

Proyecto Final

Análisis Estadístico de Datos

Evaluación Estadística de las Opiniones de Pasajeros sobre Aerolíneas

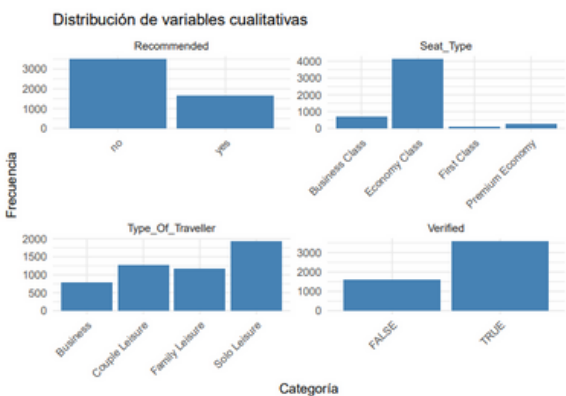
Carlos Galan, Valentina Lopez, Lucia Ardila

Introducción

Comprender las opiniones de los pasajeros es clave para que las aerolíneas mejoren su servicio. Este proyecto analiza datos de reseñas, integrando variables generadas con modelos de análisis de sentimientos como VADER y RoBERTa para enriquecer la evaluación estadística. Los resultados permiten identificar patrones y áreas de mejora en la experiencia del cliente.

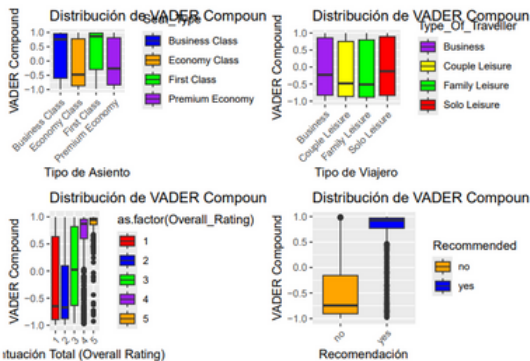
Metodología del Análisis

Variables destacadas



La mayoría de los usuarios no recomendaron el servicio, lo que indica que los modelos de predicción como VADER y RoBERTa deberían haber clasificado las reseñas como mayormente negativas.

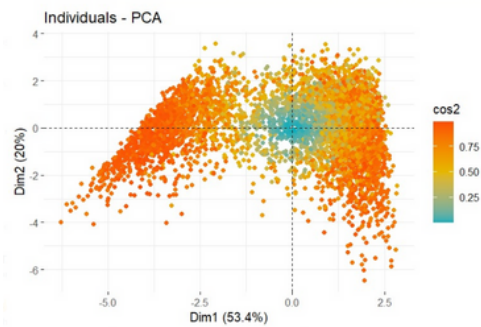
Variables cuantitativas



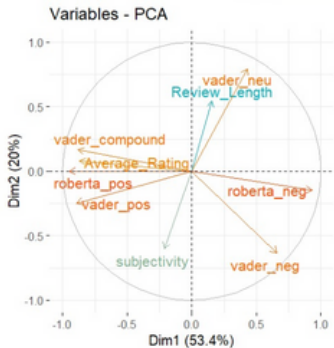
La distribución de VADER Compound por tipo de viajero muestra que Family Leisure tiene una mayor dispersión emocional, con sentimientos tanto positivos como negativos, indicando una variedad de experiencias en viajes familiares.

Reducción de Dimensionalidad y Agrupamiento de Datos: PCA, Clustering y Analisis Factorial

PCA

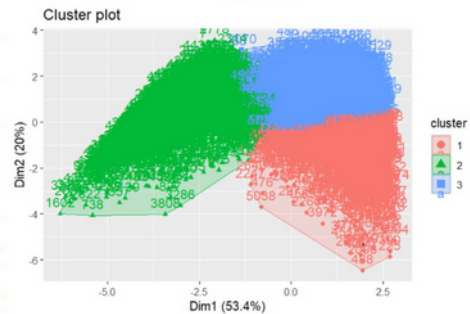


La mayoría de las reseñas se agrupan cerca del centro, lo que indica que los comentarios son en su mayoría neutrales o ligeramente negativos, alineándose con los resultados obtenidos de modelos como VADER y RoBERTa.



El eje Dim1 (53.4%) representa la mayor parte de la varianza explicada en los datos (53.4%), seguido por el eje Dim2 (20%). Esto significa que estos dos componentes juntos explican el 73.4% de la varianza total de las variables, lo cual es un buen resultado para un PCA.

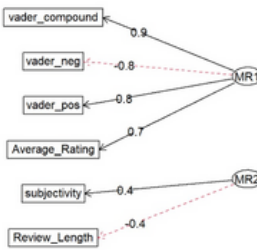
Clustering



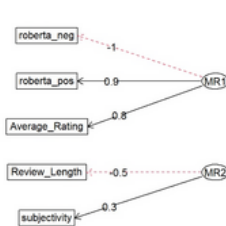
El análisis ha identificado tres grupos de sentimiento que podrían ser útiles para que las aerolíneas adapten sus estrategias de servicio y comunicación, enfocándose en mejorar las experiencias de los grupos más negativos y reforzando lo positivo en las áreas que ya están bien valoradas.

Análisis Factorial

Vader



RoBERTa



Los resultados indican que tanto VADER como RoBERTa capturan estructuras similares en los datos, aunque las cargas pueden diferir ligeramente.

MR1 en ambos casos está claramente dominado por los componentes emocionales de las reseñas (positividad, negatividad y medidas compuestas).

MR2 refleja el vínculo entre subjetividad, longitud de texto y calificaciones.

Conclusión

El análisis de las reseñas muestra que la mayoría de los usuarios no recomendarían el servicio, con sentimientos negativos y neutros predominando. VADER identifica más contenido neutral, mientras que RoBERTa detecta más emociones tanto negativas como positivas. Las reseñas provienen mayormente de pasajeros de clase económica y viajeros solitarios o de pareja, sugiriendo una mayor demanda entre turistas. Las calificaciones bajas se asocian con emociones negativas, lo que refuerza la importancia de las calificaciones como indicador de sentimiento. Estos hallazgos pueden ayudar a mejorar el servicio y guiar estrategias de segmentación de mercado.

Bibliografía

- Kaggle. (2024). Airline Reviews Dataset. <https://www.kaggle.com/datasets/juhibhojani/airline-reviews>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.0480
- Hutto, C. J., & Gilbert, E. E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM), 216-225