ser que las reseñas no son tan numerosas, o que la gente no suele compartir su opinión acerca de éstas: suelen ser sitios donde tomar algo rápido, un desayuno o una merienda, y tal vez no incitan tanto a dejar una reseña del local. Aun así, la gran cantidad de interacciones positivas que presenta esta categoría hace pensar que es un posible nicho de negocio y que, además de lo comentado, esta diferencia pueda deberse a que no existe un gran número de restaurantes dedicados a esta categoría y, por tanto, las reseñas sean menores en número. La gran percepción positiva de los usuarios de Twitter hacia esta categoría junto a la poca cantidad de reseñas que presenta es un indicador de que este tipo de restaurantes podrían tener aún margen de mejora para ser explotados y sacarles más rendimiento.

En el segundo caso observamos una situación similar a la primera comparativa. La comida valenciana y española se siguen situando en la primera posición como categorías con más restaurantes; sin embargo, es la categoría de cafeterías/heladerías la que presenta una mayor percepción positiva (Figura 32). Observando la línea de tendencia positiva, ésta parece seguir en la mayoría de los casos (quitando la categoría de comida internacional) la misma tendencia que el número de restaurantes recopilados para cada categoría. Atendiendo también a las interacciones negativas vemos como en el caso de la comida valenciana, la comida italiana, y la categoría de carne son las peor valoradas en Twitter. Esto concuerda en el caso de las categorías de comida italiana y carne con que presentan un menor número de restaurantes.

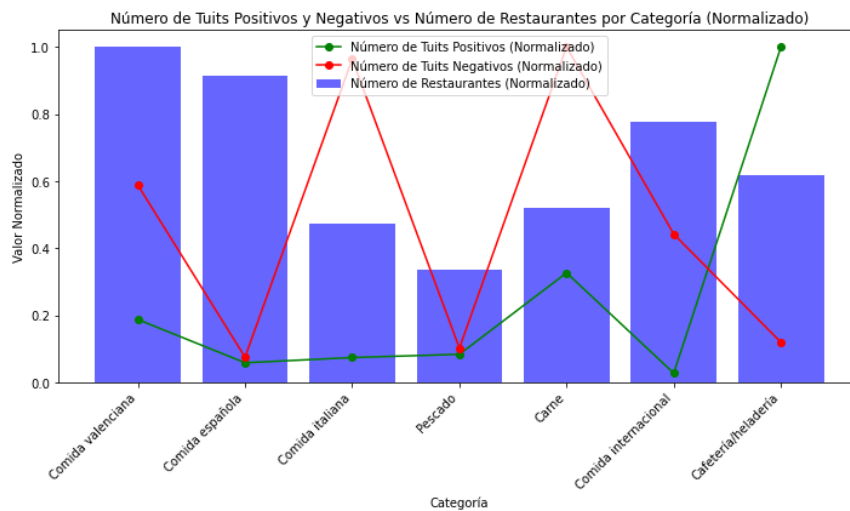


Figura 32: Representación gráfica del número de tuits positivos, tuits negativos y restaurantes según categoría.

Cabe destacar de nuevo en la Figura 32 que la mayor diferencia entre percepción positiva y número de restaurantes la encontramos en la categoría de cafeterías/heladerías, por lo que refuerza aún más nuestra suposición de que es un posible nicho de mercado.

En el tercer caso no parece haber correlación entre el sentimiento de la gente hacia esa categoría en concreto y la puntuación media de los restaurantes que ofrecen esa comida (Figura 33). Aun así, conocer la percepción de los consumidores de una categoría en concreto, junto a las estadísticas sacadas de Google Maps, podría ayudar a predecir la recepción que tendría el restaurante en cuestión. Sí que observamos una relación más clara en la categoría de cafeterías/heladerías donde coincide la percepción en Twitter con la puntuación media.

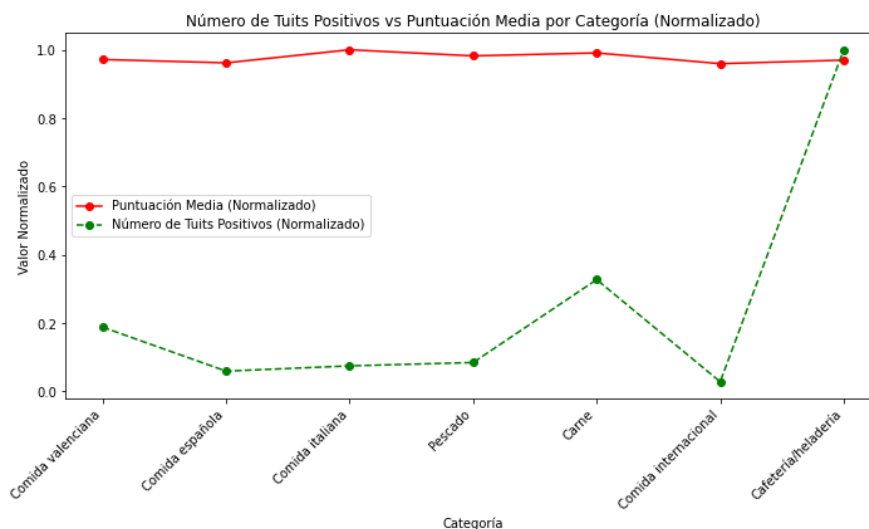


Figura 33: Representación gráfica del número de tuits positivos y de la puntuación media por categoría.

Finalmente, en la gráfica de comparación entre la percepción positiva en Twitter y precio medio de estas categorías (Figura 34), volvemos a observar una cierta tendencia entre ambas variables. Se aprecia como en las primeras categorías desciende la percepción positiva hasta repuntar en la categoría de pescado y carne, las cuales son las que mayor precio medio tienen por restaurante. Esta tendencia de tuits positivos cae en la categoría de comida internacional al igual que el precio medio de sus restaurantes, para luego repuntar en la categoría de cafeterías/heladerías.

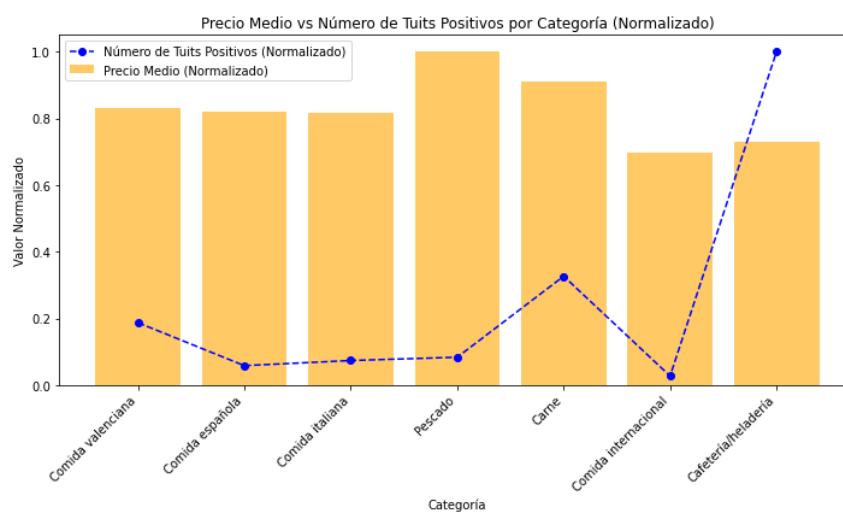


Figura 34: Representación gráfica del número de tuits positivos y el precio medio por categoría.

De la Figura 34 se puede entender que en aquellos casos donde la percepción de la gente es positiva, los restaurantes aprovechan para subir los precios y, por tanto, conseguir más beneficio. Esto coincide para prácticamente todas las categorías y tiene sentido puesto que cuando un restaurante es popular y bien percibido, es habitual que suba los precios. Encontramos sin embargo una gran diferencia de nuevo en la categoría de cafetería/heladería, la cual a pesar de tener un gran número de tuits positivos, su precio medio está muy por debajo de la percepción positiva de la gente.

Aunque en el Capítulo 4 hemos visto diferentes categorías con potencial, como las categorías de pescado y carne (especialmente esta última), las cuales presentaban una gran interacción en Twitter y contaban con un número más reducido de restaurantes en Valencia, tras analizar el sentimiento de los tuits recopilados y contabilizar las interacciones de favoritos y retuits de estos, hemos apreciado un claro despunte en la categoría de cafeterías/heladerías. Esta categoría ha resultado exitosa en tres de las cuatro comparaciones ejecutadas, las cuales han demostrado presentar ciertos patrones de conducta entre las opiniones de Twitter y los datos de Google Maps. Vemos como esta categoría ha demostrado tener gran potencial puesto que a pesar de tener una percepción del 94% positiva (muy por encima del resto), no presenta un número lo suficientemente alto de restaurantes en la ciudad como para abastecer la

demanda. Además de esto, también se posiciona como la categoría con la posibilidad de sacar más beneficio puesto que existe un gran margen entre el precio medio y la percepción en Twitter.

Así pues, atendiendo al análisis de sentimiento realizado, tomamos la categoría de cafeterías/heladerías como clara ganadora. Podemos así plantear una propuesta de negocio con los datos recopilados.

Observando la recopilación de datos de Google Maps podemos ver cómo el barrio donde más restaurantes categorizados como cafeterías/heladerías hay en Valencia es en el de Ciutat Vella. Esto podría indicar que no es buena idea abrir otro local aquí, pero observando las reseñas medias en los restaurantes de la zona se aprecia una media mucho mayor al resto de barrios, de alrededor de 1.900 reseñas por local. En concreto, la media de reseñas para esta categoría en este barrio se sitúa en 2.070 reseñas por restaurante. Esto indica que existe por tanto una gran afluencia de posibles consumidores en la zona, por lo que escogeremos este barrio como el ideal para abrir una cafetería.

Como hemos visto, existe un margen de mejora respecto al precio medio en la categoría escogida. Actualmente la media de precios en esta categoría se sitúa en un valor de 1,43. Esta media corresponde a la conversión realizada anteriormente (“asequible=0”, “moderado=1”, “caro=2”). Por lo tanto, conociendo que los rangos de precio para estas categorías corresponden respectivamente a 0 – 10€, 10 – 20€ y +20€, podemos hacer una estimación del precio medio actual, el cual resulta en aproximadamente +9,55€. Atendiendo a las gráficas anteriores, en la que se apreciaba cómo categorías con menos interacciones positivas como la de pescado contaban con un precio medio de +13€, observamos un margen de subida de los precios de casi un 37%.

Así pues, la propuesta de inversión a partir de estos datos sugiere abrir un restaurante categorizado como cafetería/heladería, el cual es probable que tenga una mayor afluencia de gente si se ubica en el barrio de Ciutat Vella, y cuyos precios pueden ser incrementados hasta un 37% respecto a los precios medios para ese tipo de locales.

# Capítulo 5. Conclusiones y trabajos futuros

En los resultados de los apartados anteriores hemos dado uso a los datos recopilados mediante “web scraping”, así como a la red neuronal diseñada, poniendo en práctica la metodología planteada para diseñar una propuesta de negocio, en este caso aplicada a la apertura de un restaurante categorizado como cafetería/heladería.

A lo largo del análisis se han identificado diversas correlaciones entre las interacciones de los usuarios en Twitter y las métricas obtenidas de plataformas como Google Maps. Aunque algunas correlaciones resultaron más evidentes que otras, los resultados en su conjunto sugieren que existe una tendencia clara en el sector de la restauración ligada a la actividad en redes sociales, a la percepción y preferencias de los consumidores. Esto no significa que los negocios sigan meticulosamente lo que se comparte en esta plataforma en cuestión, pero es evidente que Twitter es un gran reflejo de la sociedad actual a través del cual poder predecir las tendencias de consumo y sacar provecho de ello.

A pesar de los resultados satisfactorios de nuestra red neuronal (alrededor del 78%) así como de las comparaciones ejecutadas, cabe destacar algunas mejoras que se podrían implementar en este sistema de análisis y diseño de propuesta de negocio. En nuestra investigación nos hemos centrado en restaurantes especializados en unos tipos de gastronomía en concreto, que si bien son los más relevantes en Valencia, no abarcan al completo la gran variedad de ofertas gastronómicas de la ciudad.

El hecho de haber clasificado los tipos de comida por categoría, aunque haya sido de gran utilidad para simplificar el análisis, puede llevar a una generalización de los datos y omitir ciertos matices importantes en las preferencias de los consumidores. En este sentido, sería recomendable en futuras investigaciones el uso de más categorías, así como de palabras clave para conseguir un mayor número de tuits y ejecutar un análisis con mayor detalle sobre los distintos tipos de comida.

Además, se ha omitido el uso de otros parámetros relevantes como las coordenadas de los restaurantes en Google Maps. En futuras versiones de este sistema se podrían utilizar estas ubicaciones exactas de los restaurantes para, una vez encontrado el nicho de negocio, saber exactamente en qué zona convendría abrir el restaurante. Primero, analizando por barrios (como se ha hecho en este trabajo) y, posteriormente, haciendo uso de estas coordenadas para

estudiar dentro de cada barrio la distribución exacta de los restaurantes ya existentes y poder elegir una ubicación estratégica para nuestro negocio.

En cuanto a las interacciones en redes sociales, una limitación importante ha sido la imposibilidad de acceder a la ubicación geográfica de los tuits. Si bien esta información anteriormente era pública, en nuestro caso no ha sido posible encontrar estos datos dentro del código web de Twitter. Disponer de este tipo de información permitiría recopilar solo aquellos tuits que hayan sido generados por usuarios ubicados en Valencia o alrededores. Esto resultaría en un análisis del sentimiento acotado a la población de Valencia y, por tanto, mucho más exacto, ya que se entendería a la perfección la preferencia de los consumidores de la ciudad bajo estudio. Al haber recopilado tuits escritos en su mayoría en español, es muy probable que se hayan analizado opiniones de gente de otras comunidades e incluso de otros países, a pesar de que Twitter suele mostrar en sus resultados de búsqueda aquellos tuits escritos por usuarios cercanos.

Una dirección prometedora para futuros desarrollos sería la elaboración de una segunda red neuronal capaz de identificar patrones automáticamente entre todos los datos recopilados, ya sea de Twitter, Google Maps, o incluso otras plataformas sociales como Facebook o Instagram. De este modo se automatizaría la búsqueda de correlaciones y se encontrarían patrones ocultos que en un análisis estático como el realizado es posible que no se haya podido apreciar. Con un modelo entrenado continuamente con nuevos datos y procedentes de diferentes fuentes para capturar percepciones de diferentes usuarios sería posible generar propuestas de negocio más exactas y optimizar la toma de decisiones.

En resumen, en este trabajo se ha demostrado que es posible identificar tendencias relevantes para la toma de decisiones en el sector de la restauración a partir del análisis de sentimiento de Twitter y la recolección de datos de Google Maps. Sin embargo, queda margen tanto en términos de mejora como de expansión del modelo hacia otros tipos de negocios o inversiones.

# Bibliografía

- Amarante, M. F., & Ghione, R. (2019). Análisis de sentimientos en Twitter: percepciones sobre los productos ultraprocesados y su asociación con la prevalencia de obesidad (Bachelor's thesis).
- Buldur, F. E. (2023). *5 Reasons Why Python is the Ultimate Language for Data Science*. <https://medium.com/@farukbuldur/5-reasons-why-python-is-the-ultimate-language-for-data-science-3082ce2ca67b> (last accessed 06/09/2024).
- Lotfi, C., Srinivasan, S., Ertz, M., Latrous, I., & Manjushree, S. (2021). Web scraping techniques and applications: A literature review. In SCRS Conference Proceedings on Intelligent Systems (pp. 381-394).
- ECMA-404 *The JSON Data Interchange Standard*. (2024) <https://www.json.org/json-es.html> (last accessed 06/09/2024).
- Firefox, M. (2024). *Firefox DevTools User Docs*. <https://firefox-source-docs.mozilla.org/devtools-user/index.html> (last accessed 06/09/2024).
- SAURA, J. R., Reyes-Menéndez, A., & PALOS-SANCHEZ, P. (2018). Un Análisis de Sentimiento en Twitter con Machine Learning: Identificando el sentimiento sobre las ofertas de# BlackFriday. *Revista Espacios*, 39(42).
- Kirsten, S. (2024). <https://owasp.org/www-community/attacks/csrf> (last accessed 06/09/2024).
- Klaviyo developers. (2024). *Rate limits, status codes, and errors*. [https://developers.klaviyo.com/en/docs/rate\\_limits\\_and\\_error\\_handling](https://developers.klaviyo.com/en/docs/rate_limits_and_error_handling) (last accessed 06/09/2024).
- Drescher, L. S., Grebitus, C., & Roosen, J. (2023). Exploring food consumption trends on Twitter with social media analytics: The example of# Veganuary. *EuroChoices*, 22(2), 45-52.
- Vidal, L., Ares, G., Machín, L., & Jaeger, S. R. (2015). Using Twitter data for food-related consumer research: A case study on “what people say when tweeting about different eating situations”. *Food Quality and Preference*, 45, 58-69.
- Hayat, M. K., Daud, A., Alshdadi, A. A., Banjar, A., Abbasi, R. A., Bao, Y., & Dawood, H. (2019). Towards deep learning prospects: insights for social media analytics. *IEEE access*, 7, 36958-36979.
- Yadav, N., Kudale, O., Gupta, S., Rao, A., & Shitole, A. (2020, February). Twitter sentiment analysis using machine learning for product evaluation. In 2020 International conference on inventive computation technologies (ICICT) (pp. 181-185). IEEE.
- Pandey, P. (2024). *Web Scraping Using Python for Dynamic Web Pages and Unveiling Hidden Insights*. [https://medium.com/@pankaj\\_pandey/web-scraping-using-python-for-dynamic-web-pages-and-unveiling-hidden-insights-8dbc7da6dd26](https://medium.com/@pankaj_pandey/web-scraping-using-python-for-dynamic-web-pages-and-unveiling-hidden-insights-8dbc7da6dd26) (last accessed 06/09/2024).
- Rayobite. (2024). *The Ultimate Guide to Perform Ethical Google Scraping Without Getting Blocked*. <https://rayobite.com/blog/google-scraping/> (last accessed 06/09/2024).

- Gonzales, R. A. M. (2023). Análisis de sentimientos de los mensajes de Twitter respecto a la empresa KFC del primer trimestre en Hispanoamérica 2022. *Revista Campus*, 28(36).
- Sandoval Serrano, L. J. (2018). Algoritmos de aprendizaje automático para análisis y predicción de datos. *Revista Tecnológica*; no. 11.
- Soni, S., Chouhan, S. S., & Rathore, S. S. (2023). TextConvoNet: A convolutional neural network based architecture for text classification. *Applied Intelligence*, 53(11), 14249-14268.
- Sayce, D. (2022). The Number of tweets per day in 2022.
- Selenium. (2024). *Selenium.dev*. [https://www.selenium.dev/documentation/webdriver/getting\\_started/install\\_library/](https://www.selenium.dev/documentation/webdriver/getting_started/install_library/) (last accessed 06/09/2024).
- Park, S. B., Jang, J., & Ok, C. M. (2016). Analyzing Twitter to explore perceptions of Asian restaurants. *Journal of Hospitality and Tourism Technology*, 7(4), 405-422.
- Liao, S., Wang, J., Yu, R., Sato, K., & Cheng, Z. (2017). CNN for situations understanding based on sentiment analysis of twitter data. *Procedia computer science*, 111, 376-381.
- Sukheja, S., Chopra, S., & Vijayalakshmi, M. (2020, March). Sentiment analysis using deep learning—a survey. In *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)* (pp. 1-4). IEEE.
- SEPLN. (2024). Workshop on Semantic Analysis at SEPLN [www.sepln.org](http://www.sepln.org). (last accessed 06/09/2024).
- TASS. (2024). *TASS: Workshop on Semantic Analysis at SEPLN*. <http://tass.sepln.org/> (last accessed 06/09/2024).
- TIOBE. (2024). *The Python Programming Language*. <https://www.tiobe.com/tiobe-index/python/> (last accessed 06/09/2024).
- Wang, W., & Gang, J. (2018, July). Application of convolutional neural network in natural language processing. In *2018 international conference on information Systems and computer aided education (ICISCAE)* (pp. 64-70). IEEE.
- Yang, Y., Roehl, W. S., & Huang, J. H. (2017). Understanding and projecting the restaurantscape: The influence of neighborhood sociodemographic characteristics on restaurant location. *International Journal of Hospitality Management*, 67, 33-45.

# Apéndice A

Como extensión de los Capítulos 3 y 4, se adjunta a continuación la muestra tomada para comprobar la precisión de las predicciones así como parte de la información recopilada a modo de ejemplo.

## A.1 Comprobación manual de la precisión de la red neuronal para la muestra tomada

A continuación, se adjunta una tabla con los resultados del análisis manual de las predicciones realizadas por la red neuronal para una muestra de 120 tuits.

*Tabla 8: Comprobación manual de las predicciones de la red neuronal.*

Tuit	Predicción	Interpretación humana
Posiblemente los mejores calamares a la romana misrestaurants restaurants aove barcelona	POSITIVO	POSITIVO
Una cenita de tapas Gourmets Huevos revueltos con foie, jamón ibérico de bellota y patatas fritas Y pulpo a la brasa con pamentier de patata y salsa de queso midiversionenlacocina todoricorico cenadedos	POSITIVO	POSITIVO
Pues prefiero morir frente a un buen chuletón	POSITIVO	POSITIVO
Detenido un dependiente de un kebab de Bilbao por abusar de un niño en el baño	POSITIVO	NEGATIVO
Estoy antojada de sushi, comida italiana, comida mexicana, kebab, chuzo de pollo con papitas, basta yaaaaaa	POSITIVO	POSITIVO
hospitales En 2010 y para su cumpleaños, Keanu entró en una pastelería y se compró un brioche con una sola vela, se lo comió frente a la panadería y ofreció cafe a las personas que se paraban a hablar con él Tras ganar sumas astronómicas por la trilogía de Matrix, el actor donó	POSITIVO	POSITIVO
En Terrassa no sólo tenemos a @TodoJingles También hay cosas buenas como la mejor horchata del mundo mundial Hacer 1100 km pensando en mojar los fartons en ella	POSITIVO	POSITIVO
pimientos hechos Ahora a desmigajar bacalao y en un rato esgarrat hecho	POSITIVO	POSITIVO
Croquetas de Jamón @chefkoldo afuegolento recetas cocina	POSITIVO	POSITIVO

Habrá deliciosa comida latina, música en vivo, baile y mucho más para celebrar junto a tu familia Visita nuestro stand para informarte sobre voluntariado, licencias de conducir y tus derechos Ven a compartir momentos especiales y conoce organizaciones comunitarias	POSITIVO	POSITIVO
@Silv3rWitch Lit, quiero tomarme un cafe caliente agosto	POSITIVO	POSITIVO
Enero estoy a dietatgtt mi amigo enviando comida mexicana rotación y promoción de platos de pescados y mariscos Cómo lo haces	POSITIVO	POSITIVO
Hoy ha tocado barbacoa de asado argentino Me puede gustar más	POSITIVO	POSITIVO
@tsarina_nu Día Internacional de esmorzar con un tenedor	POSITIVO	POSITIVO
Queria gostar de comida japonesa	POSITIVO	POSITIVO
@GiovanonniJose Muy bien José Luis Yo acá x ahora mate solo Después una merienda x q mañana hay un cumple con asado	POSITIVO	POSITIVO
@pablomurrrr Peste a escapada a paris con ryanair y hamburguesa con crema de lotus	POSITIVO	NEGATIVO
a hacerme un cafe helado	POSITIVO	POSITIVO
ALERTA PIZZA Muy buena Pizzeria en SantAntoni Barcelona La Piccolina Sepúlveda 60 esquina Rocafort ComidaItaliana	POSITIVO	POSITIVO
Éstas son las grandes historias, recuperadas del anonimato y el olvido con las que me gusta esmorzar Muchas gracias @beanavarro	POSITIVO	POSITIVO
@jorgearmasv @4F_Mamba @SonPolemicas @RoiLopezRivas @AlexNSaab @AliErnesto32 @AmkValentina @CMonteroOficial @MonederoJC @AlbertoRodNews Definitivamente tu tienes q ser marisco	POSITIVO	POSITIVO
@Jonibarco Podríamos hacer una ensaladilla rusa gigante	POSITIVO	POSITIVO
Mimi, no te hagas el que rico pastel, que los cactus na ma son plantas y todos son suculentas	POSITIVO	NEGATIVO
Esmorzar de forquilla is the true brunch	POSITIVO	POSITIVO
@xZeus_Boy Necesitas motivación Le dijo Erik dándole un beso en la mejilla Ate sonrió Mamá, hay más carne Si, hijo, ahora te hago más Ate se puso asar más carne	POSITIVO	POSITIVO
@INCAFE70 @lovenursingmery @Kirschn48409947 @CafeteriadeHos1 Descubre la magia del acompañamiento de la horchata O fartons o danielets	POSITIVO	POSITIVO
@_Na100pre yo ni para correr cafe, si	POSITIVO	POSITIVO
Tu colega el penalista marbellí te invita a una barbacoa	POSITIVO	POSITIVO
Ese pulpo a la brasa Suave y jugoso diatrasdia en Grill Restaurante	POSITIVO	POSITIVO

Pescado fresco de temporada En mi casa solemos comer pescado congelado merluza principalmente Eso si, pagamos religiosamente IBI, impuesto de basura, IRPF, Impuestos Vehículos, ITVs, seguro del coche , hipoteca, seguro vivienda casa, comedor escolar, libros del colegio	POSITIVO	POSITIVO
Y para la merienda unos fartons rellenosmequedoencasa family en Terrassa	POSITIVO	POSITIVO
@muzeriqui @Elafrancesado De Primero Langosta al Horno y un buen Vega Sicilia, luego ya te diré	POSITIVO	POSITIVO
@DESNEURONAO pues si, a mi me gusta la madre, La langosta	POSITIVO	POSITIVO
@escarlata_2 Tal cual Metáfora preciosa de lo que pasa en realidad Y los anormales de los Desokupas, Alvises y Vitos cascándose un chuleton mientras calientan a otros	POSITIVO	NEGATIVO
@VBranchat Sí y el mejor pan con tomate y el alioli y la paella y la fabada todo lo tienen mejor	POSITIVO	POSITIVO
El dumping y el café para algunos, por Joan López Alegre	NEGATIVO	POSITIVO
@peixmariscanna Buenos días Parte de nuestra parada de hoy jueves 5 de Enero 2023 pescado peixmarisco marisco pescado-fresco pescadocalidad pececería_martorell marisco mariscos-frescos pescadofresco	POSITIVO	POSITIVO
El menú se aleja del estereotipo imperante en los último años del largoestredo Aquí el plato tiene más de un bocado 12 langosta para 2, x ej y puedes ir encontrando diferentes matices Ahora menú único que cambia cada 34 semanas Lástima no tenerlo más cerca Disfruten	POSITIVO	POSITIVO
Haciendo ensaladilla rusa para mañana	POSITIVO	POSITIVO
Pulpo a la brasa con fondant cremoso de patata santboi yummy comida bcn barcelona en Diversus Restaurant	POSITIVO	POSITIVO
@Corn292 La paella de ayer de Canet de Mar calamar y gambas	POSITIVO	POSITIVO
@MrPepito33 Entre mariscada y mariscada, igual los sindicatos ñ se animan	POSITIVO	NEGATIVO
He aprovechado los últimos albaricoques de @fvalles82 , he hecho una crema de tiramisú con mermelada de rosas y carquinyoli El albaricoque y la rosa combinan muy bien, la crema es poco dulce y los carquiños dan toque El postre se llama Atalanta por la mariposa Vanessa Atalanta	POSITIVO	POSITIVO
@abc_es Madre mía Qué riquísimo me salió hoy mi chuleton con patata fritas	POSITIVO	POSITIVO
Chicos, les advierto yo como carne, pero no la cocino Mi amiga la menos alérgica a servir	NEGATIVO	POSITIVO

@AntonioJose_EV Como churros si llega con 450km de autonomía y 200cv	POSITIVO	NEGATIVO
El arroz a banda con pulpo y allioli @ Barraca	POSITIVO	POSITIVO
@WTMpress Voy a intentar este atraco Esperen carbonara invitaciones para todo el equipo de WTM	POSITIVO	NEGATIVO
mercat de mercats glories plazadelasgloriascatalanas ceps chuleton barcelona en Plaza de las Glorias Catalanas	POSITIVO	POSITIVO
Cerraron La Jijonenca de Fabra i Puig para abrir un kebab Really	POSITIVO	NEGATIVO
menusabado fideua galicia comidacasera	POSITIVO	POSITIVO
En RestauranteColina, la carne es un arte Seleccionamos cuidadosamente los mejores cortes de carne y los preparamos a la perfección sobre las brasas Te esperamos en Ctra Del Palmar, 374, Aljucer, Murcia Haz tu reserva	POSITIVO	POSITIVO
@Catarrao1 Oh Pues tienes esta receta o también una que llamamos esgarraet en Valencia que es con berenjena y pimiento rojo asado, migas de bacalao, ajo y aceite de oliva	POSITIVO	POSITIVO
@_mariagaabriela Pues te lo compro, porque ayer cené tiramisú JAJAJAJAJ	POSITIVO	POSITIVO
@kvrpiv Los mejores los de carbonara pffffff huevito mmmm	POSITIVO	POSITIVO
De los mismos creadores de la ensaladilla rusa y el arroz a la cubana	POSITIVO	POSITIVO
@RevolucioBG Antepone el Club a cualquier otro interés particular coloca a amigos y familiares a punta pala, se lleva pasta del Club a sus bolsillospero antepone el Club, una mierda	POSITIVO	NEGATIVO
@peixmariscanna Buenos días Parada de Fin de Año 31 de Diciembre 2022 BUEN AÑO 2023 pescado peixmarisco marisco pescadofresco pescadocalidad pececería_martorell marisco mariscosfrescos pescadofresco en Pescado y Marisco Anna	POSITIVO	POSITIVO
Carne de primera el osgarmo de la caza mayor corzo en caldereta extremeña love gastronomy en restaurante cal miguelon cooking yummy gastronomia restaurant cena menucomida cuina	POSITIVO	POSITIVO
@Casfetera La culpa es de estos cojones café para todos que se han inventado, los españoles	POSITIVO	NEGATIVO
esmorzar de horquilla catalán Brunch castellano y guiri Yo lo tengo muy claro	POSITIVO	NEGATIVO
@jomivcf 11 segundos de Peter vete ya y 14 de corona es un florero Gran protesta de Mestalla que en momentos se pita a la grada de animación y si ves las caras de algunosse conforman con bocata futbol y CocaCola Mañana domingo de paella y aquí no pasa nada	NEGATIVO	NEGATIVO

@VengadorAstur @movideta @FodechinchosG Plato principal Y de postre, helado de pa amb tomàquet Que no sabemos comer dice, a ver si buscas mejor y criticas menos Los de la puta banderita sois todos igual	NEGATIVO	NEGATIVO
ALÓ ALO ES LA PIZZERÍA DE BILBAO PEDÍ UNA pizza DEL TAMAÑO DEL ATHLETIC DE BILBAO Y ME ENTREGARON UNA ACEITUNA	NEGATIVO	NEGATIVO
@oscar_puente_ Envíale una bandeja de croquetas Mientras engulle no habla	NEGATIVO	NEGATIVO
Volviendo del curro, he visto a dos guiris comiendo paella en un durum kebab	NEGATIVO	NEGATIVO
@happybichito Sabes q m pasa cn el pescao y la carne en la freidora d aire Que me sabe seco, pero seco seco jajajaja	NEGATIVO	NEGATIVO
@konypyon @MsInefable Yo si voy con los colegas a una hamburguesería y en la carta hay una ensalada de legumbres pues pata negra, pero como no la va a haber porque es una hamburguesería y quiero salir con mi amigos pues pido una hamburguesa vegana, que entiendo que para el rte es más sencillo	NEGATIVO	NEGATIVO
@europapress Esto es después de la mariscada	POSITIVO	NEGATIVO
Putá chiquillos se suspende el asado hasta nuevo aviso	NEGATIVO	NEGATIVO
@cmgorriaran Se quedan sin marisco	NEGATIVO	NEGATIVO
@lavirgenmanca Yo tenia entendido que los kebab no son muy saludables	NEGATIVO	NEGATIVO
@infarruco Las cárceles tailandesas son famosas por su elevado nivel gastronómico Allí es frecuente que los internos hagan barbacoas de marisco y embutidos Sus métodos de reinserción son legendarios	NEGATIVO	NEGATIVO
@GuillemT_ Ya no podré llevar horquilla ni cuchillo para esmorzar al trabajo	NEGATIVO	NEGATIVO
aunque te quiten el cum de la hamburguesa por qué cojones ibas a querer esa mierda es un trozo de pollo entre dos putos donuts	NEGATIVO	NEGATIVO
La Paella de Su, donde cocina Susi Bernat Algunos platillos valencianos son más difíciles de encontrar en Barcelona que un rastreador de coronavirus esgarrat, titaina Paella con carabineros, que llaman	NEGATIVO	NEGATIVO
@AyusoAgenda2030 @ArturoVilla_ Para ser coherente el muchacho debería dejar de comer carne Cualquier animal utilizado como ganado tiene un trato y una vida mucho peor que la de un toro	NEGATIVO	NEGATIVO
@dvni999 no me gusta el café, la cerveza, el nestea, la salsa barbacoa, las costillas, las alitas de pollo, el arroz al horno, las aceitunas	NEGATIVO	NEGATIVO

@xikiGz_ @ionthas @urgellensis @vacagato_ @peepolara33 OYE ESTO QUÉ ES EL SABADO TODOS SIN croquetas POR JOPUTAS	NEGATIVO	NEGATIVO
Nota mental escuchar más música y menos a la gente Tiramisú de mató bizcocho de vainilla casero, café de especialidad, crema de mató by veredabar carlesalmagroanalog	NEGATIVO	POSITIVO
@ardillakoalaoki Alguien también quiere pizza	NEGATIVO	POSITIVO
La taberna pescadora Can Maño siempre trae cola Viva sus calamares a la romana ministeriog	NEGATIVO	POSITIVO
@bichiteo Ahora mismo hago un grupo y lo cuadramos sino se irá otro año más sin asado	NEGATIVO	NEGATIVO
@nauterfley_es Estaría bien si vigilasen la inversión de este señor, que el Santuario estuviese en buenas condiciones y fuese sostenible Porque palmar pasta cada 2x3 porque un cuñao no sepa gestionar más allá de dar pena, insultar y montar barullo, no es muy sano en ningún sentido	NEGATIVO	NEGATIVO
Voy a pegarme una tripá a desayunar fartons y acabo de ver por la calle a uno de mis actores Porno favoritos No está mal para ser miércoles	NEGATIVO	POSITIVO
@vaelltyon Iauh hh Equipazo de Europa League que estamos montando Mariano, sería la guinda del pastel	NEGATIVO	POSITIVO
Jornada atípica en la que coincidieron con descanso el Senior y los SeniorB y Master a los que se añadió el SeniorA 1a Estatal retenido en la AP7 en dirección a Bordils os dejamos imágenes del gol del juvenil Cupi y la barbacoa del SMB Orgull-CHSES	NEGATIVO	NEGATIVO
@Cristin20105937 @Madre375_24_7 Lujos cuando tienes la alimentación cubierta creo yo Si pides para comer y por tanto tu dieta es deficiente no iría un día en semana a gastarme 13 por menú Me lo gasto en fruta o pescado que el banco de alimento no da	NEGATIVO	NEGATIVO
Sopar a Harrys i mojitos a la terrassa Alaire Barcelona restaurant cuisine italiano lovebcn fetuccini tiramisú LaPedrera aniversari moments @sagradafamilia	NEGATIVO	POSITIVO
roadtrip día 2 bacalao y pulpo a la brasa valença comida	NEGATIVO	POSITIVO
Los defensores de las croquetas, ignorante y pen	NEGATIVO	NEGATIVO
@Yodddio Pero no en mis zamburiñas ni en mib pulpo a la brasa	NEGATIVO	NEGATIVO
El pastel de @gon_gi_	NEGATIVO	POSITIVO
Barbacoa callejera de bidón redneck con Deedee dormida al fondo Oleo sobre lienzo	NEGATIVO	POSITIVO
@annagrauarias @salvadorilla @perearagones @CiutadansCs En liquidacioooooo,si no una mariscada en Ciutadella si sacáis más de 1 diputadoGuarda el tuit	NEGATIVO	NEGATIVO

Pudiendo ofrecerles carne a la brasa	NEGATIVO	NEGATIVO
A las 7 d la mañana buscand un bar por sagunto para acabar pidiendose uno de longaniza con esgarraet paragustoscolores	NEGATIVO	NEGATIVO
@CristianBilba12 Los mios ya estan en una edad en la que no van al comedor, así que les doy cada dia para el Dominos Pizza y todos contentos	NEGATIVO	NEGATIVO
Podría ser carbonara	NEGATIVO	NEGATIVO
Lo gracioso que es que hoy haya comido cigalas a lo fancy y me haya cenando un kebab por culpa del cual me estoy cagando, soy tan motomami, la vida es un contraste continuo	NEGATIVO	NEGATIVO
ElGranMr Ahora dilo sin llorar langosta	NEGATIVO	NEGATIVO
@tobuushi por cierto, mi padre era delgado, seguro que era para trabajar en el campo, y no para comer esmorzar de tenedor en cambio este hombre de la entrevista, está gordo porque come más de lo que necesita su cuerpo	NEGATIVO	NEGATIVO
Pulpo a la brasa	NEGATIVO	POSITIVO
@axxdriixx_ Rabiote para No es pivote y no pide poco, aunque haya pasta tampoco hay que traer a lo loco	NEGATIVO	NEGATIVO
@BoniAgain El pan empapado en sangre Ni que fuéramos hienas de la sabana A esa carne le falta minutaje a fuego medio para hacerse bien y que sepa a carne a la brasa y no a vientre de vaca abierto a bocados por las hienas	NEGATIVO	NEGATIVO
@helenag001 A ver mientras estuvo en el gobierno mejoró vida de la gente, con una pandemia terrible se prohibió el despedir y el volcán de la Palma, la guinda, pero claro el no tiene derecho a mejorar su vida y la de su familia porque es rojo, ah Y mucho menos comer una mariscada	NEGATIVO	NEGATIVO
@VidalQuadras Lo que quieren es más gente, para no dejar de tensionar los precios en vivienda, más demanda para empleos con lo que obtener mano de obra barata, es el Banco de España con sus élites que no quieren dejar repartir el pastel	NEGATIVO	NEGATIVO
@PabloIglesias Piensa en la mariscada que se pegaría después de la intervención	NEGATIVO	POSITIVO
@GxlDePaulinho No hombre no Que este tio gana pasta como estrella y lleva 3 años lesionado Encima con la N10	NEGATIVO	NEGATIVO
Nuestra Sandra te invita a leer Una novela histórica siempre es un acierto a la hora de escoger la lectura del verano Deja que esta novela te atrape desde un punto de vista distinto en la época de la Segunda Guerra Mundial mientras te tomas un helado en la piscina	NEGATIVO	POSITIVO
@Trozo1 Pero yo amo al club marronazo, vos sos un k que le entrega el culo al bolita de Riquelme, no vi que defiendan a Angelici como consejo del asado sjjsjs	NEGATIVO	NEGATIVO

Mal olor bajo tierra de la enfermedad Hay alrededor de 5 gambas escondidos en el brazo de este hombre, y estos gambas ya están en estado de descomposición	NEGATIVO	NEGATIVO
Que yo no estoy acostumbrado a comer carne ni esa salsa tan fuerte Me cago encima	NEGATIVO	NEGATIVO
Desayunar un café y un vaso de horchata significa tener un hambre a estas horas que te subes por las paredes Literalmente	NEGATIVO	NEGATIVO
@Lluis_arenay Soy tan feliz ahora podré Dedicarme plenamente a lo que me llena Comer marisco Exacto	NEGATIVO	POSITIVO
@gaceta_es Esa fundación se llama comedores de marisco de CyL	NEGATIVO	NEGATIVO
@AntonioMautor Padlla sí Fideuá no	NEGATIVO	NEGATIVO
y por la noche barbacoa	NEGATIVO	POSITIVO
@JuanbYofre Pescado podrido, es reviejo ese video	NEGATIVO	NEGATIVO
los de abajo estan haciendo barbacoa y yo tengo que comer crema de calabacin	NEGATIVO	NEGATIVO

## A.2 Muestra de resultados completos de la recopilación de datos de Google Maps

A continuación, se adjunta una tabla con los resultados de la recopilación de datos de Google Maps. Se adjuntan tan solo los resultados para un tipo de comida correspondiente a cada categoría, debido a la extensión de la tabla original.

Tabla 9: Muestra de datos completos extraídos de Google Maps.

Tipo de comida	Categoría	Nombre	Barrio	Puntuación	Reviews	Precio
arroz a banda	Comida valenciana	Restaurant El Racó de Lluís	Quatre Carreres	4,5	622	Asequible
arroz a banda	Comida valenciana	Restaurante la Zarandona	Quatre Carreres	4,6	939	Moderado
calamares a la romana	Comida española	Casa Mundo desde 1953	Ciutat Vella	3,7	2148	Moderado
calamares a la romana	Comida española	La taberna Casera. Arrocería y freiduría.	Poblats Marítims	4,6	2520	Moderado
calamares a la romana	Comida española	Cerveceria Erajoma	El Pla del Real	4,3	1166	Moderado
calamares a la romana	Comida española	Casa Miriam Tapas Y Bocadillos	Benicalap	4,2	522	Asequible

calamares a la romana	Comida española	Restaurante Levante Valencia	Campanar	4,5	1361	Moderao
calamares a la romana	Comida española	Restaurante Augusto Di Carmen	L'Eixample	4,3	748	Asequible
calamares a la romana	Comida española	bella milano	Camins al Grau	4	524	Asequible
calamares a la romana	Comida española	J.M. Restaurant	Quatre Carreres	4,4	2097	Moderao
calamares a la romana	Comida española	Marisquería Civera	Ciutat Vella	4,4	2183	Caro
calamares a la romana	Comida española	Taberna Jaizkibel	Benicalap	3,5	1122	Asequible
calamares a la romana	Comida española	Restaurant Ferro	Ciutat Vella	4,4	663	Moderao
calamares a la romana	Comida española	Bolos	Camins al Grau	3,8	2133	Moderao
calamares a la romana	Comida española	La Casa del Tapeo	Desconocido	4	3153	Asequible
calamares a la romana	Comida española	La Maruja	Ciutat Vella	2,5	607	Moderao
calamares a la romana	Comida española	Masía del Vño	Ciutat Vella	4	1079	Asequible
calamares a la romana	Comida española	Restaurante La Murciana	Poblats Marítims	4,2	4879	Moderao
calamares a la romana	Comida española	Cervecería Cruz Blanca	Extramurs	4,1	2861	Asequible
calamares a la romana	Comida española	Restaurante San Miguel	Ciutat Vella	2,7	786	Moderao
calamares a la romana	Comida española	Mini Rioja	Benicalap	4	1086	Asequible
calamares a la romana	Comida española	Adelina Bar	Poblats Marítims	4,4	525	Asequible
calamares a la romana	Comida española	Entretiempos	Extramurs	3,9	282	Asequible
calamares a la romana	Comida española	Gran Azul	El Pla del Real	4,3	1635	Moderao
calamares a la romana	Comida española	12 INTENCIONES	El Pla del Real	4,5	873	Asequible
calamares a la romana	Comida española	CASA CLEMÈNCIA	Benimaclet	3,9	1902	Moderao

lasaña	Comida italiana	Da Antonella	Benicalap	4,3	496	Asequible
lasaña	Comida italiana	Restaurante La Tagliatella   Valencia Sur	Calle Ep Sector PRR-9 Patraix	4,4	895	Moderado
lasaña	Comida italiana	Restaurante La Tagliatella   Arena Multiespacio, Valencia	Rascanya	4,1	1221	Moderado
lasaña	Comida italiana	Pizzería La Trattoria di Fede	La Saïdia	4,3	511	Asequible
lasaña	Comida italiana	Paffuto (Cortes)	Campanar	4,6	718	Moderado
lasaña	Comida italiana	Sorsi e Morsi Alameda	Camins al Grau	4,1	1852	Moderado
lasaña	Comida italiana	Gastro-Pizza IL Vulcano	El Pla del Real	4,4	1306	Moderado
lasaña	Comida italiana	Tagomago Valencia	Ciutat Vella	4,3	4429	Moderado
lasaña	Comida italiana	IL GHIOTTONE NAPOLETANO	La Saïdia	4,5	1702	Moderado
lasaña	Comida italiana	Saona Cortes Valencianas	Benicalap	4	4821	Moderado
lasaña	Comida italiana	Pizzería La Fantástica	Extramurs	3,7	72	Asequible
lasaña	Comida italiana	Las Lunas Soul Kitchen	L'Eixample	4,6	857	Asequible
lasaña	Comida italiana	Restaurant La Tagliatella   CC Gran Túria, València	Desconocido	4,2	1439	Moderado
lasaña	Comida italiana	Saona Alameda	Camins al Grau	4,1	5330	Moderado
lasaña	Comida italiana	Oragù LasagnaBar	Algirós	4,6	142	Asequible
lasaña	Comida italiana	Los Gómez, Correos	Ciutat Vella	4,5	9139	Moderado
lasaña	Comida italiana	Turqueta	L'Eixample	4,2	6607	Moderado
lasaña	Comida italiana	Restaurante Submarino	Quatre Carreres	4	4701	Caro
lasaña	Comida italiana	Restaurante No es lo Mimmo	Extramurs	4,4	562	Asequible

pescado	Pescado	Puerta del Mar Valencia	Ciutat Vella	4,4	9252	Mo- de- rado
pescado	Pescado	Ostras Pedrín Centro	Ciutat Vella	4,6	2774	Mo- de- rado
pescado	Pescado	Restaurante Bocamada	L'Eixample	4,6	402	Mo- de- rado
pescado	Pescado	Rausell	Extramurs	4,6	2447	Mo- de- rado
pescado	Pescado	Som de Mar Restaurante	Poblats Marítims	4,4	822	Mo- de- rado
pescado	Pescado	Taska La Reina	Poblats Marítims	4,6	7864	Ase- quible
pescado	Pescado	Ostrabar Valencia	L'Eixample	4,6	832	Mo- de- rado
pescado	Pescado	El Trocito del Medio	Ciutat Vella	4,4	1246	Ase- quible
pescado	Pescado	Aladroc taverna de peix	L'Eixample	4,5	1137	Ase- quible
pescado	Pescado	Restaurante Casa Eulogio	Ciutat Vella	4,3	1372	Ase- quible
pescado	Pescado	Restaurante Navarro	Ciutat Vella	4,6	2763	Mo- de- rado
pescado	Pescado	Restaurante Eladio	Patraix	4,5	344	Mo- de- rado
asado	Carne	Asador Aurora	El Pla del Real	4,4	644	Caro
asado	Carne	ASADOR BUFALO ( Армянский Шашлык)Restaurante Armenio.	Extramurs	4,8	380	Mo- de- rado
asado	Carne	Asador Criollo	Benimaclet	4,4	321	Ase- quible
asado	Carne	Asados Carmen	Desconocido	4,4	170	Ase- quible
asado	Carne	Asador Boro	Camins al Grau	4	315	Ase- quible
asado	Carne	A la broaster	La Saïdia	3,8	1723	Ase- quible
asado	Carne	LA GRANDAÇA POLLOS ASADOS Y CO- MIDAS PARA LLEVAR	Patraix	4,5	78	Ase- quible
asado	Carne	Asador Casa Vidal	Camins al Grau	4,4	880	Mo- de- rado

asado	Carne	Asados Express Gil Comidas para Llevar Valencia	Algirós	4,3	279	Asequible
comida china	Comida internacional	Restaurante Chino Min Dou	Extramurs	4,3	2061	Asequible
comida china	Comida internacional	Mey-Mey	Extramurs	4,1	957	Asequible
comida china	Comida internacional	Restaurant Pato Beikin	Extramurs	4,5	606	Asequible
comida china	Comida internacional	King Baos - Auténtica Comida China	El Pla del Real	4,3	211	Asequible
comida china	Comida internacional	Tang City	Benicalap	3,1	81	Asequible
comida china	Comida internacional	Restaurante Dynasty	Campanar	4,3	423	Asequible
comida china	Comida internacional	Restaurante Chino Amistad	Extramurs	4,2	1262	Asequible
comida china	Comida internacional	Restaurante Casa Ru	L'Eixample	4,2	712	Asequible
comida china	Comida internacional	Restaurant Mey Chen	Benimaclet	4,3	2598	Asequible
comida china	Comida internacional	Restaurante Chang Fu S.L.	Extramurs	3,8	716	Asequible
comida china	Comida internacional	BAR SUECA - Comida Auténtica China	L'Eixample	4,2	152	Asequible
comida china	Comida internacional	Restaurante Wei Feng	La Saïdia	4	463	Asequible
comida china	Comida internacional	La Gran Muralla	Ciutat Vella	4,3	960	Moderado
comida china	Comida internacional	MEI-HUA	Campanar	4,2	1030	Asequible
comida china	Comida internacional	Chino Festin	El Pla del Real	3,7	161	Asequible
comida china	Comida internacional	MEIHUA AMISTAT / Restaurante Chino - GRUPO MEIHUA	Algirós	4	611	Asequible
comida china	Comida internacional	Restaurante chino Chin Chin	L'Olivereta	4,3	611	Asequible
comida china	Comida internacional	Gran dragon Restaurante Chino	Quatre Carreres	4,1	513	Asequible
comida china	Comida internacional	Lu Cheng Restaurante Chino	L'Olivereta	3,9	582	Asequible
comida china	Comida internacional	Restaurante CHE KIANG	Extramurs	4,3	498	Asequible
comida china	Comida internacional	Laifu	L'Eixample	4,2	297	Asequible
comida china	Comida internacional	Restaurante Los Amigos	Rascanya	3,9	394	Asequible
comida china	Comida internacional	Restaurante YUN NAN 餐厅 雲南	L'Eixample	4,6	437	Moderado

comida china	Comida internacional	Restaurant Bar Frenazo	Extramurs	4,3	1979	Asequible
comida china	Comida internacional	Hang Zhou	Ciutat Vella	2,7	348	Asequible
comida china	Comida internacional	Rong Hua	Ciutat Vella	3,8	533	Asequible
comida china	Comida internacional	XiaoGeZi	Extramurs	4	278	Asequible
comida china	Comida internacional	Guo An	L'Olivereta	4,1	237	Asequible
comida china	Comida internacional	Tapas Chinas Wei Wei	Extramurs	4,1	936	Asequible
comida china	Comida internacional	May Lin	Camins al Grau	4,1	130	Asequible
comida china	Comida internacional	restaurante chino 103	Benimaclet	4,1	140	Asequible
comida china	Comida internacional	Restaurante Gran Hong Kong	Benimaclet	4,3	474	Asequible
comida china	Comida internacional	Casa Dragón 龍记美食	Extramurs	3,8	744	Asequible
comida china	Comida internacional	Restaurante Chino Sur	Camins al Grau	4	634	Asequible
comida china	Comida internacional	Restaurante Asia	Camins al Grau	3,9	367	Asequible
comida china	Comida internacional	Restaurante Nuevo Siglo	Quatre Carreres	3,5	328	Asequible
comida china	Comida internacional	Rui Restaurante Chino	Benimaclet	4,5	111	Asequible
comida china	Comida internacional	Restaurante Chino Oulí	Patraix	3,6	287	Asequible
comida china	Comida internacional	Changlong Bar Restaurant	Extramurs	4,1	186	Asequible
comida china	Comida internacional	Restaurante ZEN	Campanar	4,4	1092	Moderado
comida china	Comida internacional	Restaurante Exquisito	Patraix	4	301	Asequible
comida china	Comida internacional	Restaurante Singapur	Poblats Marítims	3,9	684	Asequible
comida china	Comida internacional	Nuevo Chang An	Rascanya	4,1	452	Asequible
comida china	Comida internacional	Restaurante Asiático - COCINA ASIÁTICA HAI S.L	La Saïdia	4,2	860	Moderado
comida china	Comida internacional	Restaurant Felisano	Extramurs	4,1	1568	Asequible
comida china	Comida internacional	Beijing Palace Restaurante Chino	Patraix	4	577	Asequible

comida china	Comida internacional	Spicy Soul Hot Pot	Extramurs	4,4	1759	Mod-erado
comida china	Comida internacional	Restaurante Hang Zhou 2	Extramurs	4,4	426	Ase-quible
comida china	Comida internacional	Restaurante China Town	Desconocido	4,2	351	Ase-quible
comida china	Comida internacional	Restaurante Tiananmen	Camins al Grau	3,7	401	Ase-quible
comida china	Comida internacional	Wok Gourmet	Campanar	2,9	423	Ase-quible
comida china	Comida internacional	Restaurante China XINTAI	Benimaclet	3,8	275	Ase-quible
comida china	Comida internacional	Chun Ly Noodles	La Saïdia	4,6	316	Ase-quible
comida china	Comida internacional	Casa de Oro II	Patraix	3,8	349	Ase-quible
comida china	Comida internacional	Restaurante Chino Oriental	Quatre Carreres	3,1	311	Ase-quible
comida china	Comida internacional	Guang Zhou	Desconocido	3,3	104	Ase-quible
comida china	Comida internacional	Mil Grullas	Algirós	4,2	1202	Ase-quible
comida china	Comida internacional	Casa Tafu	Extramurs	4,9	539	Ase-quible
comida china	Comida internacional	Gran china	Quatre Carreres	4	74	Ase-quible
comida china	Comida internacional	HUME Russafa	L'Eixample	4,5	165	Ase-quible
comida china	Comida internacional	Restaurante chino Duke	Pobles de l'Oest	3,9	695	Ase-quible
comida china	Comida internacional	Asian Té	Campanar	3,5	488	Ase-quible
comida china	Comida internacional	Restaurante Chino Asiático Mislata	Desconocido	4	613	Ase-quible
comida china	Comida internacional	Mei Hao	Poblats Marítims	3,8	324	Ase-quible
cafe	Cafete-ria/heladería	Café Valencia Bar	Extramurs	3,7	152	Ase-quible
cafe	Cafete-ria/heladería	Restaurante & Cafetería Beluga	Ciutat Vella	4,2	713	Ase-quible
cafe	Cafete-ria/heladería	Café Rialto	Ciutat Vella	4,2	433	Ase-quible
cafe	Cafete-ria/heladería	Federal Café	Ciutat Vella	3,9	3217	Mo-de-rado
cafe	Cafete-ria/heladería	Bar Restaurante Iván	Campanar	4,6	1460	Ase-quible
cafe	Cafete-ria/heladería	Cafeteria Suizo	Ciutat Vella	3,7	596	Ase-quible

cafe	Cafete- ría/heladería	Babel Cafe Bar	Campanar	3,6	19	Ase- quible
cafe	Cafete- ría/heladería	Café Café	Extramurs	3,3	102	Ase- quible
cafe	Cafete- ría/heladería	Valencia Pá	El Pla del Real	4,5	123	Ase- quible
cafe	Cafete- ría/heladería	Restaurante Atmosphère - Restau- rante francés Valencia	Extramurs	4,5	1372	Ase- quible
cafe	Cafete- ría/heladería	Café Infanta	Ciutat Vella	4,2	3165	Mo- de- rado
cafe	Cafete- ría/heladería	CANDILEJAS - Cafe & Restaurante	Campanar	4,3	382	Ase- quible
cafe	Cafete- ría/heladería	BAR KRAMER	Av. de Campanar	4,3	1255	Ase- quible
cafe	Cafete- ría/heladería	Café Barón	Ciutat Vella	4,4	387	Ase- quible
cafe	Cafete- ría/heladería	Bar Restaurante Rojas Clemente	Extramurs	4,4	1398	Ase- quible
cafe	Cafete- ría/heladería	Café Madrid Valencia	Ciutat Vella	4,4	1381	Mo- de- rado
cafe	Cafete- ría/heladería	Café de Las Horas	Ciutat Vella	4,5	7908	Mo- de- rado
cafe	Cafete- ría/heladería	Bar & Kitchen	Ciutat Vella	4,3	1496	Mo- de- rado
cafe	Cafete- ría/heladería	Cafe Bar El Valenciano	Patraix	3,7	74	Ase- quible
cafe	Cafete- ría/heladería	Cafetería Paniacos	Campanar	3,8	894	Ase- quible
cafe	Cafete- ría/heladería	Restaurante Vinoteca La Dehesa	Benicalap	4,3	1776	Mo- de- rado
cafe	Cafete- ría/heladería	Restaurante La Ceba Vieja (Valencia)	Desconocido	4,5	726	Mo- de- rado
cafe	Cafete- ría/heladería	Cafeteria	Campanar	4	37	Mo- de- rado
cafe	Cafete- ría/heladería	Maunaloa Café Bar - Billar	Extramurs	4,1	280	Ase- quible
cafe	Cafete- ría/heladería	MESTIZO - Café de especialidad • Brunch • Tapas	Ciutat Vella	4,9	1640	Ase- quible
cafe	Cafete- ría/heladería	Restaurante los Alpes	Campanar	4,5	445	Ase- quible
cafe	Cafete- ría/heladería	Cafè & Tapas	Ciutat Vella	2,4	436	Mo- de- rado

cafe	Cafetería/heladería	Cafetería Rodolfo/La Pecera	Extramurs	4,4	241	Asequible
cafe	Cafetería/heladería	Restaurant Camp de Túria	Pobles de l'Oest	4,5	555	Asequible
cafe	Cafetería/heladería	Bar Madrid	Av. de Campanar	4	242	Asequible
cafe	Cafetería/heladería	Restaurante MoN	Ciutat Vella	2,9	795	Moderado
cafe	Cafetería/heladería	Bar Nuevo Oslo	Extramurs	4,6	1851	Asequible
cafe	Cafetería/heladería	Restaurante Chikito	Desconocido	4,3	414	Asequible
cafe	Cafetería/heladería	Bar Nevada Gv53	Extramurs	3,6	551	Asequible
cafe	Cafetería/heladería	LIA VALENCIA	Ciutat Vella	4,5	1550	Moderado
cafe	Cafetería/heladería	Vera Restaurante	L'Eixample	3,9	1171	Asequible
cafe	Cafetería/heladería	A huevo - Restaurante en Valencia	Eixample	4,6	1530	Moderado
cafe	Cafetería/heladería	Café Madrigal	L'Eixample	4,8	1147	Asequible
cafe	Cafetería/heladería	Ubik Café Cafetería Librería	L'Eixample	4,3	4087	Asequible
cafe	Cafetería/heladería	Café Bar Pegaso	Extramurs	3,7	330	Asequible
cafe	Cafetería/heladería	MIKENGO	Poblats Marítims	4,6	1050	Asequible
cafe	Cafetería/heladería	The Ceramic Bar	Campanar	4,4	438	Asequible
cafe	Cafetería/heladería	Ateneo Restaurant * 6ª Planta *	Ciutat Vella	4,1	1015	Moderado
cafe	Cafetería/heladería	Petit Bistró Virgen - Restaurante Mediterraneo	Ciutat Vella	4,2	580	Moderado
cafe	Cafetería/heladería	Restaurante El Encuentro	Ciutat Vella	4,4	1123	Moderado
cafe	Cafetería/heladería	JULIANA Restaurante (Valencia)	Extramurs	4,1	477	Asequible
cafe	Cafetería/heladería	Central Bar by Ricard Camarena	Ciutat Vella	4,5	2203	Moderado
cafe	Cafetería/heladería	Palau 11	Ciutat Vella	3,9	378	Asequible

cafe	Cafete- ría/heladería	Café ArtySana - Ruzafa	L'Eixample	4,5	1401	Ase- quible
cafe	Cafete- ría/heladería	CAFETERÍA DON NELO	Benicalap	4	140	Ase- quible
cafe	Cafete- ría/heladería	Bar Esmas	Ciutat Vella	4,2	68	Ase- quible
cafe	Cafete- ría/heladería	Restaurante Espinosa	Desconocido	4,4	1065	Ase- quible
cafe	Cafete- ría/heladería	Bar Los Navarros	L'Eixample	4,2	546	Ase- quible
cafe	Cafete- ría/heladería	Haus Navellos	Ciutat Vella	4	1492	Mo- de- rado
cafe	Cafete- ría/heladería	deli-rant vlc	Ciutat Vella	4,1	635	Mo- de- rado
cafe	Cafete- ría/heladería	Bar restaurante Carmen De Ronda	L'Olivereta	4,3	479	Ase- quible
cafe	Cafete- ría/heladería	Brunch Corner - Sant Bult	Ciutat Vella	4,5	1484	Ase- quible
cafe	Cafete- ría/heladería	RESTAURANTE MARZAL	Extramurs	4,3	584	Ase- quible
cafe	Cafete- ría/heladería	Restaurant LIENZO VALÈNCIA	Ciutat Vella	4,6	1129	Mo- de- rado
cafe	Cafete- ría/heladería	Bar Ricardo	Extramurs	4,4	4066	Mo- de- rado
cafe	Cafete- ría/heladería	ESPAI SEDA RESTAURANT	Ciutat Vella	4,2	462	Mo- de- rado
cafe	Cafete- ría/heladería	Bar restaurante Serrania Campanar	Campanar	4,4	307	Ase- quible
cafe	Cafete- ría/heladería	Bar-Restaurante Patujú	Extramurs	4	613	Ase- quible
cafe	Cafete- ría/heladería	Sabor A Ti	El Pla del Real	3,9	107	Ase- quible
cafe	Cafete- ría/heladería	Restaurante Bar Biosca, Ruzafa	L'Eixample	4,1	796	Ase- quible
arroz a banda	Comida va- lenciana	Arrocería Maribel	Pobles del Sud	4,7	4145	Mo- de- rado
arroz a banda	Comida va- lenciana	El Racó del CUINER	Benicalap	4,6	46	Ase- quible
arroz a banda	Comida va- lenciana	El Trompo	Poblats Marítims	4,5	1161 6	Mo- de- rado
arroz a banda	Comida va- lenciana	Paellas Velarte	Quatre Carreres	4,4	1350	Ase- quible

arroz a banda	Comida valenciana	Restaurante Bon Aire	Pobles del Sud	4,6	5087	Moderado
arroz a banda	Comida valenciana	CASA COMER COMER	Desconocido	4,2	509	Asequible
arroz a banda	Comida valenciana	L'espardenya	Algirós	4,6	634	Asequible
arroz a banda	Comida valenciana	Restaurante Mala Hierba	El Pla del Real	4,3	1657	Moderado
arroz a banda	Comida valenciana	El Nuevo Colmao	Desconocido	4,8	56	Asequible
arroz a banda	Comida valenciana	Alqueria El Brosquil	Pobles del Sud	4,6	3431	Moderado
arroz a banda	Comida valenciana	La Lonja Del Pescado	Poblats Marítims	4,2	1884	Moderado
arroz a banda	Comida valenciana	Restaurante Planta Azul	Desconocido	4,4	1903	Moderado
arroz a banda	Comida valenciana	Saona Casino	Ciutat Vella	4,5	6043	Moderado
arroz a banda	Comida valenciana	Restaurante Los Viñedos	L'Eixample	4,1	278	Asequible
arroz a banda	Comida valenciana	Restaurante Mediterráneo	Desconocido	3,9	2056	Asequible
arroz a banda	Comida valenciana	El Camarote & The Roof	Desconocido	3,5	2012	Moderado
arroz a banda	Comida valenciana	El quinto flamenco	El Pla del Real	4	86	Asequible
arroz a banda	Comida valenciana	El Temple Restaurant	L'Eixample	4,4	755	Moderado
arroz a banda	Comida valenciana	Restaurante Vuelve Carolina	Ciutat Vella	4,1	2422	Moderado
arroz a banda	Comida valenciana	Restaurante Casa Zaragoza	Poblats Marítims	2,2	2545	Moderado
arroz a banda	Comida valenciana	Saona Viveros	El Pla del Real	4,3	2034	Moderado
arroz a banda	Comida valenciana	Restaurante Espinosa	Desconocido	4,4	1065	Asequible
arroz a banda	Comida valenciana	Gorgos 25	Algirós	3,9	1445	Asequible

arroz a banda	Comida valenciana	Carrión	Poblats Marítims	4,2	1212	Asequible
arroz a banda	Comida valenciana	Cervecería Juan	Camins al Grau	4,5	962	Asequible
arroz a banda	Comida valenciana	Restaurante L'Estimat	Poblats Marítims	4	997	Moderado
arroz a banda	Comida valenciana	Olrait	Quatre Carreres	4,3	941	Moderado
arroz a banda	Comida valenciana	Restaurante Malvarrosa	Poblats Marítims	3,7	1867	Moderado
arroz a banda	Comida valenciana	180 Restaurante Cientochenta Places	Patraix	4,4	825	Asequible
arroz a banda	Comida valenciana	Restaurante los Mejillones	Patraix	4	1140	Asequible
arroz a banda	Comida valenciana	Restaurante Vinoteca La Dehesa	Benicalap	4,3	1776	Moderado
arroz a banda	Comida valenciana	Restaurante Pizzaiolo	L'Eixample	4,3	1033	Asequible
arroz a banda	Comida valenciana	Bar la Amistad	La Saïdia	4,4	340	Asequible
arroz a banda	Comida valenciana	La Alquería del Jamón.	Desconocido	4,2	1853	Asequible
arroz a banda	Comida valenciana	El Nuevo Virrey	El Pla del Real	4,1	1632	Asequible