

Informe de Aprendizaje Automático

Chavez, Mauro
a@gmail.com

Lewkowicz, Iván
a@gmail.com

Drelewicz, Santiago
a@gmail.com

Torrez, Matías
matiastorrez157@gmail.com

Culaciatì, Dante
a@gmail.com

20 de abril de 2025

Índice

1. Ejercicio 1

En este inciso se pide separar los datos en conjuntos de entrenamiento y evaluación, donde no se debe utilizar la librería `train_test_split` de `sklearn`.

Primero se realizó una exploración de los datos, donde se observa que el dataset posee 200 features, todas numéricas, y 500 filas. Se observa que el dataset no tiene valores nulos y que además se trata de un problema desbalanceado, donde el 70% de los datos pertenecen a la clase 1 y el 30% restante pertenece a la clase 0, por lo que no es necesario realizar un preprocesamiento de los datos. Se decide entonces utilizar el 80% de los datos para entrenamiento y el 20% restante para evaluación.

Como la proporción de los datos es desbalanceada, realizamos un `stratified split` en la separación de los datos, procurando mantener la proporción del dataset original para los datos de entrenamiento y evaluación.

2. Ejercicio 2

Para la primera parte de este ejercicio, entrenamos un árbol de decisión con altura máxima 3 y estimamos la performance del modelo con K fold cross validation para distintas métricas.

Permutación	Accuracy (training)	Accuracy (validación)	AUPRC (training)	AUPRC (validación)	AUC ROC (training)	AUC ROC (validación)
1	0.8125	0.6375	0.6710	0.3226	0.8058	0.5298
2	0.840625	0.5875	0.7337	0.3337	0.8458	0.5246
3	0.825	0.6875	0.6431	0.3437	0.7513	0.5811
4	0.81875	0.7	0.6573	0.3626	0.7877	0.5938
5	0.84375	0.65	0.6958	0.4144	0.8085	0.5967
Promedios	0.828125	0.6525	0.6802	0.3554	0.7998	0.5651
Global	(NO)		(NO)		(NO)	

Cuadro 1: Resultados por permutación y métricas

3. Ejercicio 3

Describe los algoritmos y técnicas de aprendizaje automático utilizados. Incluye ecuaciones relevantes y justifique sus elecciones.

4. Ejercicio 4

Presente los resultados obtenidos, como métricas de evaluación, gráficos de desempeño, etc.

5. Ejercicio 5

Analice los resultados, las limitaciones del modelo y posibles mejoras.

6. Conclusión

Resuma los hallazgos principales y las conclusiones del informe.

Referencias

Incluya las referencias bibliográficas utilizadas en el informe.