Marco Jurado 20308
Oscar Lopez 20679

Laboratorio no. 3

Detección de Malware

Comparación de modelos implementados

Fueron implementados los modelos RandomForest y del MLPClassifier para las cuales discutiremos de las diferencias y las métricas obtenidas de los mismos.

El RandomForest mostró una superioridad en términos de Accuracy, Precision y ROC AUC con valores de 0.964, 0.987, y 0.992, respectivamente, frente a los valores de 0.949, 0.944, y 0.982 del MLPClassifier. Esto sugiere que el RandomForest es generalmente más efectivo en la correcta clasificación de las muestras, con una alta confiabilidad en sus predicciones positivas (malware), y una excelente capacidad para distinguir entre clases a lo largo de todos los umbrales de decisión. La superioridad en precisión indica que RandomForest tiene una tasa más baja de falsos positivos.

Por otro lado, el MLPClassifier mostró un mejor desempeño en Recall con un valor de 0.955 frente a 0.941 del RandomForest. Esto muestra que el MLPClassifier es ligeramente más capaz de identificar todas las instancias positivas reales (malware).

Las diferencias entre las métricas pueden atribuirse a las naturalezas y enfoques al aprender de los datos que utilizan. RandomForest, al ser un ensamble de árboles de decisión, es eficaz para manejar datos heterogéneos y puede capturar fácilmente relaciones no lineales y complejas, lo que podría explicar su alto desempeño en Precision y Accuracy. Mientras tanto, el MLPClassifier, como una red neuronal, puede requerir una configuración y entrenamiento más cuidadosos para alcanzar su máximo potencial. Sin embargo su estructura le permite modelar interacciones muy complejas entre las features, lo cual puede ser la razón detrás de su alto Recall.

Por lo tanto podemos decir que en cuanto al uso práctico el modelo de RandomForst es preferible pues tiene una mejor detección y una menor tasa de falsos positivos que MLPClassifier.