




# **Clasificación de textos según los ODS**

Daniel Arango Cruz · Javier Cerino Pulido · Marco Zuliani Escalante



# Tabla de contenidos

01 →



**Introducción**

02 →



**Técnicas**

03 →



**Modelos**

04 →



**Resultados**

05 →



**Roles**

06 →



**Conclusiones**



# 01 **Introducción**



# Objetivo

- Apoyar al **Fondo de Poblaciones de las Naciones Unidas (UNFPA)** que, junto con entidades públicas, busca identificar problemas y evaluar soluciones utilizando herramientas de participación ciudadana.
- Desarrollo de un **modelo de clasificación**, con técnicas de aprendizaje automático, que permita relacionar de manera automática un texto según los ODS.



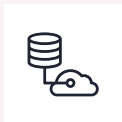


## 02 **Técnicas**



# Técnicas de analítica de textos

Implementadas por la librería de *sklearn*



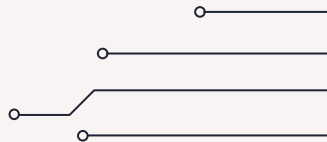
## BoW

Bag of Word (BoW) determina la relevancia de cada palabra realizando un conteo de la frecuencia de aparición de cada palabra en todos los textos analizados.



## TF-IDF

Text Frequency - Inverse Document Frequency (TF-IDF) realiza una media numérica que expresa la relevancia de cada palabra según la cantidad de veces que aparece en cada texto y el número de textos en los que aparece.





# 03 Modelos



# HERE ARE THREE IMPORTANT IDEAS



**Random Forest  
+ BoW**



**Random Forest  
+ TF-IDF**

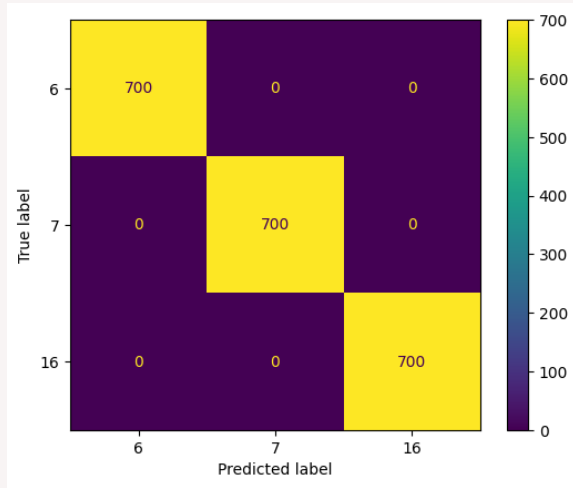


**Logistic Regression  
+ TF-IDF**



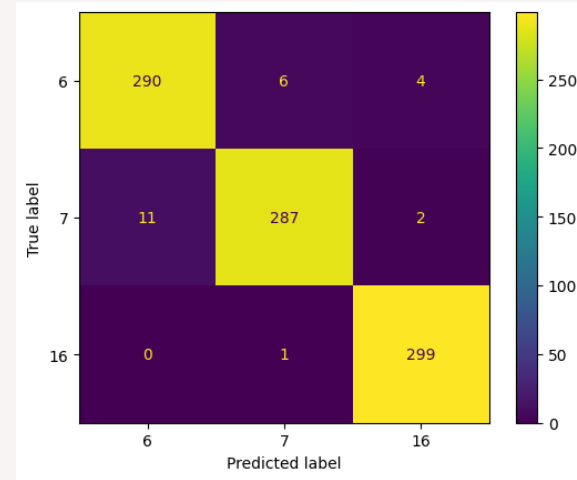
# Random Forest + BoW

Train



Precision	Recall	F1-score
1	1	1

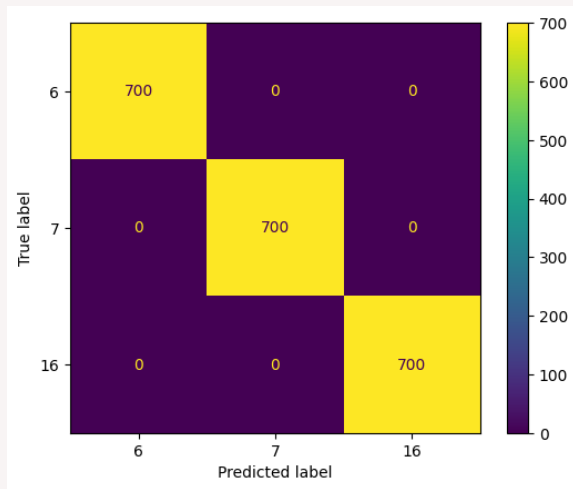
Test



Precision	Recall	F1-score
0.97332449..	0.97333333..	0.97327265..

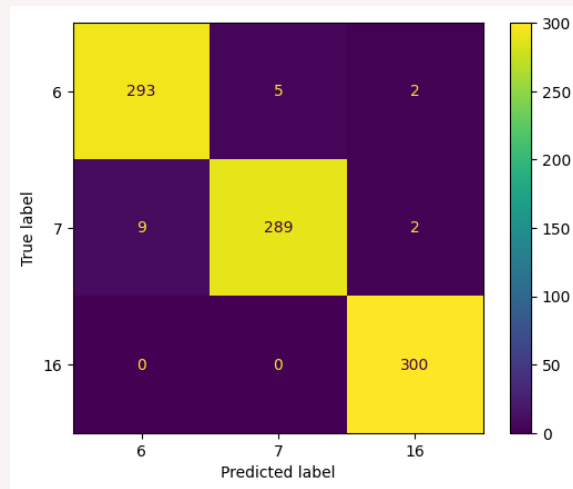
# Random Forest + TF-IDF

Train



Precision	Recall	F1-score
1	1	1

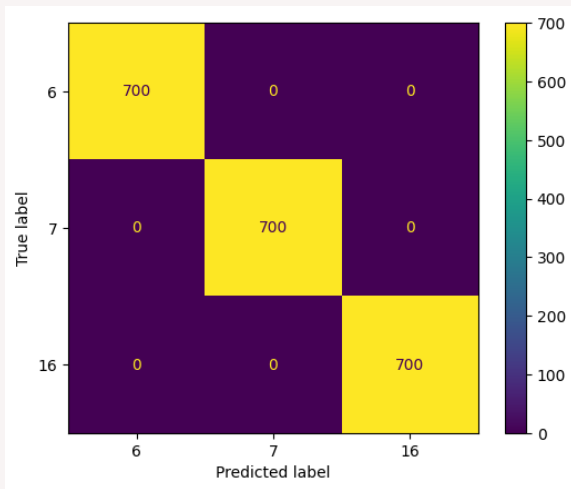
Test



Precision	Recall	F1-score
0.98001132..	0.98	0.9799544...

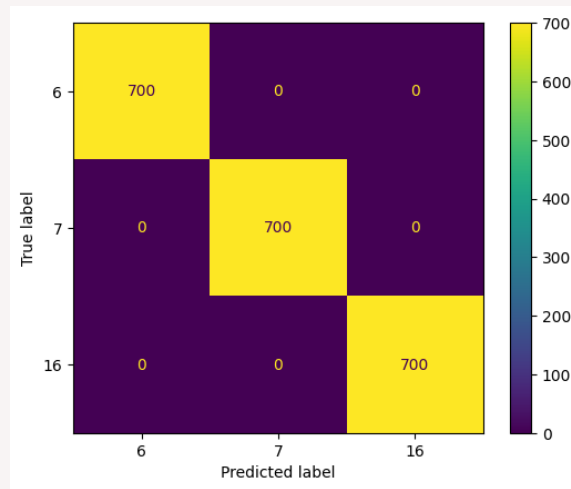
# Logistic Regression + TF-IDF

Train



Precision	Recall	F1-score
0.99571223..	0.99571428..	0.99571224..

Test



Precision	Recall	F1-score
0.98453546..	0.98444444..	0.98441935..



# 04 Resultados

# Comparación

	Train			Test		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Random Forest + BoW	1	1	1	0.97332449..	0.97333333..	0.97327265..
Random Forest + TF-IDF	1	1	1	0.98001132..	0.98	0.9799544...
Logistic Regression + TF-IDF	0.99571223..	0.99571428..	0.99571224..	0.98453546..	0.98444444..	0.98441935..

- El modelo que da mejores predicciones en los modelos es TF-IDF.
- Entre el algoritmo de Random Forest y Logistic Regression, el que da mejores predicciones es Logistic Regression aumentando en un 0.4% todas las métricas analizadas en el Testing, por este motivo se eligió este algoritmo con los hiperparámetros establecidos para realizar las predicciones de los datos no etiquetados.



# 05 **Roles**



# Beneficio – Riesgo



→ **Analista de  
textos de  
UNFPA**

**Beneficio:** Acelera clasificación para informes y proyectos potenciales.

**Riesgo:** Mala predicción afecta trabajo y recursos incorrectos, desorganización.



→ **Trabajador  
de la ONU**

**Beneficio:** Ahorro de costos y tiempo, mejora perspectiva en proyectos.

**Riesgo:** Fallo del modelo derrocha inversión, posible contratación más efectiva.



→ **Comunidad  
publicadora  
del artículo.**

**Beneficio:** Facilita retroalimentación y apoyo eficiente de ONG y entidades.

**Riesgo:** Demora en financiación por proyectos mal clasificados, expertos escasos.



# 06 **Conclusiones**



# Conclusiones



Técnica analítica de textos  
TF-IDF



Algoritmo Logistic  
Regression



Clasificación con precisión de  
0.9845354622503447



**Gracias**