

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

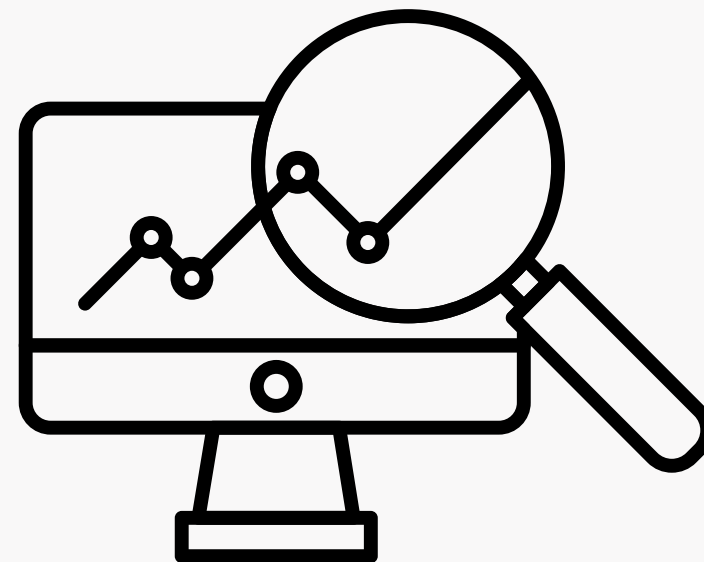
- 03 Contexto del problema
- 04 Objetivos y oportunidades de negocio
- 05 Organización y sus beneficiados
- 06 Enfoque analítico
- 07 Análisis de resultados
- 08 Conclusiones e insights

Quienes somos

Equipo interdisciplinario de expertos en
ingeniería y ciencia de datos



Ronald Pardo
Lider de marketing



Samuel Goncalves
Lider de datos



Jefferson Hernandez
Lider de negocio

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.



Objetivos y oportunidades

Goal # 1

Comprender que hace un destino turistico mas atractivos o poco recomendados.

Goal # 2

Establecer un mecanismo para evaluar la calidad de los sitios según las reseñas de los turistas

Goal # 3

Aumentar el numero de visitas y niveles de satisfaccion de clientes por medio de insights respecto a patrones encontrados en las reviews de sitios exitosos o no exitosos.



Beneficiados

Actores de la organizacion que se beneficiarian por medio de este modelo



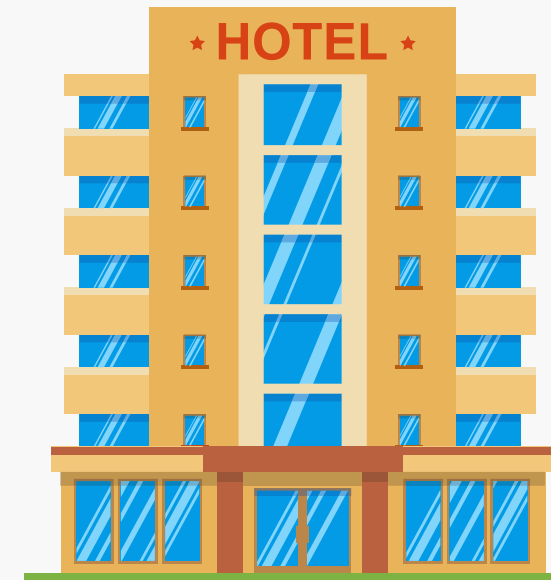
Turistas

Personalizacion basado en predicciones del modelo



Inversores

Retorno de inversiones y fortalecimiento de la compañía



Hoteles, restaurantes, operadores turísticos

Mayor numero de clientes



Comunidad local

Aumento del turismo y gastos turisticos

Enfoque analitico



1. Preprocesamiento de los datos

Puntuación, stop words & lematización

2. Selección de algoritmos

Regresión Logística Multinomial, Árboles de decisión & máquinas de vector de soporte

3. Creación del modelo

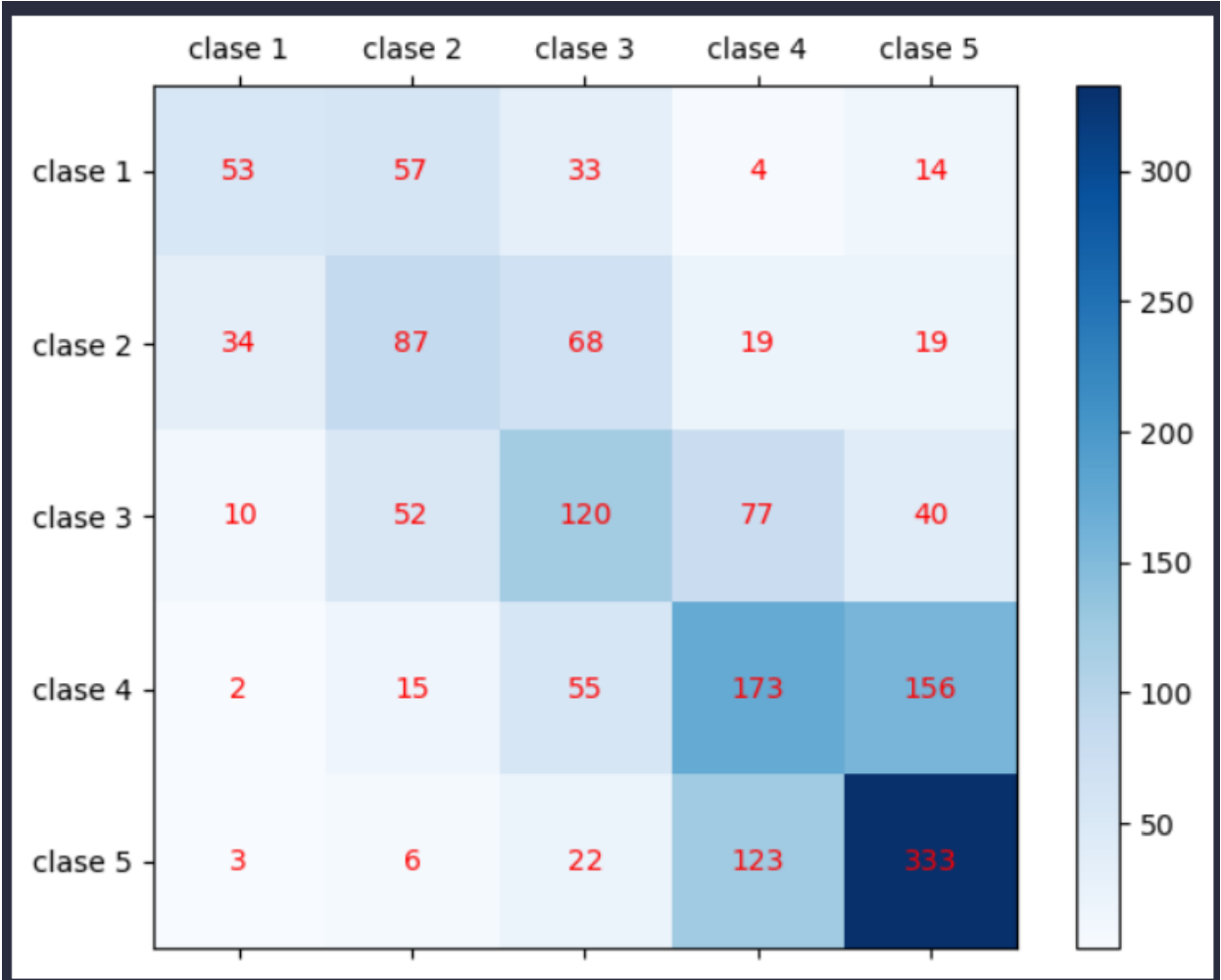
Datos de entrenamiento y prueba

4. Validación del modelo

Presión, recall & F1-score

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



Resultados

Palabras mas relevantes

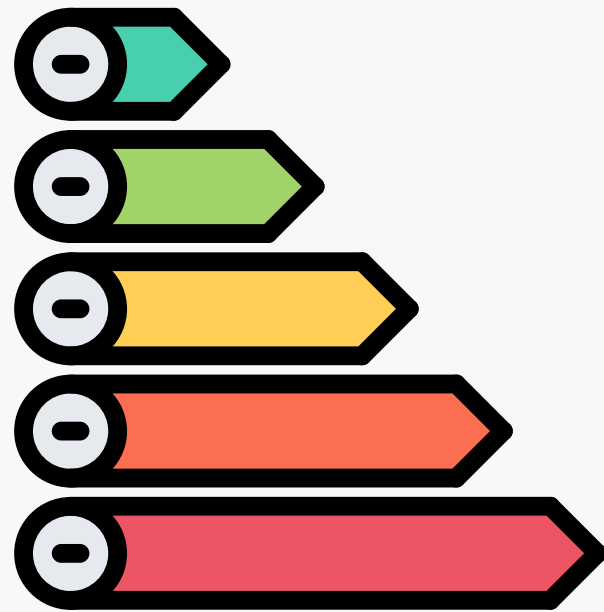
Para una review de 5 estrellas

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.

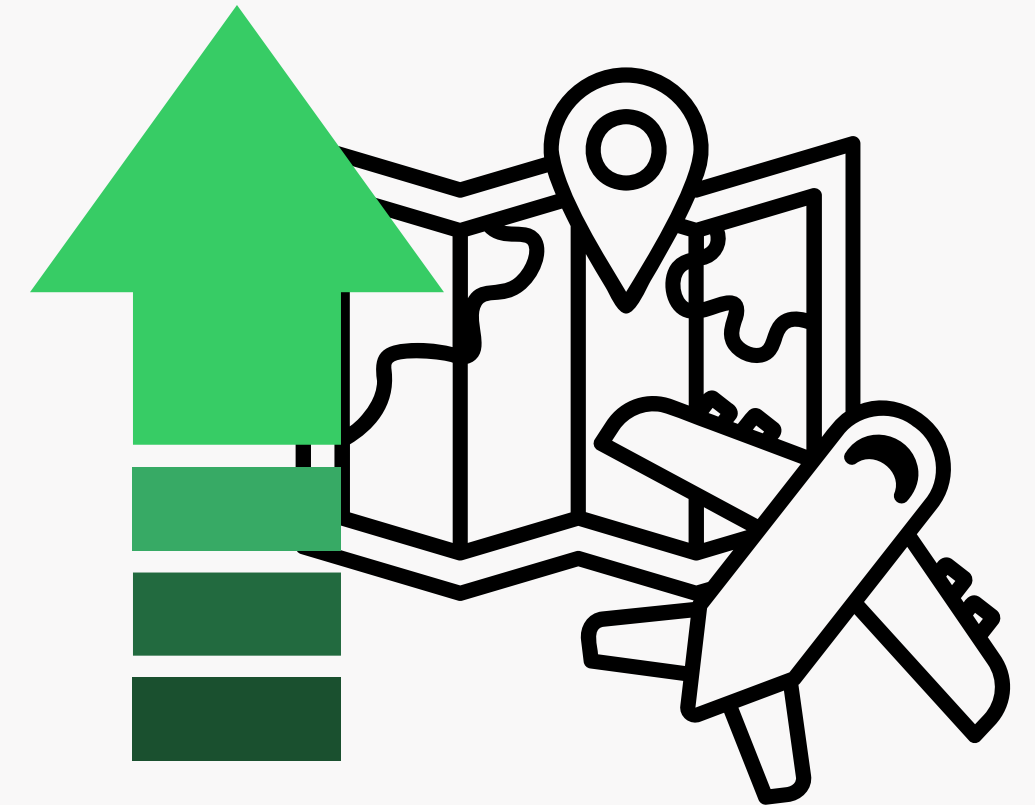
Conclusiones



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.



¿Preguntas?

Ronald Pardo
Lider de marketing

Samuel Goncalves
Lider de datos

Jefferson Hernandez
Lider de negocio