

Análisis de Opiniones de Clientes en Ropa para Mujer: Procesamiento de Texto y Modelos Predictivos

Diana Cordero

Marzo 2025

1. Introducción

En la era del comercio electrónico, las opiniones de los clientes desempeñan un papel fundamental en la toma de decisiones de compra. El análisis de reseñas de productos permite extraer información valiosa sobre la satisfacción del cliente y las características más relevantes de los productos.

Este estudio se enfoca en el análisis de un conjunto de datos de reseñas de ropa para mujer comprada en línea. Se aplican diversas técnicas de procesamiento de lenguaje natural (NLP), como la eliminación de stopwords, lematización, stemming, generación de bigramas y análisis de sentimientos. Posteriormente, se entrenan y comparan varios modelos de clasificación para predecir la satisfacción del cliente a partir de las reseñas, incluyendo Random Forest y XGBoost, con y sin balanceo de clases utilizando SMOTE.

El objetivo de este trabajo es evaluar la efectividad de estos modelos y determinar cuál ofrece la mejor precisión en la clasificación de reseñas positivas y negativas, proporcionando así información útil para la optimización de estrategias comerciales en el sector de la moda en línea.

2. Descripción del Conjunto de Datos

El conjunto de datos utilizado en este estudio proviene de Kaggle y contiene reseñas de clientes sobre ropa para mujer comprada en línea. Este dataset proporciona información valiosa sobre la percepción de los clientes en relación con diversos productos, lo que permite analizar patrones de satisfacción y detectar tendencias en las opiniones.

El dataset incluye las siguientes variables principales:

- **Review Text:** Texto de la reseña escrita por el cliente.
- **Rating:** Calificación otorgada al producto en una escala del 1 al 5.

Variable	Count	Mean	Std	Min	25 %	50 %
Age	23486	43.20	12.28	18	34	41
Rating	23486	4.20	1.11	1	4	5
Recommended IND	23486	0.82	0.38	0	1	1
Positive Feedback Count	23486	2.54	5.70	0	0	1

Cuadro 1: Resumen estadístico de las principales variables del dataset.

- **Recommended IND:** Variable binaria que indica si el cliente recomienda el producto (1: Sí, 0: No).
- **Positive Feedback Count:** Número de clientes que encontraron útil la reseña.
- **Division Name, Department Name y Class Name:** Categorías que identifican el tipo de prenda reseñada.

Para este estudio, el análisis se centra en la variable Review Text como fuente de datos textuales y en Recommended IND como la variable de respuesta para modelado predictivo. Se llevaron a cabo procesos de limpieza y transformación del texto antes de entrenar los modelos de clasificación.

3. Metodología

3.1. Análisis Exploratorio de Datos

El primer paso en la construcción del modelo fue realizar un análisis exploratorio de datos (EDA, por sus siglas en inglés) para comprender la estructura y distribución de las variables en el dataset. A continuación, se presentan algunas estadísticas descriptivas clave:

Como se observa en la Tabla 1, la calificación promedio de los productos es de 4.2, con una desviación estándar de 1.11, lo que indica que la mayoría de las calificaciones tienden a ser positivas. Además, el 82.2% de las reseñas recomiendan el producto.

A continuación, se presentan las distribuciones de las calificaciones y la longitud de las reseñas para un análisis más detallado.

Los histogramas en las Figuras 1 y 2 muestran que la mayoría de los clientes otorgan calificaciones altas y que la longitud de las reseñas varía significativamente.

3.2. Preprocesamiento de Datos

Antes de aplicar modelos de aprendizaje automático, se realizó un preprocesamiento del texto en las reseñas con el objetivo de mejorar la calidad de los datos. Este proceso incluyó los siguientes pasos:

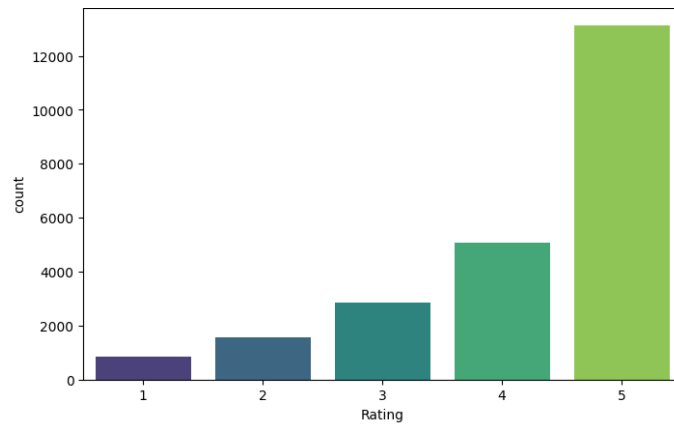


Figura 1: Distribución de calificaciones

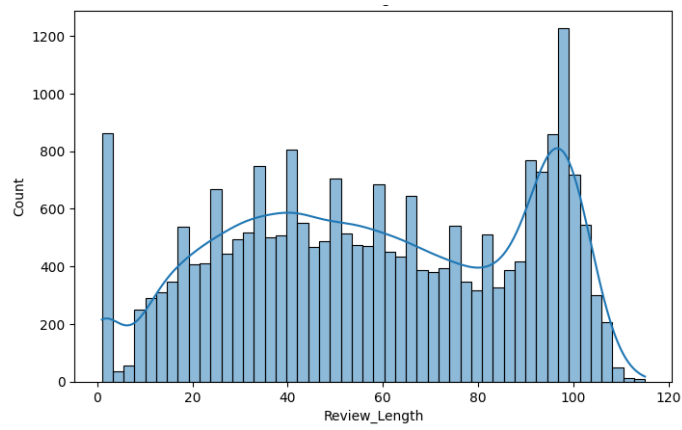


Figura 2: Distribución de longitud de reseñas

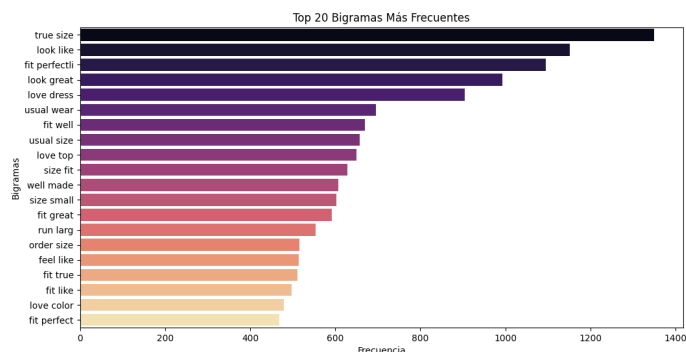


Figura 3: Top 20 de bigramas más frecuentes

- **Eliminación de stopwords:** Se eliminaron palabras comunes en el idioma que no aportan significado relevante al análisis, como artículos, preposiciones y pronombres.
- **Lematización:** Se redujeron las palabras a su forma base o raíz, lo que permite agrupar términos similares y reducir la dimensionalidad del texto.
- **Stemming:** Se aplicó un recorte de palabras a su forma más simple, eliminando sufijos y prefijos para mejorar la representación de los datos.
- **Creación de bigramas:** Se identificaron pares de palabras consecutivas más frecuentes en las reseñas para capturar relaciones clave entre términos.

Como resultado de este proceso, se obtuvo el siguiente top 20 de bigramas más frecuentes en las reseñas.

3.3. Análisis de Sentimientos

Para comprender mejor la opinión de los clientes, se realizó un análisis de sentimientos sobre las reseñas procesadas. La Figura 4 muestra la distribución de los valores de sentimiento obtenidos.

La mayoría de las reseñas presentan un sentimiento positivo, con un pico principal alrededor de 0.25, lo que indica que los clientes suelen expresar opiniones levemente favorables. La distribución es asimétrica hacia la izquierda, lo que sugiere que hay más reseñas positivas que negativas. También se observa un grupo de reseñas con valores cercanos a 0, lo que indica una presencia de opiniones neutrales.

Existen pocas reseñas con valores extremadamente negativos (menores a -0.5), lo que sugiere que las experiencias altamente desfavorables son menos comunes. Del mismo modo, hay pocas reseñas con valores cercanos a 1, lo que indica que los clientes no suelen expresar emociones extremadamente positivas.

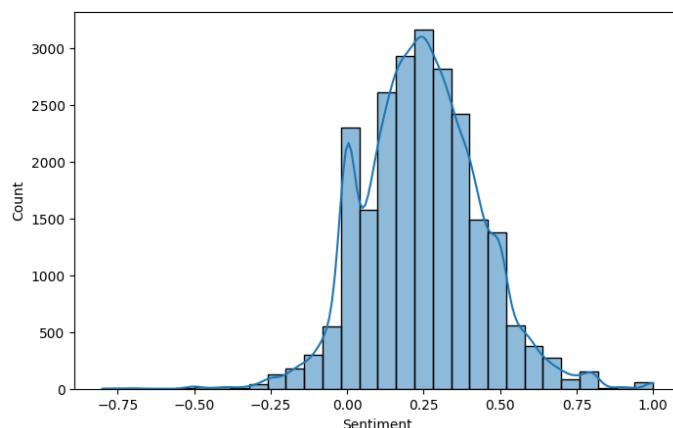


Figura 4: Distribución de sentimientos

En general, estos resultados sugieren que los productos reseñados suelen ser bien recibidos, con la mayoría de los clientes dejando opiniones positivas o neutrales.

3.4. Modelo Multinomial Naïve Bayes

Para la clasificación de las reseñas, se implementó el modelo *Multinomial Naïve Bayes* (MNB), ampliamente utilizado en el procesamiento de lenguaje natural debido a su eficiencia y buen desempeño en tareas de clasificación de texto. Este modelo se basa en la suposición de independencia condicional entre las características y utiliza la probabilidad de aparición de palabras para predecir la categoría de una reseña.

En este caso, el modelo fue entrenado utilizando las reseñas preprocesadas, donde las palabras más relevantes fueron seleccionadas mediante la representación de texto con *Term Frequency-Inverse Document Frequency* (TF-IDF).

Los resultados obtenidos con este modelo muestran un desempeño satisfactorio en la clasificación de las reseñas. La precisión global alcanzada fue del 84.95 %, con una mejor capacidad para identificar reseñas positivas que negativas. A continuación, se presentan las métricas de evaluación:

- **Precisión:** 0.86 para la clase negativa y 0.85 para la clase positiva.
- **Recall:** 0.42 para la clase negativa y 0.98 para la clase positiva.
- **F1-score:** 0.57 para la clase negativa y 0.91 para la clase positiva.

El modelo mostró una alta capacidad para identificar correctamente las reseñas positivas (*recall* de 0.98), pero tuvo dificultades en la clasificación de reseñas negativas, lo que sugiere un desbalance en las predicciones. A pesar de ello, el desempeño general fue adecuado para esta tarea de clasificación de texto.

3.5. Modelo Random Forest

Además del modelo *Multinomial Naïve Bayes*, se implementó un clasificador basado en *Random Forest* (RF), el cual es un algoritmo de ensamble que combina múltiples árboles de decisión para mejorar la precisión y reducir el sobreajuste. Este modelo es robusto ante datos ruidosos y maneja bien la clasificación de texto al considerar la importancia de las características en la toma de decisiones.

El modelo fue entrenado con las reseñas preprocesadas, representadas mediante la técnica *TF-IDF*, y los resultados obtenidos reflejan un desempeño competitivo en la tarea de clasificación. La precisión global alcanzada fue del 86.33 %, con un mejor desempeño en la detección de reseñas positivas. A continuación, se presentan las métricas de evaluación:

- **Precisión:** 0.79 para la clase negativa y 0.87 para la clase positiva.
- **Recall:** 0.33 para la clase negativa y 0.98 para la clase positiva.
- **F1-score:** 0.47 para la clase negativa y 0.92 para la clase positiva.

El modelo mostró un alto rendimiento en la clasificación de reseñas positivas, con un *recall* de 0.98, lo que indica que identifica correctamente la mayoría de las reseñas recomendadas. Sin embargo, la capacidad para detectar reseñas negativas fue más limitada (*recall* de 0.33), lo que sugiere una menor sensibilidad a esta categoría. A pesar de esto, el rendimiento general fue bueno, superando ligeramente al modelo *Naïve Bayes* en términos de precisión global.

Para visualizar el desempeño del modelo, se presenta la matriz de confusión obtenida:

3.6. Mejoras con SMOTE en el Modelo Random Forest

Para abordar el desbalance en las clases y mejorar el rendimiento del modelo, se aplicó la técnica de *Synthetic Minority Over-sampling Technique* (SMOTE). Esta técnica genera nuevas instancias sintéticas de la clase minoritaria para equilibrar la distribución de clases en el conjunto de entrenamiento, lo que permite al modelo aprender mejor las características de las reseñas negativas.

Después de aplicar SMOTE, se observó una mejora en las métricas de clasificación. La precisión global del modelo aumentó ligeramente a 86.42 %, con una mejora notable en la detección de reseñas negativas. A continuación, se presentan las métricas de evaluación:

- **Precisión:** 0.66 para la clase negativa y 0.90 para la clase positiva.
- **Recall:** 0.51 para la clase negativa y 0.94 para la clase positiva.
- **F1-score:** 0.58 para la clase negativa y 0.92 para la clase positiva.

En comparación con el modelo sin SMOTE, el *recall* de la clase negativa aumentó de 0.33 a 0.51, lo que indica que el modelo es ahora más efectivo en la identificación de reseñas negativas. Esto se debe a que, al equilibrar las clases

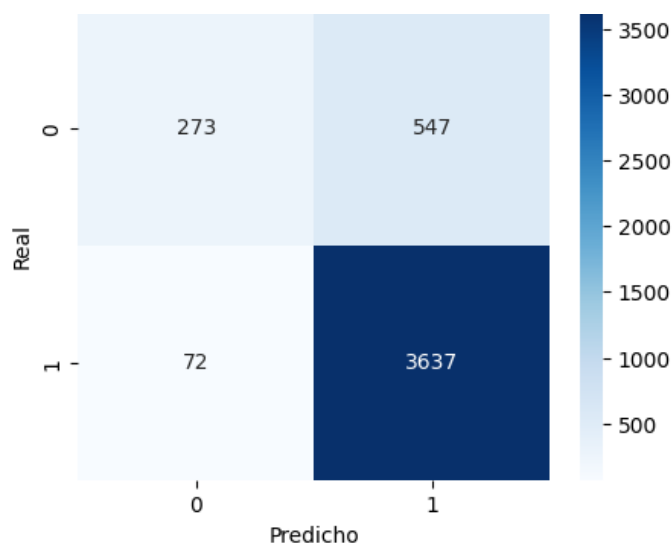


Figura 5: Matriz de confusión - Random Forest

Cuadro 2: Matriz de confusión del modelo Random Forest con SMOTE

	Predicción: Negativo	Predicción: Positivo
Real: Negativo	418	402
Real: Positivo	223	3486

en el entrenamiento, el modelo pudo aprender mejor las características de las reseñas con opiniones negativas, reduciendo la cantidad de falsos positivos.

Como se observa en la matriz de confusión, el número de reseñas negativas correctamente clasificadas aumentó a 418, mientras que el número de falsos positivos disminuyó. Sin embargo, aún se presentan errores en la clasificación de reseñas negativas, lo que sugiere que pueden explorarse otros enfoques de balanceo de datos o ajustes en los hiperparámetros del modelo.

Estos resultados demuestran que la aplicación de SMOTE permitió mejorar el desempeño del modelo, especialmente en la detección de la clase minoritaria, sin comprometer significativamente la precisión global.

3.7. Ajuste de Hiperparámetros para el Modelo Random Forest

Para mejorar el rendimiento del modelo Random Forest, se realizó un ajuste de hiperparámetros utilizando *RandomizedSearchCV*. Esta técnica permite explorar un espacio de búsqueda definido de manera aleatoria, seleccionando las combinaciones de hiperparámetros más prometedoras y reduciendo el tiempo de cómputo en comparación con una búsqueda exhaustiva.

Cuadro 3: Matriz de confusión del modelo Random Forest ajustado

	Predicción: Negativo	Predicción: Positivo
Real: Negativo	361	459
Real: Positivo	148	3561

El espacio de búsqueda incluyó los siguientes hiperparámetros:

- **Número de árboles en el bosque** (*n_estimators*): 100, 200, 300.
- **Profundidad máxima** (*max_depth*): 10, 20, 30 o sin restricción.
- **Mínimo número de muestras para dividir un nodo** (*min_samples_split*): 2, 5, 10.
- **Mínimo número de muestras en una hoja** (*min_samples_leaf*): 1, 2, 4.
- **Uso de *bootstrap***: Verdadero o Falso.

Se realizaron un total de 60 iteraciones, evaluando distintas combinaciones de parámetros mediante validación cruzada de 3 pliegues. Como resultado, se logró una precisión optimizada del 86.31 %. A continuación, se presentan las métricas de evaluación del modelo ajustado:

- **Precisión**: 0.69 para la clase negativa y 0.89 para la clase positiva.
- **Recall**: 0.44 para la clase negativa y 0.96 para la clase positiva.
- **F1-score**: 0.54 para la clase negativa y 0.92 para la clase positiva.

La optimización de los hiperparámetros permitió una mejora en la detección de reseñas negativas en comparación con el modelo base. A pesar de que el *recall* de la clase negativa aún presenta oportunidades de mejora, se logró un equilibrio entre la precisión y la capacidad del modelo para generalizar.

Como se observa en la matriz de confusión, la cantidad de falsos positivos se redujo en comparación con modelos previos, lo que indica una mejor discriminación entre reseñas negativas y positivas. Sin embargo, aún existe un margen de error en la detección de reseñas negativas, lo que sugiere que podrían explorarse técnicas adicionales como el uso de modelos más complejos o el ajuste fino de parámetros adicionales.

Estos resultados confirman la efectividad del ajuste de hiperparámetros en la mejora del modelo Random Forest, permitiendo una clasificación más precisa de las reseñas.

Cuadro 4: Reporte de clasificación del modelo con reducción de características

Clase	Precisión	Recall	F1-score	Soporte
0	0.75	0.39	0.51	820
1	0.88	0.97	0.92	3709

3.8. Reducción de Características y Evaluación del Modelo

En este paso, se aplicó una técnica de reducción de características con el objetivo de mejorar el rendimiento del modelo. Tras esta reducción, se obtuvieron las siguientes métricas de evaluación:

- **Precisión del modelo:** 0.8655, lo que indica que el modelo logró una tasa de predicciones correctas del 86.55 %.

Reporte de clasificación:

Accuracy general: 0.87, lo que muestra un buen desempeño del modelo en términos generales. A continuación, se presentan las métricas específicas de cada clase:

- Para la clase 0 (negativa): la precisión fue de 0.75, el recall de 0.39 y el F1-score de 0.51.
- Para la clase 1 (positiva): la precisión fue de 0.88, el recall de 0.97 y el F1-score de 0.92.

La **media ponderada (weighted avg)** también muestra buenos resultados:

- **Precisión:** 0.85
- **Recall:** 0.87
- **F1-score:** 0.85

Estos resultados indican que, aunque el modelo presenta una ligera desventaja en la clasificación de la clase negativa, es muy efectivo en la clasificación de la clase positiva. Este modelo muestra una ligera ventaja sobre el modelo de Random Forest con Smoote.

3.9. Modelo XGBoost

Para comparar el desempeño del modelo Random Forest con el de otros modelos, se implementó el algoritmo XGBoost. Este modelo es conocido por su capacidad para manejar grandes volúmenes de datos y su efectividad en tareas de clasificación. A continuación se presentan las métricas de evaluación del modelo XGBoost:

Cuadro 5: Reporte de clasificación del modelo XGBoost

	Precision	Recall	F1-score	Support
0	0.76	0.29	0.42	820
1	0.86	0.98	0.92	3709
Accuracy	0.86			
Macro avg	0.81	0.64	0.67	4529
Weighted avg	0.84	0.86	0.83	4529

- **Precisión:** 0.76 para la clase negativa y 0.86 para la clase positiva.
- **Recall:** 0.29 para la clase negativa y 0.98 para la clase positiva.
- **F1-score:** 0.42 para la clase negativa y 0.92 para la clase positiva.

La precisión general del modelo fue de 0.8552, y el reporte de clasificación mostró los siguientes resultados:

A pesar de que el modelo XGBoost mostró una alta precisión para la clase positiva (0.86) y un recall excelente (0.98), el modelo tiene un desempeño deficiente en la detección de la clase negativa, con un recall bajo de 0.29 y un F1-score de 0.42 para esa clase. Esto indica que el modelo tiende a clasificar incorrectamente muchas de las reseñas negativas, lo que podría mejorar con el ajuste de parámetros adicionales o la implementación de técnicas de balanceo de clases.

4. Conclusiones

En este estudio se evaluaron diversos modelos para la clasificación de reseñas, buscando identificar el que mejor se adapta a los datos. A continuación, se comparan los resultados obtenidos de cada uno de los modelos implementados.

4.1. Modelo Multinomial Naive Bayes (NB)

El modelo NB multinomial presentó una precisión general de 0.8495, destacándose en la clasificación de reseñas positivas con un alto recall de 0.98. Sin embargo, la clasificación de reseñas negativas mostró una limitación considerable, con un recall de solo 0.42 y un F1-score de 0.57. A pesar de esto, la precisión general del modelo es competitiva, pero la capacidad para detectar reseñas negativas podría mejorar con un ajuste adecuado.

4.2. Modelo Random Forest

El modelo Random Forest mostró una precisión de 0.8633, con un recall de 0.98 para la clase positiva, similar al modelo NB. No obstante, su desempeño en la clase negativa fue inferior, con un recall de 0.33 y un F1-score de 0.47, lo que indica que este modelo tiene dificultades para clasificar correctamente

las reseñas negativas. Aunque el modelo tiene un buen desempeño general, sus problemas de clasificación para la clase negativa lo limitan en su capacidad de predicción.

4.3. Modelo Random Forest con SMOTE

El modelo Random Forest ajustado con la técnica de SMOTE (Synthetic Minority Over-sampling Technique) mejoró significativamente en cuanto a la clasificación de reseñas negativas. Con una precisión de 0.8642 y un recall de 0.51 para la clase negativa, el modelo mostró un mayor balance entre las clases, con un F1-score de 0.58. Sin embargo, el modelo sigue teniendo un desempeño superior en la clase positiva, con un recall de 0.94, lo que refleja que la clase positiva está más representada en el modelo.

4.4. Ajuste de Hiperparámetros en Random Forest

El ajuste de hiperparámetros para el modelo Random Forest resultó en una ligera mejora en la precisión general (0.8631). Aunque hubo una pequeña mejora en el desempeño, el recall de la clase negativa sigue siendo bajo (0.44), lo que indica que el ajuste no resolvió completamente los problemas de clasificación de reseñas negativas. Sin embargo, se observó una ligera mejora en el equilibrio entre las clases, con un macro promedio de recall de 0.70.

4.5. Reducción de Características

El modelo con reducción de características mostró una precisión ligeramente superior a los anteriores, alcanzando un 0.8655. Aunque la precisión para la clase positiva sigue siendo alta (0.88), la clasificación de reseñas negativas mejoró, con un recall de 0.39. Este modelo tiene un buen desempeño general, con un F1-score de 0.51 para la clase negativa, lo que sugiere que la reducción de características ayuda a mejorar el balance entre clases.

4.6. Modelo XGBoost

El modelo XGBoost presentó una precisión del 0.8552, similar a la de los modelos anteriores. Aunque el recall para la clase positiva fue alto (0.98), el recall para la clase negativa fue muy bajo (0.29), lo que resultó en un F1-score de solo 0.42. Esto indica que, aunque el modelo es muy eficiente para clasificar reseñas positivas, su desempeño en la clase negativa es deficiente, lo que limita su aplicabilidad en este caso.

4.7. Comparación y Modelo Más Eficiente

Comparando los modelos evaluados, el modelo con reducción de características se destacó como el que logró un mejor equilibrio entre la clasificación de reseñas positivas y negativas, con una precisión general de 0.8655. Este modelo

mostró una mejora en el recall de la clase negativa (0.39) en comparación con los demás modelos. Aunque su desempeño para la clase negativa aún no es ideal, su capacidad para mantener una alta precisión y mejorar el balance general lo convierte en el modelo más prometedor.

Por otro lado, el modelo XGBoost, a pesar de su alta precisión para la clase positiva, tuvo un rendimiento deficiente en la detección de reseñas negativas, lo que hace que no sea el modelo más adecuado para este conjunto de datos.

5. Referencias

- Women's E-Commerce Clothing Reviews. (2018, 3 febrero). Kaggle. Recuperado de <https://www.kaggle.com/datasets/nicapotato/womens-ecommerce-clothing-reviews>