

Universidad Tecnológica de Chihuahua
Tecnologías de la Información



**Universidad Tecnológica
de Chihuahua**

Reporte de Métricas de Evaluación

Alumno:

Jatzel Israel Cruz Castruita

Grupo:

IDGS91N

Materia:

Extracción de Conocimiento en Bases de Datos

Docente:

Enrique Mascote

Introducción.....	3
Investigación de métricas.....	4
Sección 2 – Solución de caso con KNN.....	7
Conclusiones.....	9
Referencias.....	10

Introducción

En el análisis de datos y la construcción de modelos de aprendizaje automático, la evaluación del desempeño es un proceso fundamental para garantizar que los modelos generados sean confiables, precisos y útiles en la práctica. Cada tipo de modelo requiere métricas específicas que permitan interpretar su comportamiento y determinar si las predicciones realizadas son adecuadas de acuerdo con el objetivo del estudio. En particular, los modelos de regresión y clasificación utilizan diferentes métodos de evaluación debido a la naturaleza de sus salidas: valores numéricos en el caso de la regresión y categorías o etiquetas en el caso de la clasificación.

El presente documento tiene como finalidad identificar, describir y comprender diversas métricas de evaluación utilizadas comúnmente en modelos de regresión y clasificación, analizando su definición, fórmula matemática, interpretación práctica, ventajas y limitaciones. Asimismo, se aplican estas métricas a un caso práctico utilizando el algoritmo de clasificación K-Nearest Neighbors (KNN), con el objetivo de reforzar su comprensión y observar su utilidad en un escenario real de análisis de datos.

Investigación de métricas

Métricas de Clasificación

Accuracy (Exactitud)

Definición:

Proporción de predicciones correctas sobre el total de predicciones

Fórmula.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Interpretación práctica:

Indica qué tan frecuentemente el modelo acierta. Útil cuando las clases están balanceadas.

Ventajas:

- Fácil de interpretar.
- Útil como primera métrica general.

Limitaciones:

Engañosa cuando las clases están desbalanceadas por ejemplo, 95% de una clase y 5% de otra.

Precisión

Definición:

De todas las predicciones positivas del modelo, cuántas realmente son positivas.

$$Precision = \frac{TP}{FP+TP}$$

Interpretación práctica:

Mide qué tan confiable es el modelo cuando predice la clase positiva. Importante cuando los falsos positivos son costosos.

Ventajas:

Útil en problemas donde es crítico evitar falsos positivos (ej: diagnóstico médico erróneo).

Limitaciones: Si se aumenta la precisión, usualmente disminuye el recall.

Recall

Definición:

El recall es una métrica que indica la proporción de casos realmente positivos que el modelo logra identificar correctamente. En otras palabras, mide qué tan bien el modelo es capaz de detectar todos los ejemplos que pertenecen a la clase positiva, considerando cuántos de ellos fueron clasificados correctamente frente a los que no logró reconocer.

Fórmula

$$Recall = \frac{TP}{FP + FN}$$

Interpretación práctica:

Indica la capacidad del modelo para encontrar casos positivos. Importante cuando los falsos negativos son muy costosos.

Ventajas:

Útil cuando no detectar un positivo tiene consecuencias graves, por ejemplo detección de enfermedades.

Limitaciones:

Aumentar el recall suele disminuir la precisión.

F1-Score

Definición:

El F1-Score es una métrica que representa la media armónica entre la precisión y el recall. Su función es equilibrar ambas medidas en un solo valor, especialmente útil cuando existe un desbalance entre clases o cuando es importante considerar tanto los falsos positivos como los falsos negativos. Al combinar precisión y recall, ofrece una visión más completa del rendimiento del modelo en la predicción de la clase positiva.

Fórmula

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Interpretación práctica:

Combina precisión y recall en un solo valor. Útil cuando las clases están desbalanceadas.

Ventajas:

- Fácil de interpretar.
- No penaliza fuertemente los outliers.

Limitaciones:

- No detecta grandes errores; todos pesan igual.

Sección 2 – Solución de caso con KNN

Preparación de datos:

Se dividió el conjunto en 70 % entrenamiento y 30 % prueba.

Se realizó escalado sobre las variables.

k	F1-score
3	0.75
5	0.8571
7	0.8571

Mejor modelo: $k = 5$ o $k = 7$

	Precision	Recall	F1-score	Support
0	0.83	1.00	0.91	5
1	1.00	0.75	0.86	4
accuracy			0.89	9
macro avg	0.92	0.88	0.89	9
weighted avg	0.91	0.89	0.89	9

Matriz de confusión

[[5,0] [1,3]

En este no hubo falsos positivos, pero sí hubo un falso negativo.

Análisis de resultados

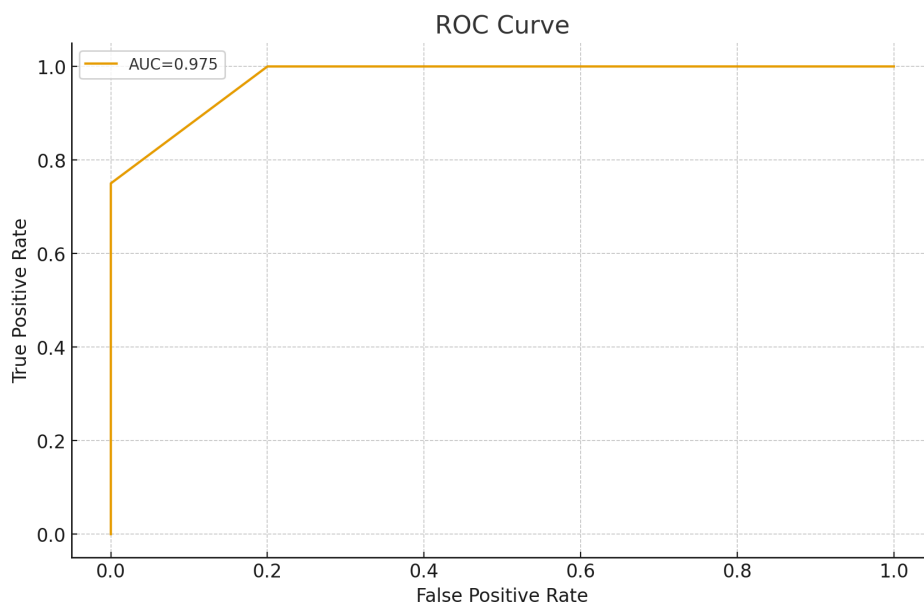
Comparación de los k

- $k=3 \rightarrow$ modelo más ruidoso, menor F1.
- $k=5$ y $k=7 \rightarrow$ mejor equilibrio, máximo F1-score.

El modelo final clasifica muy bien AUC alto y 89 % accuracy.

Posibles mejoras

- Probar más valores de k (1–20).
- Usar más variables predictoras si tu dataset lo permite.
- Balancear clases (si estuvieran desbalanceadas).
- Probar modelos alternativos:
- Logistic Regression
- SVM
- Random Forest



Conclusiones

Al realizar este ejercicio, entendí de manera más clara cómo se implementa y evalúa un modelo de clasificación usando K-Nearest Neighbors (KNN). Pude comprobar que la preparación de los datos es un paso fundamental, especialmente el escalado, ya que el rendimiento del modelo depende directamente de que todas las variables estén en la misma escala.

También aprendí la importancia de probar diferentes valores de k , porque noté que la elección del número de vecinos puede cambiar notablemente las métricas del modelo. En mi caso, descubrí que valores como $k = 5$ y $k = 7$ ofrecieron el mejor equilibrio, lo cual me permitió entender mejor cómo se comporta el modelo con diferentes configuraciones.

Además, al analizar las métricas, la matriz de confusión y la curva ROC, comprendí cómo interpretar el desempeño real del clasificador. El valor del AUC me permitió ver que el modelo tiene una muy buena capacidad de distinguir entre clases, y esto me dio más confianza en la calidad del resultado.

Referencias

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Geron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.