

# What Makes an Online Review More Helpful: An Interpretation Framework Using XGBoost and SHAP Values

---

María Gabriela Delgado  
Diego Pando  
María Belén Toledo  
Christian Torres

# • Tabla de Contenidos

01

Contexto

02

Estado del Arte

03

Problema que ataca la  
investigación

04

Metodología

05

Resultados más  
interesantes

06

Conclusiones

# | 01

## : CONTEXTO DEL PROBLEMA

# Contexto del problema

- Las reseñas de productos en línea desempeñan un papel importante siendo las reseñas útiles las influyen en las decisiones de compra de los clientes.
- Los métodos de aprendizaje son útiles pero aún difíciles de utilizar.
- El rápido desarrollo del comercio electrónico ha provocado un crecimiento explosivo en las reseñas de productos en línea.
- Cuando los consumidores compran en línea, prestan especial atención a las evaluaciones de productos de otros consumidores.
- Sitios web de comercio electrónico configuran la función de "votos útiles" en cada reseña de producto.



| 02

ESTADO DEL ARTE

# Estado del arte

2011

Chen

Revisión como  
un elemento de  
información

2015

Krishnamoorthy

Características  
lingüísticas

2017

Hu

Cantidad de  
palabras es un  
predictor

2018

Akbarabadi

Efecto de las  
características  
del título

# | 03

PROBLEMA QUE ATACA LA  
INVESTIGACIÓN

# Problema que ataca la investigación

- Aliviar la presión de sobrecarga causada por un gran número de reseñas en línea para consumidores y empresas.
- No se ha comprendido bien la utilidad de las reseñas.
- El propósito de interpretar la utilidad de las reseñas es identificar las características clave y medir cómo estas características clave afectan la utilidad de la revisión.



# | 04

## ⋮ METODOLOGÍA USADA

# Metodología

→ Se eligió dos tipos diferentes de revisiones de productos de Amazon



## Auriculares

2406 revisiones útiles  
2594 revisiones inútiles



## Limpiadores faciales

2515 revisiones útiles  
2485 revisiones inútiles

# Metodología

Partes de la metodología

- 1 Preprocesamiento de datos.
- 2 Extracción de características.
- 3 Modelado.

# Metodología

## 1. Preprocesamiento de datos

Incluye la segmentación de oraciones, palabras y eliminación de palabras vacías.

## 2. Extracción de características

Se incluyeron tres dimensiones sobre las características del texto:

- Legibilidad: medida en que el texto de revisión es fácil de entender
- Confiabilidad: medida en que se puede confiar en el texto de revisión
- Relevancia: medida en que una reseña es relevante para el producto en sí

# Metodología

## 3. Modelamiento

Se emplearon 3 modelos

- GBDT
- XGBoost
- LightGBM

→ Al combinar el método de valores SHAP con los modelos GBT, se logró dar una explicación detallada de por qué es útil una revisión.



| 05

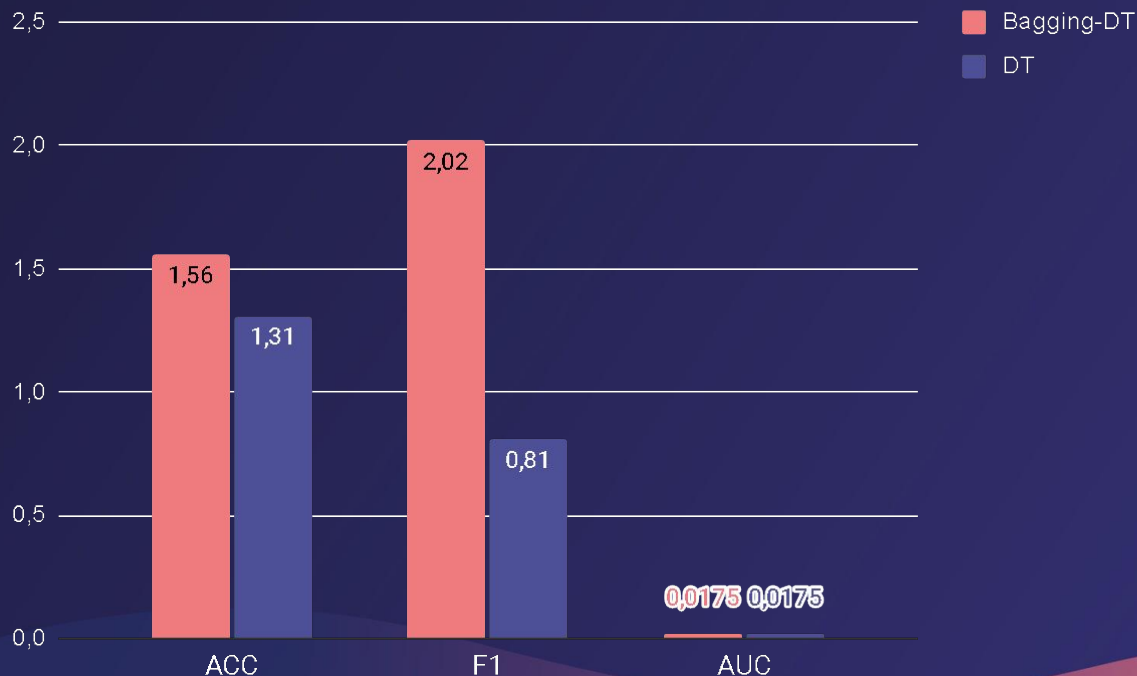
⋮

RESULTADOS MÁS  
INTERESANTES

# Resultados más interesantes

- Comparaciones de rendimiento se utilizó:: Adaboost y Bagging con DT , LR y SVM, RF y ExtraTrees.
- El modelo de RF presenta el resultado más consistente entre todos los modelos de conjunto de línea de base.
- XGBoost se desempeñó mejor en el grupo de modelos GBT.
- Resultados de SVM en el modelo ordinario y conjuntos fueron peores que las de DT o LR.

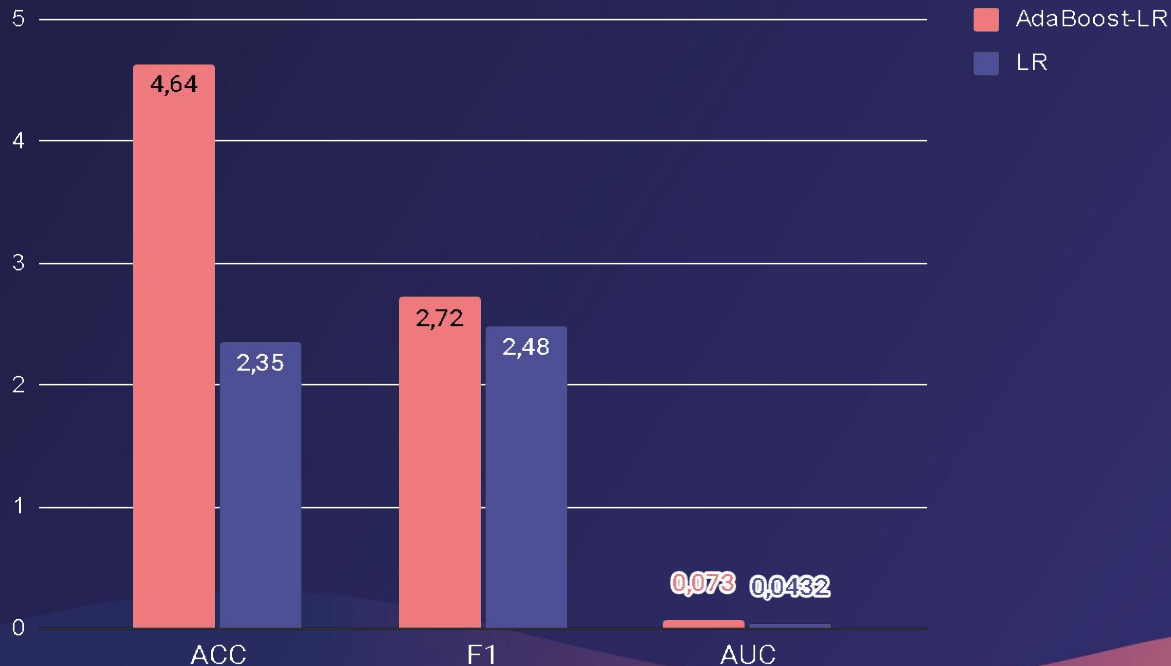
# Resultados más interesantes



- Algunas técnicas de conjunto funcionaron mejor que sus correspondientes .
- Bagging-DT superó a DT .



# Resultados más interesantes



El rendimiento de la regresión logística ordinaria (LR) se puede mejorar en gran medida cuando se combina como AdaBoost-LR.



# | 06

CONCLUSIONES

# Conclusiones

- Para el primer conjunto de datos de los auriculares la cantidad de palabras (o cantidad de oraciones) y la frecuencia con las que aparezcan, contribuye más a la utilidad de las reseñas.
- En los limpiadores faciales, las revisiones están influenciadas por el número de adjetivos de forma interactiva.

# Conclusiones

- Por lo tanto, para la muestra de las revisiones las contribuciones de las características individuales varían según la revisión (conjunto de datos). Pero las contribuciones conjuntas individuales (n-grama) disminuyen gradualmente cuando se aumentan las características.
- Este estudio revela la influencia del valor de la característica en la utilidad de la revisión, al visualizar las relaciones entre el valor de la característica, la contribución de la característica y la contribución conjunta de la característica para una revisión individual.



GRACIAS