

# Identificación de Objetivos de Desarrollo Sostenible

---

JUAN CAMILO BELTRÁN

LEIDY LOZANO

ANDRÉS REYES

INTELIGENCIA DE  
NEGOCIOS

ISIS 3301



# Enfoque Analítico

---



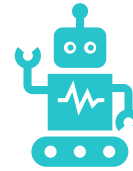
Predictivo – Determinar  
ODS con el texto dado



Aprendizaje Supervisado  
– Datos etiquetados



Clasificación – Clasificar  
entradas de texto en un  
ODS



Algoritmos: Árbol de  
Decisión, KNN y Bayes  
Ingenuo

# Entendimiento de Datos

---

4049 registros en total

Cada fila del archivo es un comentario o fragmento de texto hecho con relación a los ODS 3, 4 o 5

Unicidad – No se encuentran registros duplicados

Compleitud – Todos los registros estaban completos

Consistencia – No hay columnas que puedan afectar la consistencia

Validez – Se eliminaron algunos registros que estaban en inglés y en francés

# Preparación de Datos

---

## Limpieza

Remoción de caracteres no ASCII

Conversión a minúsculas

Remoción de *stopwords*

Remoción de puntuación

## Normalización

Eliminación de prefijos y sufijos con Stemmer de NLTK Snowball

## Vectorización

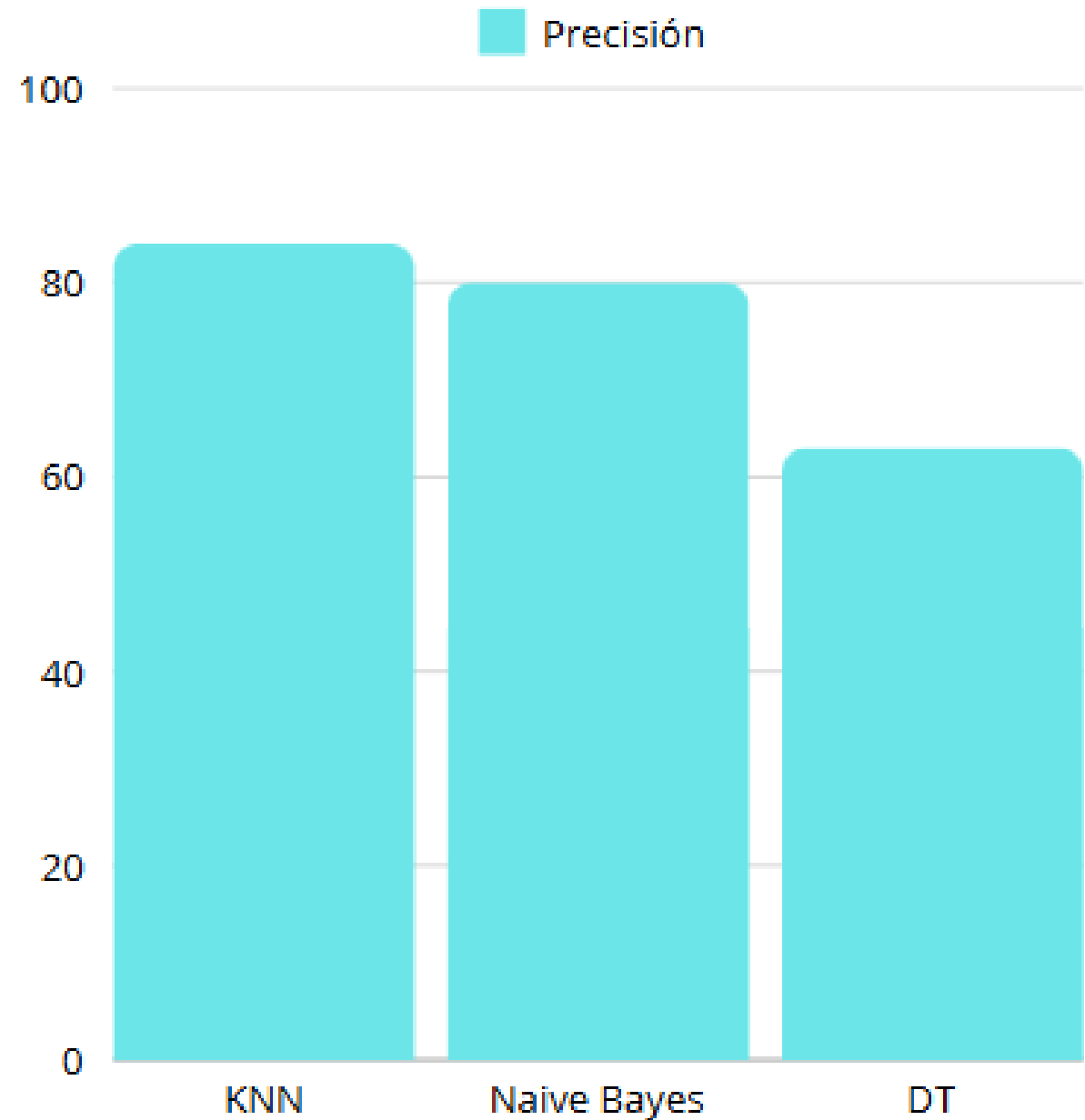
Uso de Bag of Words de la librería SKLearn

# Resultados - Precisión

---

KNN y Bayes tienen precisión igual o mayor a 80%

Sin embargo, es necesario ver más métricas para concluir correctamente

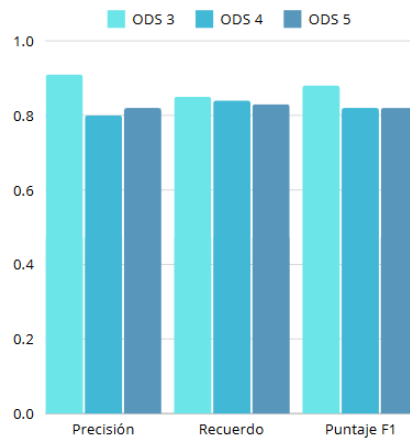


# Resultados - Comparación de Métricas

---

Los resultados más comparables son KNN y Bayes, puesto que DT da métricas bajas en comparación. Además, Bayes disminuye sus puntajes en el ODS 5. KNN resulta ser más estables con sus métricas más bajas.

**Métricas de KNN**



**Métricas de Naive Bayes**



**Métricas de Arbol de Decisión**

