

IMMUNE
TECHNOLOGY INSTITUTE

Análisis de sentimientos de reseñas - Disneyland

Master en Data Science
2406VDSO

Nombre de los autores



Keilor Fallas
Lindsay López
Wendy Rodríguez
Allan Vargas

>OUTLINE_





1. INTRODUCCIÓN_

<MOTIVACIÓN>

Objetivo general:

Analizar el sentimiento de las reseñas de visitantes de Disneyland (California, París y Hong Kong) utilizando NLP, DL y ML.

<OBJETIVOS>

> OBJETIVO 1:

Preparar y limpiar los datos

> OBJETIVO 2:

Aplicar técnicas de EDA

> OBJETIVO 3:

Clasificar las opiniones usando ML, DL y NLP

> OBJETIVO 4:

Comparar desempeño de modelos

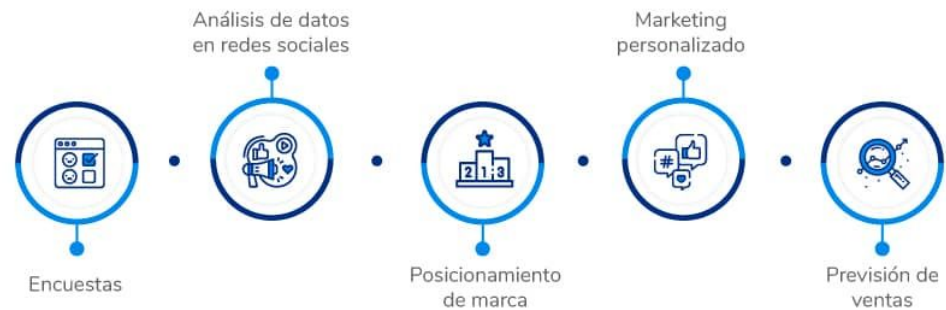
2. ESTADO DEL ARTE

Fundamentos del análisis de sentimiento

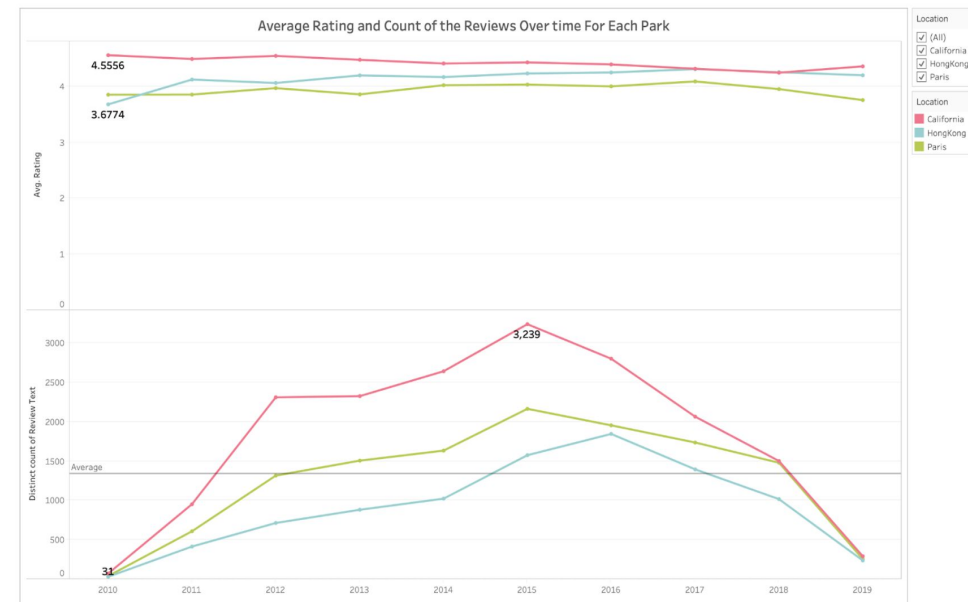
- > Procesamiento del Lenguaje Natural (NLP)
 - >> Representación del Texto
 - >> Clasificación
 - >> Métricas de Rendimiento



Usos de la herramienta de análisis de sentimientos



Trabajos relacionados





3. DESARROLLO Tratamiento de datos

<Variables>

- > review_id
- > rating
- > year_month
- > reviewer_location
- > review_text
- > branch

<Limpieza>

- > Eliminar duplicados
- > Estandarizar fechas
- > Limpieza de texto
- > Consistencia de variables
- > Normalización de texto de reseñas

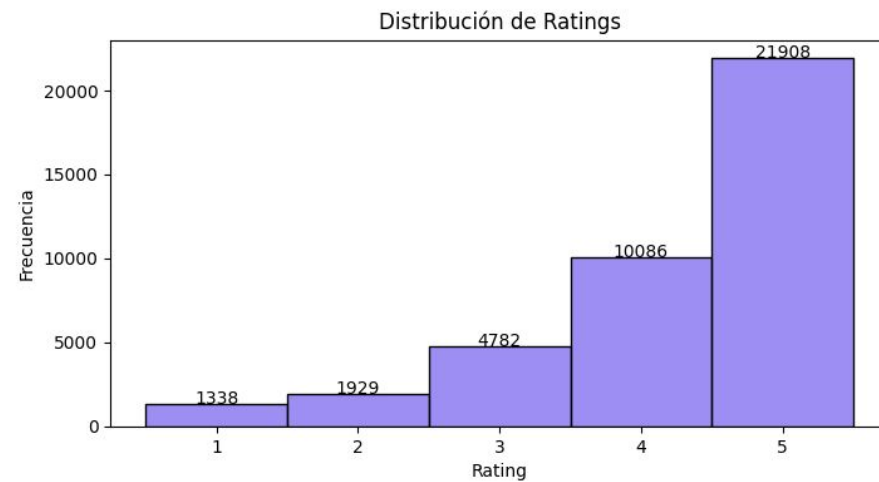
<Variable target>

- > 1 – 3: Negativa
- > 4 – 5: Positiva

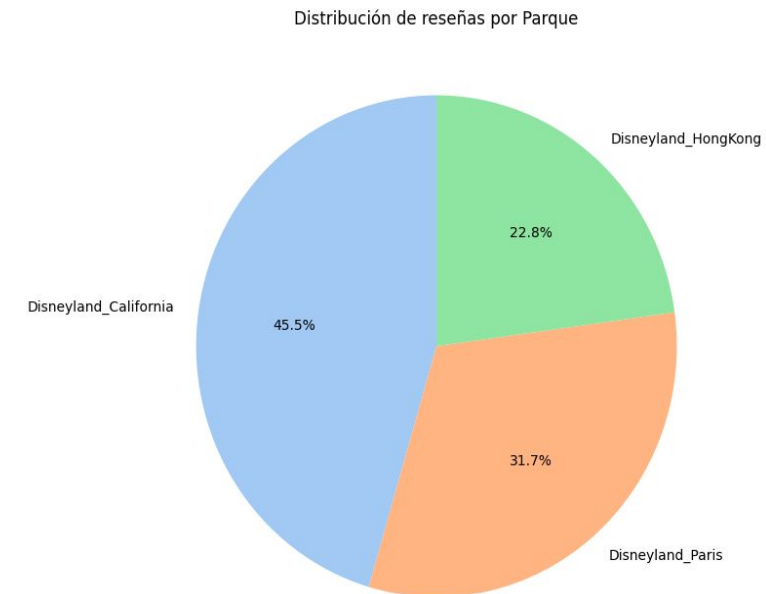


3. DESARROLLO_EDA

>> EDA: Histograma de la Distribución de los Ratings

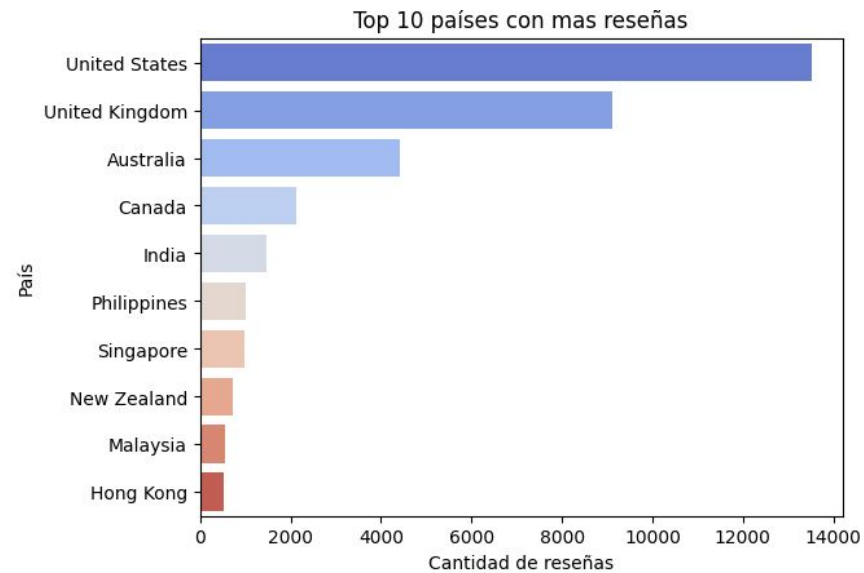


>> EDA: Gráfico de Pastel de la Distribución por Parque



3. DESARROLLO_EDA

>> EDA: Gráfico de Barras Horizontal de los Top 10 Países con Más Reseñas

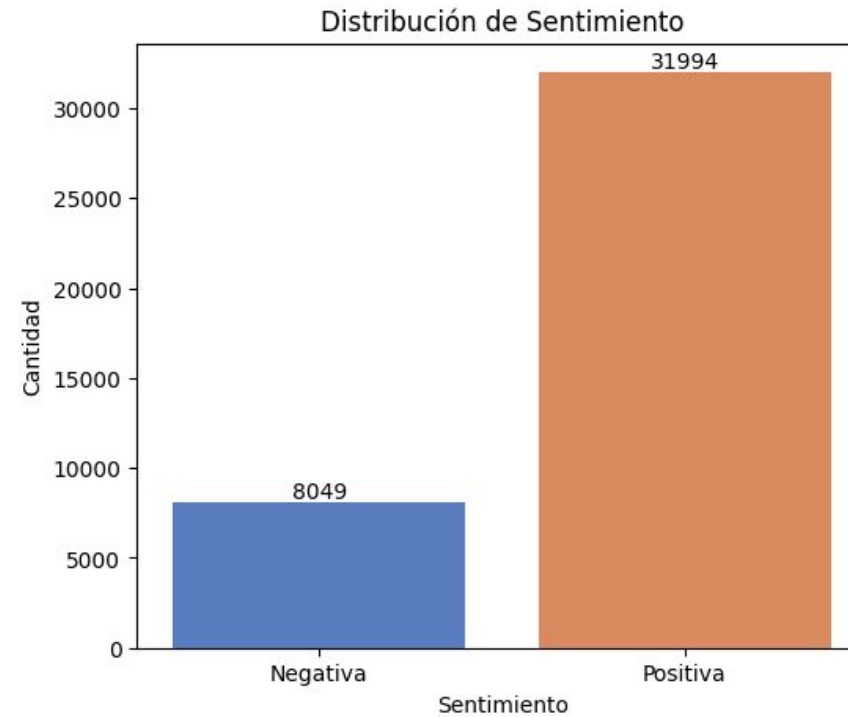


>> EDA: Gráfico de líneas: Reseñas a lo Largo del Tiempo



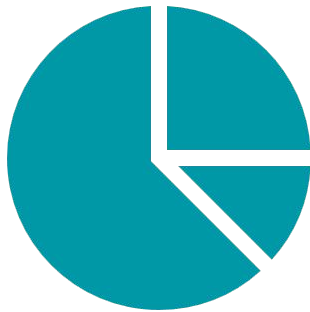
3. DESARROLLO_EDA

>> EDA: Gráfico de Barras de la Distribución de Sentimiento



3. DESARROLLO_EDA

Split Train – Test



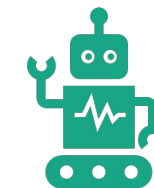
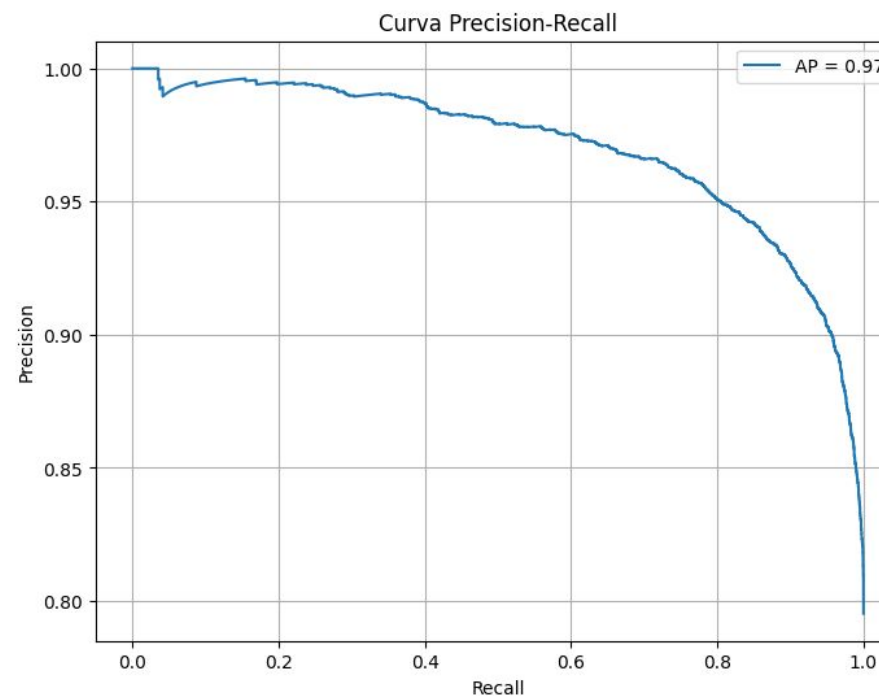
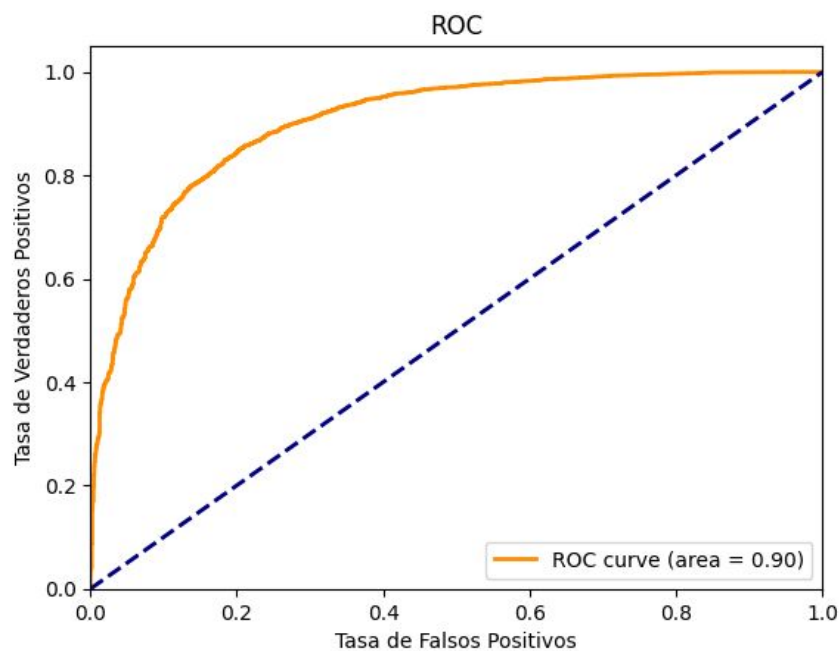
> 80 % Entrenamiento

> 20 % Pruebas



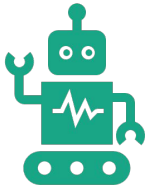
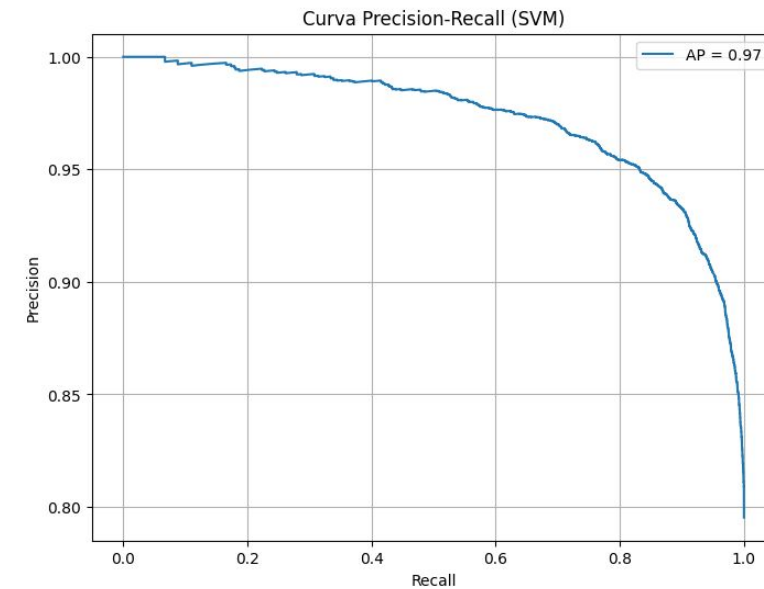
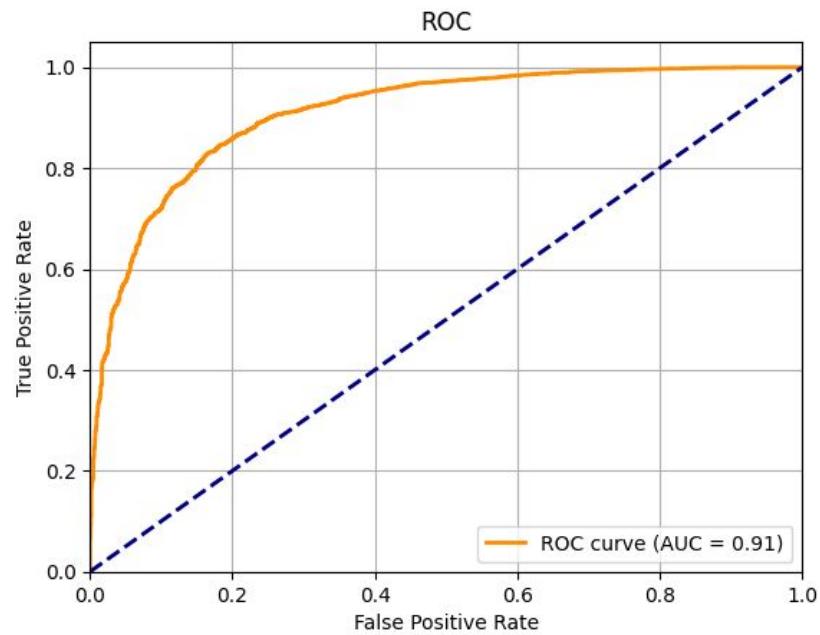
3. DESARROLLO_ML, DL y NLP

>>ML: Regresión Logística



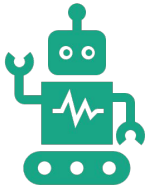
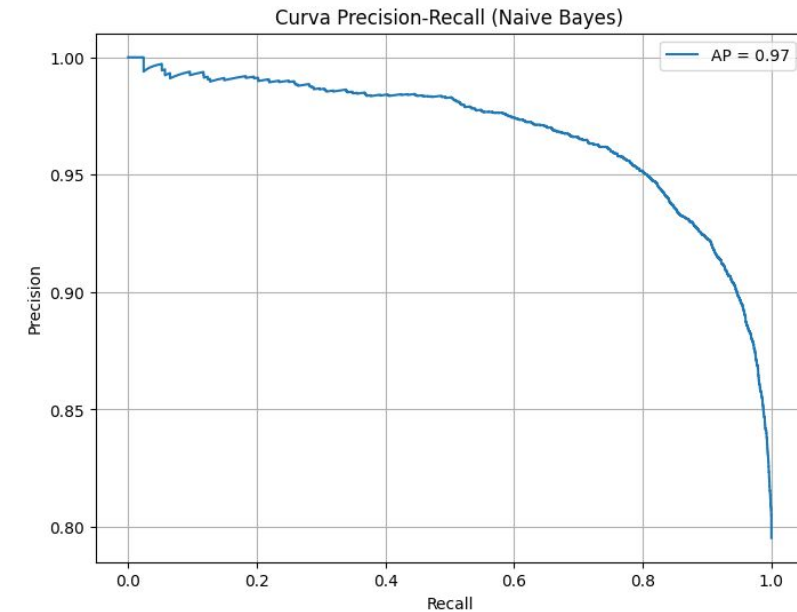
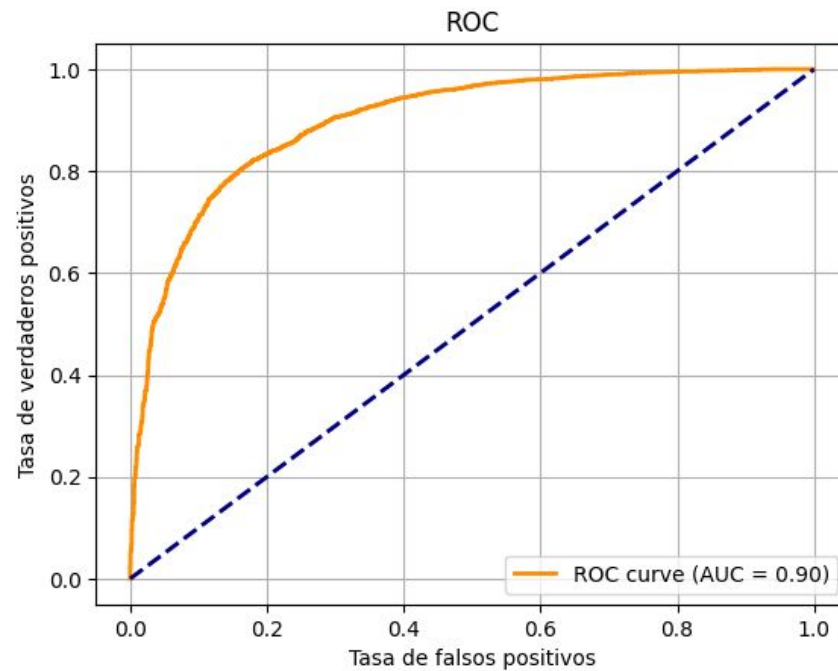
3. DESARROLLO_ML, DL y NLP

>>ML: Support Vector Machine



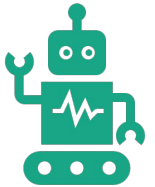
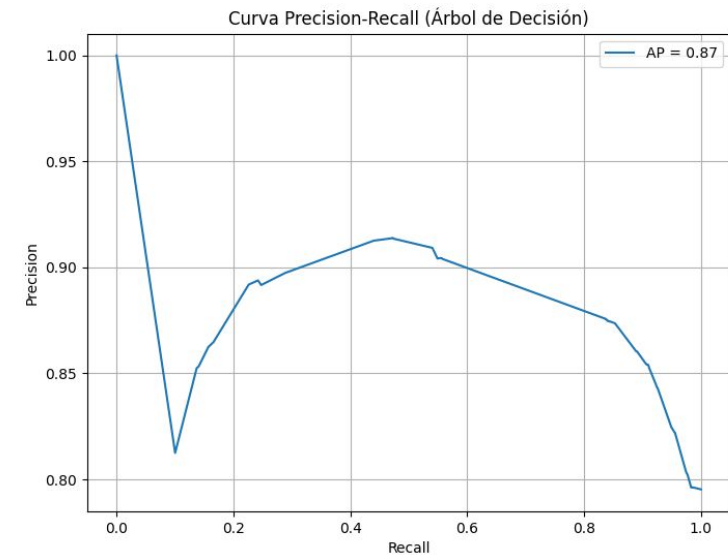
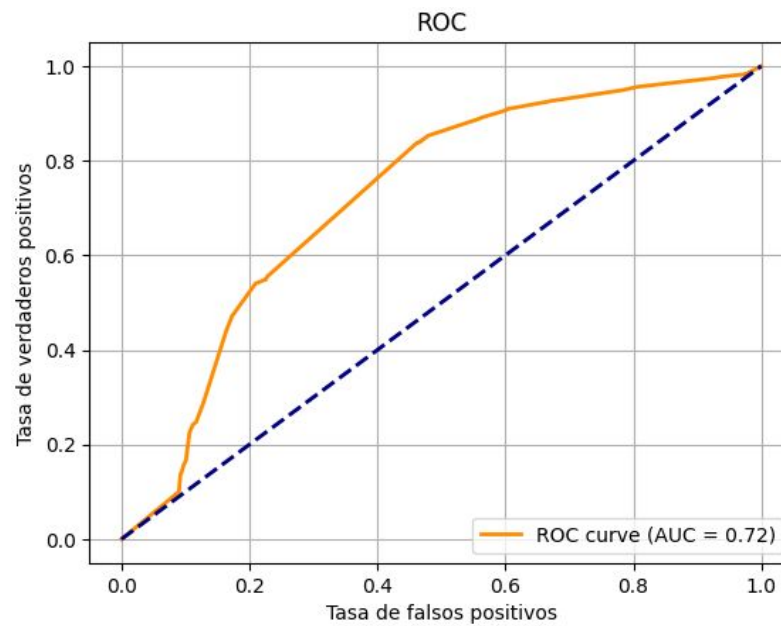
3. DESARROLLO_ML, DL y NLP

>>ML: Naive Bayes



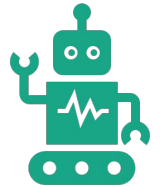
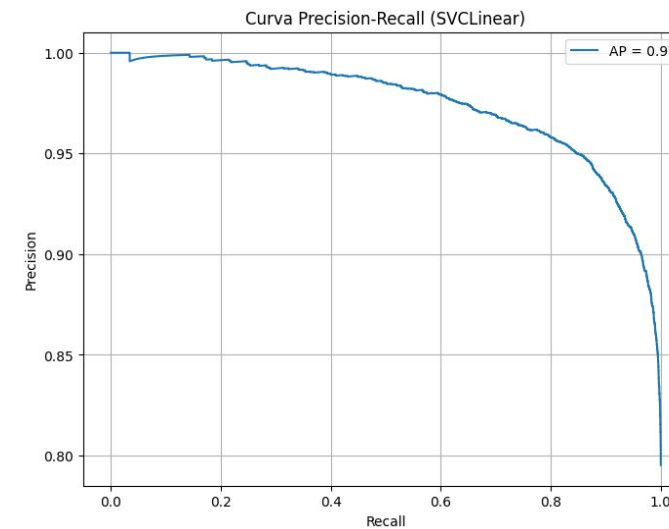
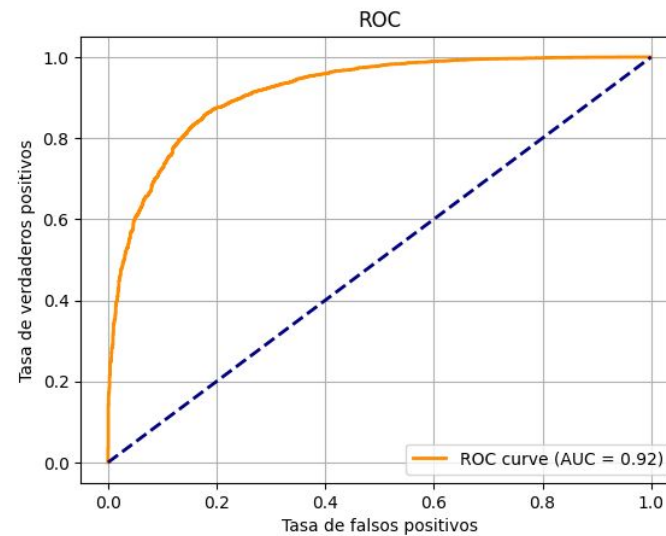
3. DESARROLLO ML, DL y NLP

>>ML: Árbol de Decisión



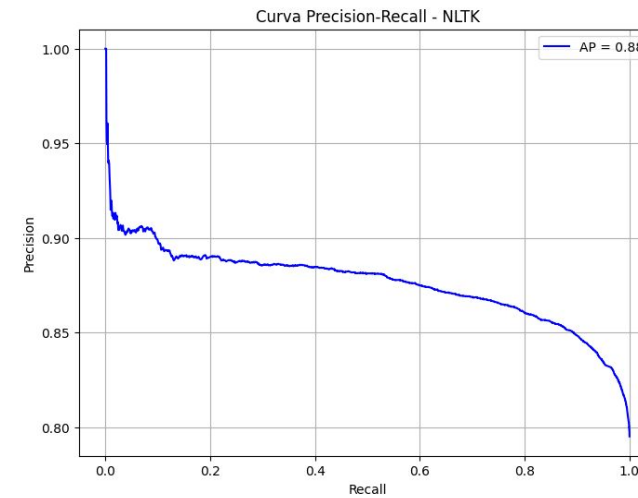
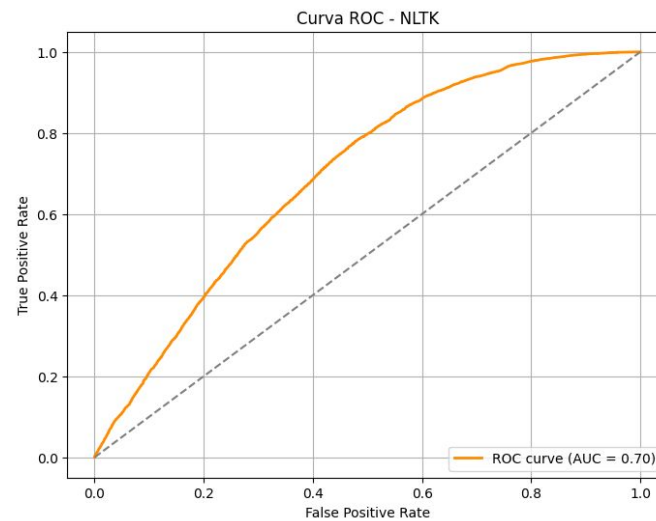
3. DESARROLLO_ML, DL y NLP

>>DL_NLP: SVC Linear



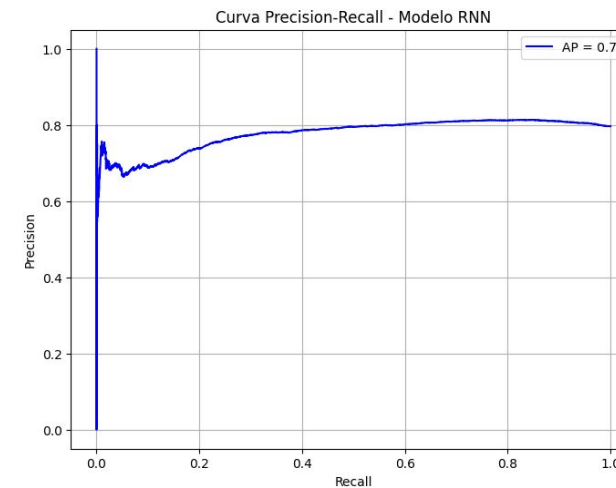
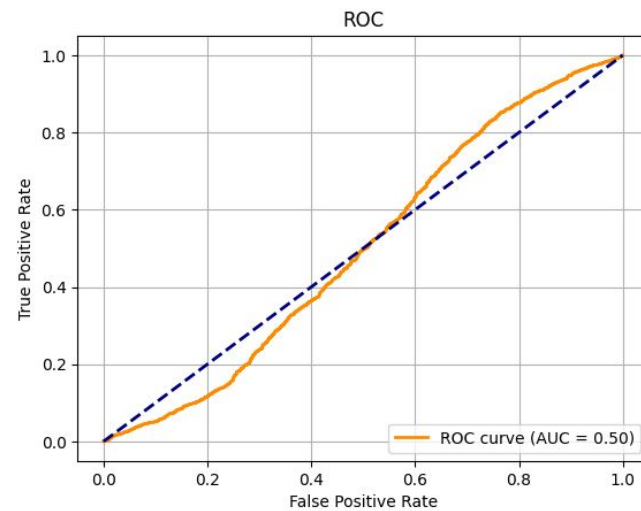
3. DESARROLLO_ML, DL y NLP

>>DL_NLP: Natural Language Toolkit



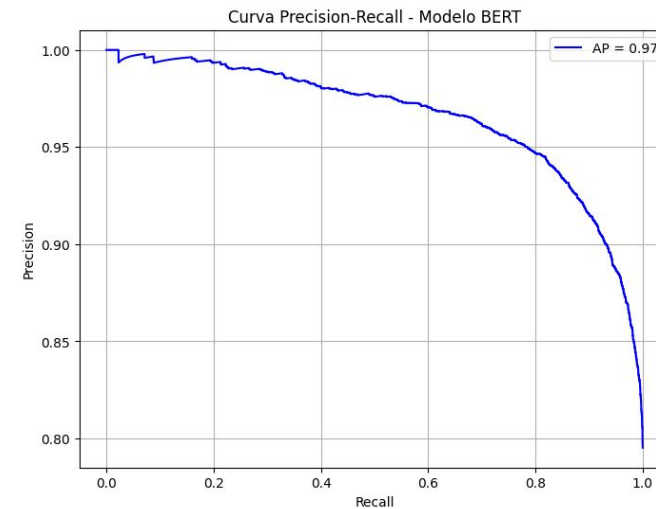
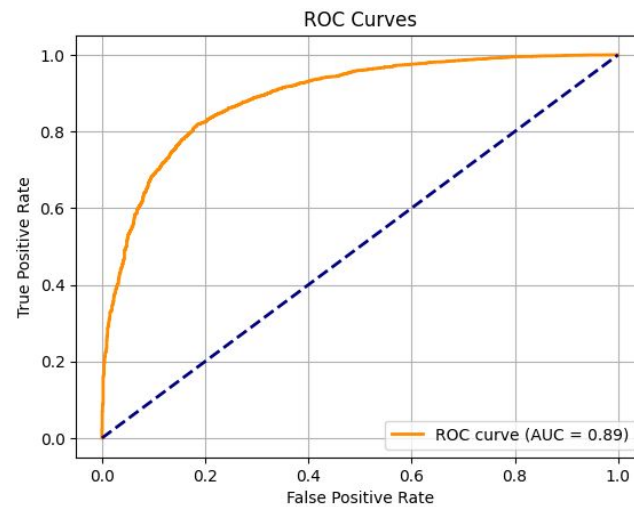
3. DESARROLLO_ML, DL y NLP

>>DL_NLP: Recurrent Neural Network (RNN)



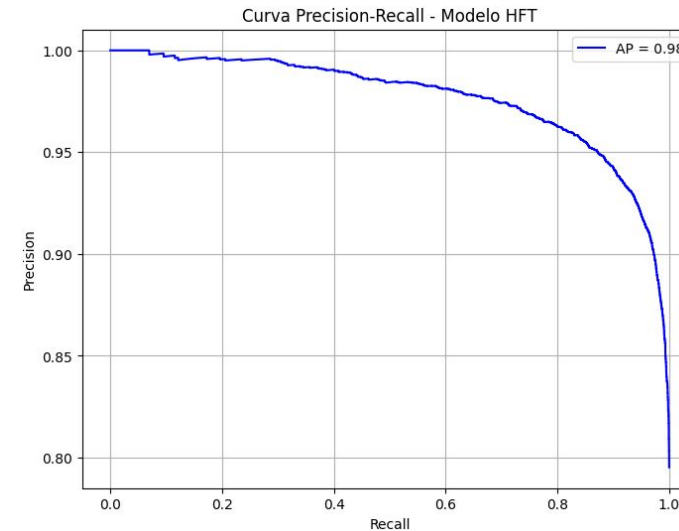
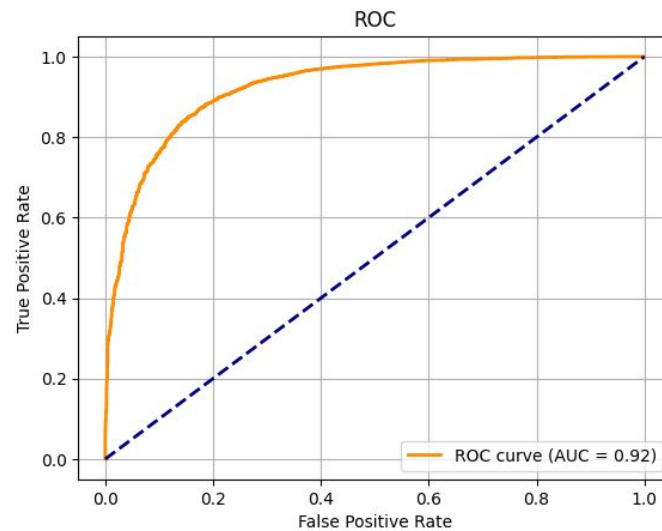
3. DESARROLLO_ML, DL y NLP

>>DL_NLP: BERT



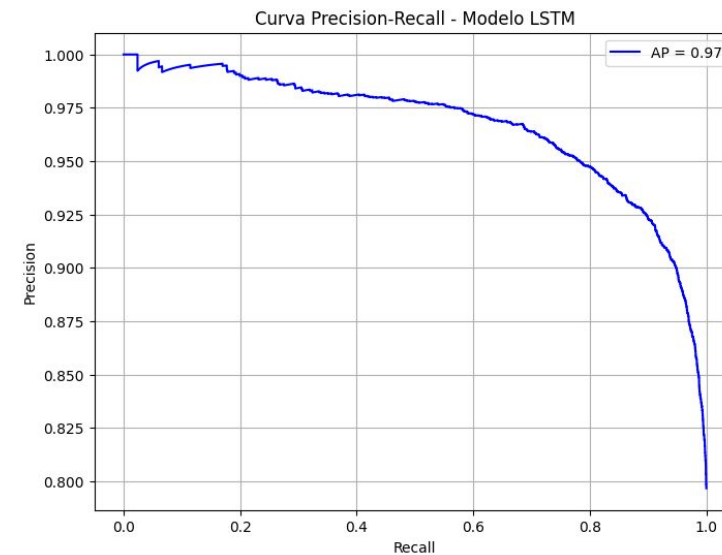
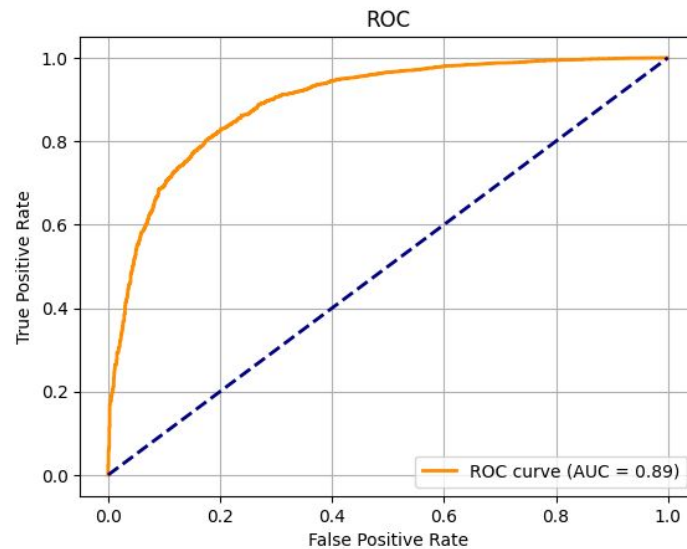
3. DESARROLLO_{ML}, DL y NLP

>>DL_NLP: Hugging Face Transformers (DistilBERT)

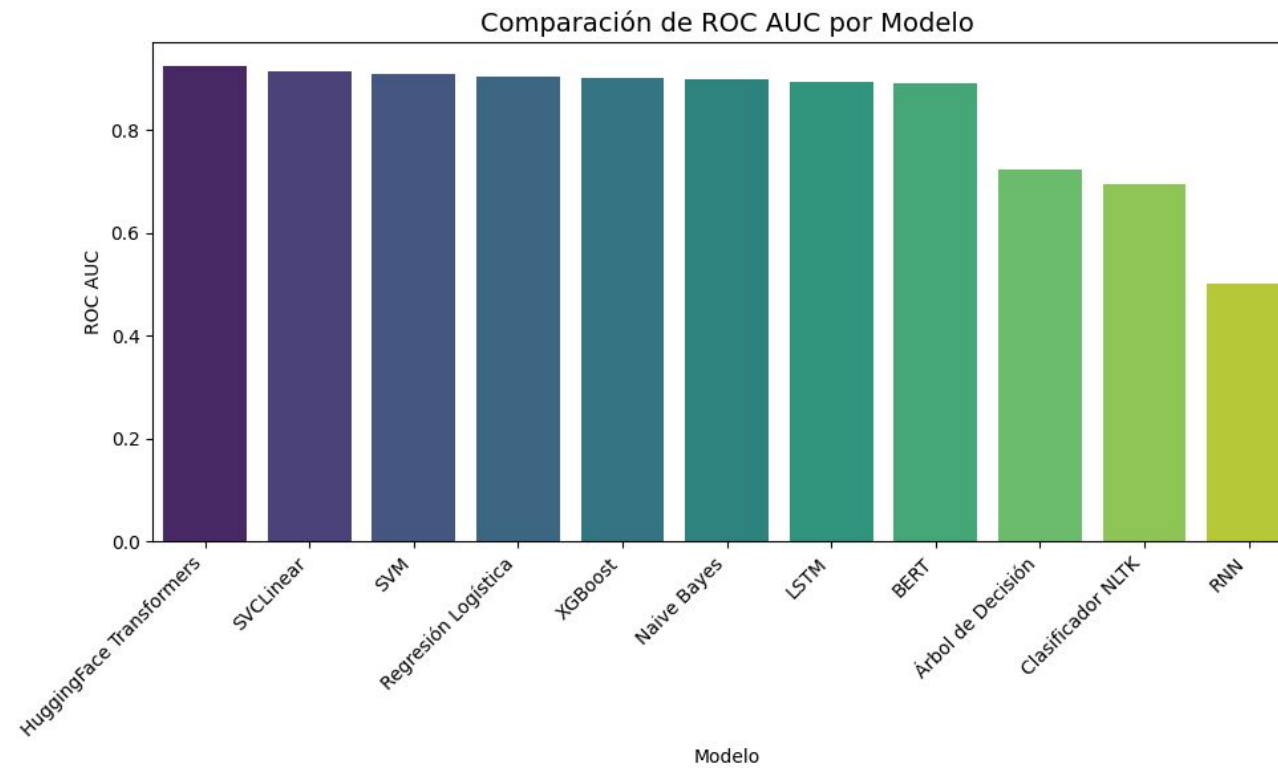


3. DESARROLLO ML, DL y NLP

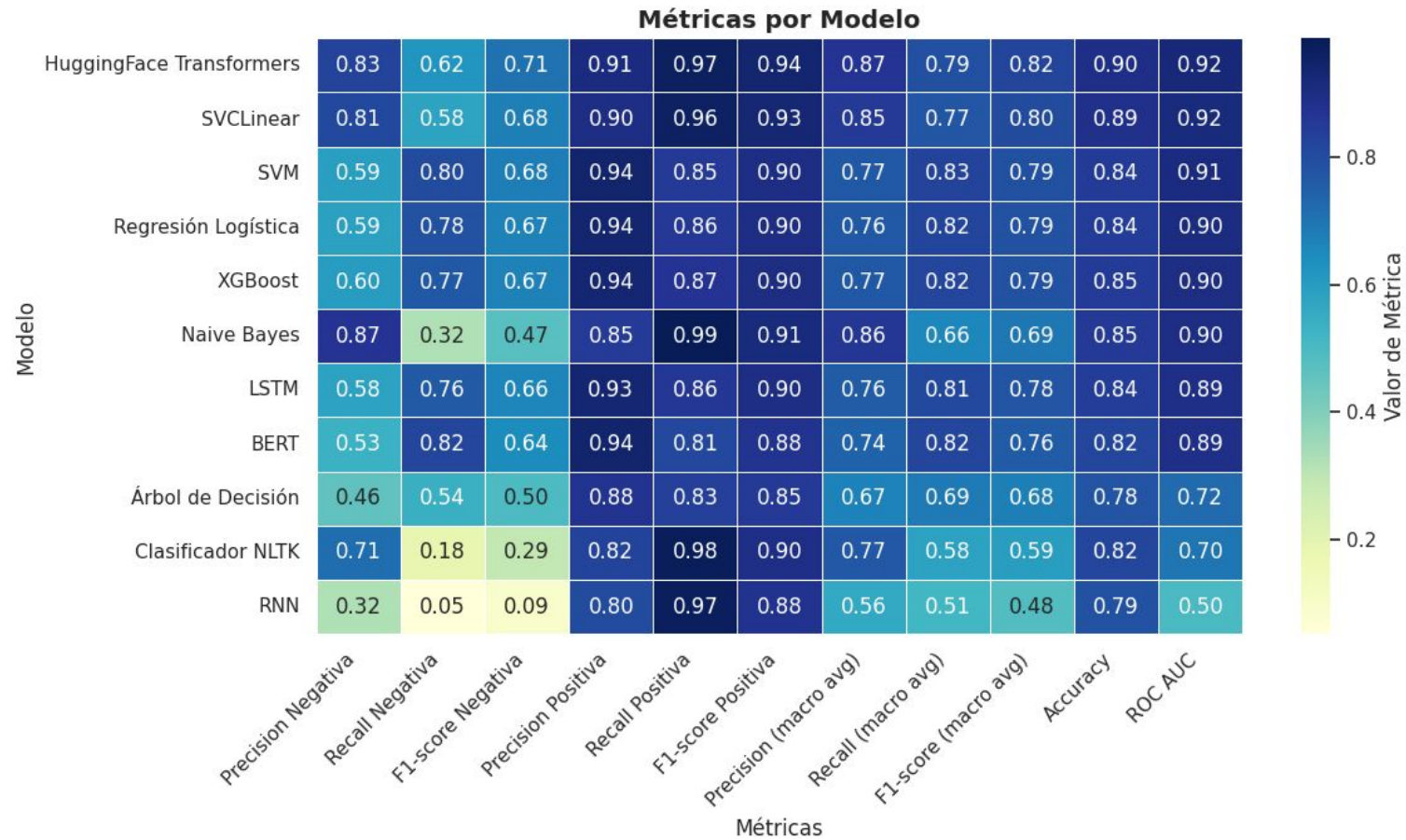
>>DL_NLP: Long Short-Term Memory (LSTM)



4. PRUEBAS Y RESULTADOS



4. PRUEBAS Y RESULTADOS



5. CONCLUSIONES



> CONCLUSIÓN_1

El EDA permitió comprender mejor el comportamiento de los datos e identificar patrones en la distribución temporal y geográfica.



> CONCLUSIÓN_2

Mejor modelo predictivo Hugging Face Transformers (SVCLinear y SVM mostraron buenos resultados).



> CONCLUSIÓN_3

Se recomienda recolectar más datos de la clase negativa para mejorar la predicción.



> CONCLUSIÓN_4

Se recomienda utilizar técnica *Ensemble* para generar sinergias entre los modelos.

IMMUNE
TECHNOLOGY INSTITUTE

Muchas gracias a IMMUNE por
la oportunidad_

<https://immune.institute>