

Estudiantes y roles

- Leonardo Laureles Olmedo
A01659241
- Mariana Rincón Flores
A01654973
- Carlos Mateos Pérez
A01654085
- Daniel Núñez López
A01654137
- Juan Manuel Cantú
A01284053
- PM y Científico de
datos
- Analista de datos y
Científico de datos
- Desarrollador web y
Científico de datos
- CDO y Científico de
datos
- Ingeniero de datos y
Científico de datos

Profesoras

- Angelina Alarcón
- Rubi Gutiérrez

Reto: Modelo predictivo para disminuir caídas y retención de materiales recubiertos en base a indicadores de comportamiento de cada equipo.



Introducción

Problemática

Se busca predecir los defectos probables con los que sale el material en línea de Recubiertos.

Solución propuesta

Armado de propuesta para determinación de un predictivo de Defecto en base a indicadores de comportamiento de cada equipo.

Objetivo

Tener un modelo matemático basado en Machine Learning que pueda identificar las causas de los defectos más significativos con base en los datos brindados por parte de la empresa para lograr líneas de producción más eficientes, mitigando los errores y dando solución a las posibles causantes de los mismos.

●● Exploración y visualización de datos

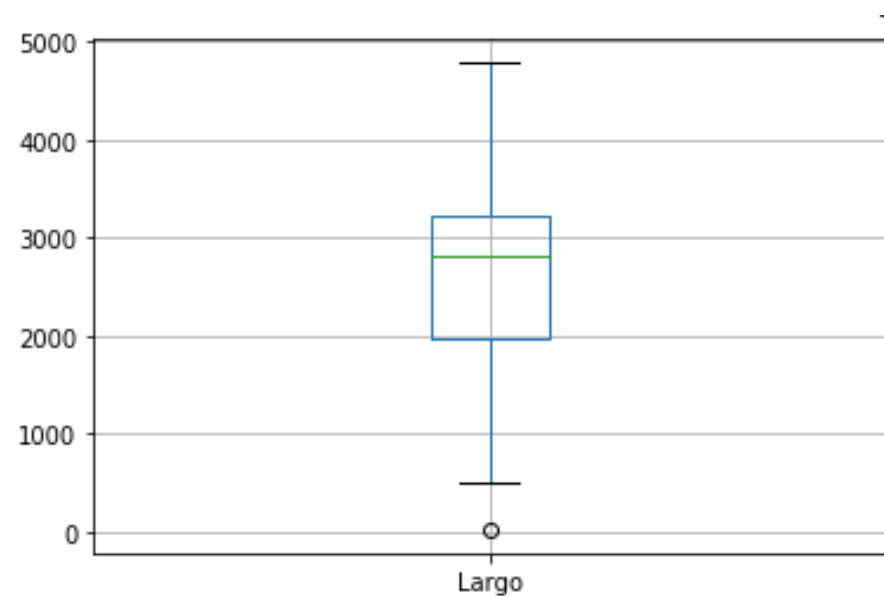


Diagrama boxplot del largo del metal



Diagrama de Pastel de la resolución de los materiales

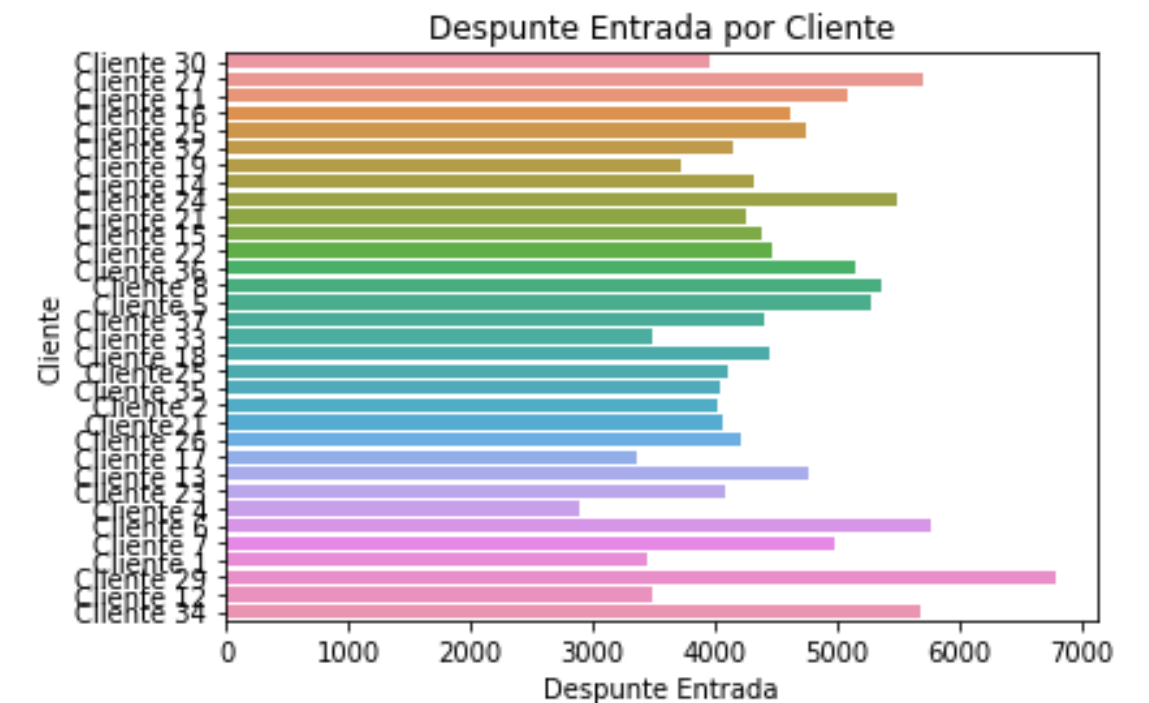


Diagrama de barras de despunte por cliente

Histograma de la velocidad de decapado

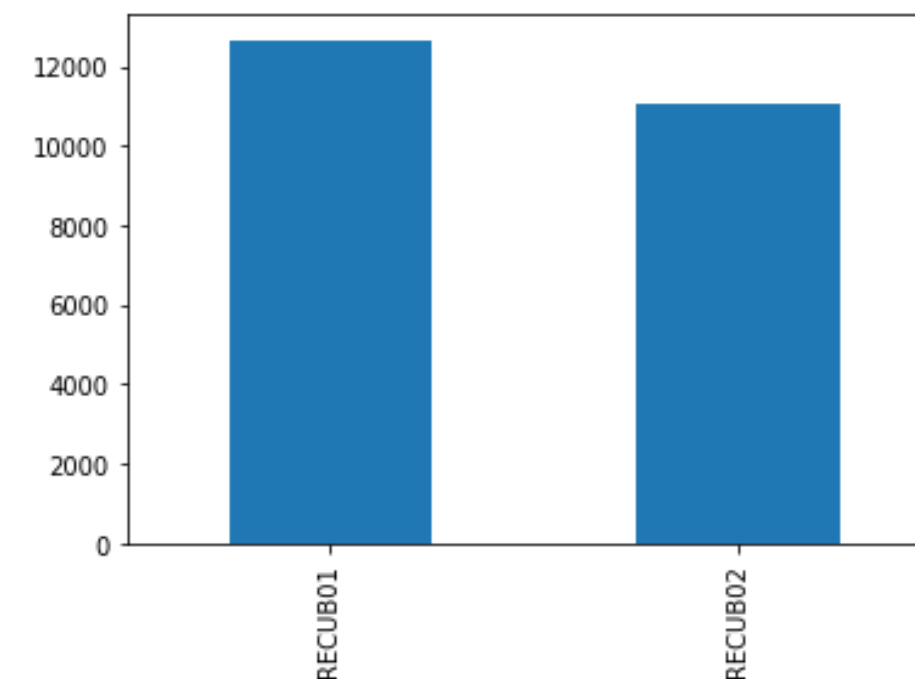
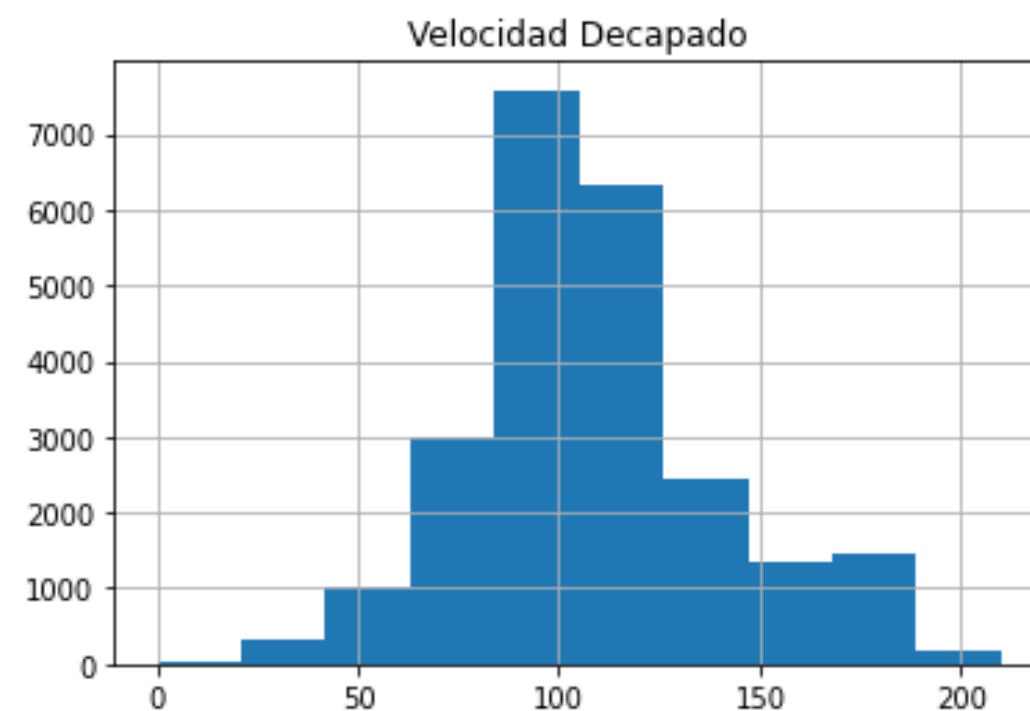


Diagrama de barras de los tipos de recubierto

Comprensión y limpieza de los datos

Funciones relevantes utilizadas para el entendimiento y la limpieza de los datos :





Transformación de datos

- Se eliminaron los datos duplicados en las filas.
- Se eliminaron las columnas en donde todas las filas tienen el mismo valor.
- Se eliminaron las columnas con correlación mayor al 95%, ya que se vuelve redundante.
- Se eliminaron las filas que no tuvieran un defecto importante.
- Se eliminaron las columnas que contenían datos teóricos.
- Se cambiaron las variables categóricas a numéricas.
- Se redujo el tamaño de datos de 8,895,750 a 3,155,026

Hipótesis

Al utilizar técnicas de Machine Learning, se determinará si se encuentran defectos en la producción de productos de acero de Ternium, con esto, se podrá dar información de utilidad a la empresa para que reduzcan o mitiguen dichos defectos, por lo tanto, se verán beneficios económicos para Ternium y una mayor eficiencia y precisión al momento de realizar las órdenes de sus clientes a través de las distintas líneas de producción por las que pasa el acero.

Modelación

Variable Target

CodigoDefectoPrincipal LineaB

Modelos de Clasificación

- Linear SVC
- Decision Tree Classifier
- Random Forest

Modelos de Clustering

- K Means
- Spectral Clustering

Breve énfasis de cada modelo

- Linear SVC : Categorizar los datos que sean proporcionados en un hiperplano.
- Decision Tree Classifier : Busca clasificar los datos con base en reglas de decisión con respuestas binarias.
- Random Forest : Varios clasificadores de árboles de decisión y el promedio de los mismos.
- K Means : Se enfoca más en hacer un número k de agrupaciones.
- Spectral Clustering : Se enfoca más en la descomposición espectral de los datos.

Métricas

1

Librería sklearn.metrics

Implementa varias funciones de pérdida, puntuación y utilidad para medir el rendimiento de la clasificación.

2

accuracy_score()

Esta función calcula la precisión, así como la fracción o la cuenta de predicciones correctas.

3

confusion_matrix()

Esta función evalúa la precisión de la clasificación calculando la matriz de confusión con cada línea correspondiente a la clase True

4

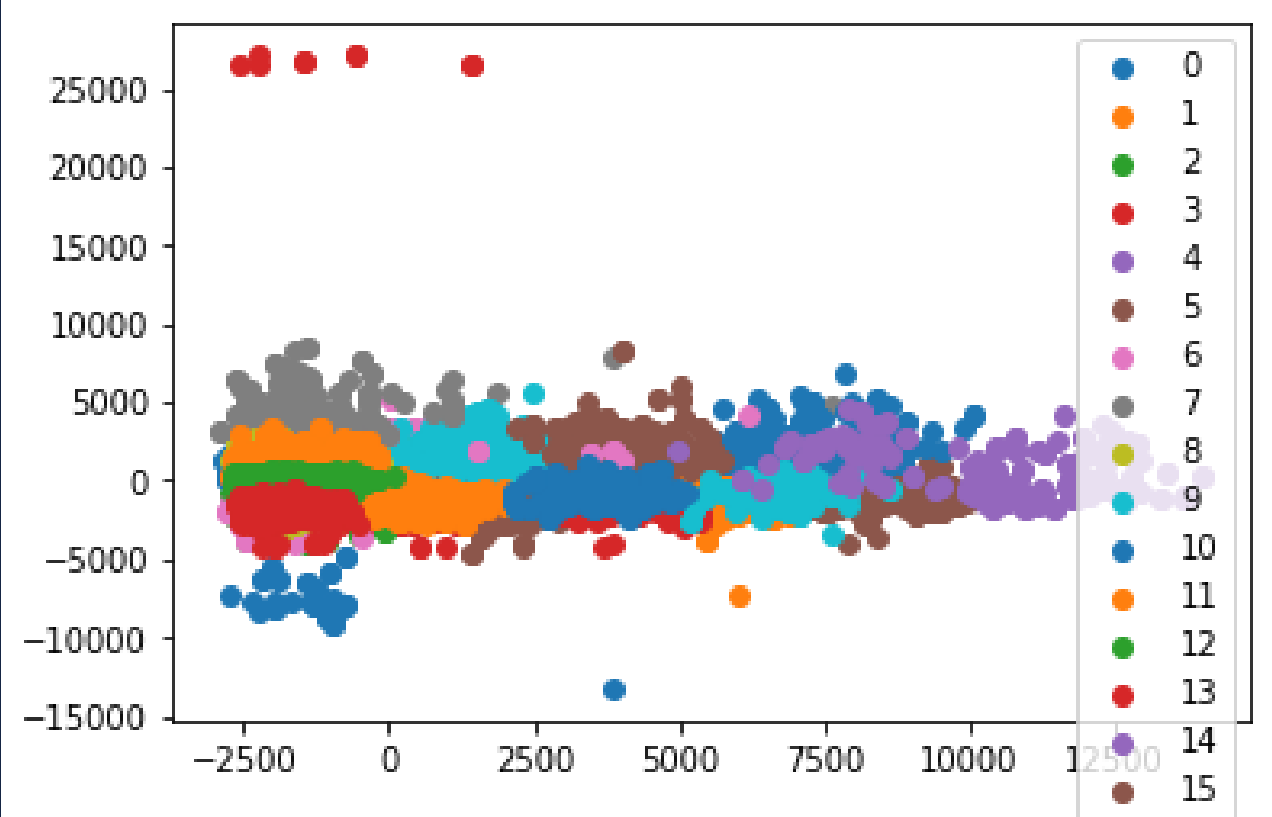
classification_report

Esta función crea un reporte de texto mostrando las métricas de la clasificación principal.

Evaluación y selección del modelo de acuerdo a las métricas

Modelo	SVM	Decision Tree Classifier	Random Forest Classifier	KMeans	Spectral Clustering
	Test size 0.2	Test size 0.2	Test size 0.2	35 KMeans	3 Spectral Clusterings
Accuracy Test	78.6655	97.5658	94.9026	2.1081	73.9688
Test de precisión	78.6655	97.5658	94.9026	2.1081	73.9688
Sensibilidad	0.9996	1.0	1.0	0.7151	1.0
Especificidad	0.0	1.0	0.9959	NaN	0.0

K Means



Gráfica Decision Tree Classifier





- Tanto el modelo de Decision Tree Classifier como el de Random Forest Classifier obtuvieron un porcentaje mayor en el test de precisión.
- El modelo de Decision Tree será el mejor para realizar la predicción, ya que su porcentaje e mantuvo entre 91% y 97% para todas las pruebas.

Resultados



Impacto y utilidad

La empresa Ternium será beneficiada no solo en un impacto económico, sino también de calidad, al reducir los defectos en sus productos.

Además de elevar su reputación como empresa en el mercado internacional.



Prototipo funcional



Modelo de clasificación para predicción de defectos

Arrastra un archivo o [selecciona uno](#)

Nombre del archivo:
testdf2.csv

index	Cliente	Norma	Grado	Código Acabado	Grado Acero Molino	Cara Garantizada	MUESTRA	PROVEEDOR	C_CLASE_PRODTO	'SUPERFICIE'	PRAM	C_CLASE_PRODTO.1	TIPO_RECUBRIMIENTO	Peso Báscula Entrada	Despunte Entrada	Descole Entrada	Largo	Velocidad Tandem	Velocida Promedio Mrh	Velocidad Decap
0	0	0	39	1	19	2	2584	6	0	1	3	1	0	25360	6350	3277	2492	221	76.9826179029234	
1	0	0	39	1	19	2	2596	6	0	1	3	1	0	24810	6590	3272	2482	210	76.9826179029234	

Predicciones:

0	627
1	627

Recomendaciones

Técnicas

Creación de un diccionario con la descripción de cada variable.



Ambiental

Un enfoque más ecológico, implementación de las ODS 7, 11, 13 y 17



Defectos

Énfasis en los defectos con código 2 y 627. (DEF_0007 y DEF_00051), ya que son los que predominan más en el algoritmo de predicción.

Gráfica de defectos en algoritmo de predicción

```
[ ] 1 print(pred_df.groupby("Defectos").size())
```

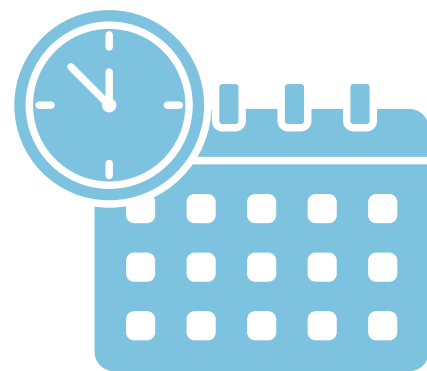
Defectos	
0	2593
2	262
4	1
8	63
19	18
22	3
25	39
36	1
40	1
94	87
101	10
111	28
115	6
224	12
304	10
308	72
414	1
576	2
593	6
627	123
665	1
674	13
705	11
1069	9
1096	44
1171	23
1345	52
1346	1

dtype: int64



Siguientes pasos

Buscar crear un algoritmo que busque la forma de manipular las variables de tal forma que el modelo de predicción proyecte en su mayoría o totalidad, los rollos que no tengan ningún defecto.



Agradecimientos y aprendizajes obtuvimos

Ternium

Empresa Socio Formadora

Angelina y Rubí

Profesoras de bloque

- La ciencia de datos es una rama de la tecnología computacional con un futuro prometedor.
- Se puede lograr un buen análisis a pesar del poco tiempo que se trabajó.
- El modelo CRISP-DM fue de mucha utilidad para seguir paso a paso cada etapa del reto.
- Todos aprendimos mucho y nos interesó más esta rama de Ciencia de Datos.