Customer Relationship Prediction

By Flipping -A -Coin
Shambhavi Aggarwal, Lakshay Chawla,
Devansh Shrestha, Rommel Jalasutram

Univ.AI

About the project

Orange, a French Telecom Company has collected data of their customer activities. Based on the data, we are predicting the propensity of customers to switch providers (churn), buy new products or services (appetency), or buy upgrades or add-ons proposed to them to make the sale more profitable (up-selling).

Data Available:

- Both training and test sets contain 50,000 examples.
- Data sets have numerical and categorical variables.
- The first 190 variables are numerical and the last 40 are categorical.

Labels:

There are three target values to be predicted:

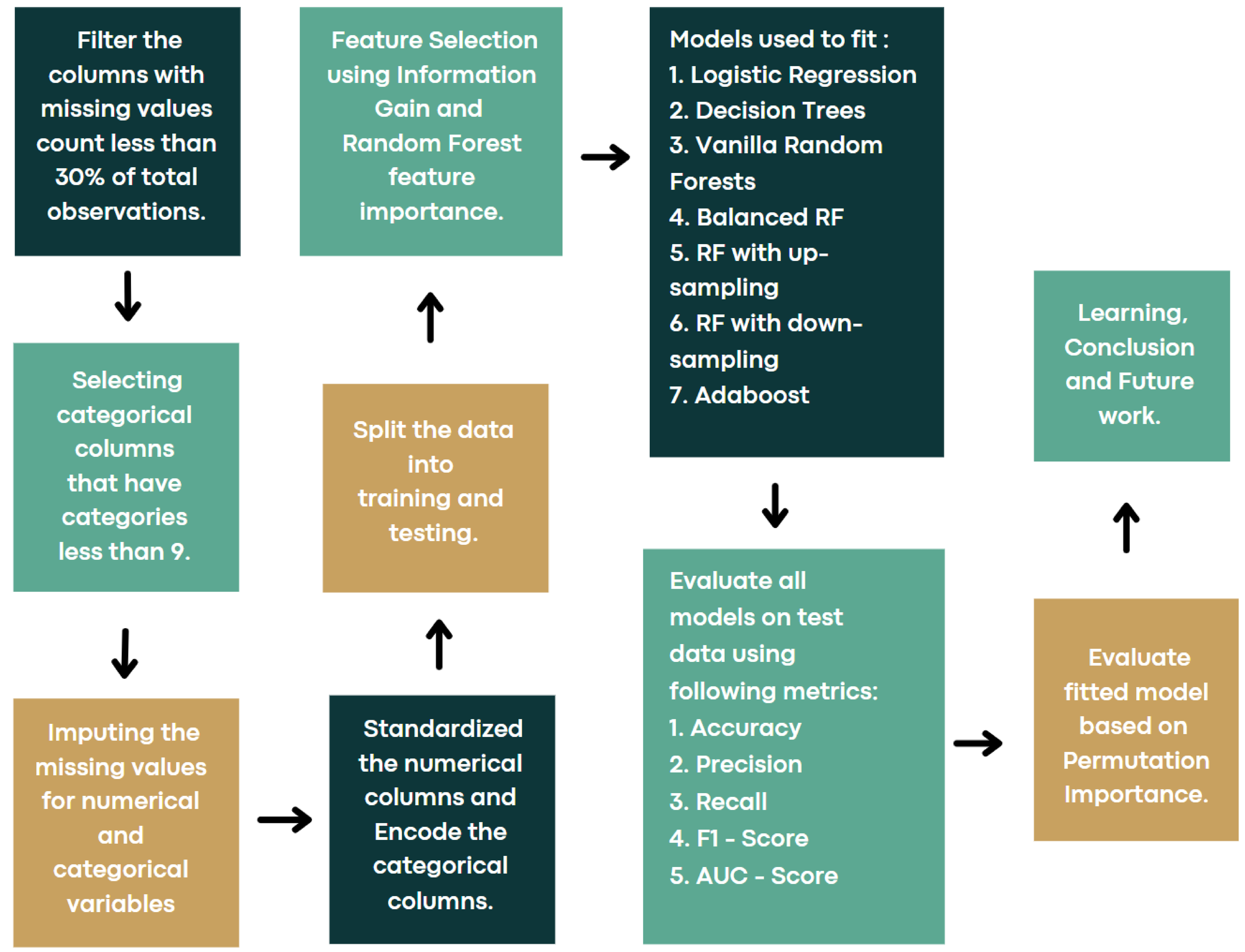
- Churn
- Appetency
- Upselling

Results/Findings:

- Upselling
  - Logistic Regression: Lowest F1 and AUC score.
  - AdaBoost: Highest F1 and high AUC score.
- Appetency
  - Logistic Regression: Lowest accuracy, low F1 & highest AUC score.
  - AdaBoost: Highest accuracy, zero F1, and lowest AUC score.
- Churn
  - Logistic Regression: Low accuracy, low F1, and lowest AUC score
  - AdaBoost: Highest accuracy, zero F1, and high AUC score.

IMPORTANT!

Target Variabl;e is heavily imbalanced.



Model Name	Accuracy	Recall	Precision	F1 - Score	AUC-Score
Logistic Regression	0.7005	0.6752	0.1527	0.2491	0.7829
Decision Tree	0.9471	0.4266	0.7458	0.5427	0.8386
Vanilla RF	0.9472	0.3668	0.8132	0.5056	0.8373
Balanced RF	0.872	0.5679	0.3028	0.3950	0.8427
RF Smote	0.8915	0.5122	0.3417	0.4100	0.8282
RF Downsample	0.7464	0.7119	0.1839	0.2924	0.8404
Adaboost	0.9492	0.4266	0.7850	0.5528	0.8394

Conclusion

- Tree models performed well.
- Data highly imbalanced.
- Models highly biased towards majority class as a result.
- Grid search while useful is computationally expensive.

Future scope

- Train with fewer predictor variables.
- Improve performance by balanced datasets, resampling.
- Try SVM and XGBoost on these datasets.

References:

The data: https://kdd.org/kdd-cup/view/kdd-cup-2009/Data