|  |  |
| --- | --- |
| Model dataset size: 1e5 / Original: 7e6 | |
| Imbalanced dataset (“as is”) | |
| Correlation Heatmap | |
|  | |
| One Hot Encoding (220 columnas) | [Additive-Smoothing Target Encoding](https://maxhalford.github.io/blog/target-encoding/) (27 columnas)  [probar [K-Fold Target Encoding](https://medium.com/@pouryaayria/k-fold-target-encoding-dfe9a594874b)?] |
| Model fitting time: 7h 50m 28s | Model fitting time: 1h 53m 17s |
| Best scorer: | Best scorer: |
| -+-+-+ TRAINING dataset +-+-+-  Confusion matrix:  [[41567 15293]  [ 3400 9740]]  Normalized confusion matrix:  [[0.73104115 0.26895885]  [0.2587519 0.7412481 ]]  precision recall f1-score support  on-time 0.92 0.73 0.82 56860  delayed 0.39 0.74 0.51 13140  accuracy 0.73 70000  macro avg 0.66 0.74 0.66 70000  weighted avg 0.82 0.73 0.76 70000  F-beta (ß=2) = 0.628  F1 = 0.510  Recall = 0.741  Precision = 0.389  Accuracy = 0.733  -------------------------------------------------------  -+-+-+ TEST dataset: +-+-+-  Confusion matrix:  [[8403 3735]  [1158 1704]]  Normalized confusion matrix:  [[0.69228868 0.30771132]  [0.40461216 0.59538784]]  precision recall f1-score support  on-time 0.88 0.69 0.77 12138  delayed 0.31 0.60 0.41 2862  accuracy 0.67 15000  macro avg 0.60 0.64 0.59 15000  weighted avg 0.77 0.67 0.71 15000  F-beta (ß=2) = 0.505  F1 = 0.411  Recall = 0.595  Precision = 0.313  Accuracy = 0.674  -------------------------------------------------------  -+-+-+ VALIDATION dataset: +-+-+-  Confusion matrix:  [[8358 3734]  [1163 1745]]  Normalized confusion matrix:  [[0.69120079 0.30879921]  [0.39993122 0.60006878]]  precision recall f1-score support  on-time 0.88 0.69 0.77 12092  delayed 0.32 0.60 0.42 2908  accuracy 0.67 15000  macro avg 0.60 0.65 0.59 15000  weighted avg 0.77 0.67 0.70 15000  F-beta (ß=2) = 0.510  F1 = 0.416  Recall = 0.600  Precision = 0.318  Accuracy = 0.674 | -+-+-+ TRAINING dataset +-+-+-  Confusion matrix:  [[54229 2442]  [ 318 13011]]  Normalized confusion matrix:  [[0.95690918 0.04309082]  [0.02385775 0.97614225]]  precision recall f1-score support  on-time 0.99 0.96 0.98 56671  delayed 0.84 0.98 0.90 13329  accuracy 0.96 70000  macro avg 0.92 0.97 0.94 70000  weighted avg 0.97 0.96 0.96 70000  F-beta (ß=2) = 0.946  F1 = 0.904  Recall = 0.976  Precision = 0.842  Accuracy = 0.961  -------------------------------------------------------  -+-+-+ TEST dataset: +-+-+-  Confusion matrix:  [[10721 1378]  [ 1899 1002]]  Normalized confusion matrix:  [[0.88610629 0.11389371]  [0.65460186 0.34539814]]  precision recall f1-score support  on-time 0.85 0.89 0.87 12099  delayed 0.42 0.35 0.38 2901  accuracy 0.78 15000  macro avg 0.64 0.62 0.62 15000  weighted avg 0.77 0.78 0.77 15000  F-beta (ß=2) = 0.358  F1 = 0.379  Recall = 0.345  Precision = 0.421  Accuracy = 0.782  -------------------------------------------------------  -+-+-+ VALIDATION dataset: +-+-+-  Confusion matrix:  [[10612 1442]  [ 1892 1054]]  Normalized confusion matrix:  [[0.88037166 0.11962834]  [0.64222675 0.35777325]]  precision recall f1-score support  on-time 0.85 0.88 0.86 12054  delayed 0.42 0.36 0.39 2946  accuracy 0.78 15000  macro avg 0.64 0.62 0.63 15000  weighted avg 0.76 0.78 0.77 15000  F-beta (ß=2) = 0.369  F1 = 0.387  Recall = 0.358  Precision = 0.422  Accuracy = 0.778 |
|  |  |
| Demasiadas filas… (220 features) | Model’s built-in Feature Importance |
| Demasiadas filas… (220 features) | sklearn.inspection / permutation\_importance |
| Demasiadas filas… (220 features) | xgboost / plot\_importance / importance\_type='weight' |
| Demasiadas filas… (220 features) | xgboost / plot\_importance / importance\_type=’gain’    *(Parece ser el mismo que el primero: sklearn.inspection / permutation\_importance)* |