

EXAMEN 01

TEMAS DE INGENIERÍA MECATRÓNICA H

Nombre: Armando Arturo Custodio Díaz

Código: 20196355

Generación de Conclusiones

1. Análisis Exploratorio de Datos (EDA): 2 puntos • Identificación de tendencias, patrones y valores nulos para los datos AOI y SPI. Mostrar gráficas y comentarios.

Variables Categoricals

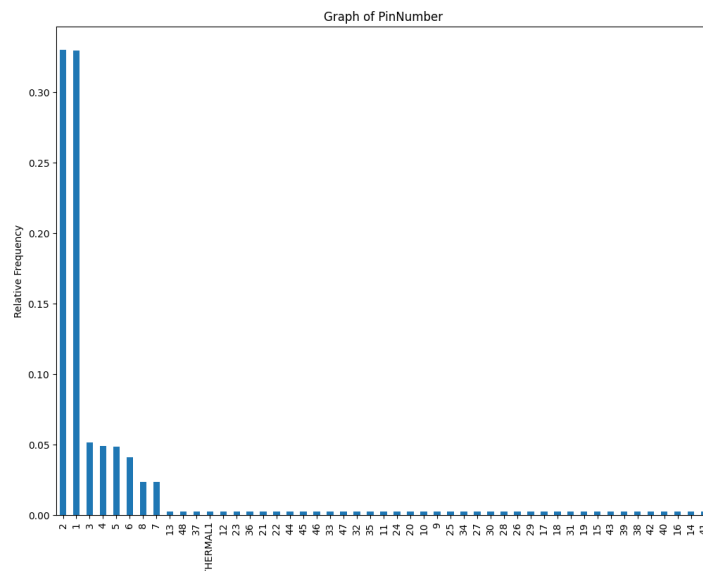
- Variables Nominales: ComponentID, PanelID, FigureID, PinNumber, PadID, PadType, FigureID, Result
- Variables Ordinales:

Variables Numericas

- Variables Continuas: Volume(%), Height(um), Area(%), OffsetX(%), OffsetY(%), SizeX, SizeY, Volume(um3), Area(um2), Shape(um), PosX(mm), PosY(mm)
- Variables Discretas:

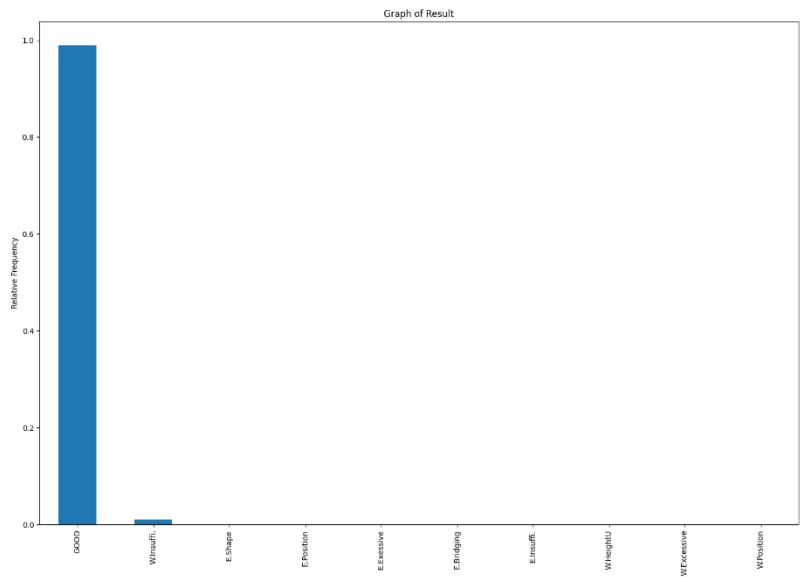
Variables de Tiempo: Date, Time

La distribución observada en el gráfico de frecuencia relativa para PinNumber revela patrones significativos. Los valores 1 y 2 destacan con aproximadamente un 33%, posiblemente indicando una frecuencia común para los primeros pines en el conjunto de datos. Los valores 3, 4, 5 y 6 muestran una frecuencia más baja, alrededor del 5%, sugiriendo que estos pines son menos comunes o específicos para ciertos casos. Sin embargo, llama la atención que los valores desde 33 hasta 48, junto con THERMAL1, tienen una frecuencia considerablemente menor al 1%.

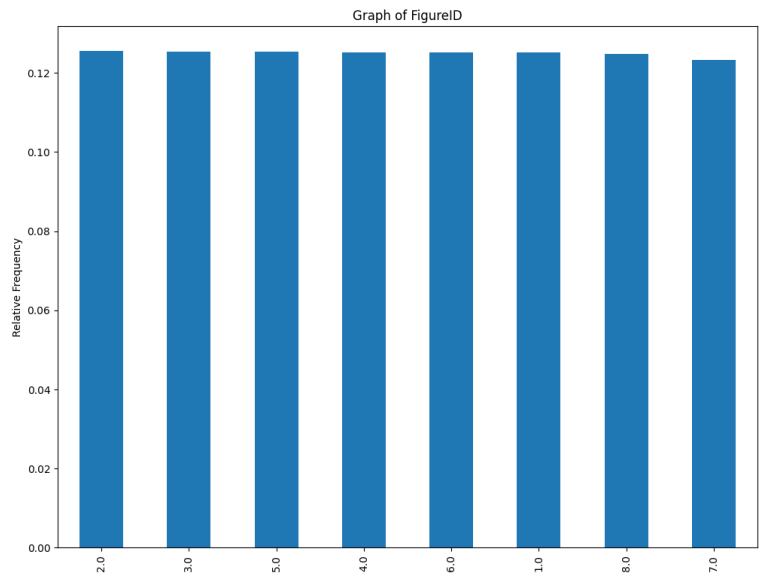


La predominancia abrumadora del valor "GOOD" en más del 95% en la gráfica de frecuencias de la variable RESULT indica que la gran mayoría de los componentes inspeccionados no presentan

defectos según la máquina de inspección. Este resultado es positivo y refleja la eficacia del proceso de fabricación. La presencia de un valor ligeramente significativo, como "W.Insuff" después de "GOOD", sugiere que existe una proporción pequeña pero no despreciable de componentes con insuficiencias de soldadura. Las demás denominaciones se pueden despreciar.



La distribución equitativa del FigureID del 1 al 8 en el conjunto de datos SPI sugiere que cada panel de producción contiene ocho PCBs individuales. Esta estructura organizativa facilita un etiquetado único y específico para cada componente en el proceso de fabricación de placas de circuito impreso.



Por otro lado, se tiene la limpieza de datos y análisis se ha eliminado los outliers de SPI_data y los valores nulos de AOI_data.

Resultados de predicciones: Como se analizó en clases, tener un accuracy alto no es el único indicador para concluir que nuestro modelo predictivo está funcionando de manera correcta. Dado el análisis exhaustivo, que incluyó la limpieza de datos, la integración de datos y la ingeniería de características, utilizando estadísticas descriptivas como la media, el mínimo, el máximo, la desviación estándar y la mediana, se ha obtenido un valor F1 del 82%. Este resultado indica que nuestro modelo tiene un buen equilibrio entre precisión y exhaustividad, lo que es crucial cuando las clases están desbalanceadas. Además, se ha logrado un valor de accuracy del 76%, lo que significa que el 76% de las predicciones realizadas por nuestro modelo fueron correctas.

```
✓ Realización de predicciones

[34] Y_test_pred = model.predict(X_test)

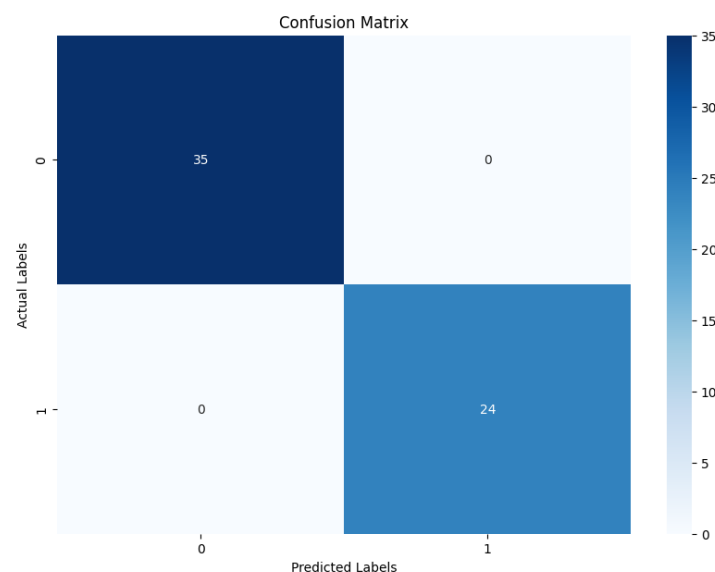
[35] f1_score(Y_test,Y_test_pred)

0.8201438848920863

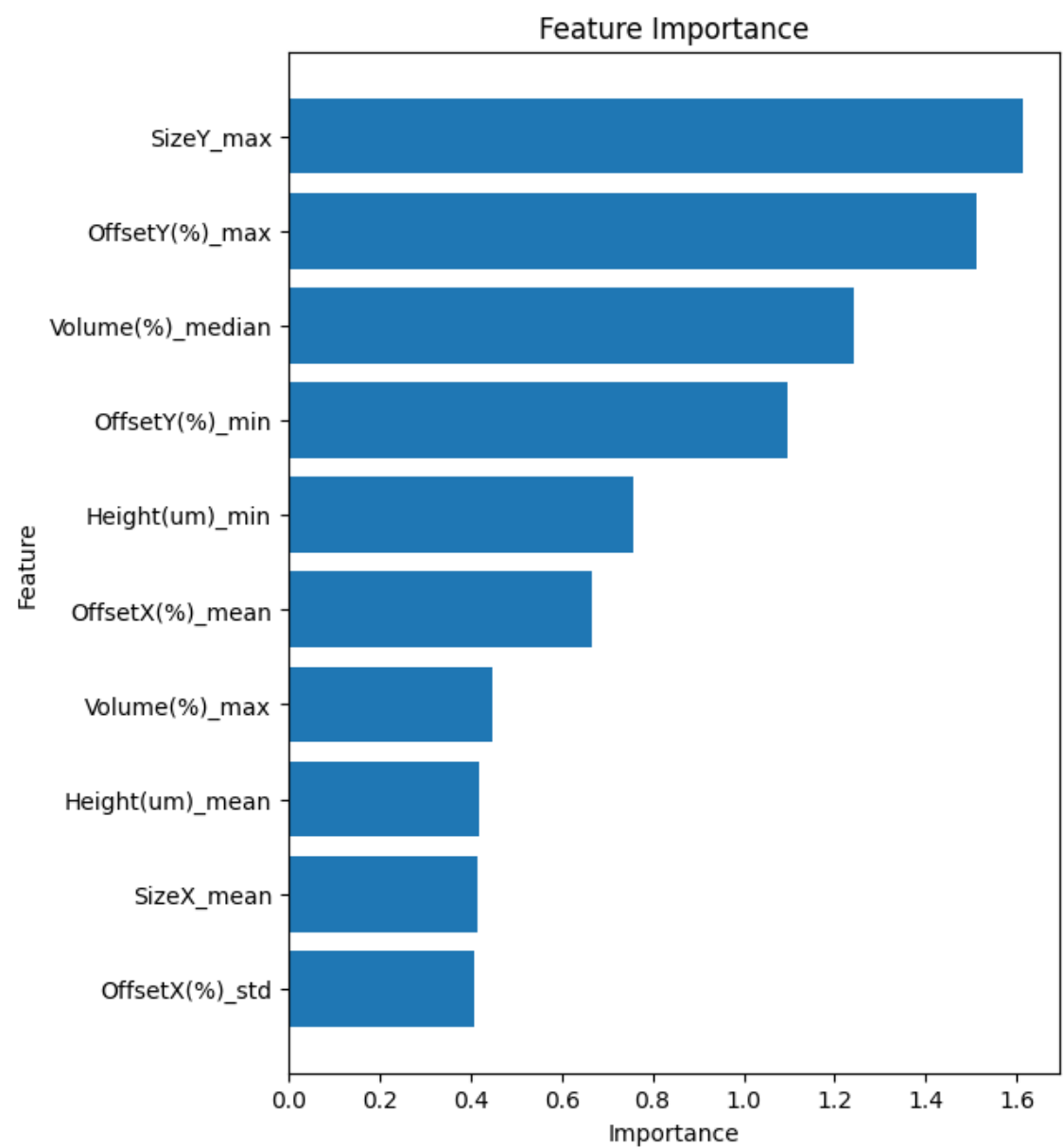
▶ accuracy_score(Y_test,Y_test_pred)

📄 0.7619047619047619
```

Basándonos en los resultados de la matriz de confusión proporcionados, observamos que nuestro modelo presenta un buen desempeño en la clasificación. Con 23 verdaderos positivos, 57 verdaderos negativos, 10 falsos negativos y 15 falsos positivos, hemos logrado un equilibrio entre la capacidad de identificar correctamente casos positivos y negativos. Este equilibrio se refleja en un sólido valor de F1 del 82%, indicando que nuestro modelo mantiene una notable precisión y exhaustividad. Aunque el accuracy es del 76%, la evaluación integral de la matriz de confusión sugiere que el modelo es efectivo en su tarea específica de clasificar circuitos defectuosos como buenos o irreparables.



Analizar la importancia de las características (Feature Importance) es crucial para comprender qué variables están contribuyendo significativamente a la capacidad predictiva del modelo. En nuestro estudio, hemos identificado que variables como 'SizeY_max', 'OffsetY(%)_max', y 'Volume(%)_median' destacan como las más importantes. La relevancia de estas variables se refleja en su capacidad para influir de manera significativa en las predicciones del modelo. 'SizeY_max' sugiere que la altura máxima del componente es un predictor esencial, mientras que 'OffsetY(%)_max' indica la importancia de la máxima desviación vertical. 'Volume(%)_median' resalta la mediana del volumen como un factor clave.



Dada la importancia de estas variables, existe la oportunidad de explorar transformaciones entre ellas en una segunda iteración del modelo. Al examinar combinaciones o transformaciones de estas características, podríamos mejorar aún más el rendimiento del modelo en términos de F1 y accuracy. Esta estrategia permitiría capturar relaciones más complejas entre las variables, potencialmente refinando la capacidad del modelo para realizar predicciones más precisas y robustas. La iteración centrada en estas variables clave podría ser un paso clave para optimizar y perfeccionar la capacidad predictiva del modelo.