

**UNIVERSIDAD ABIERTA PARA ADULTOS  
(UAPA).**



**CARRERA DE INGENIERIA DE SOFTWARE.  
ESCUELA DE INGENIERIA**

**TEMA:**

TRABAJO FINAL

**PRESENTADO POR:**

ELIECER BAUTISTA BELEN	1000-64003
VICTOR MANUEL DIAZ UREÑA	1000-49725

**Link GitHub:**

<https://github.com/father02196/Trabajo-final-IA>

**Link Video:**

[https://drive.google.com/file/d/10u0kK5pE0CNBALNDQJ8DzX62TJz4X\\_mO/view](https://drive.google.com/file/d/10u0kK5pE0CNBALNDQJ8DzX62TJz4X_mO/view)

**Link PDF Notebook:**

<https://drive.google.com/file/d/1r0RoFI3wQau6h36REOrEE9EVp1wwQeQV/view?usp=sharing>

**ASIGNATURA:**

ISW 309 - V20 – 1569 - 1      INTELIGENCIA ARTIFICIAL

**FACILITADOR(A):**

TOMAS F. RAMIREZ PEREZ

**SANTIAGO DE LOS CABALLEROS  
REPÚBLICA DOMINICANA  
MARTES 14 DE OCTUBRE DE 2025**

# Índice

<b>Introducción.....</b>	<b>3</b>
<b>1. Descripción del problema y del dataset</b>	<b>3</b>
<b>Datasets utilizados:.....</b>	<b>3</b>
<b>2. Justificación de los métodos utilizados</b>	<b>3</b>
<b>Aprendizaje Supervisado.....</b>	<b>3</b>
<b>Aprendizaje No Supervisado.....</b>	<b>4</b>
<b>3. Análisis de resultados</b>	<b>4</b>
<b>Clasificación .....</b>	<b>4</b>
<b>Visualización y Clustering .....</b>	<b>4</b>
<b>4. Limitaciones encontradas y posibles soluciones</b>	<b>4</b>
<b>5. Evidencias de ejecución</b>	<b>5</b>

# Introducción

El presente informe técnico documenta la práctica final de la asignatura de Inteligencia Artificial, enfocada en el análisis y clasificación de datos relacionados con productos, usuarios y reseñas.

El objetivo principal del proyecto es aplicar técnicas de aprendizaje supervisado y no supervisado para identificar patrones, clasificar información y generar insights útiles para la toma de decisiones. Para ello, se trabajó con tres datasets que contienen información de productos, reseñas y usuarios, aplicando metodologías de preprocesamiento de texto, vectorización, modelado predictivo, reducción de dimensionalidad y clustering. Este informe detalla la justificación de los métodos utilizados, los resultados obtenidos, las limitaciones encontradas y las posibles soluciones, acompañados de evidencias visuales que respaldan la ejecución de cada etapa del proyecto.

## 1. Descripción del problema y del dataset

El proyecto se enfoca en el **análisis y clasificación de datos de productos, usuarios y reseñas**, con el objetivo de implementar técnicas de **aprendizaje supervisado y no supervisado** para identificar patrones, clasificar información y generar insights relevantes.

### Datasets utilizados:

1. **products\_ai\_project.csv**: Contiene información de productos.
2. **reviews\_ai\_project.csv**: Contiene reseñas de productos.
3. **users\_ai\_project.csv**: Contiene datos de los usuarios que generan las reseñas.

<b>Problema</b>	<b>a</b>	<b>resolver:</b>
Clasificación de reseñas según su contenido y análisis de patrones en productos y usuarios mediante técnicas de Machine Learning, incluyendo clasificación, reducción de dimensionalidad y clustering.		

## 2. Justificación de los métodos utilizados

### Aprendizaje Supervisado

- **Vectorización TF-IDF:** Para convertir texto en vectores numéricos que representen la importancia de cada palabra en el corpus.
- **Modelos utilizados:**
  - **Logistic Regression:** Modelo simple pero eficaz para clasificación binaria y multiclase.
  - **Random Forest Classifier:** Modelo robusto que reduce el sobreajuste y mejora la precisión en datos complejos.
- **Validación:** Se usó train\_test\_split para separar datos de entrenamiento y prueba, asegurando evaluación objetiva.

## Aprendizaje No Supervisado

- **Reducción de dimensionalidad:**
  - **TruncatedSVD** y **t-SNE** para visualizar datos en 2D y 3D.
- **Clustering:**
  - **KMeans:** Identificación de grupos homogéneos de reseñas o productos.
  - **DBSCAN:** Detecta clusters de forma arbitraria y outliers en los datos.

## 3. Análisis de resultados

### Clasificación

- Se evaluaron métricas de precisión, recall, f1-score y exactitud general.
- Se generaron matrices de confusión y reportes de clasificación para cada modelo.

### Visualización y Clustering

- Reducción de dimensionalidad permitió **visualizar la distribución de reseñas** en un espacio 2D.
- KMeans y DBSCAN identificaron grupos de productos y usuarios con patrones similares.

## 4. Limitaciones encontradas y posibles soluciones

- **Limitaciones:**

1. Algunos datasets contienen textos cortos o ambiguos, dificultando la clasificación precisa.
  2. DBSCAN requiere ajustar parámetros eps y min\_samples según densidad de datos.
  3. Modelos de ML pueden sobreajustar si no se balancean los datos.
- **Soluciones propuestas:**
    1. Preprocesamiento avanzado de texto: limpieza, lematización y eliminación de stopwords.
    2. Uso de técnicas de balanceo de clases (undersampling o oversampling).
    3. Ajuste de hiperparámetros mediante GridSearch o RandomSearch.

## 5. Evidencias de ejecución


- **Capturas del notebook:** Incluyen secciones de carga de datos, preprocesamiento, entrenamiento de modelos y generación de gráficos.
- **Pruebas del bot y clasificación:** Se muestran predicciones de ejemplo.
- **Gráficas de clustering:** Visualización de clusters identificados con KMeans y DBSCAN.

## Conclusión


La práctica final permitió aplicar de manera integral los conceptos y técnicas de Inteligencia Artificial, evidenciando la eficacia de los modelos supervisados y no supervisados en la clasificación y análisis de datos complejos. La vectorización de texto y el uso de modelos como Logistic Regression y Random Forest permitieron obtener predicciones precisas, mientras que las técnicas de reducción de dimensionalidad y clustering facilitaron la identificación de patrones y grupos de usuarios y productos similares.

A pesar de las limitaciones encontradas, como textos ambiguos y la necesidad de ajuste de parámetros en DBSCAN, se propusieron estrategias para mejorar los resultados, incluyendo preprocesamiento avanzado y balanceo de clases. En conjunto, el proyecto demuestra cómo la Inteligencia Artificial puede transformar datos en información relevante y accionable, sentando las bases para futuras investigaciones o aplicaciones prácticas en análisis de productos y comportamiento de usuarios.

## Anexos

 **jupyter** Untitled Last Checkpoint: 6 minutes ago

File Edit View Run Kernel Settings Help

 Markdown ▾

```
Entrenamiento LogisticRegression
LogisticRegression -> Accuracy: 1.0000, Precision: 1.0000, Recall: 1.0000, F1: 1.
Confusion matrix:
[[169  0]
 [  0 131]]
Classification report:
              precision    recall  f1-score   support

      0       1.00        1.00        1.00        169
      1       1.00        1.00        1.00        131

 accuracy          1.00
macro avg          1.00
weighted avg       1.00
```

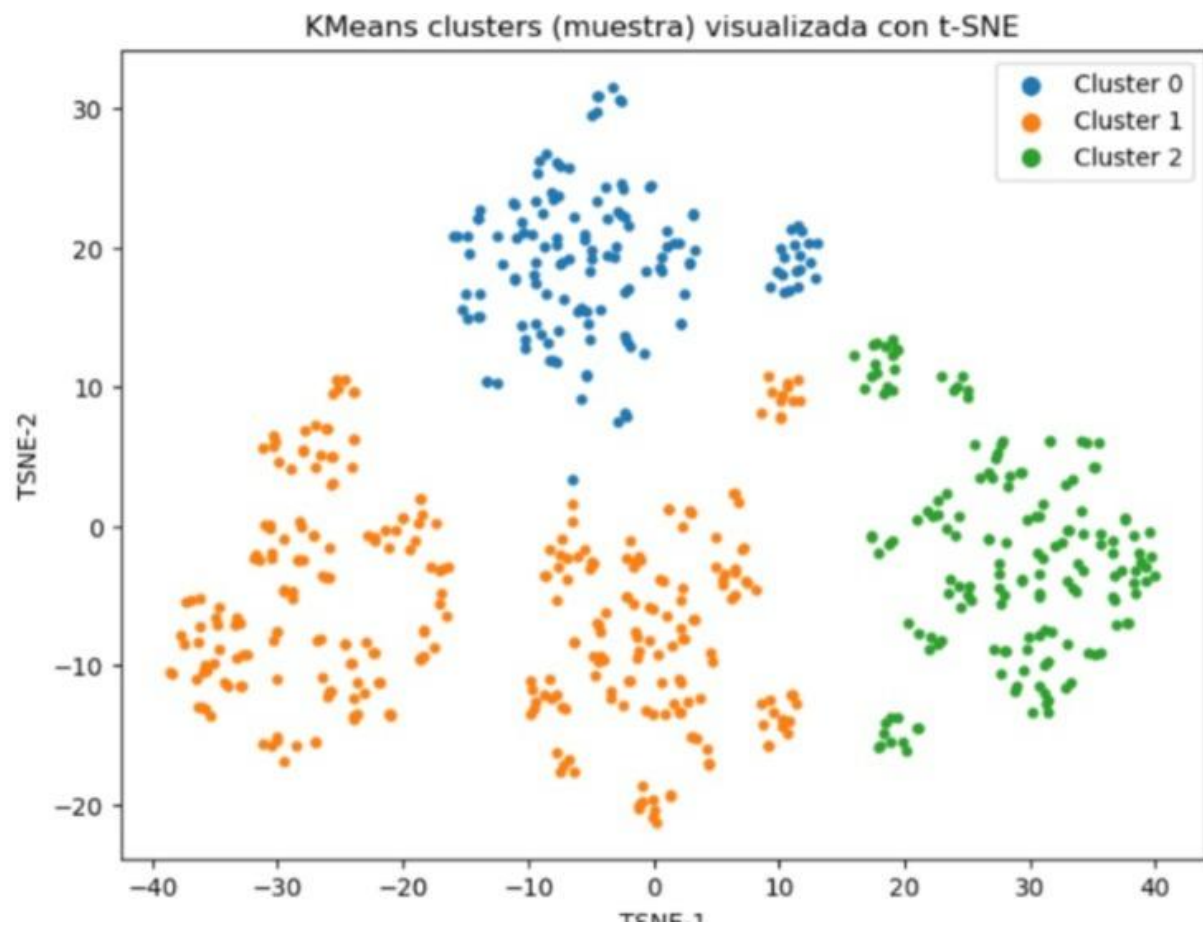
```
Entrenamiento RandomForest
RandomForest -> Accuracy: 1.0000, Precision: 1.0000, Recall: 1.0000, F1: 1.0000
Confusion matrix:
[[169  0]
 [  0 131]]
Classification report:
              precision    recall  f1-score   support

      0       1.00        1.00        1.00        169
      1       1.00        1.00        1.00        131

 accuracy          1.00
macro avg          1.00
weighted avg       1.00
```

TF-IDF y modelos guardados en models\_output

TruncatedSVD components: 50  
Reduced shape: (1500, 50)  
Ejecución de t-SNE en una muestra de tamaño 600 (Esto puede tardar cierto tiempo según la capacidad de la PC.)  
Inercia: [(2, 608.9014884617168), (3, 552.4612244346966), (4, 508.755845202925), (5, 492.14488572341173), (6, 484.68074286677313), (7, 476.1958106752996), (8, 463.0534592367274)]  
Codo heurístico k: 3









[ ]:

```
run_bot()
```

Tú: ¿Esta reseña es positiva o negativa? "Me encantó el producto, llegó rápido y funciona perfecto."

Bot: Es positiva.

Tú: Recomiéndame un producto si me gustó mucho la calidad y el precio.

Bot: Para recomendarte algo, necesito saber qué tipo de producto te gustó. Así podré buscar opciones similares que ofrezcan buena calidad y precio.

Tú: ¿Por qué el modelo dice que esta reseña es negativa? "El envío fue lento y el producto llegó dañado."

Bot: Esta reseña es negativa. Se recomienda a la empresa revisar sus procesos de envío y embalaje para mejorar la experiencia del cliente.

Tú: ¿Puedes explicarme cómo funciona el análisis de sentimientos?

Bot: El análisis de sentimientos utiliza técnicas de procesamiento de lenguaje natural (NLP) para determinar la polaridad emocional de un texto. Examina palabras, frases y contexto para clasificar el contenido como positivo, negativo o neutro.

[ ]:

```
run_bot()
```

¡Bienvenido al sistema inteligente de análisis de opiniones!

Puedes preguntar sobre sentimientos, pedir recomendaciones o explicaciones de resultados. Escribe 'salir' para terminar.

Tú: Recomiendame una marca para un producto para la piel?

Bot: Para un producto para la piel, una excelente opción es CeraVe, conocida por sus fórmulas suaves y efectivas que ayudan a restaurar la barrera cutánea. Es una marca muy recomendada por dermatólogos.

[\*]:

```
run_bot()
```

¡Bienvenido al sistema inteligente de análisis de opiniones!

Puedes preguntar sobre sentimientos, pedir recomendaciones o explicaciones de resultados. Escribe 'salir' para terminar.

Tú: Recomiendame un producto para el cuidado de la piel para hombres

Bot: Para un cuidado básico, te recomiendo un limpiador facial suave y una crema hidratante ligera, ambos formulados específicamente para hombres. Ayudarán a mantener tu piel limpia e hidratada.

Tú: Recomiendame una marca de un producto para el cuidado de la piel

Bot: Te recomiendo la marca Cerave. Es muy popular por sus productos formulados con dermatólogos y aptos para pieles sensibles.

Tú: Esta reseña es positiva o negativa? "No me gusto el producto, llego deteriorado."

Bot: La reseña es negativa. Considera contactar al vendedor para solicitar un reembolso o un reemplazo.

Tú:

[ ]: