# Checkpoint 3 - Grupo 39

## **Introducción**

En el preprocesamiento hicimos unas funciones para mostrar métricas y para exportar, así no repetimos código. También hicimos normalización y balanceo del dataset para mejorar los scores que vamos a entrenar. Luego entrenábamos modelos con parámetros por default para ver cual es el score base y luego buscábamos hiperparametros. Con respecto al dataset utilizamos One Hot Encoding para procesar las variables cualitativas.

**Construcción del modelo**

Para KNN optimizamos:

* Weights: ‘distance’
* n\_neighbors: 19
* Metric: ‘manhattan’
* Algorithm: ‘brute’

Para SVM Lineal:

* C: 10

Para SVM Polinomico:

* Gamma: 0.1
* degree: 1
* coef0: 0.5
* C: 1

Para SVM Radial:

* Probability: False
* gamma: 0.1
* C: 10

Para RF:

* min\_samples\_split: 10
* n\_estimators: 200
* min\_samples\_leaf: 1
* max\_features: log2
* max\_depth: None
* criterion: gini
* class\_weight: balanced
* ccp\_alpha: 0.0
* bootstrap: False

Para XGBoost:

* subsample: 0.8
* n\_estimators: 200,
* min\_child\_weight: 2,
* max\_depth: 8,
* learning\_rate: 0.2,
* gamma: 0.2,
* colsample\_bytree: 0.9

Stacking:

* base models: RF, SVM radial, XGBoost
* meta modelo: logistic regression

Voting:

## **Cuadro de Resultados**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Modelo | F1-Test | Presicion Test | Recall Test | Accuracy | Kaggle |
| KNN | 0.840 | 0.845 | 0.840 | 0.840 | 0.759 |
| SVM | 0.835 | 0.840 | 0.835 | 0.840 | 0.827 |
| Random Forest | 0.910 | 0.905 | 0.910 | 0.910 | 0.864 |
| XGBoost | 0.890 | 0.885 | 0.885 | 0.890 | 0.777 |
| Voting |  |  |  |  |  |
| Stacking |  |  |  |  |  |

Para KNN utilizamos el modelo con los parámetros optimizados ya que el accuracy era mayor que con parámetros default.

Para SVM usamos el Kernel radial que es especialmente útil para conjuntos no linealmente separables y optimizamos sus parámetros para tener el mayor f1 posible.

En cambio para Random Forest utilizamos los parámetros default ya que dio mejor score que con parámetros optimizados.

Para el XGBoost utilizamos optimizados ya que el f1 score era mayor que con parámetros default.

**Matriz de Confusion**

Mostrar la matriz de confusión de su mejor modelo y comentar brevemente lo que se puede observar.

## **Tareas Realizadas**

|  |  |
| --- | --- |
| Integrante | Tarea |
| Cattaneo Ariadna Antonella | Modelos: KNN, SVM  Preprocesamiento, Reporte |
| Agustin Ezequiel Sanchez Decouflet | Modelos: Voting, Stacking  Preprocesamiento, Reporte |
| Franco Dario Mazzaro | Modelos: XGBoost, Random Forest  Preprocesamiento, Reporte |