



Reporte de Desempeño de Clasificadores (TC3007C)

Clasificación de Correos Spam

Este informe resume el desempeño de varios modelos de clasificación entrenados con el archivo “baseline_clasificacion_spam_email.ipynb”, donde se realizaron las siguientes etapas:

- Lectura y pre-procesamiento de datos
- Procesamiento de datos
- Entrenamiento y evaluación de modelos de clasificación

Se probaron cuatro tipos de clasificadores de acuerdo con su complejidad, cuyos resultados se describen a continuación:

Complejidad de Modelos (de menor a mayor):

1. Regresión Logística: Modelo lineal básico para clasificación binaria.
2. Support Vector Machine: Modelo que optimiza márgenes, más sofisticado y no necesariamente lineal.
3. Random Forest: Ensamblado de árboles de decisión, más robusto y preciso.
4. Gradient Boosting Machine: Modelo secuencial de árboles, que mejora iterativamente.

Rendimiento de Modelos (ordenados por score de mayor a menor):

1. Regresión Logística:	0.9884
2. Random Forest con GridSearch:	0.9866666667
3. Gradient Boosting Machine:	0.985
4. Random Forest:	0.984
5. Support Vector Machine:	0.94

El modelo de Regresión Logística obtuvo el mejor desempeño, superando a los modelos más complejos como GBM, donde nos indica que en este conjunto de datos *los modelos más complejos no necesariamente son los que dan los mejores resultados*.