

# Clasificación de Opiniones Ciudadanas en ODS



- Ximena Lopez - 202312848
- Juliana Ferreira - 202312785
- María Juliana Ballesteros - 202313216

# Objetivo del proyecto

Desarrollar un modelo de aprendizaje automático capaz de clasificar opiniones ciudadanas en los ODS 1, 3 y 4 de manera automática, precisa y balanceada, apoyando el análisis de grandes volúmenes de información para la toma de decisiones.

- ODS 1 - Fin de la pobreza: Busca erradicar la pobreza extrema y reducir la pobreza en todas sus formas en el mundo.
- ODS 3 - Salud y bienestar: Garantizar una vida sana y promover el bienestar para todo.
- ODS 4 - Educación de calidad: asegurar una educación inclusiva, equitativa y de calidad.



Sobre nosotros

# Perfilamiento de datos

01

## Estructura del data set

Dos columnas

- **Textos:** Contiene las opiniones de la ciudadanía
- **Labels:** Indica el ODS al que pertenece la opinion

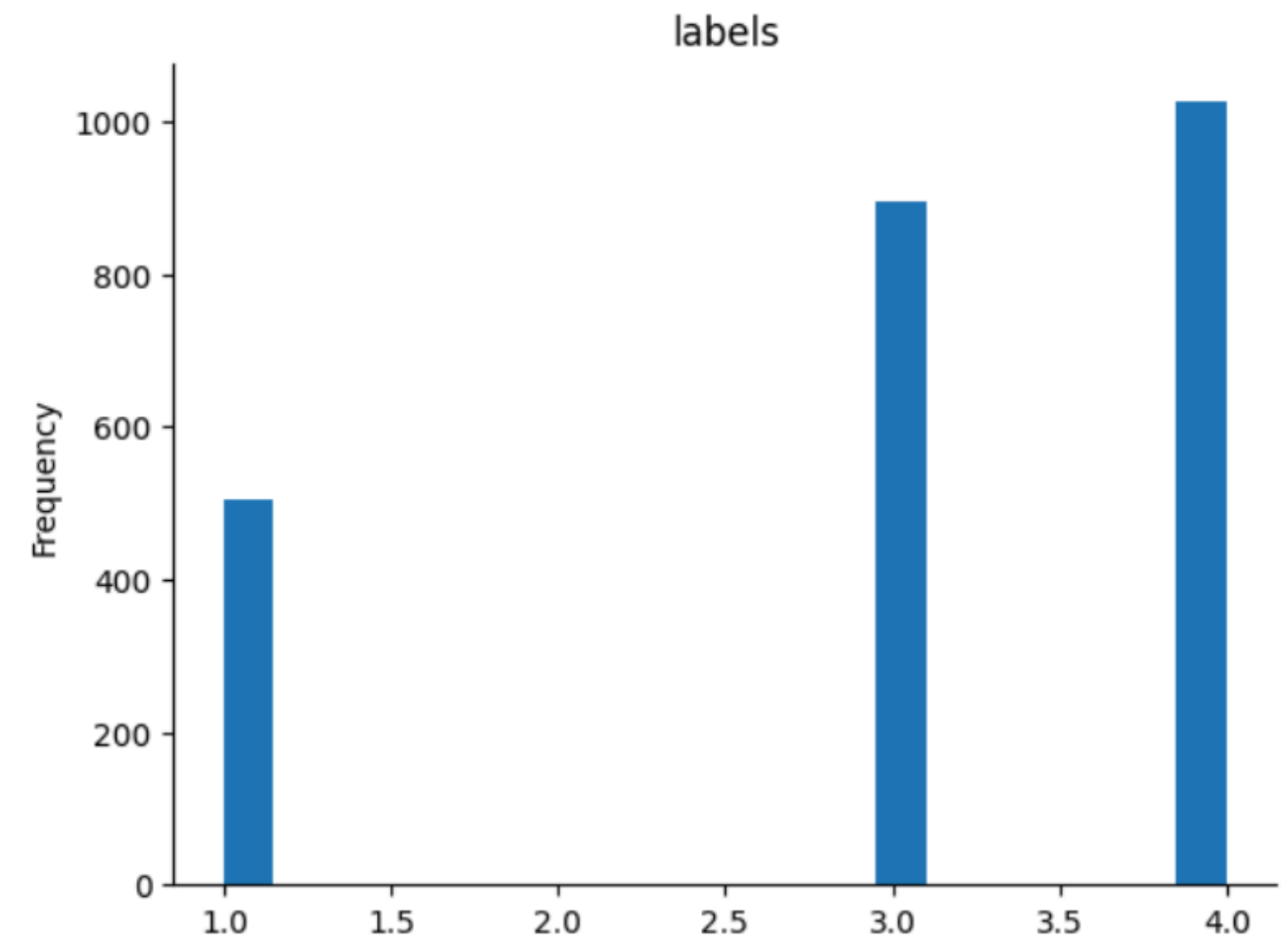
02

## Distribución de clases

Como se observa en el histograma, la **clase 4** tiene mayor concentración que la **clase 1**.

03

- **Longitud promedio:** ~500 palabras (textos medianos)
- **Datos completos:** sin nulos ni duplicados



Sobre nosotros

# Limpieza de datos

- **Limpieza:** Se paso todo a minúsculas, eliminación de acentos, puntuación y espacios extra.
- **Tokenización:** división en palabras.
- **Stopwords:** eliminación de palabras vacías (ej. el, la, de).
- **Vectorización:** representación numérica con TF-IDF.

## Texto original

"La Educación es clave para el futuro"



## Texto limpio y vectorizado

["educacion", "clave", "futuro"] → [0.21, 0, 0.08, ...]

# Modelos probados

## Regresión logística

Modelo lineal que estima la probabilidad de pertenencia a cada clase. Se adapta muy bien a datos de texto con TF-IDF y logra alta precisión.

## K-Vecinos

Clasifica un texto comparándolo con sus vecinos más cercanos en el espacio vectorial. La categoría más común entre esos vecinos es la asignada.

## Naive bayes

Se basa en la probabilidad de que un texto pertenezca a una clase, asumiendo independencia entre palabras. Es simple, rápido y funciona bien en texto.

# Resultados de cada modelo



## Regresión Logística

F1-Score: 0.9622.

Modelo más robusto y equilibrado, con resultados sobresalientes en todas las clases.



## K-vecinos

F1-Score: 0.945882

Aunque su desempeño es ligeramente inferior al de la Regresión Logística, KNN ofrece resultados consistentes y confiables, siendo una alternativa válida.

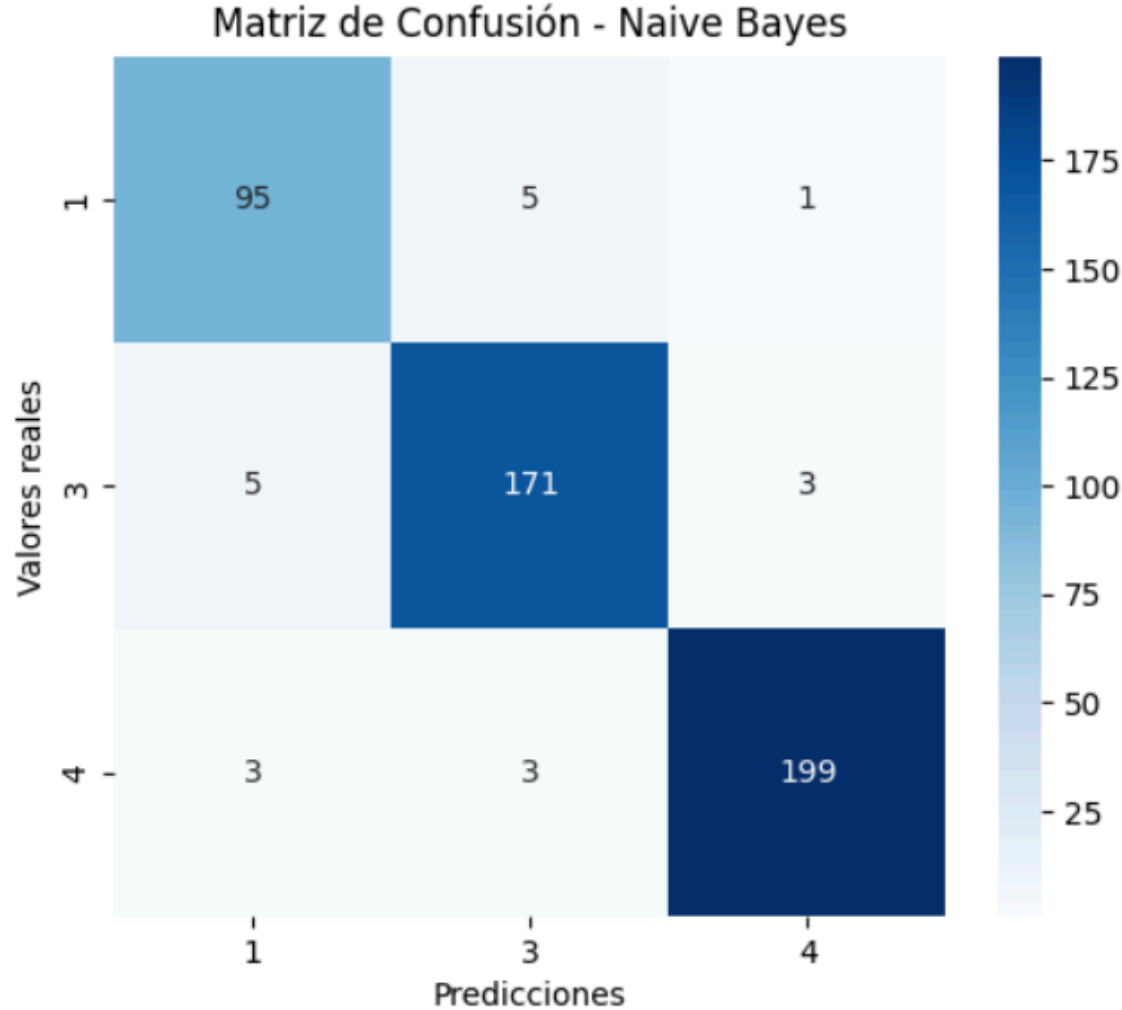
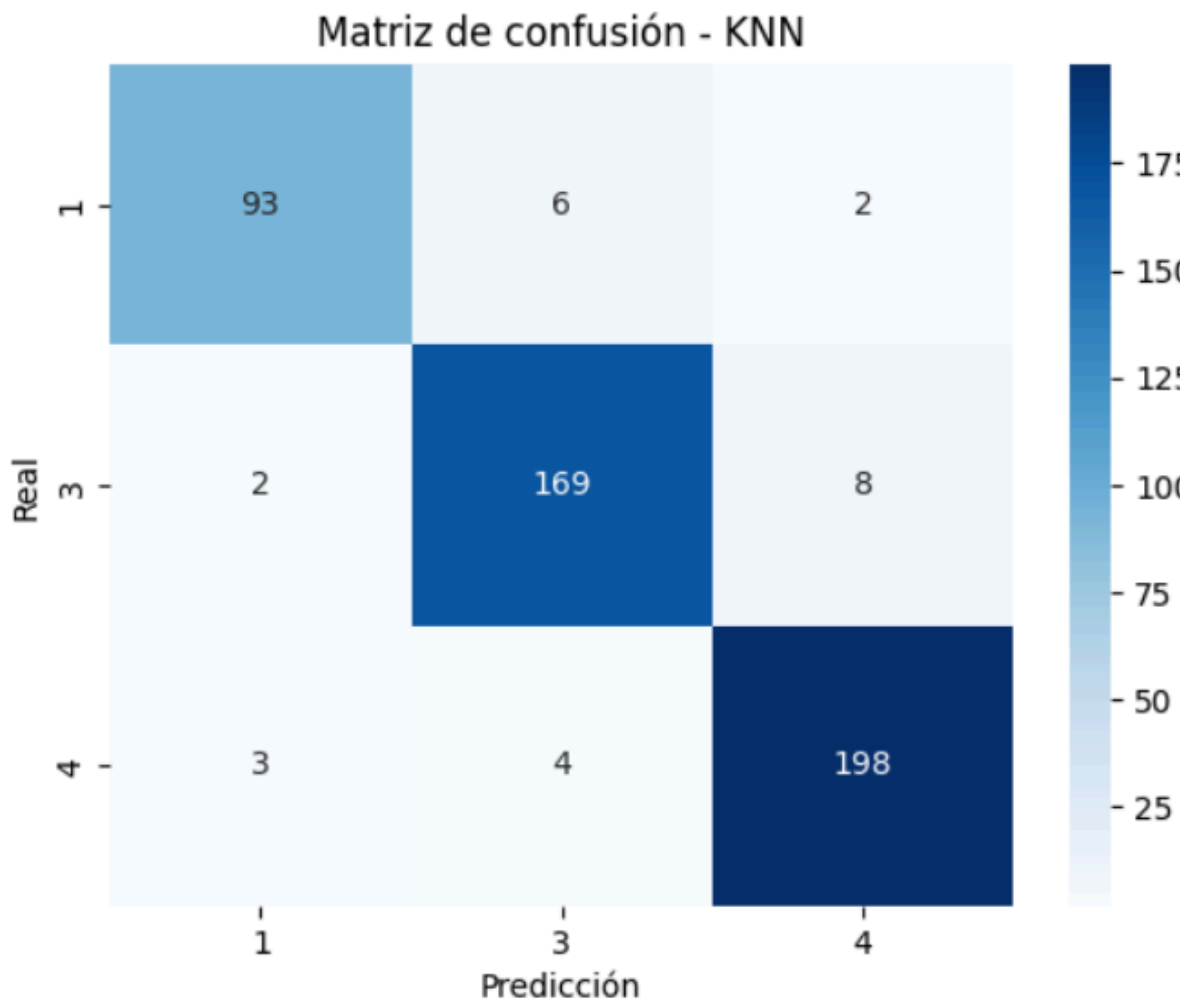
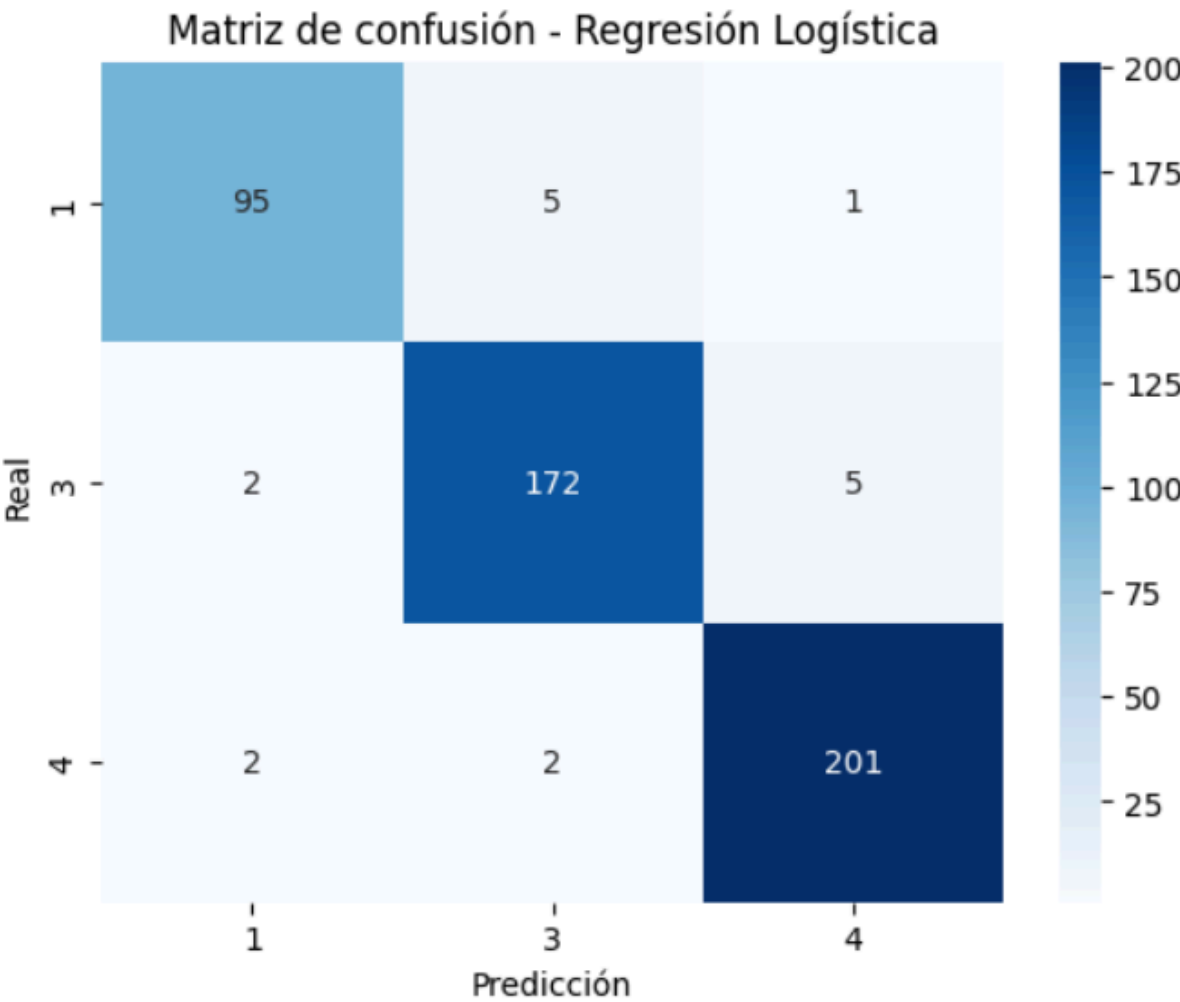


## Bayes Ingenuo

F1-Score: 0.954057

Modelo simple pero muy eficaz, con predicciones consistentes y pocos errores incluso en la clase menos representada

# Modelos probados



# Modelos probados

Modelo	Exactitud	Precisión (macro)	Recall (macro)	F1 (macro)	Observaciones
Regresión Logística	~96,5%	~96.5%	~96%	0,962	Mejor desempeño general. Balanceado y robusto, pocos errores entre ODS1 y ODS3.
Naive Bayes	~95%	~95%	~95%	0,954	Simple y eficiente. Muy buen rendimiento, especialmente en ODS4 (Educación).
KNN	~94-95%	~94%	~94%	0,946	Estable y consistente, sin sesgo hacia una clase. Ligeramente inferior a Logística.



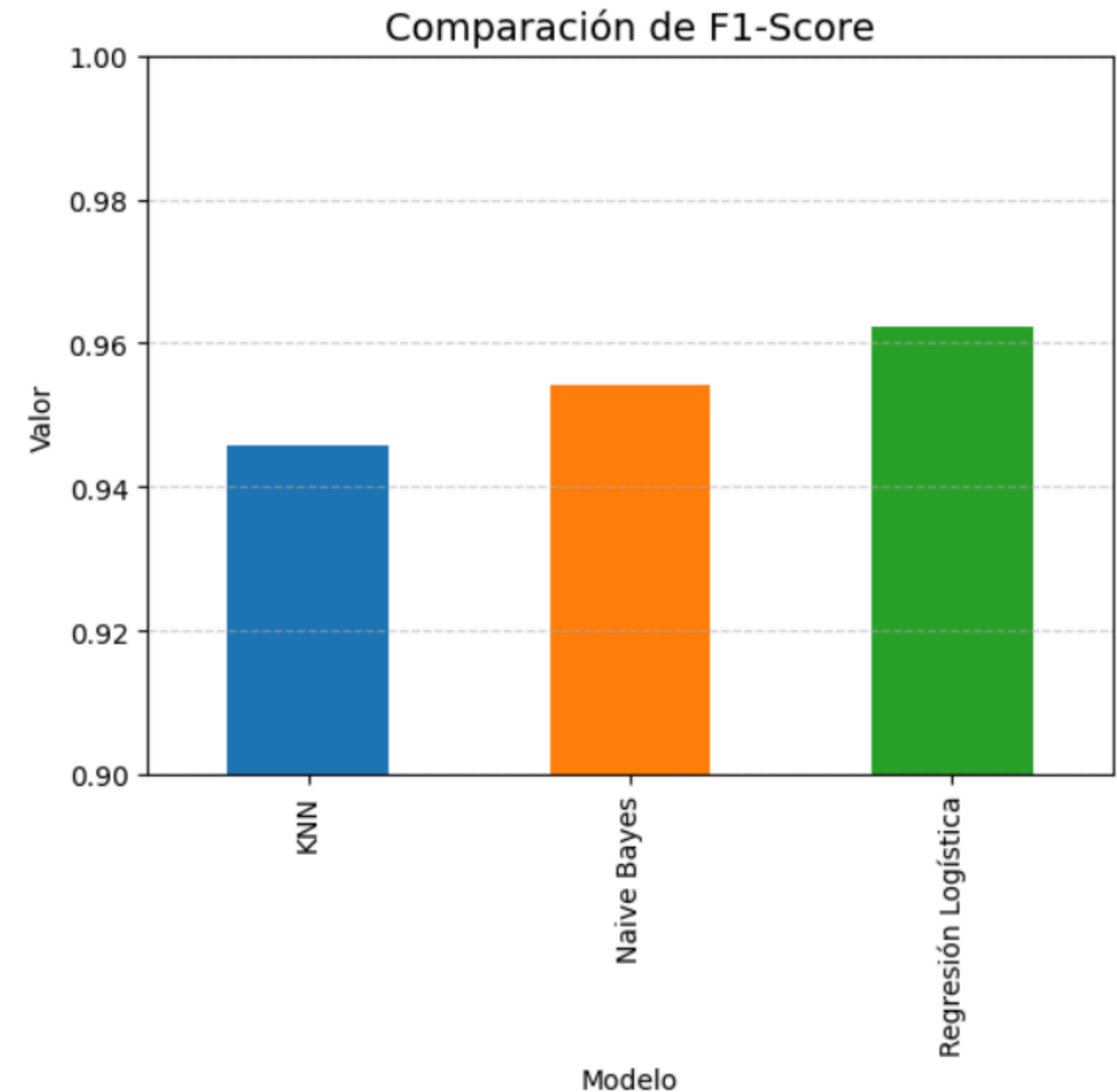
Modelo seleccionado

# Regresión logística

Mostró un equilibrio sólido entre las clases ODS 1, 3 y 4, evitando sesgos hacia la clase mayoritaria.

Logró el mayor desempeño en todas las métricas (Exactitud, Precisión y F1-Score).

Es un modelo interpretativo y confiable, lo que permite a la organización entender los patrones lingüísticos detrás de cada predicción.



# Intuición del modelo y estrategias

## Palabras representativas por ODS

### ODS 1 – Fin de la pobreza:

- pobreza, ingresos, empleo, economía, oportunidades

### ODS 3 – Salud y bienestar:

- salud, hospital, médico, enfermedad, bienestar

### ODS 4 – Educación de calidad:

- educación, aprendizaje, escuela, profesor, conocimiento

## Monitoreo en tiempo real

- Uso del modelo para clasificar opiniones al instante y detectar tendencias emergentes (ej. salud, educación).

## Campañas dirigidas

- Identificar palabras clave (ej. empleo, pobreza) y enfocar acciones y programas en necesidades reales.

# Conclusiones



En conclusión, nuestro mejor modelo fue Regresión Logística, con un F1-macro de 0.962, demostrando ser el más confiable para la clasificación de opiniones ciudadanas en los ODS 1, 3 y 4.

Desde la perspectiva del negocio, estos resultados significan que la organización ahora cuenta con una herramienta robusta y automática para procesar grandes volúmenes de opiniones ciudadanas, reduciendo tiempos de análisis y facilitando la toma de decisiones estratégicas. En una siguiente etapa, este modelo puede integrarse en una plataforma web de monitoreo en tiempo real, permitiendo identificar tendencias emergentes y responder de forma ágil en temas críticos como pobreza, salud y educación.

Gracias!