

8

# Natural Language Classifier of Ternium NLC Ternium

DATA MANIACS

Profesores:

- Rubi Gutierrez
- Angelina Alarcon

Desarrollador web



Valeria Lugo

Data scientist



Data Scientist

Sabrina Carsellé

Data Engineering



Donnet Hernández

PM



Pablo Monzón

CDO



Yuu Akachi

Data Engineering



Juan Sala

Diseñador UX



Itzel Beltrán





# Introducción

Ternium es un Centro Industrial productor de aceros, los cuales ofrecen sus productos a millones de mexicanos para el uso en sus hogares. Es por esto, que sus productos y maquinarias pueden llegar a tener algunos errores o defectos. Para esto es que registran miles de datos con los avisos de falla, así como también, llevan el registro de lo que ha salido mal y cómo se podría resolver.



## Problemática

- Datos no ordenados ni estructurados correctamente.
- Registros con faltas de ortografía, de información, valores nulos y renglones sin llenar.
- Difícil sacar conclusiones, procesar los datos y analizarlos.
- La base no proporciona información significativa - fallas de los equipos.

## Solución

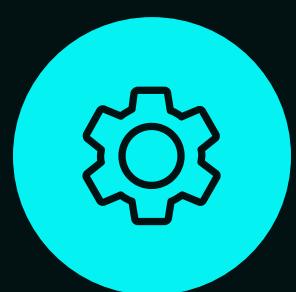
- Propuesta de modelo de clasificación de datos.
- Tomar en cuenta las variables SÍNTOMA y CAUSA.
- Usar métodos de limpieza y transformación de datos.

## Objetivo

- Mejorar la organización, para facilitar el análisis de los procesos.
- Implementación de tecnologías y métodos como el ML, que contribuyan en una mejor toma de decisiones.

# Datos proporcionados por Ternium

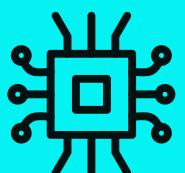
La base de datos contiene 64,719 registros y cuenta con 16 variables.



Categóricas  
**15**



Numéricas  
**1**



Se realizó un proceso de limpieza y transformación para poder aplicar el modelo



# Hipótesis



## Previo al modelo:

"Si realizamos una clasificación eficiente sobre las variables SÍNTOMA y CAUSA utilizando técnicas de aprendizaje supervisado y no supervisado, árboles de decisión y algoritmos NLPs, nos ayudarán a reducir el ruido, dimensionalidad y clasificar datos que serán de ayuda para Ternium".

## Posterior al modelo:

"Se lograron obtener modelos tanto de aprendizaje supervisado como no supervisado y de NLP (Naive Bayes, Regresión Logística, SVM, Random Forest, KNN y K-means, tomando como variables target SÍNTOMA y CAUSA, es por esto que se confirma la hipótesis, añadiendo que dos de los modelos tuvieron un accuracy de 61% ".

# Modelación

Diferentes algoritmos utilizados para la modelación



**KNN**

Accuracy: 52%



**Random  
Forest**

Accuracy previo: 61%

Accuracy  
hiperparámetros: 86%



**K-Means**

Clusters: 35



**SVM**

Accuracy: 45%



**Naive  
Bayes**

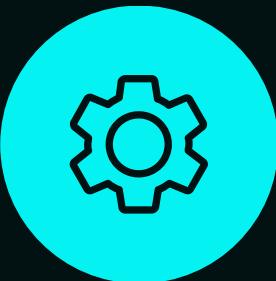
Accuracy: 61%



**Logistic  
Regression**

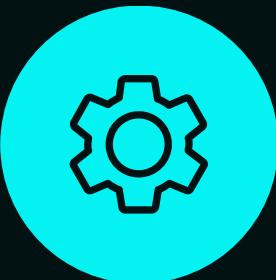
Accuracy: 57%

# Modelo elegido



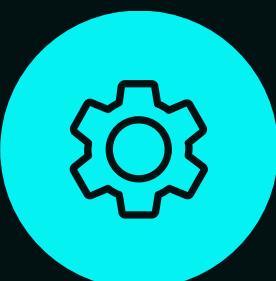
## Previo al ajuste de hiperparámetros

El modelo tenía un accuracy del 61%



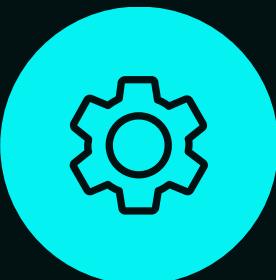
## Ajuste de hiperparámetros

```
{'rf_max_depth': None, 'rfmin_samples_split':  
5, 'rfn_estimators': 200, 'tfidf_max_df': 0.5}
```



## Posterior al ajuste de hiperparámetros

Mediante **grid\_search** se obtienen los mejores parámetros a utilizar en el modelo



## Accuracy obtenido

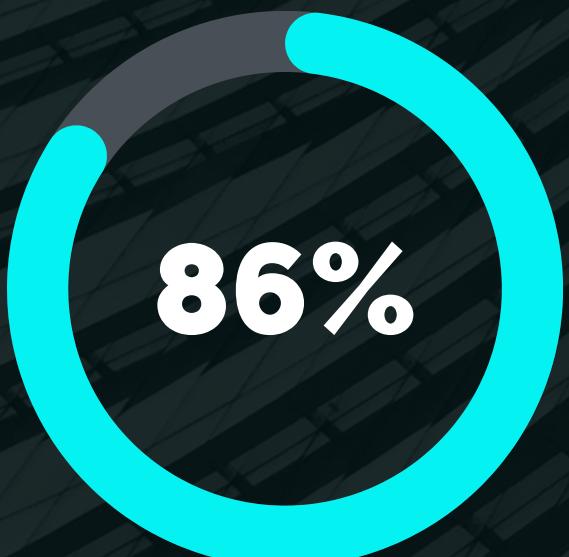
El modelo arrojó un accuracy de 86%



# Resultados

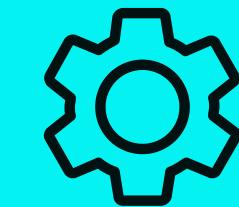


Accuracy:



## Criterio de éxito de la empresa

“Interpretar y clasificar automáticamente textos compuestos en categorías definidas de un proceso sin tener un set de datos inicialmente clasificados”



**Aplicación al negocio**

El modelo tiene un gran impacto y utilidad considerando los criterios de éxito de Ternium.

## Clasificacion para Ternium

Metodos de clasificacion para avisos en la fabrica

Introduce tu descripcion:

Introduce que paso:

Introduce por que paso:

Introduce que se hizo:

Clasificar

Registrar

Revisar 5 sintomas mas frecuentes

Los top 5 sintomas son:

# Prototipo funcional

## Web app

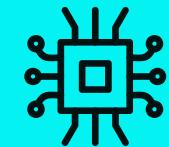
A través de streamlit y anaconda, se pudo realizar un prototipo funcional de un modelo de aplicacion, el cual permitirá a los trabajadores de ternium, introducir las variables independientes de forma que se pueda hacer la clasificacion correcta.

## Demostración



# Recomendaciones

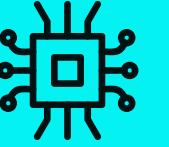
Algunas de las recomendaciones para Ternium



Columnas más descriptivas y significativas.  
Cuidar ortografía



Nombre de variables más cortos pero informativas



Desarrollar un glosario para el mejoramiento en la identificación y entendimiento de variables



Buscar técnicas de recopilación de los datos para disminuir el tiempo de preparación de datos



# Agradecimientos

- Agradecemos a Ternium por la oportunidad de realizar este proyecto, donde pudimos desarrollar habilidades que generen un impacto en el mundo laboral.
- Agradecemos a la profesora Angelina Alarcón y Rubí Gutiérrez por su guía durante el desarrollo del proyecto.
- El desarrollo del proyecto permitió comprender sobre la implementación del clasificador de lenguaje natural y modelos de predicción en el mundo laboral