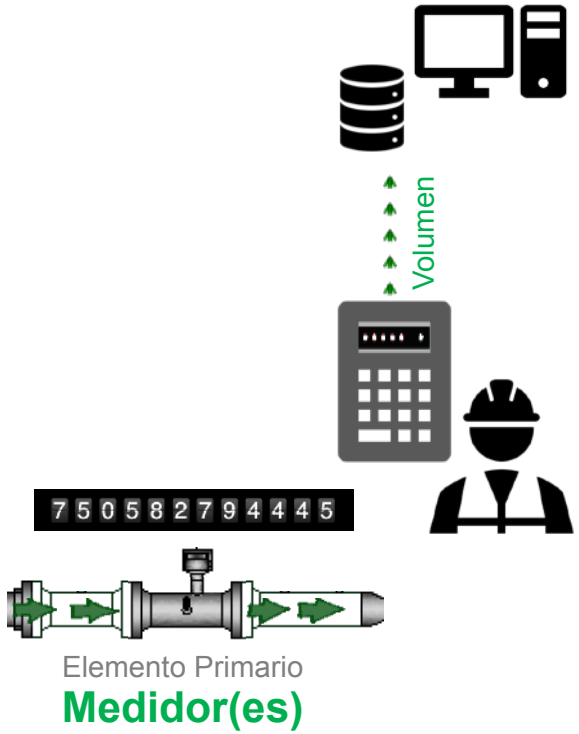
MODELO PREDICTIVO: VERIFICACIÓN DE MEDIDORES

Unidad de Gestión Perdidas y Servicio Técnico
Gerencia Técnica Gas
Vicepresidencia Gas

Contexto

¿Creen que es posible predecir la probabilidad de que un medidor residencial falle?



PILOTO 3

IMPACTO EN LA CAÍDA DEL CONSUMO

Despues de realizar el cambio de 2207 medidores, se evidenció que la hipotesis inicial de recuperar un volumen significativo no sucede,
Por lo que se procede a buscar un nuevo enfoque al piloto

PROBABILIDAD DE FALLA EN EL MEDIDOR

Ahora, se busca darle prioridad de cambio a medidores que a partir de diferentes variables tengan mayor probabilidad de fallar en la medición del volumen

CANTIDAD DE SALIDAS DE GAS

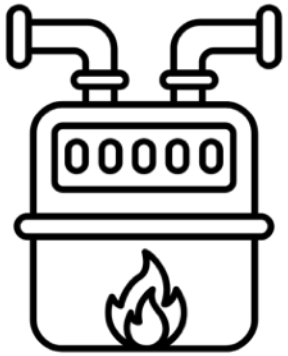
Medidores que alcanzan fácilmente el Qmax debido a la alta demanda de gas

LOCALIDAD

Factores meteorológicos, de orden público, antigüedad, estrato, etc

TIEMPO EN USO

La vida util de cada medidor es uno de los factores de mayor peso que incide en el daño de estos



REVOLUCIONES POR VOLUMEN CICLICO

Desgaste de las piezas mecánicas por las revoluciones del sistema desde su inicio hasta la última lectura

MARCA MEDIDOR

Cada proveedor tiene una vida útil diferente, materiales diferentes y diferentes procesos de fabricación

EXTERNO O INTERNA

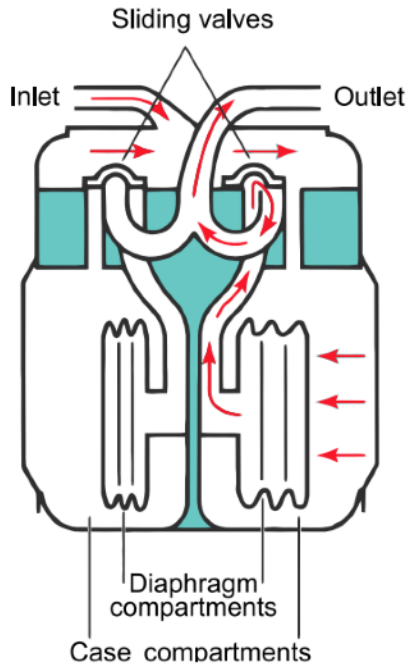
Factores de humedad, temperatura, polvo y suciedad en general podrían afectar el correcto funcionamiento de los medidores

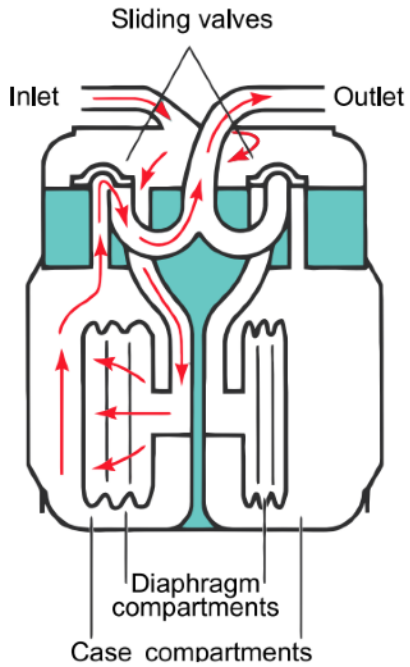
RELACION CON OTROS SERVICIOS

Comparar el comportamiento del gas frente al consumo de energía y agua a traves del tiempo.

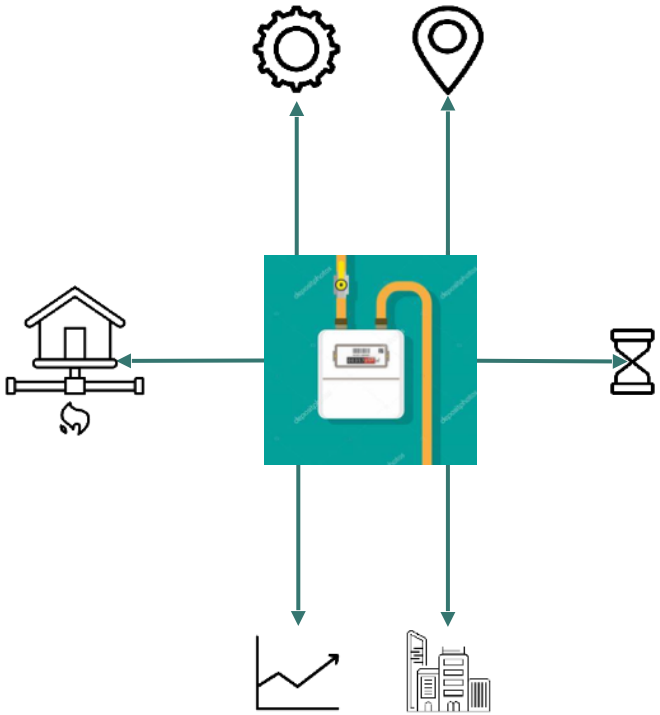
REVOLUCIONES POR VOLUMEN CICLICO

Funcionamiento de un medidor de gas tipo diafragma



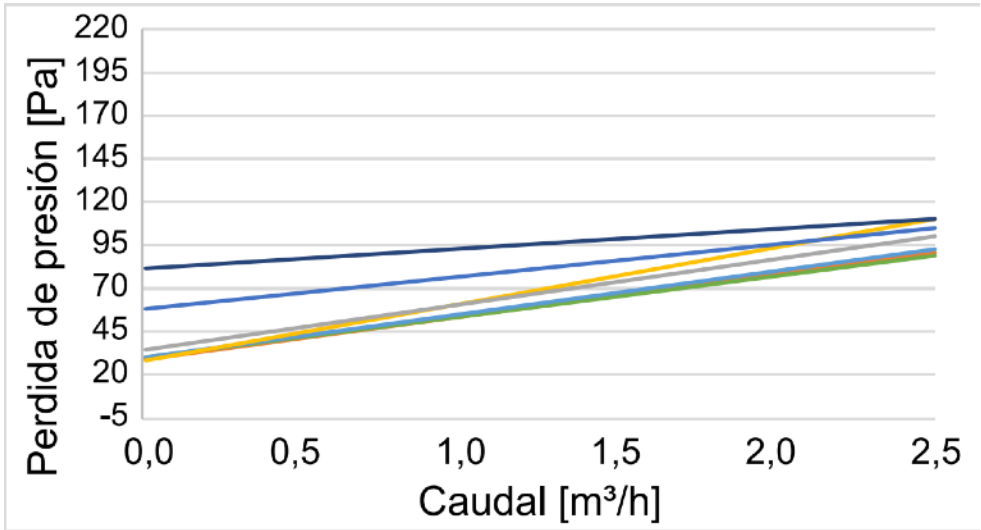


NUEVO ENFOQUE



CANTIDAD DE SALIDAS DE GAS

Perdidas de presión en medidores G1.6 después de más de 20 años en uso



MEDIDORES RESIDENCIALES 20 AÑOS

Flujo de trabajo, pasos a seguir

Contextualización

Estudiar la información que se tiene acerca del piloto, de donde nace este proyecto y bajo que estudios sustentados existe una viabilidad para este.

Datos

Recolección y el pre-procesamiento de los datos con los que se van a trabajar para el modelo predictivo

Entrenar

Realizar los diferentes modelos de machine learning en python para ser entrenados y cumplir con los objetivos propuestos

Predecir

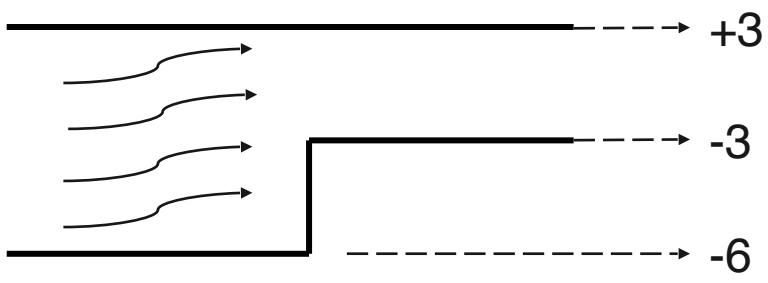
Extrapolar y/o predecir valores que sean determinantes bajo criterios asociados para cumplir con el objetivo de este piloto

Preprocesamiento de datos

Rangos_Qma	Rangos_Qt	Rangos_Qmin	RESULTADO_PR
a probabilidad de fa	a probabilidad de fal	Alta probabilidad de fallar	NO CONFORME
a probabilidad de fa	a probabilidad de fal	Alta probabilidad de fallar	NO CONFORME
Bueno	Bueno	Bueno	CONFORME
Bueno	Bueno	Bueno	CONFORME
a probabilidad de fa	Bueno	Bueno	NO CONFORME
a probabilidad de fa	robabilidad de fallar	Alta probabilidad de fallar	NO CONFORME
Bueno	Bueno	Bueno	CONFORME
a probabilidad de fa	Bueno	Alta probabilidad de fallar	NO CONFORME
Bueno	Bueno	Bueno	CONFORME
Bueno	Bueno	Bueno	CONFORME
Bueno	Bueno	Bueno	CONFORME
a probabilidad de fa	robabilidad de fallar	Alta probabilidad de fallar	NO CONFORME

RANGOS CAUDALES	VALOR
66% ≤ Medida ≤ 100%	Bueno
33% ≤ Medida < 66%	Regular
0% ≤ Medida < 33%	Aceptable
Medida ≠	No_Conforme

RANGOS CAUDALES	VALOR
20% ≤ Medida	Bueno
Medida < 20%	Alta probabilidad de fallar



La medida se debe encontrar dentro de los límites máximos permitidos.
De encontrarse por fuera de este rango o frenados, se catalogarán como “Alta probabilidad de fallar”

Comparativa entre modelos

Modelos Clasificatorios y Métodos estadísticos

Modelo	Pred. Qmax	Pred. Qt	Pred. Qmin
Random Forest	76.02%	80.39%	71.96%
XGBRandomForest	76.17%	81.00%	73.00%
SVM	75.11%	68.78%	70.74%
LGBMClassifier	73.60%	77.83%	68.64%
XGBoost	71.04%	76.23%	70.29%

```
Acc del modelo intermedio Rangos_Qt: 80.39%
Precision del modelo intermedio Rangos_Qt: 83.75%
Acc del modelo intermedio Rangos_Qmax: 76.02%
Precision del modelo intermedio Rangos_Qmax: 77.17%
```

- **Simulaciones/Método de Monte Carlo** → Usa muestreo aleatorio para estimar resultados.
- **Test Chi-cuadrado** → Evalúa si dos variables categóricas son independientes.
- **Test de Correlación** → Mide la relación entre variables numéricas
- **Matriz de Confusión Combinada** → Analiza el rendimiento de modelos de clasificación.
- **Índice V de Cramer** → Mide la fuerza de asociación entre variables categóricas basándose en Chi-cuadrado.

```
#Exactitud del modelo Rangos_Qmax:
76.02%
#Exactitud del modelo Rangos_Qmin:
71.96%
#Exactitud del modelo Rangos_Qt:
 80.39%
```

RANDOM FOREST

(Modelo de decisión)

```
#Exactitud del modelo Rangos_Qmax:
76.17%
#Exactitud del modelo Rangos_Qmin:
73.00%
#Exactitud del modelo Rangos_Qt:
81.00%
```

XGBRandom Forest

(Modelo Optimizado)

```
#Exactitud del modelo Rangos_Qmax:
75.11%
#Exactitud del modelo Rangos_Qmin:
70.74%
#Exactitud del modelo Rangos_Qt:
68.78%
```

SVM

Resultados

Métricas

La exactitud mide la proporción de predicciones correctas (tanto positivas como negativas) sobre el total de predicciones.

$$\text{Exactitud (Accuracy)} = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{Total de casos}}$$

$$\text{Total de casos} = \text{True Positives (TP)} + \text{True Negatives (TN)} + \text{False Positives (FP)} + \text{False Negatives (FN)}$$

- **True Negatives (TN):** Predicciones correctas de la clase negativa.
- **False Negatives (FN):** Predicciones incorrectas que deberían ser positivas pero se clasificaron como negativas.

La precisión mide qué proporción de las predicciones positivas realizadas por el modelo son realmente correctas.

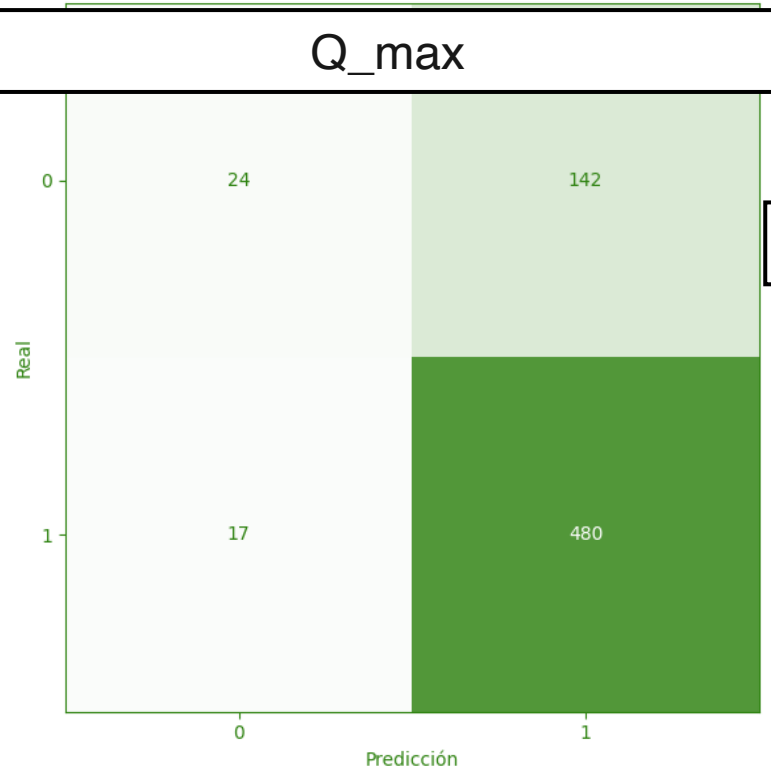
$$\text{Precisión} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

- **True Positives (TP):** Predicciones correctas de la clase positiva.
- **False Positives (FP):** Predicciones incorrectas que fueron clasificadas como positivas.

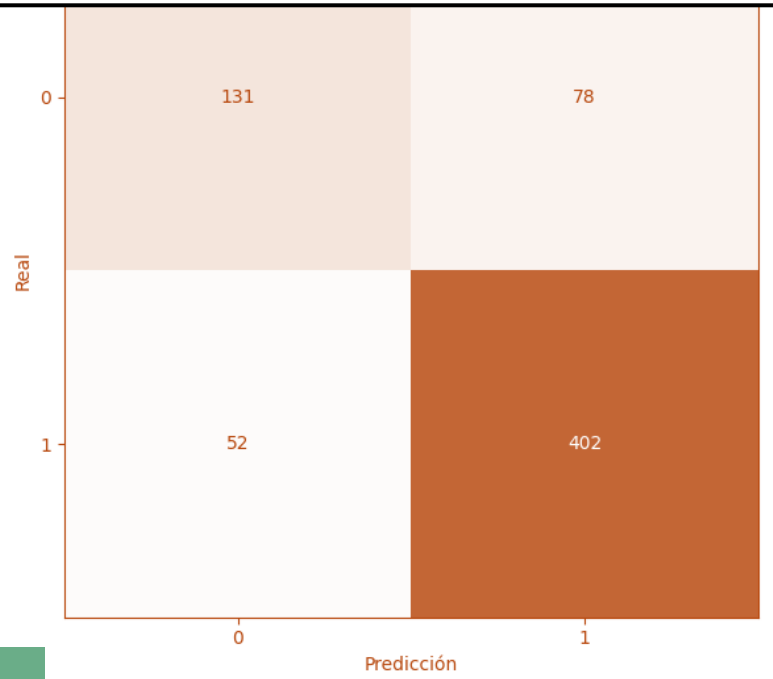
Resultados

Matrices de confusión

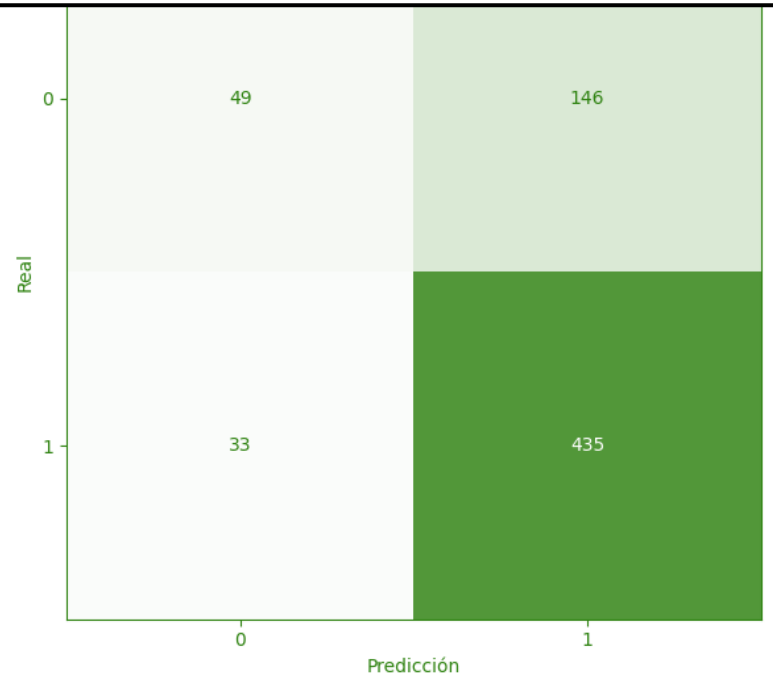
Q_max



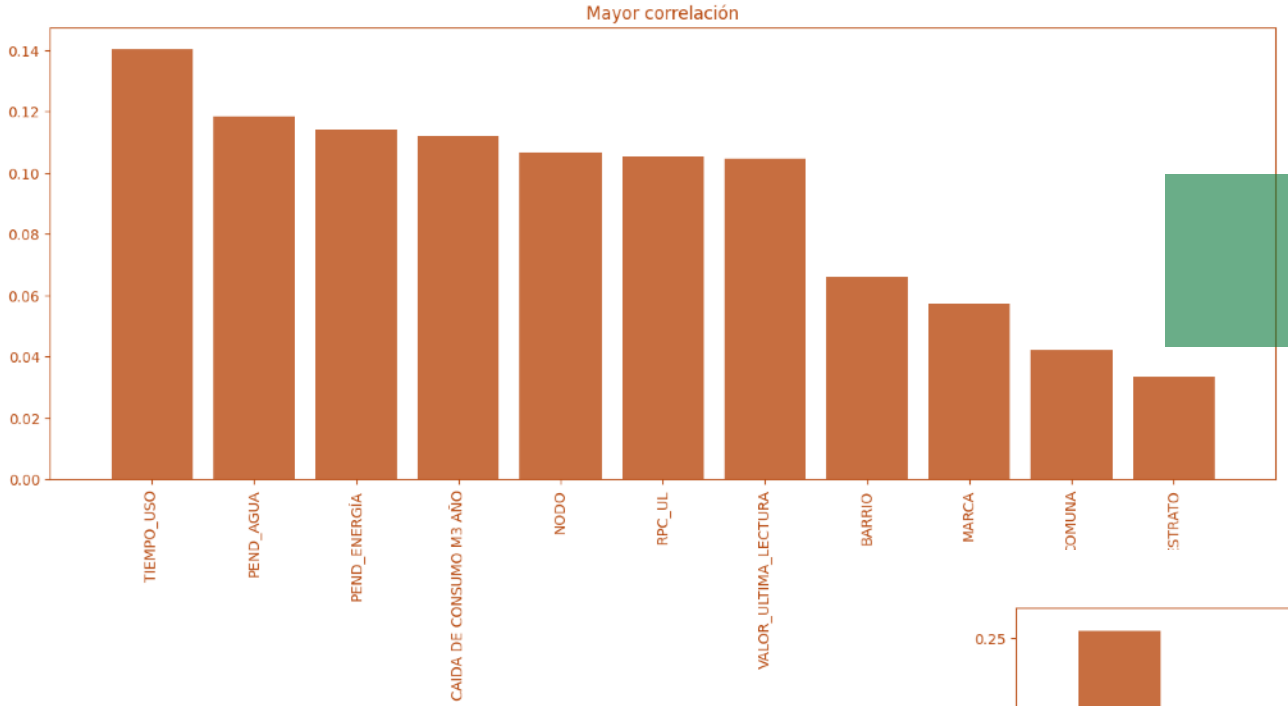
Q_t



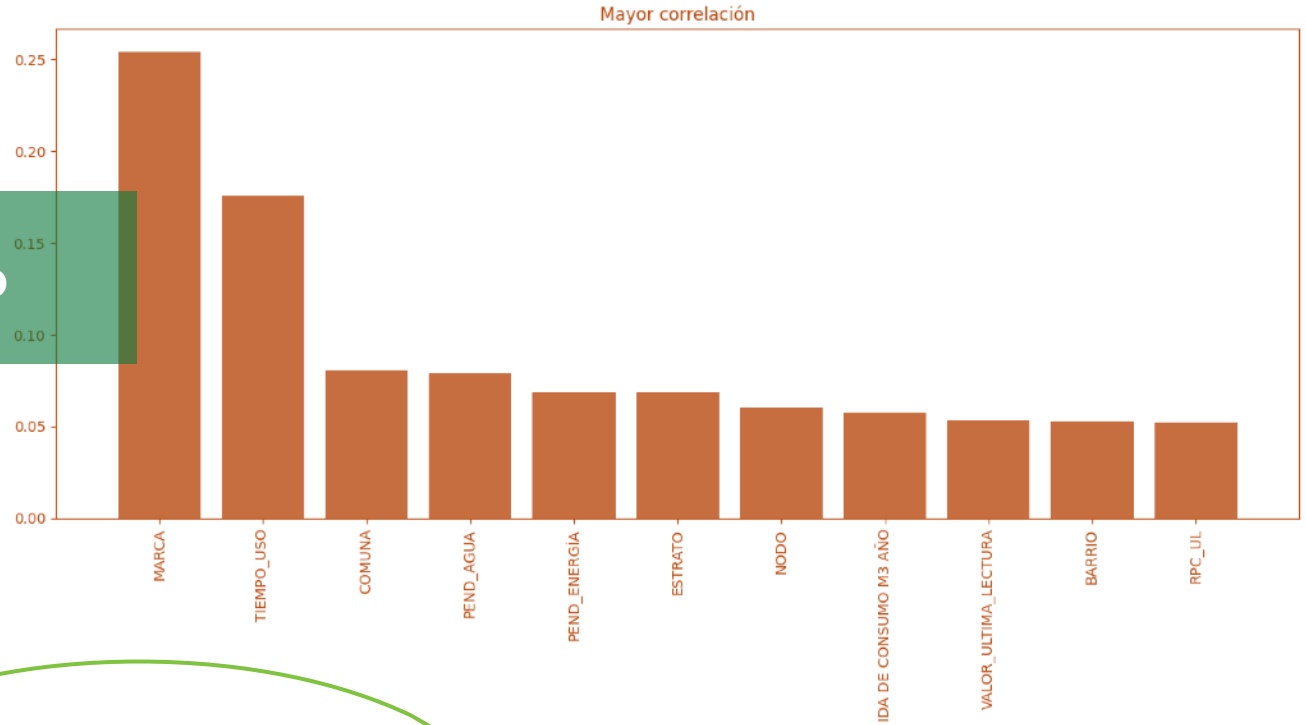
Q_min



0 = Alta probabilidad de fallar
1 = Baja probabilidad de fallar



Modelo de decisión



Modelo optimizado

Sustentación resultados

“Si uno de los caudales Q resulta no conforme en la prueba, es criterio suficiente para indicar la No Conformidad del Medidor.”

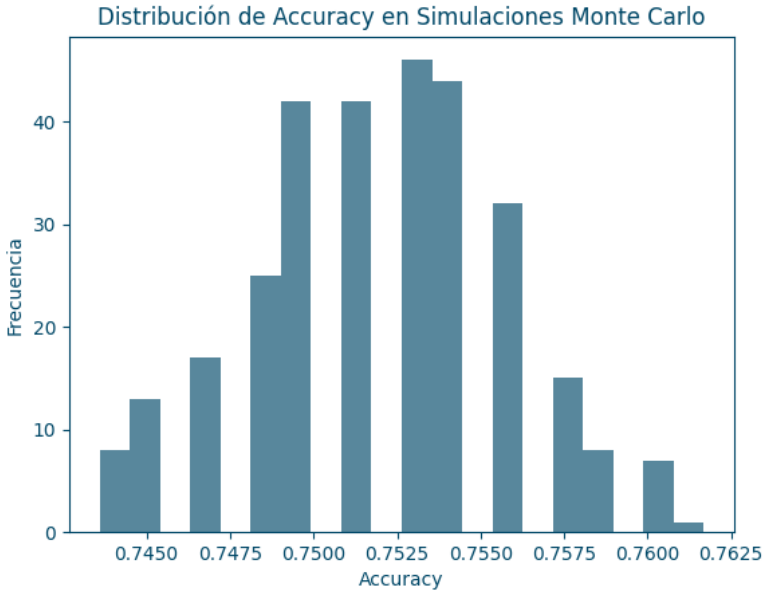
```
Correlación entre fallos de Qt y Qmax: 0.13

Estadístico Chi-Cuadrado: 10.77, p-valor: 0.0010
Hay una relación significativa entre los fallos de los modelos.
Índice V de Cramer: 0.1274
Coeficiente de Contingencia: 0.1264
```

```
Correlación entre fallos de Qt y Qmin: 0.31

Estadístico Chi-Cuadrado: 62.59, p-valor: 0.0000
Hay una relación significativa entre los fallos de los modelos.
Índice V de Cramer: 0.3073
Coeficiente de Contingencia: 0.2937
```

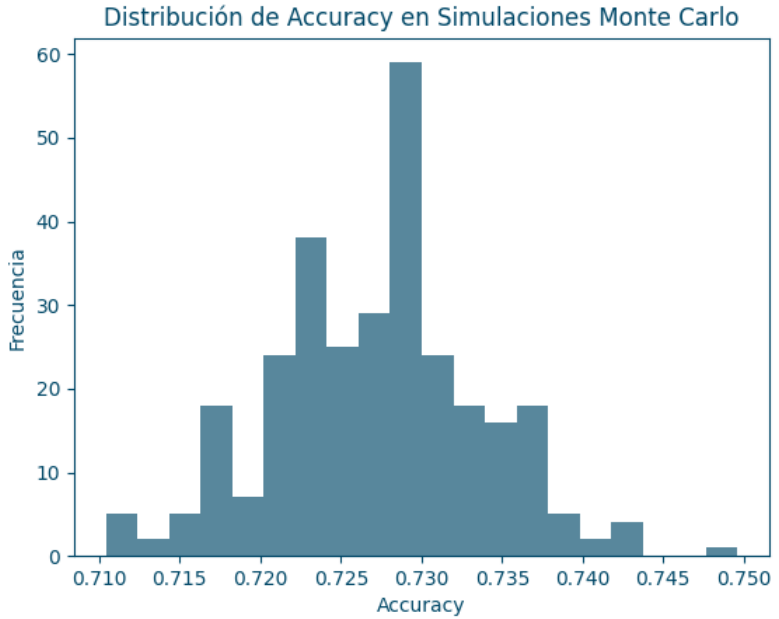
```
Correlación entre fallos de Qmax y Qmin: 0.15

Estadístico Chi-Cuadrado: 13.60, p-valor: 0.0002
Hay una relación significativa entre los fallos de los modelos.
Índice V de Cramer: 0.1432
Coeficiente de Contingencia: 0.1418
```

1. **Correlación entre fallos de Qt y Qmax: 0.13**
   - Esto indica que existe una correlación muy baja entre los fallos de los modelos para las variables Qt y Qmax. En términos prácticos, los errores de uno no están fuertemente relacionados con los errores del otro.
2. **Estadístico Chi-Cuadrado: 10.77, p-valor: 0.0010**
   - El estadístico Chi-Cuadrado mide si existe una relación significativa entre las categorías de fallos de Qt y Qmax.
   - Un p-valor de 0.0010 (< 0.05) indica que hay evidencia estadísticamente significativa para afirmar que existe una relación entre los fallos de ambos modelos.
3. **Conclusión sobre la relación significativa**
   - Dado el p-valor bajo, podemos concluir que los fallos de los modelos Qt y Qmax no son completamente independientes; es decir, hay algún grado de interacción entre ellos, aunque sea leve.
4. **Índice V de Cramer: 0.1274**
   - Este índice mide la intensidad de la relación entre dos variables categóricas. Los valores están entre 0 (sin relación) y 1 (relación perfecta).
   - Un valor de 0.1274 indica una relación muy débil entre los fallos de ambos modelos.
5. **Coeficiente de Contingencia: 0.1264**
   - Similar al V de Cramer, este coeficiente mide la asociación entre variables categóricas.
   - Un valor de 0.1264 refuerza que la relación entre los fallos de Qt y Qmax es débil.

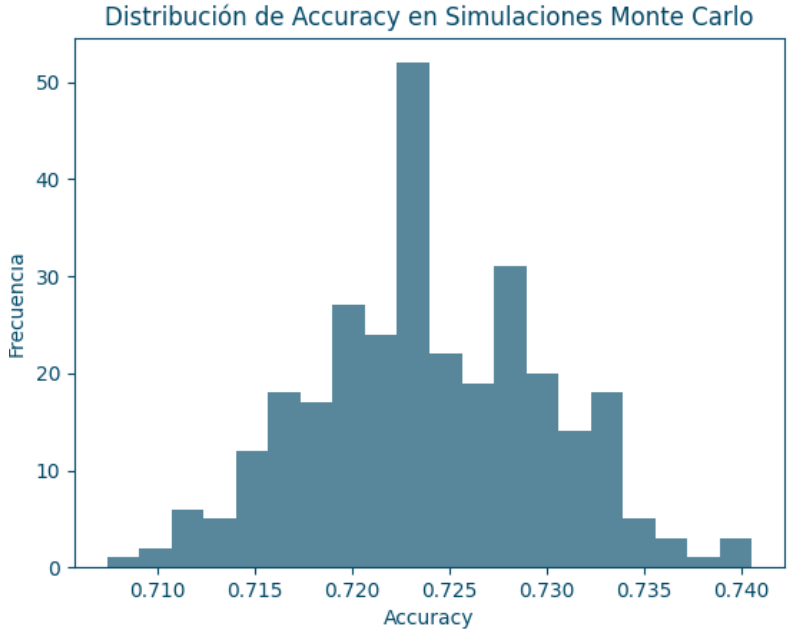
Sustentación resultados



Q_max



Q_t



Q_min

300 Simulaciones

180.000 posibles datos

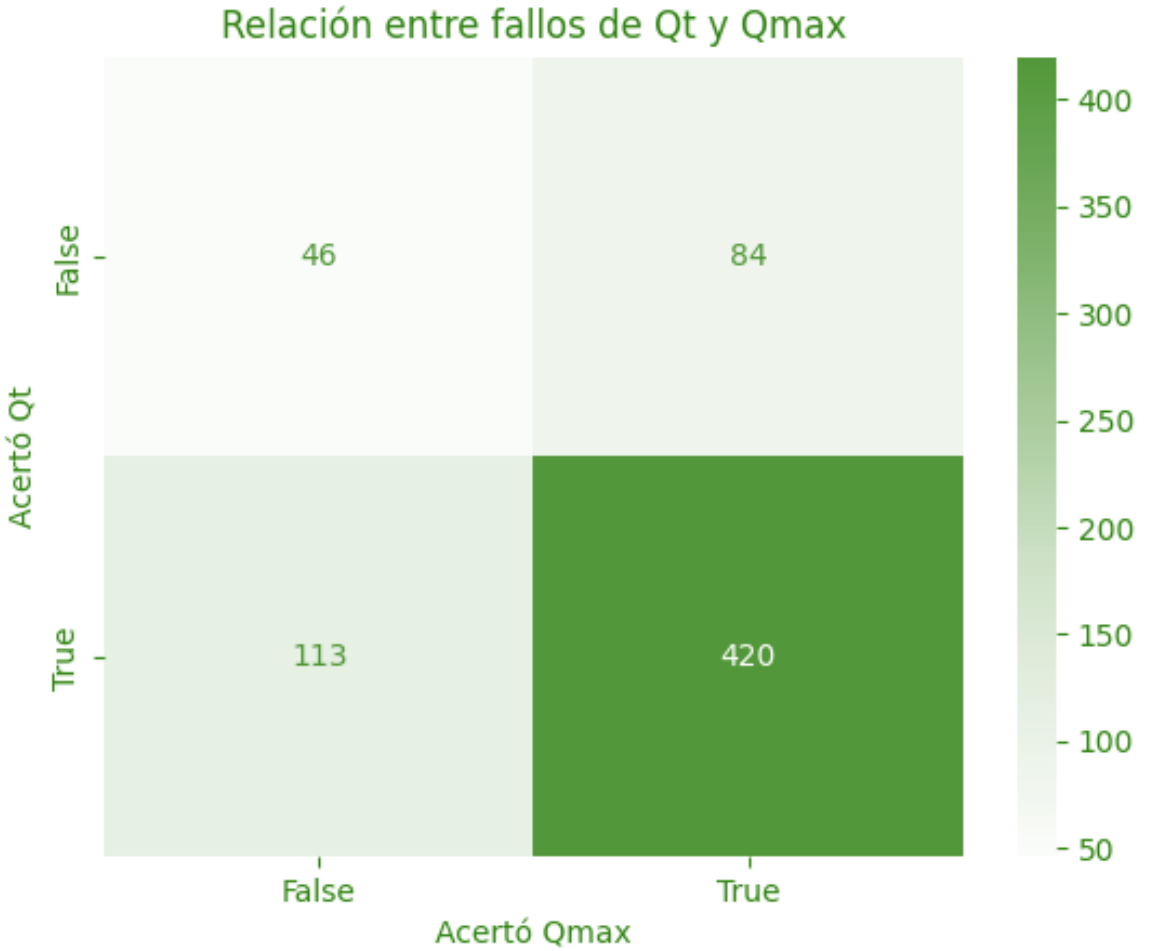
Conclusiones

Correlación entre Qt y Qmax

Qt	Qmax	Predic_Qt	Predic_Qmax
Baja probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo
Baja probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo
Alta probabilidad de fallo	Alta probabilidad de fallo	Alta probabilidad de fallo	Baja probabilidad de fallo
Baja probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo
Baja probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo
Alta probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo
Alta probabilidad de fallo	Baja probabilidad de fallo	Alta probabilidad de fallo	Baja probabilidad de fallo
Baja probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo
Baja probabilidad de fallo	Baja probabilidad de fallo	Baja probabilidad de fallo	Alta probabilidad de fallo

Valores reales

Valores de la predicción



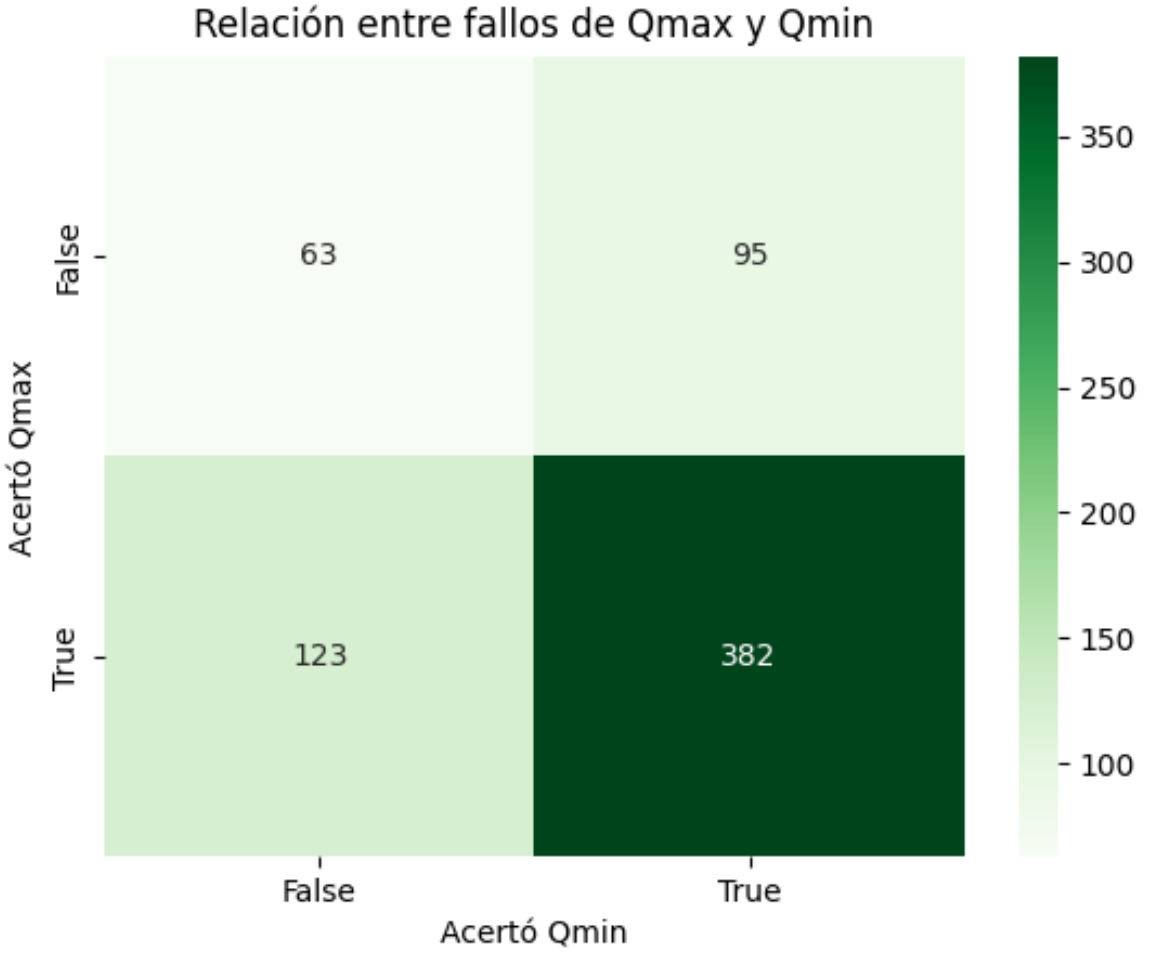
Conclusiones

Correlación entre Qmin y Qmax

Qmax	Qmin	Predic_Qmax	Predic_Qmin
Baja probabilidad de falla	Alta probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Alta probabilidad de falla	Alta probabilidad de falla	Baja probabilidad de falla	Alta probabilidad de falla
Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Alta probabilidad de falla
Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Alta probabilidad de falla	Alta probabilidad de falla	Alta probabilidad de falla	Alta probabilidad de falla
Alta probabilidad de falla	Alta probabilidad de falla	Alta probabilidad de falla	Alta probabilidad de falla
Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Alta probabilidad de falla	Alta probabilidad de falla	Baja probabilidad de falla	Alta probabilidad de falla

Valores reales

Valores de la predicción



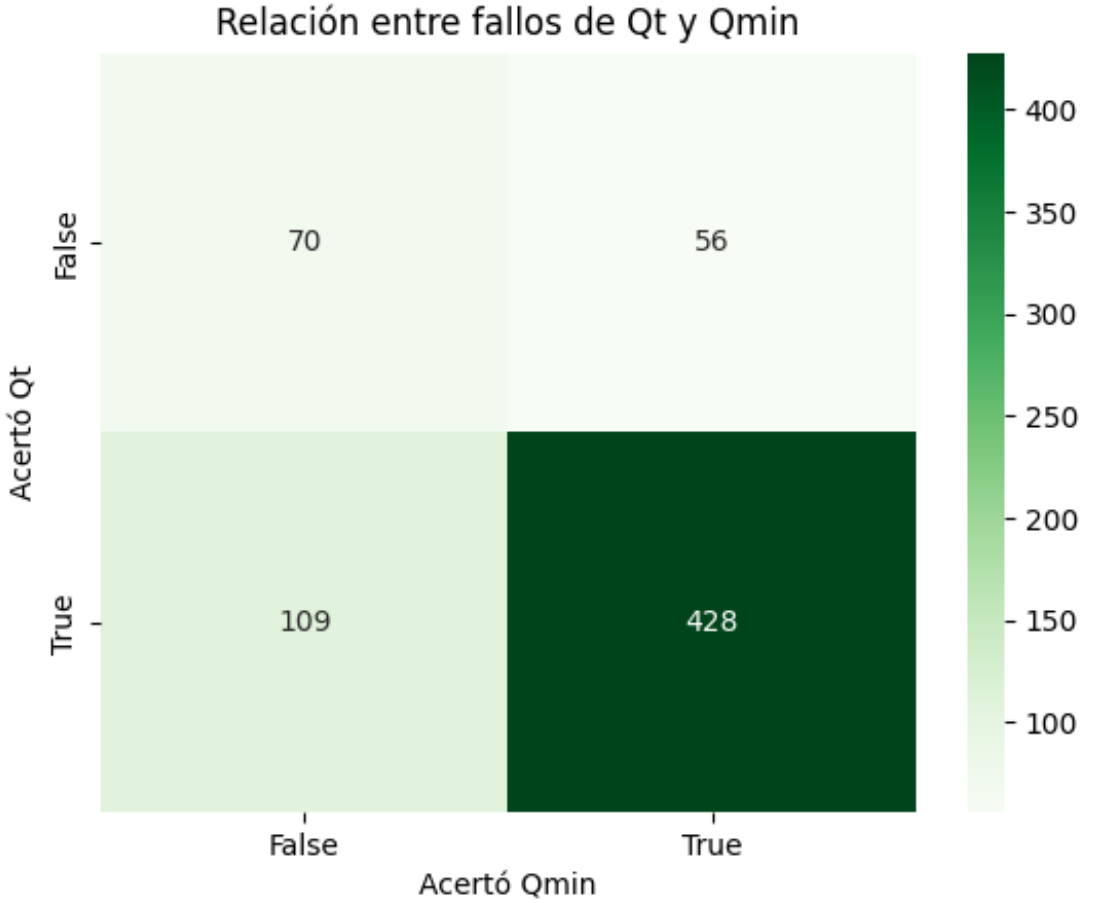
Conclusiones

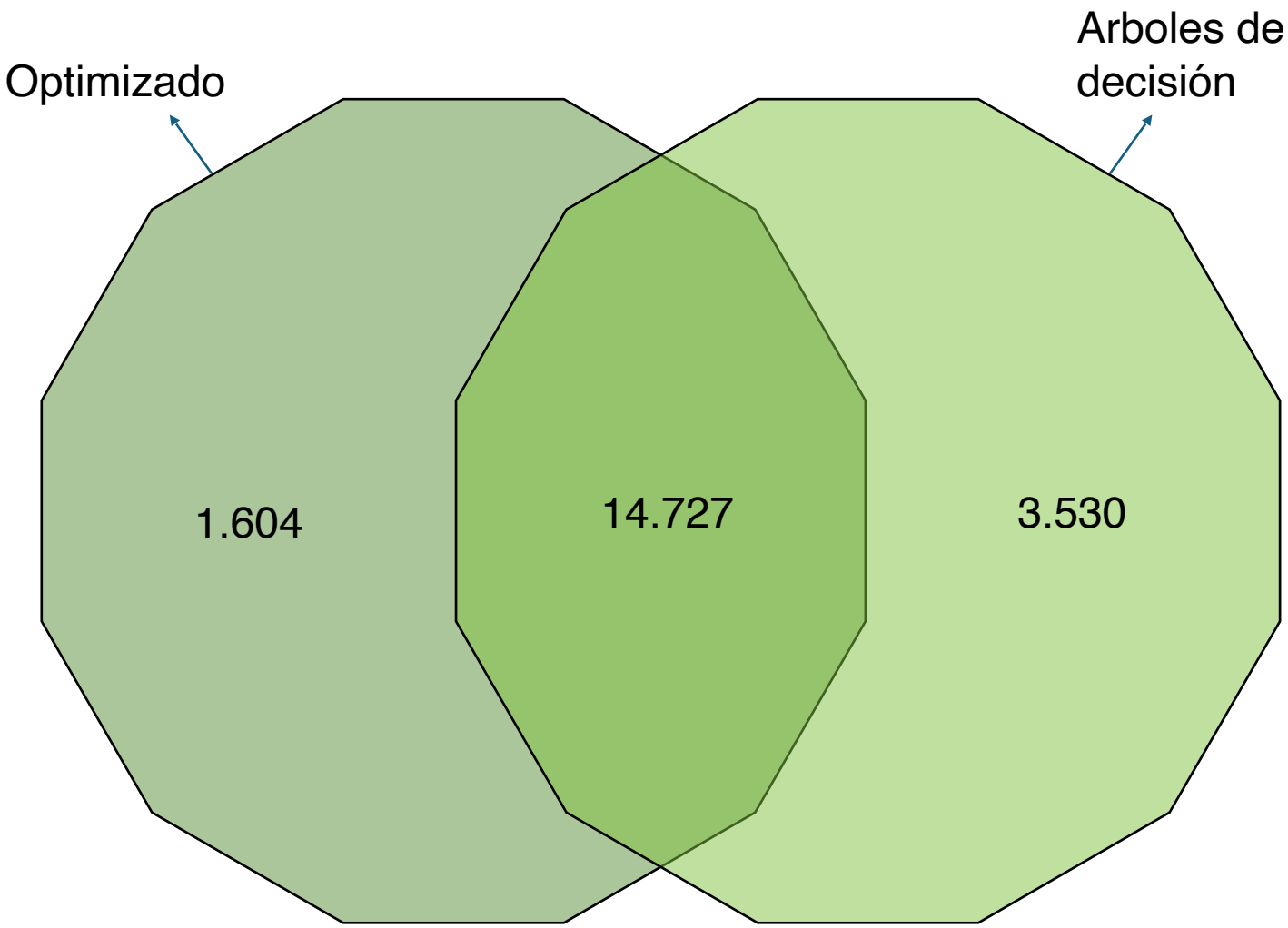
Correlación entre Qmin y Qt

Qt	Qmin	Predic_Qt	Predic_Qmin
Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Alta probabilidad de falla	Baja probabilidad de falla	Alta probabilidad de falla	Alta probabilidad de falla
Baja probabilidad de falla	Alta probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Alta probabilidad de falla	Alta probabilidad de falla	Baja probabilidad de falla	Alta probabilidad de falla
Alta probabilidad de falla	Baja probabilidad de falla	Alta probabilidad de falla	Alta probabilidad de falla
Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Alta probabilidad de falla
Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Baja probabilidad de falla	Alta probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Alta probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla	Baja probabilidad de falla
Alta probabilidad de falla	Baja probabilidad de falla	Alta probabilidad de falla	Alta probabilidad de falla

Valores reales

Valores de la predicción

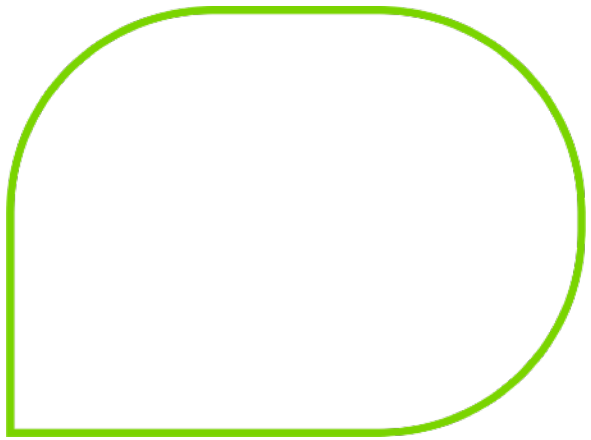
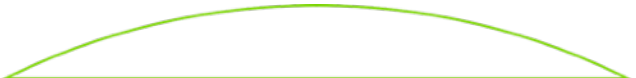




64.713 Medidores

59.579 Medidores en los que ambos modelos coinciden en la predicción

44.852 medidores en los que ambos modelos coinciden en una Baja probabilidad de falla

Ambos modelos coinciden en que 14.727 medidores tienen una alta probabilidad de fallar

¡Gracias!