



Universidad Nacional Autónoma de México



Facultad de Ciencias

“Clasificadores”

Programación

Proyecto 2

Profesores:

- Allison Odette Merino Trejo
- Fernanda Sánchez-Puig
- José Fernando Méndez-Torres

Elaborado por

Alan Ledezma Puente

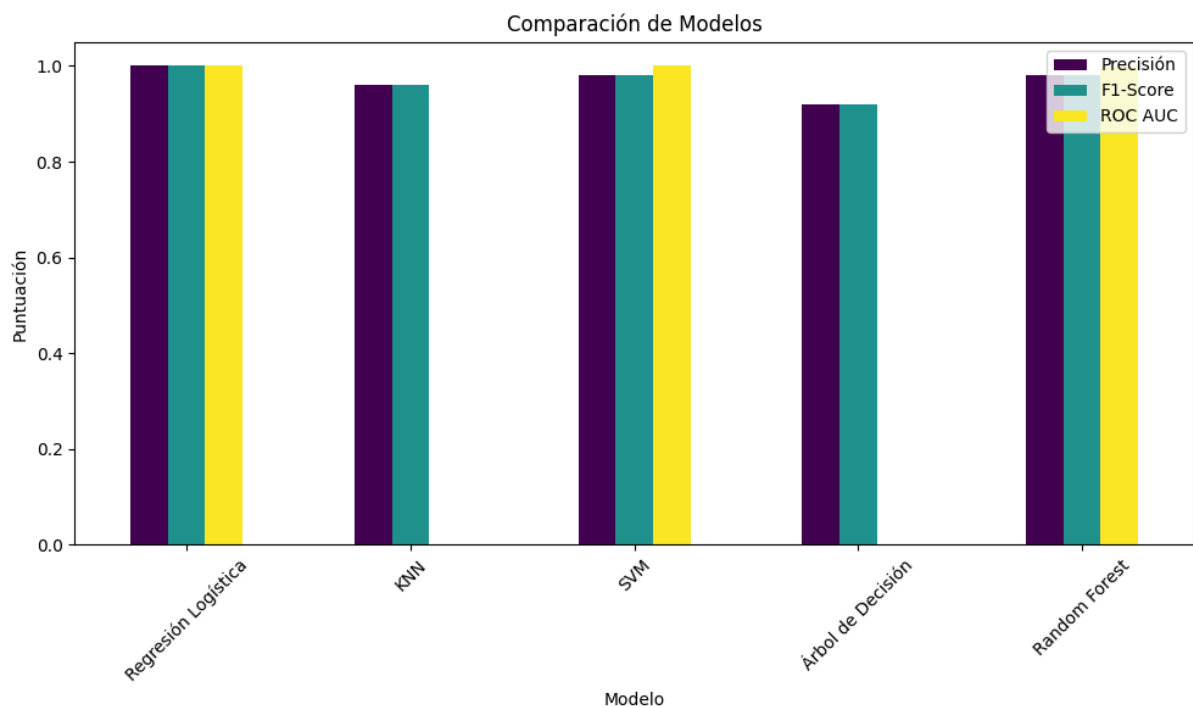
Este código lo que hace es recibir un archivo llamado "cáncer" que contiene 499 datos en el cual cada fila de datos determina si cada paciente es un paciente con cáncer de seno de acuerdo a las características, siendo "B" ó "M" y en el código se entrenan dos conjuntos (entrenamiento (80%) y de prueba(20%)), del cual cada clasificador tomará los conjuntos para entrenarse con los datos de entrenamiento y se evalúa usando los datos de prueba. Las métricas de evaluación incluyen precisión, ROC, AUC, matriz de confusión y un reporte clasificación, adicionalmente si las clases en el conjunto están desbalanceadas, el código lo detecta y avisa (no da técnicas para manejar ese desbalance).

De acuerdo a los datos obtenidos por los métodos de clasificación.

Comparación de Modelos:

Modelo	Precisión	F1-Score	ROC AUC
Regresión Logística	1.00	1.000000	1.000000
KNN	0.96	0.959514	NaN
SVM	0.98	0.979888	1.000000
Árbol de Decisión	0.92	0.920000	NaN
Random Forest	0.98	0.979888	0.999576

Gráficamente se vería de esta forma comparando los cada modelo



Dado a los resultados obtenidos en base a la precisión (con un 1.00), f1-Score (con un 1.000) y ROC AUC (1.000) el mejor modelo sería Regresión Logística, ya que tiene el mejor desempeño a comparación de los otros modelos.

Bibliografias (formato appa)

- GeeksforGeeks. (2023, 4 diciembre). *Logistic Regression using Python*. GeeksforGeeks. <https://www.geeksforgeeks.org/ml-logistic-regression-using-python/>
- *Arboles de decision python*. (s. f.). https://cienciadedatos.net/documentos/py07_arboles_decision_python
- *W3Schools.com*. (s. f.). https://www.w3schools.com/python/python_ml_knn.asp