**Estudiante 1:** Nicolas Perez Teran - 202116903

**Estudiante 2:** David Zamora - 202113407

**Profesor:** Maria del Pilar Villamil.

**Proyecto 1 - Procesamiento de text**o.

**Tabla de contenidos**

[**1. Entendimiento del negocio y enfoque analítico. 2**](#_bmxlrq29eru5)

[**1.2 Criterios de aceptacion. 2**](#_dh502p1hixgv)

[**1.3 Descripción de la organizacion: 2**](#_r10c5p3yp0e)

[**2. Preparacion de los datos. 3**](#_a0efsi4fbvwn)

[**2.1 Estadisticas descriptivas. 3**](#_tteh351oirhm)

[**2.2 Preparacion de los datos. 4**](#_ypcptk7ibjk4)

[**2.2.1 Limpieza de los datos. 4**](#_j9eay1rs6nom)

[**2.2.2 Tokenización. 5**](#_7eitm1qlie0)

[**2.2.3 Normalizacion: 5**](#_jz5dffv9rh48)

[**2.2.3 Perfilamiento final de las categorias. 6**](#_aj9cc5lp8fis)

[**3. Modelado y evaluacion. 6**](#_3gm708ngyoe5)

[**3.1 Red neuronal con dos LSTM bidireccionales David Zamora. 6**](#_dv651mb2fhpe)

[**3.2 Red neuronal simple: David Zamora: 7**](#_hfo8ei10toio)

[**3.3 GPT2 fine tuning: David Zamora 8**](#_8pnl5p7npw8a)

[**3.4 Support Vector Classifier: 9**](#_7d8eisq4kt03)

[**3.5 Regresion Logistica: 10**](#_g2cv213qm5b2)

[**​​3.6. Propuesta final del modelo a utilizar. 11**](#_oermq35rjsl8)

[**4. Trabajo en equipo. 11**](#_d7ptjspnbpc7)

[**4.1 Roles. 11**](#_k6hkbpuwx49p)

[**4.2 Programacion de reuniones. 11**](#_kkbwaq3fjrsz)

[**4.3 Tareas importantes asignadas a cada estudiante. 12**](#_larupxqhiblz)

# Entendimiento del negocio y enfoque analítico.

****

# 1.2 Criterios de aceptacion.

Para poder determinar que el analisis fue satisfactorio de deben cumplir los siguientes objetivos:

1. La tarea de clasificacion debe evidenciar porcentajes altos en tus metricas de calidad.
2. Cotelco podra clasificar satisfactoriamente la mayoria de los sentimientos que sintieron los turistas en sus estadias.
3. Se logran encontrar los patrones que determinan que aspectos influyen en las calificaciones que dan los turistas.

# 1.3 Descripción de la organizacion:

COTELCO es una entidad gremial que agrupa y representa a los hoteles y otros proveedores de servicios turísticos en Colombia. Su objetivo principal es fomentar el desarrollo sostenible del turismo en Colombia, mejorar la competitividad del sector y asegurar la máxima calidad en los servicios ofrecidos a los turistas. El modelo analítico proporcionaría a COTELCO una herramienta valiosa para comprender mejor las preferencias y sentimientos de los turistas respecto a los distintos sitios turísticos de Colombia. Esta información permitiría a la asociación y a sus miembros identificar tendencias de satisfacción o insatisfacción, permitiendo dirigir esfuerzos en mejoras, promoción y desarrollos estratégicos enfocados en aumentar la popularidad de los sitios y mejorar la experiencia turística.

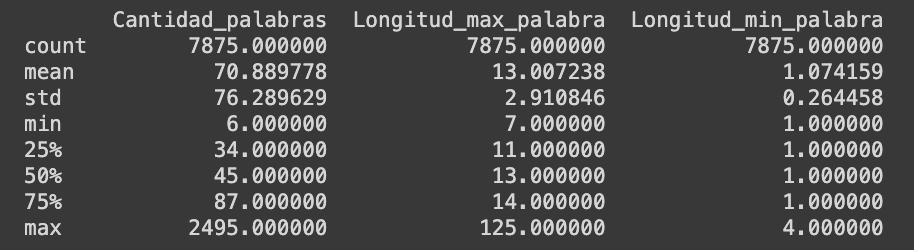
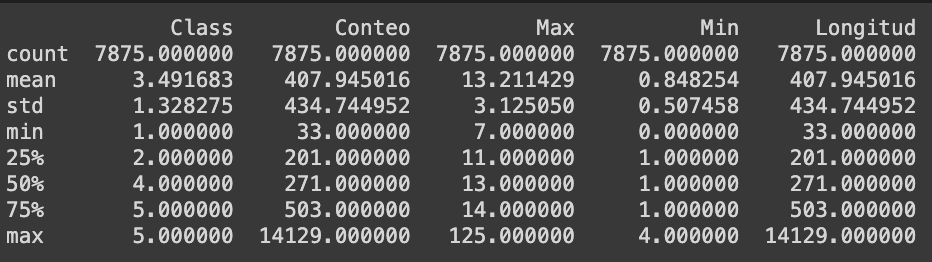
| Rol dentro de la empresa | Tipo de actor | Beneficio | Riesgo |
| --- | --- | --- | --- |
| Dirección de marketing | Usuario-  cliente | Identificación de puntos clave que atraen a los turistas, permitiendo enfocar mejor las campañas publicitarias. | Si el modelo no representa adecuadamente las preferencias de los turistas, la campaña puede no ser efectiva y no atraer a los turistas deseados. |
| Dirección de desarrollo de producto | Financiador | Desarrollo de ofertas y servicios basados en preferencias actuales de turistas, lo que podría incrementar la satisfacción y las ventas. | Desarrollo de productos que no coincidan con las tendencias actuales o futuras, resultando en inversiones malgastadas. |
| Gerencia de operaciones | Proveedor | Mejorar la eficiencia en la asignación de recursos a las áreas más valoradas por los turistas. | Malinterpretación de los datos podría llevar a una asignación ineficiente de recursos y pérdida de oportunidades. |
| Servicio al cliente | Beneficiado | Mejora en la calidad del servicio ofrecido al turista gracias al entendimiento de sus preferencias. | El centro del servicio al cliente recomendara de manera incorrecta a los usuarios los lugres turisticos al no identificar bien la calidad. |

# 2. Preparacion de los datos.

En este caso, los datos se tratan de una colección de reseñas respecto a las experiencias de usuarios de restaurantes u hoteles con su respectiva calificación, que en este caso es una escala de uno a cinco. A grandes rasgos, no podemos entender bien las justificaciones porque cada texto está asignado a cada clase (etiqueta), por esta razón es importante realizar un análisis para determinar más a fondo qué relación existe entre el texto y la variable objetivo.

# 2.1 Estadisticas descriptivas.

El análisis estadístico revela que los textos en el conjunto de datos exhiben una diversidad en su longitud, con una media de alrededor de 408 palabras. Esta variabilidad se manifiesta en la presencia de textos de longitud mínima de 33 palabras y máxima de 14,129 palabras. No obstante, la mayoría de las reseñas se encuentran dentro de un rango moderado, con una longitud media de alrededor de 407 palabras y una mediana de 271 palabras. Respecto a la longitud de las palabras, se observa que estas tienden a ser relativamente largas, con una media de aproximadamente 13 caracteres, y predominan las palabras cortas, con una longitud mínima media de alrededor de 1 caracter. Lo anterior, nos permite ver que el entrenamiento sera enfocado principalmente en caracterizar textos con una considerable longitud, mientras que las reseñas cortas podrian carecer de aspectos determinantes al contar con un numero de palabras mucho menor que las que conforman la mayoria de los datos.



Por otro lado, al realizar el análisis del conteo por palabras, se identificó que las 10 palabras más comunes en las reseñas son "de", "y", "la", "que", "en", "el", "es", "a", "un" y "no", con frecuencias que van desde aproximadamente 25,000 hasta 7,000 veces. Algunas de estas palabras, aunque son muy frecuentes en el texto, no aportan información relevante para la clasificación, ya que son términos comunes en el idioma español. Por lo tanto, se considera que podrían ser Stop Words y se sugiere eliminarlas del conjunto de datos, ya que su presencia no contribuye significativamente al proceso de clasificación y puede afectar la eficacia del modelo al introducir ruido innecesario.

# 2.2 Preparacion de los datos.

Para esta etapa, se opto por realizar tres pasos para la preparacion de los datos: limpieza de datos, tokenizacion y normalizacion.

# 2.2.1 Limpieza de los datos.

**Tecnicas y transformaciones hechas:**

**Limpieza de texto:** Inicialmente, es necesario que todo el texto quede en minuscula y se eliminen los caracteres especiales para facilitar el procesamiento al reducir el numero de palabras (Recordemos que es posible que una palabra se escriba de varias manera pero siga significando lo mismo). Ademas, sabemos que los caracteres especiales no aportan nada a la toma de decisiones.

**Expandir contracciones:** Hay palabras en español que son formas cortas de decir una cosa. Por esta razon, es importante hacer que la palabra se entienda para poder hacer que el modelo pueda entenderlo y agruparlo con palabras que puedan denotar el significado que requieren. Ademas, en esta etapa se aprovecha para estandarizar palabras (Recordemos que existen los sinonimos). Por otro lado, tambien se puede utilizar para cambiar palabras por un sinonimo mas extenso para evitar que se elimine como si de un StopWord se tratara (Ejemplo, en este caso se borra la palabra NO y es realmente importante).

**Filtrar Substrings:** Existen palabras que podrian no ser contadas como StopWords, a pesar de que lo son, ademas estas palabras pueden tener varriaciones que inician o tienen cierta subcadena en comun. Por esta razon, cree una funcion para borrar esas palabras en caso de ser necesario.

**Eliminar StopWords:** Como el nombre lo indica, este paso buscara las StopWords de la lengua española y las eliminara. De esta manera, estaremos reduciendo el campo de busqueda y al mismo tiempo se eliminan palabrar que posiblemente no aporten nada al modelo.

**Eliminacion de risas:** En texto es muy comun que las personas manden risas, sin embargo estas risas no estan siendo tomadas en cuenta para lo que se quiere decir. En otras palabras, las risas no aportan a la clasificacion porque podrian utilizarse de igual manera en mas de una cateogira.

# 2.2.2 Tokenización.

La tokenización es util porque permite dividir frases u oraciones en palabras. Con el fin de desglozar las palabras correctamente para el posterior análisis. Ademas, la tokenización permite contar la frecuencia de cada palabra en un texto y determinar su importancia para el sentimiento general expresado. En este caso, vamos a utilizar la tokenizacion para representar las palabras, es decir, las caracteristicas de. Usamos distintos algoritmos de tokenizacion.

* En esta parte, utilizamos parse, metodo extraido de spacy para los modelos de redes neuronales, mientras que para regresion logistica y SVC se utiliza la libreria Patterns.
* Tokenizacion de Spacy: Para los modelos de secuencia LSTM junto con redes neuronales usamos la libreria Spacy que se usa para trabajo de texto para tokenizar. Probamos los modelos de espaniol de spaCy pequenio y grande, teniendo una mejoria en los resultados con el grande. Este modelo se puede encontrar como nlp = spacy.load('es\_core\_news\_lg').
* Tokenizacion de GPT2: Para el modelo que usaba GPT2 usamos el tokenizador de gpt2. Este lo trajimos de la libreria transformers. De esa libreria obtuvimos GPT2Tokenizer y GPT2ForSequenceClassification. Mas adelante en este documento explicamos como se uso torch para el modelo.

# 2.2.3 Normalizacion:

Posterior a la tokenizacion, pasaremos a la etapa de normalizacion. En esta parte, necesitamos hacer que las palabras empiecen a importar mas por su significado y relevancia que por la forma en la que esta escrito. Por esta razon, lo mejor es estandarizar las palabras que representan algo en especifico, las cuales por temas de reglas gramaticales cambian su forma dependiendo del contexto. Para ello, utilizaremos las tecnicas de lemmatizacion y stemming.

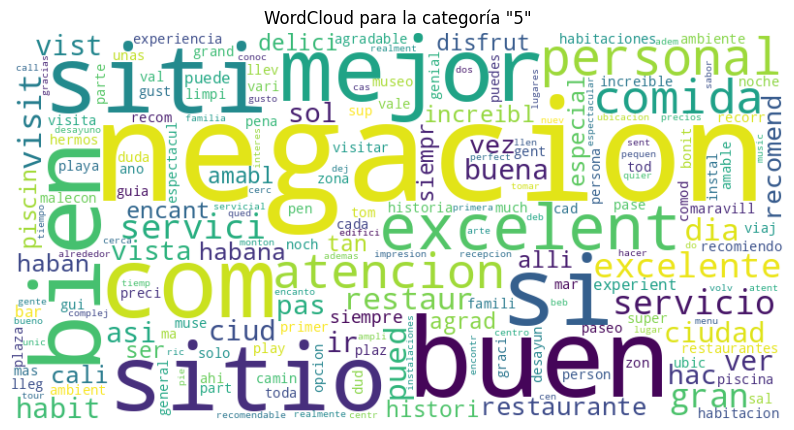
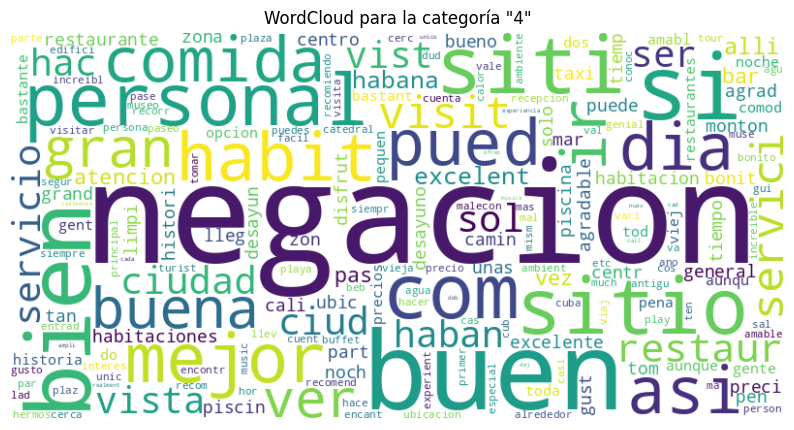
**Lemmatizacion**: Es buena idea utilizarlo porque permite reducir una palabra a su version base, de tal manera que se quitan las conjugaciones y se empieza a agrupar por conceptos. Esto es realmente util para el modelo porque le permite identificar las similitudes de palabras que significan lo mismo pero estan representadas con diferente palabras.

**Stemming**: Al igual que la lematizacion, permite que las palabras sean interpretadas por su significado mas que por su forma de escribirse. En este caso, no se busca llevar a su forma raiz, sino que se eliminan los prefijos y sufijos de la palabra.

Con ayuda de los dos procesos anteriores, podremos garantizar que el modelo podra agrupar palabras con significados semejantes y entender las relaciones que puede llegar a tener.

# 2.2.3 Perfilamiento final de las categorias.

Por ultimo, debemos tener en cuenta que en el lenguaje muchas palabras pueden usarse en varios contexto y significar cosas diferentes, al mismo tiempo que no significa que esa palabra sea realmente adecuada para ese contexto. Por esta razon, decidimos que debemos encargarnos de buscar e identificar que palabras son las que mas pesan en cada una de las clases de las reseñas. Para esta parte, realizamos un analisis a las palabras relacionandolas con su frecuencia en cada una de las clases.

​​****

De esta manera, podemos observar mas claramente como estan siendo perfilados cada uno de las categorias. Sin embargo, como podemos evidenciar en el ejemplo anterior, hay muchas palabras en comun en dos categorias y con una frecuencia similar (Revisar codigo fuente para ver exactamente las frecuencias, se muestra graficamente para facilitar la explicacion). Esos comportamientos son bastante perjudiciales para el modelo, razon por la cual es necesario tomar decisiones y decidir con base a los datos cual categoria es mas apropiada para ser descrita con esas palabras. En otras palabras, se van a eliminar palabras en ciertas categorias dependiendo de la importancia que representan para una u otra categoria. Por otro lado, tambien podemos evidenciar palabras que realmente no aportan nada y podrian estar causando ruido, por esta razon tambien procedemos a eliminarlas.

Al completar este paso, el ruido se ha disminuido al minimo y las categorias ya tienen caracteristicas mas solidas y definidas para el entrenamiento.

Es asi como, tras analizar los datos y estas palabras por categoria, nos dimos cuenta que las categorias son perfiladas de esta manera:

**Categoria 1:** Esta esta mas enfocada en expresar desagrado extremo, ya sea temas relacionados con asco, sucio, no recomendado, problemas y adjetivos negativos como horrible, desastre. Ademas, es mas propenso a utilizar palabras intensificadoras como Mas, Peor y etc.

**Categoria 2**: Esta mas enfocada en hablar de quejas enfocadas en el ambiente y quizas algun aspecto rescatable: amabilidad casi nula, humedad, agua sucia, inconformidades con pedidos, comida fria o dura. Ademas, utiliza adjetivos negativos pero no al punto de equipararse con la categoria 1, es decir utiliza palabras mas similares a Malo, mal y raro.

**Categoria 3:** Esta la mas complicada de perfilar. Recordemos que en este pueden haber tanto aspectos negativos como positivos. Sin embargo, la diferencia mas grande es que posee mas palabras de contraste, tales como Aunque, A pesar, Al menos, Eso si y Pero. Ademas, no posee palabras tan fuertes como Mejor, Peor y asi, sino que se enfoca en terminos mas similares a Regular, No estuvo mal, Agradable, Caro y Normal.

**Categoria 4:** Esta es muy similar a lo que pasa con la categoria 2 y 1, es decir, es muy similar a la categoria 5. Sin embargo, las palabras que mas abundan son positivas pero sin llegar a adjetivos tan exagerados. Es decir, se optan por palabras positivas como Bien, Bueno, Agradable, Buena, Rico, Positivo y Bonito. Ademas, tiene un indice de palabras de contraste mas alto en que en la Categoria 5.

**Categoria** **5:** Esta se destaca por utilizar palabras mas creativas relacionada a sentimientos positivos, tales como Excelente, Buenisimo, Precioso, Perfecto, Emocionante, Precioso, Delicioso, Espectacular y Extraordinario. Como podemos ver, es la que mas se aleja al perfil de las categorias porque, segun los datos, las personas mas se enfocan enn los aspectos positivos y dejan casi completamente de lado palabras de contraste como Falta, Aunque y Pero.

Teniendo en cuenta lo anterior, tras un arduo trabajo revisando las palabras, se definio un conjunto para cada categoria de palabras que NO deberian estar en cada una de esas categorias.

# 3. Modelado y evaluacion.

# 3.1 Red neuronal con dos LSTM bidireccionales David Zamora.

La primera con 64 unidades y la segunda con 32. La primera devuelve secuencias que es lo que se le pasa a la segunda. La segunda secuencia no devuelve secuencias y sus resultados se pasan a una red neuronal de 3 capas de 16 unidades cada una y por último una ultima capa densa de 5 unidades con activacion relu. Todas las capas ocultas usan activación relu. Entre capa y capa pusimos dropout al 50% para evitar el sobreajuste. También pusimos batch normalization entre capa y capa luego del dropout para acelerar el aprendizaje. Batch normalization soluciona el problema que la distribución de las salidas de una capa es diferente a la de sus entradas lo que hace más lento el aprendizaje.

La preparación que se usó para este modelo especificamente tenia la limpieza de datos, la expansión de contracciones, los filtros de subsitrngs y las stopwords, risas, stemming y lematizacion, los resultados que obtuvimos no tuvieron una accuracy alto.

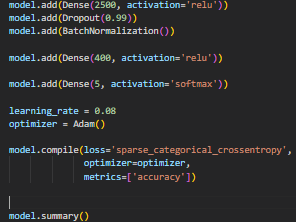
El motivo de la prueba de este modelo era lograr capturar patrones en las secuencias de texto que permitieran entender y clasificar los reviews. Usualmente se usan LSTM porque logran tener memoria de puntos relevantes en los comentarios que estan atras, no solamente teniendo en cuenta la ultima palabra. En una lstm, en cada paso de tiempo se le hace input de un item de la secuencia. Una LSTM bidireccional tiene en cuenta no solamente los tokens que se han pasado de tiempos anteriores sino del futuro. Esos tokens en realidad estan representados por embeddings. En este caso se uso los embeddings de spacy. 

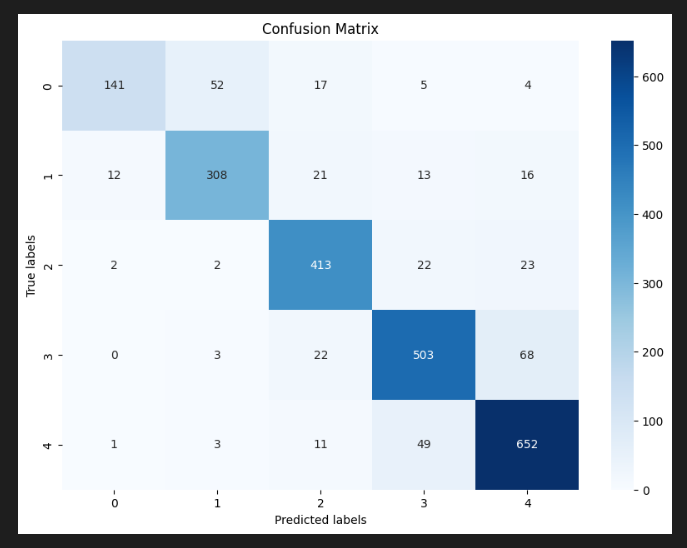
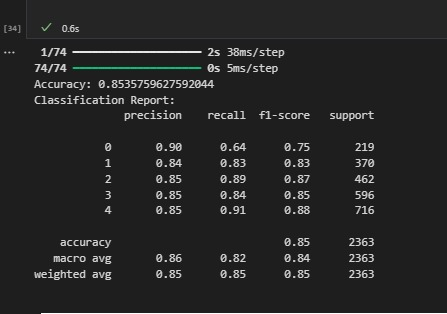
Tener en cuenta que estas metricas fueron calculadas con macro, o sea promedia sin tener en cuenta desbalances.

# 3.2 Red neuronal simple: David Zamora:

Las redes neuronales se organizan en capas y procesan la información recibida mediante la ponderación de las entradas, aplicando una función de activación para determinar si la señal se debe pasar a la siguiente capa. La primera capa, conocida como capa de entrada, recibe los datos iniciales. Las capas intermedias, o capas ocultas, extraen características y patrones relevantes. Finalmente, la capa de salida produce la predicción o clasificación final. Durante el entrenamiento, las redes neuronales ajustan iterativamente los pesos de las conexiones entre neuronas mediante un proceso llamado retropropagación, minimizando la diferencia entre las predicciones y los valores reales para mejorar la precisión de las predicciones. En este tipo de modelo probamos diferentes arquitecturas. Principalmente nos pudimos dar cuenta que redes neuronales profundas no lograban una buena precision. Por ejemplo una red neuronal de 4 capas, relu en cada una, dropout y batchnormalization, con 16 neuronas por capa y finalmente una caa de 5 con softmax solo lograba tener un accyracy del 40%. Tambien nos dimos cuenta que en general, las que se comportaban mejor eran las de 1 o dos capas antes de la de softmax, pero con un gran numero de unidades en la primera capa y un dropout muy alto para evitar el sobreajuste. Para todos los modelos se usa la función de perdida sparse categorical cross entropy.

La arquitectura que logró un mejor resultado fue:



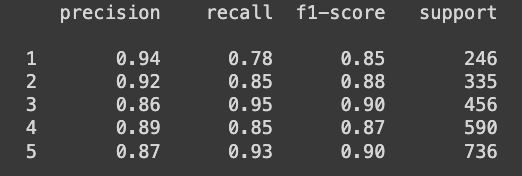


Este modelo fue entrenado por 580 epocas.Para acelerar el aprendizaje se utilizaron batches de 4096 muestras de datos Los batches funcionan al agrupar los datos al momento de la propagacion hacia delate, calculo de la funcion de perdida y propagacion hacia atras dentro de la red neuronal. De esta forma, la red propaga todas las muestras del mimi batch en un solo momento y con base en la funcion de perdida calculada para los datos del mini batch actualiza los pesos asignados en cada neurona. En general, entre mayor sea el tamaño del batch, el tiempo que toma cada etapa es menor. Sin embargo, eso tambien hace que se consuma más memoria. Nosotros en cada uno de los modelos de redes neuronales intentamos entrenarlas con el minibatch mas grande que permitiera la memoria disponible.

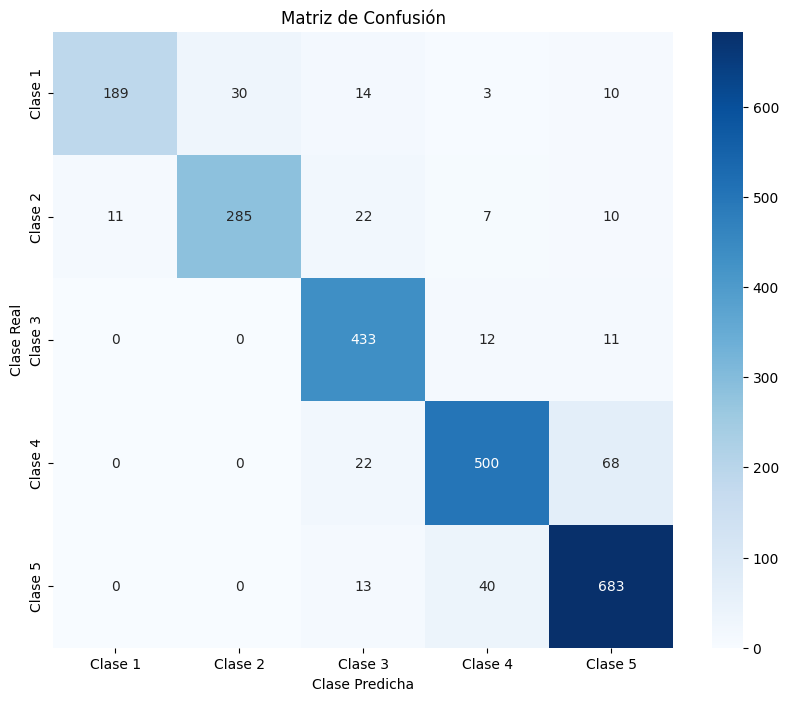
# 3.5 Support Vector Classifier:

Este es un algoritmo de aprendizaje supervisado en el cual se busca crear un hiperplano que sea el encargado de separar por una linea los datos que sean de un tipo u otro. El entrenamiento consiste en lograr diferenciar los datos de cada categoria trazando un hiperplano que separe equitativamente los datos de cada categoria; mas especificamente, la distancia entre los puntos mas cercanos de una categoria y el hiperplano sera maxima. Lo importante de este modelo es que una vez se define el hiperplano optimo que separe las categorias, identificar la categoria de un nuevo dato resulta evidente al ver en que parte quedo ubicado.

En terminos de hiperparametros, en los casos donde la linea siempre queda encima de algun dato, causando una ambiguedad durante el entrenamiento, es posible ajustar un parametro C que se encargara de definir que tantos errores puede permitirse, valores mas altos de C hace que sea mas propenso a permitir errores, mientras que menores que 1 hace que la tolerancia a errores se haga menor. Ademas, tambien se puede elegir el Kernel a utilizar.Para este caso, optamos por utilizar un Kernel lineal por la naturaleza de los datos y un C igual a 1.Dado que los datos de entrenamiento son las palabras vectorizadas con su respectiva frecuencia, es natural pensar que este modelo puede encargarse de clasificar teniendo en cuenta la forma de los datos.

**Resultado.**

El modelo SVM presenta una precisión general del 88.53%, lo que significa que es capaz de clasificar correctamente la mayoría de las instancias en el conjunto de datos de prueba. Este alto nivel de precisión sugiere que el modelo tiene una capacidad confiable para identificar y clasificar las reseñas de los sitios turísticos en sus respectivas categorías. Además, al analizar las métricas de precisión, recall y puntuación F1 para cada clase, observamos valores consistentemente altos, lo que indica que el modelo es capaz de predecir con precisión tanto las instancias positivas como las negativas para cada clase. Esta capacidad equilibrada de predecir correctamente tanto los verdaderos positivos como los verdaderos negativos es esencial para garantizar la calidad general del modelo y su capacidad para realizar predicciones confiables en diferentes escenarios.

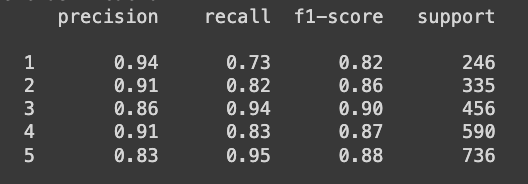


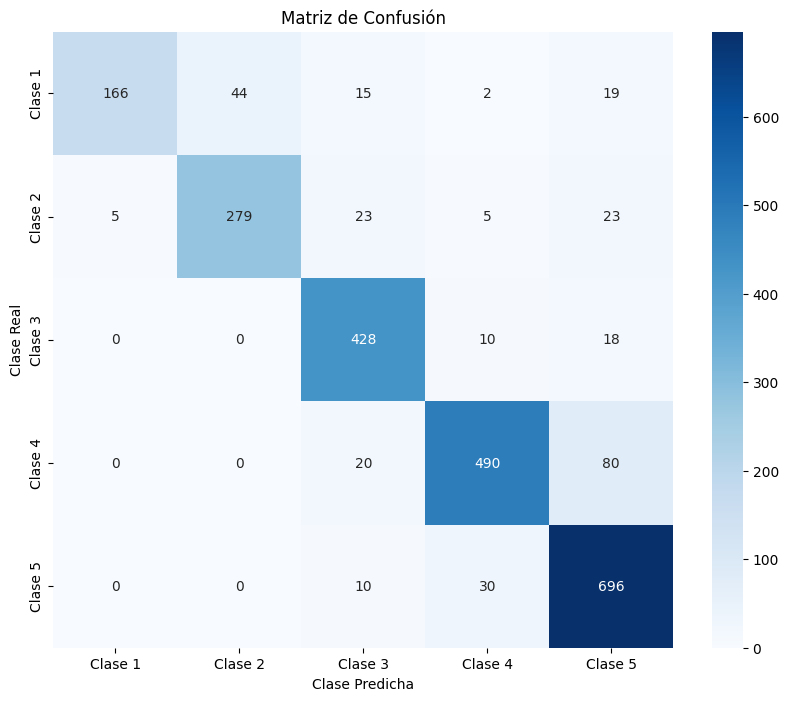
Implementado por Nicolas Perez Teran

# 3.6 Regresion Logistica:

La regresión logística es un método utilizado para predecir la probabilidad de que una instancia pertenezca a una de dos o más clases. Funciona modelando esta probabilidad utilizando una función sigmoide, que mapea cualquier valor real en un rango entre 0 y 1.

Durante el entrenamiento, el algoritmo ajusta sus parámetros para maximizar la probabilidad de observar los datos de entrenamiento dados esos parámetros. En otras palabras, trata de encontrar la función logística que mejor se ajuste a los datos y que pueda separar eficazmente las diferentes clases. Una vez que el modelo está entrenado, puede usarse para predecir la clase de nuevas instancias calculando la probabilidad de pertenencia a cada clase y eligiendo la clase con la probabilidad más alta. Este enfoque probabilístico permite interpretar las predicciones del modelo en términos de probabilidades, lo que proporciona una mayor comprensión de la incertidumbre asociada con cada predicción.

**Resultado.**

El modelo de regresión logística muestra una precisión general del 87.56%, lo que indica su capacidad para clasificar correctamente la mayoría de las instancias en el conjunto de datos de prueba. Si bien esta precisión es ligeramente inferior a la del modelo SVM, sigue siendo un indicador sólido de su capacidad para realizar clasificaciones precisas. Al analizar las métricas de precisión, recall y puntuación F1 para cada clase, observamos que el modelo de regresión logística muestra valores considerablemente altos en estas métricas para cada clase. Esto sugiere que el modelo es capaz de predecir con precisión tanto las instancias positivas como las negativas para cada categoría, lo que contribuye a su capacidad general para realizar predicciones confiables.

Implementado por Nicolas Perez Teran

# ​​3.6. Propuesta final del modelo a utilizar.

Teniendo en cuenta los resultados anteriores, nuestro modelo final a utilizar sera SVC. Como bien sabemos, se busca tener un alto indice en terminos de recall, F1-Score y precision, siendo este el que mejor se desempeño en esas tres metricas. En terminos del negocio, podemos decir que las metricas reflejan que si sera posible con un 88% de probabilidad de asignar correctamente las calificaciones a las reseñas de los turistas. Sin embargo, cabe resaltar que el modelo en este punto es propenso a asignar incorrectamente calificaciones negativas o regulares (De 1 a 3), lo cual puede ser un poco peligroso para los interesados porque lo que se busca identificar es una relacion que muestre cuando una reseña refleja un sentimiento positivo o negativo en una escala de 1 a 5. Sin embargo, nuestro modelo muestra con sus metricas de calidad que es el que mejor se ha desempeñado en los aspectos importantes.

# 4. Trabajo en equipo.

# 4.1 Roles.

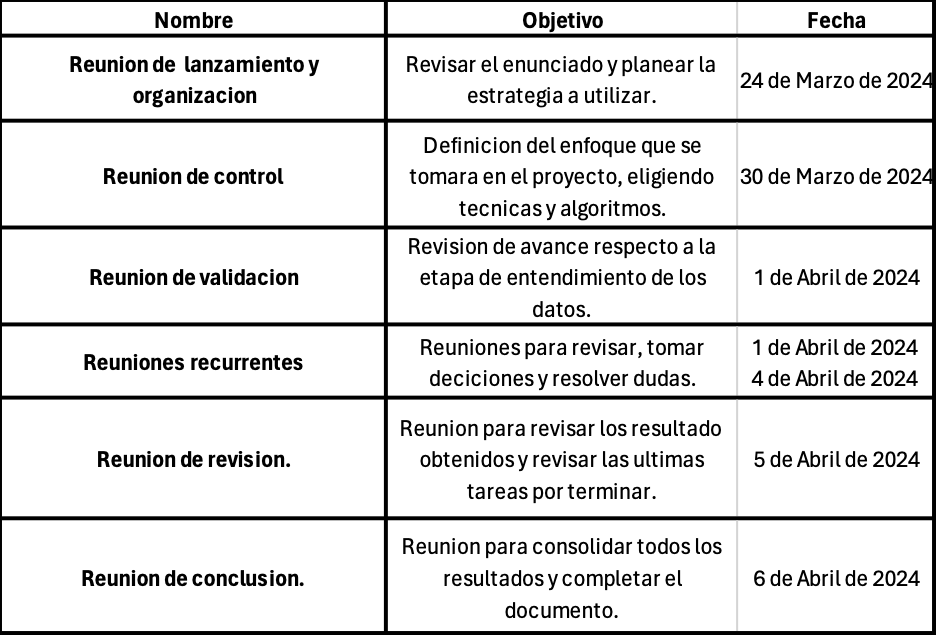
**Lider de proyecto:** El estudiante 2 fue el encargado de revisar los avances y determinar que la carga sea equitativa. Ademas, cumplio con el objetivo de realizar control y llevar registro del avance del proyecto.

**Lider de negocio:** El estudiante 1 fue el encargado de dialogar con los estudiantes de estadistica y revisar que lo implementado sea acorde a las necesidades del negocio.

**Lider de datos:** El estudiante 1 fue el encargado de asumir este rol. En esta parte, se asignaron las tareas de transformacion a aplicar en los datos. Mas especificamente, asignar las etapas descritas en la **seccion 2.2.**

**Lider de analitica:** El estudiante 2 fue el encargado de revisar que el analisis se haya realizado correctamente. Mas especificamente, se encargo de verificar que los algoritmos tengan sus respectivas metricas que permitan comparar y decidir cual podria ser el mejor.

# 4.2 Programacion de reuniones.

Como es bien sabido, este proyecto debe ser realizado utilziando varias tecnicas, sin embargo es necesario que el enfoque que vayan a tomar los involucrados sea el mismo. Con el fin de mantener el equipo controlado y en el mismo camino, se han estipulado las siguientes reuniones de control.

Cabe resaltar, la reunion con los expertos en estadistica corresponde al 4 de Abril de 2024, sin embargo los muchachos no respondieron a través del grupo que creamos para dichas reuniones, por esa razon, no se toma en cuenta dicha reunion.

# 4.3 Tareas importantes asignadas a cada estudiante.

A continuacion, explicaremos las tareas realizadas. Los tiempos que teniamos estipulados resultaron no ser los esperados para analisis de categorias, vectorizacion e implementacion de modelos de redes neuronales, sin embargo esto se debia mas a problemas sobre la utilizacion de tecnologias o de los datos en si; es decir, lo mas demorado fue la solucion de las dificultades.

****