

Turismo de los Alpes | 7 de abril de 2024



Proyecto de analítica de textos

Clasificación de reviews de turismo por medio de modelos de aprendizaje automático

Presentado a:
Turismo de los Alpes



Agenda

- 03 Contexto del problema
- 04 Automatizacion de preparacion de datos
- 05 **Desarrollo de modelo y API**
- 06 Demo de uso de la solucion web
- 07 Analisis de resultados
- 08 Conclusiones e insights

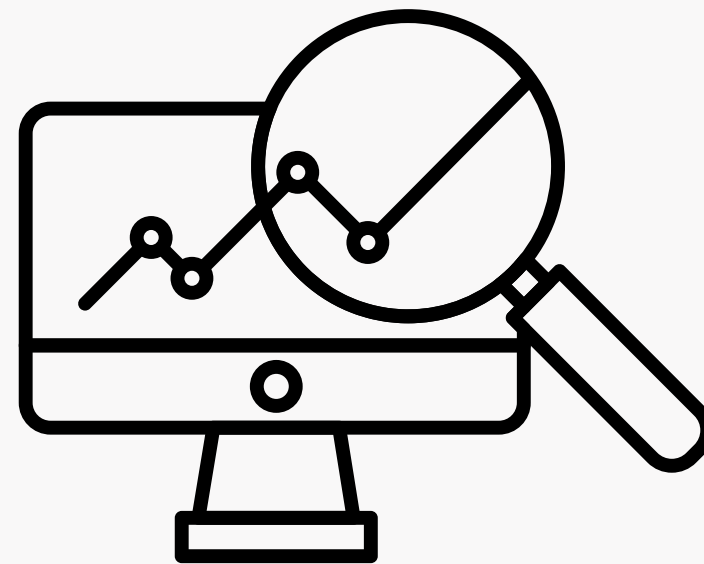
Quienes somos

Equipo interdisciplinario de expertos en
ingeniería y ciencia de datos



Ronald Pardo

Lider de marketing
e ingeniero de
datos



Samuel Goncalves

Ingeniero de datos e
ingeniero de software



Jefferson Hernandez

Lider de negocio e
ingeniero de software

Contexto

El Ministerio de Comercio, Industria y Turismo de Colombia, COTELCO y varias cadenas hoteleras están colaborando para analizar y entender las características que hacen atractivas a las zonas turísticas. Utilizando análisis de reseñas de turistas, buscan evaluar la calidad de estos sitios, con el objetivo de aumentar la visita de turistas, mejorar su satisfacción y potencialmente incrementar los ingresos por turismo.



Enfoque analitico



1. Preprocesamiento de los datos

Puntuacion, caracteres especiales, stop words & lematizacion

2. Selecccion de algoritmos

Regresión Logística
Multinomial, Arboles de desicion & máquinas de vector de soporte

3. Creacion del modelo

Datos de entrenamiento y prueba e implementación de algoritmos

4. Validacion del modelo

Presicion, recall & F1-score

Enfoque analitico



5. Automatizacion de procesamiento de datos

Se hizo un pipeline que permite la automatizacion de ellos, junto a funciones de python que lo hacen para el API.

6. Diseño de API

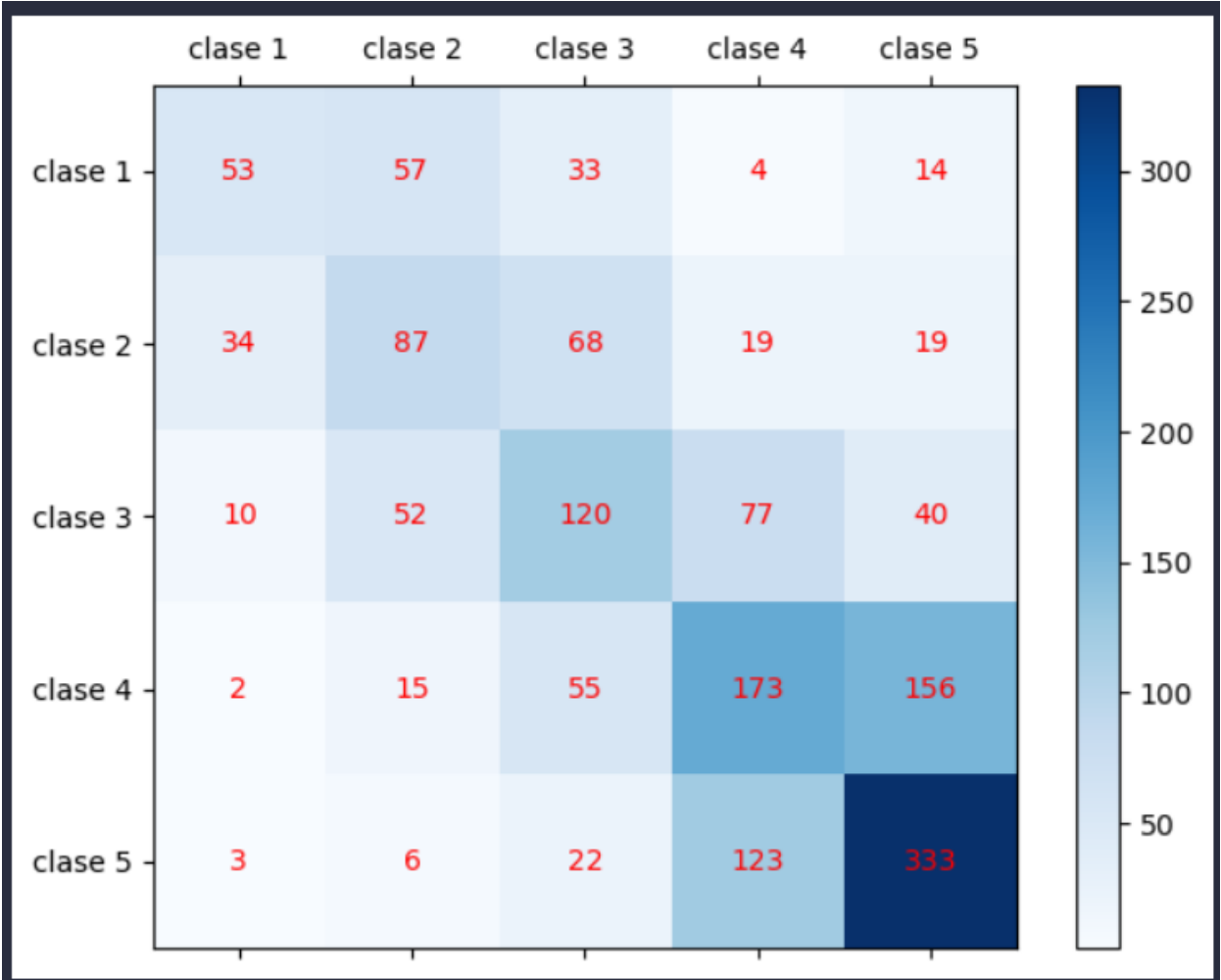
Se hizo con la libreria FastApi de python.

7. Creacion de aplicacion web

Una pagina web de React que permite ingresar reviews y obtener predicciones en tiempo real de la calificacion.

Resultados

Integrante	Modelo	Precision	Recall	F1-Score
Jefferson Hernandez	Regresión logística multinomial	0.47	0.47	0.46
Samuel Goncalves	Arbole de decisión	0.34	0.34	0.34
Ronal Pardo	Máquinas de vector de soporte (SVM)	0.48	0.49	0.48



Presentacion de solucion de analitica web



Resultados

Palabras mas relevantes

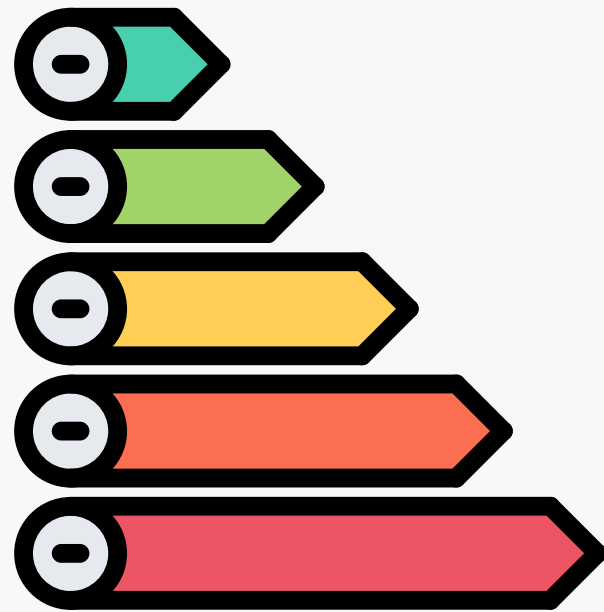
Para una review positiva

Bien, gente, subir, gran, bueno, buena, historia, agradable, viejo, desayuno, precio, siquiera, mojitos, vendedores, dentro, demasiado, control, salón, mojito, ser, habitación, piso, caro, dos, guía, bastante, visita, interesante, igual, pareció.

Para una review negativa

Peor, horrible, asco, pésima, cover, pésimo, ayer, recorrido, chico, jamás, robo, rota, nunca, cubano, nardo, persona, sucias, vuelta, tardo, noma, llegada, sábado, respuesta, cero, sucia, asqueroso, quemas, lamentable, terminar, seguridad.

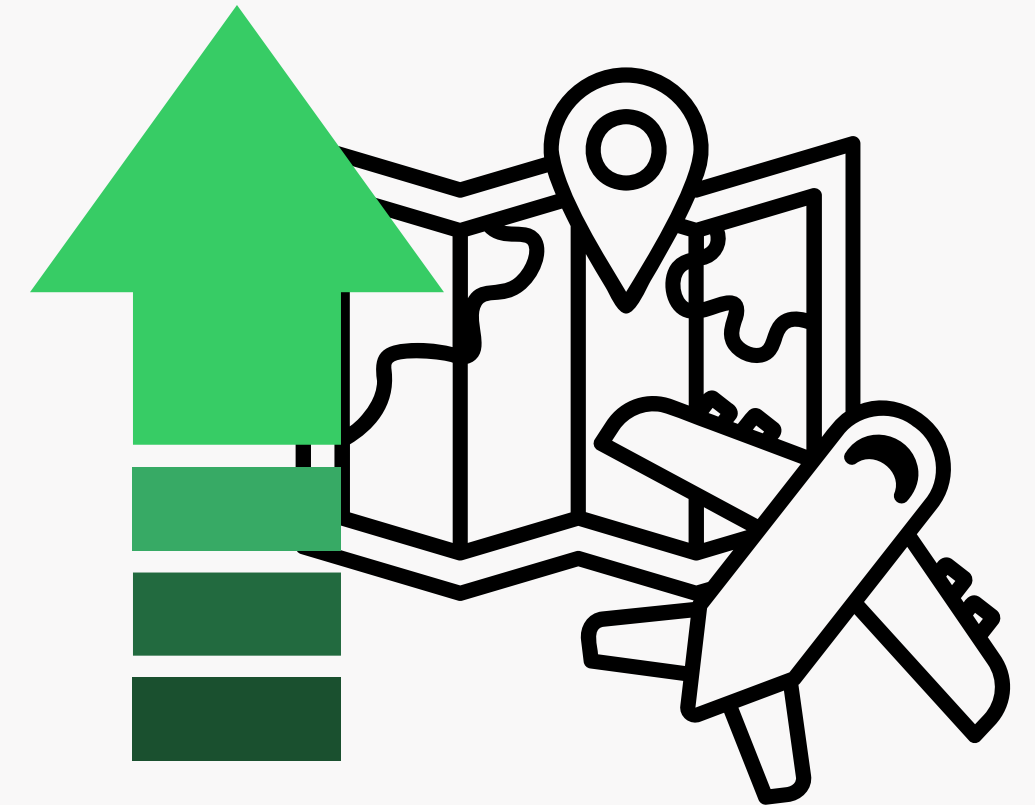
Conclusiones



Se concluye que el mejor modelo para la clasificación de calificaciones a sitios turísticos mediante reviews es el de máquinas de vector de soporte, ya que este obtuvo mejores puntajes en las variables de calidad.



Mediante la aplicación de técnicas más avanzadas de preparación de datos en el ámbito del procesamiento del lenguaje natural, existe un gran potencial para mejorar el rendimiento de los modelos de clasificación de reseñas turísticas



Estas mejoras podrían conducir a una mejor comprensión y clasificación de las opiniones de los turistas, lo que a su vez podría facilitar la identificación de áreas de mejora en los destinos turísticos y el desarrollo de estrategias más efectivas para promover el turismo.



Gracias por su atención

Ronald Pardo

Lider de marketing

Samuel Goncalves

Lider de datos

Jefferson Hernandez

Lider de negocio

