Submission No.	Date	Leaderboard position	Local train set score (cross validation)	Prediction score on test set	Data Preprocessing	Algorithm	Parameters of Algorithm
1	29/12/2020, 14:00	Last place	0.4576	0.2010	1. Drop column „id“ and „Descuento“ 2. Remove chars from numerical attributes of dtype string (Motor CC, Consumo, Potencia). E.g. 74 bhp -> 74. 3. Impute missing values of numerical attributes with mean value 4. Impute missing values of categorical attribute with most frequent value 5. Binary Encoding of categorical attributes 6. Normalizing numerical attributes	Random Forest	max_features=0.5,max_depth=10,min_samples_leaf=2, random_state=42
2	29/12/2020, 16:00	Last place	0.8594	0.20362	Same as in submission 1, except for **One Hot Encoding** instead of Binary Encoding in step 4	Random Forest	max_features=0.5,max_depth=10,min_samples_leaf=2, random_state=42
3	30/12/2020, 10:00	Last place	0.8470	0.23123	Same as in submission 2 (One Hot Encoding)	MLP Classifier (Multi-layer Perceptron)	hidden_layer_sizes = [100]*3
4	30/12/2020, 14:00	Last place	0.40133	0.22001	Same as in submission 1, except for Label Encoding of **all** the variables in step 4	XGB Classifier (Extreme Gradient Boosting from xgboost library)	default parameters
5	31/12/2020, 00:23	23	0.8319	0.76790	1. Drop columns „id“ & „Descuento“ 2. Impute missing values of **all** attributes using the most frequent value 3. Use Label Encoding on **all** attributes	XGB Classifier	default parameters

Submission No.	Date	Leader board position	Local train set score (cross validation)	Prediction score on test set	Data Preprocessing	Algorithm	Parameters of Algorithm
6	31/12/2020, 11:00	23	0.8373 on full data 0.8360 on hold out validation set with early stopping	0.76445	Same as in submission 5	XGB Classifier Create hold out set from training set, to use early stopping with the hold out set as the validation set during model fitting.	objective="multi:softmax", min_child_weight=1, learning_rate=0.5, colsample_bytree = 0.7, n_estimators=70, max_depth=4
7	31/12/2020, 12:52	11	0.8373	0.79378	Same as in submission 5	XGB Classifier without early stopping: fit on full train set	Same as in submission 6
8	31/12/2020, 14:38	8	0.8424	0.80155	Same as in submission 5 + Feature Selection: drop attributes „Kilometros“ and „Mano“	XGB Classifier	Same as in submission 6
9	01/01/2021, 22:00	11	0.8448	0.79982	Same as in submission 8	XGB Classifier	objective="multi:softmax", scale_pos_weight=1, **learning_rate=0.1**, colsample_bytree = 0.8, subsample = 0.8, **n_estimators=510**, reg_alpha = 0.3, max_depth=4, gamma=0.6

Submission No.	Date	Leader board position	Local train set score (cross validation)	Prediction score on test set	Data Preprocessing	Algorithm	Parameters of Algorithm
10	01/01/2021, 00:30	10	0.8425	0.80241	Same as in submission 8 + Class Balancing: Oversampling **the minority class** using SMOTE algorithm (to 400 samples instead of 269)	XGB Classifier	XGB: same as in sub. 9, n_estimators = 400 SMOTE: k_neighbors=2 , sampling_strategy = samples -> samples = {**1:400**, 2: …}
11	01/01/2021, 00:43	8	0.8392	0.80845	Same as in submission 8 + Class Balancing: Oversampling **the minority class** using **BorderlineSMOTE** algorithm (to 400 samples instead of 269)	XGB Classifier	XGB: same as in sub. 9, n_estimators = 400 BorderlineSMOTE: k_neighbors=3, sampling_strategy = samples -> samples = {**1:400**, 2: …}