



Encoding

Encoding of categorical variables such as country names (za, USA,...), product categories (jewelry, clothings,...)

Standardisation

Bring our data back to a scale that's conducive, for our algorithms to train and analyze better. Techs : StandardScaler()

Train/Test data partitioning.

Defining evaluation function including metrics, confusion matrix, classification report and Learning curves.

Creating process pipelines.

Oversampling

Only 5% of robotic auction samples, so we used the SMOTE technique to make our class equal.

Pre-Processing

Predictive modeling & evaluation

scikit learn

encoder_merchan-dise.pkl
encoder_country.pkl
scaler.pkl
model.pkl

Model generation for deployment

Pickle

Selecting high-performance models

Metrics taken into consideration :
-Model Tuning
-MCC
-F1-score

Optimizing algorithms

-RandomizedSearchCV
-GridSearchCV

Building and Testing Models

- Logistic Regression, LR with PCA, RandomForest, SVM, KNN, DNN, ...