Capstone Presentation

Customer Churn (E-commerce)

Business Problem Summary

- A **leading e-commerce company** wishes to be able to:
  - **predict which of their customers are likely to churn,**
  - so as to be able to take suitable measures to prevent the customers from leaving.

- The given dataset contains attributes of the company's customers, and aspects of their behaviour on the company's e-commerce website and app:
  - The records contain demographic details, such as gender, age, marital status, address etc.
  - The records also reflect the customer's buying behaviour such as their preferred category, preferred login device, frequency of purchase, recency of transaction, satisfaction scores etc.
  - There are 5630 unique customer records, and 20 features per record, in the dataset.

- The **objective** of the capstone project is to:
  - mine patterns in the customer data and
  - analyse what factors typically lead to customer churn, and
  - build a model that will predict churn based on the important features thus identified.

Modeling

Part 1

- A range of models were built on the training data, and validated on the test data. And the performances were tabulated and compared.
- The principal objective of the exercise is to identify customers who are likely to churn. So our **focus was on the Recall and Precision metrics for Class Label 1 (Churn).**
- **Artificial Neural Network was the best performing model** on all counts, and it was seen that almost all the other models performed poorly on Recall and Precision metrics for Churn.
- We had noted the **imbalance in the dataset** in the course of EDA: Churn labeled records account for only 17% of the dataset. And hence this performance wasn't unexpected.

Evaluation Metrics on Test Data for Churn

	Recall (Churn)	Precision (Churn)	Accuracy	AUC
Logistic Regression	0.52	0.78	0.89	0.74
Linear Discriminant Analysis	0.52	0.79	0.89	0.74
K Nearest Neighbour	0.54	0.79	0.9	0.76
Naive Bayes	0.73	0.36	0.73	0.73
SVM	0.54	0.88	0.91	0.76
Decision Tree	0.57	0.7	0.88	0.76
Artificial Neural Network	0.79	0.87	0.94	0.88

Modeling

Part 2

- To improve model performance:
  - **tuned the models** to mainly account for class imbalance,
  - applied **Bagging and Boosting** techniques,
  - **optimised hyper parameters**
- Tried synthetically oversampling the minority class using SMOTE, but the technique didn't prove particularly useful in producing better models in this instance.
- Based on the performance metrics and cross-validation scores, the following 3 models have proved to be the best predictors of customer churn in this case:
  - **Support Vector Machine (tuned)**
  - **XGBoost (tuned)**
  - **ANN (tuned)**

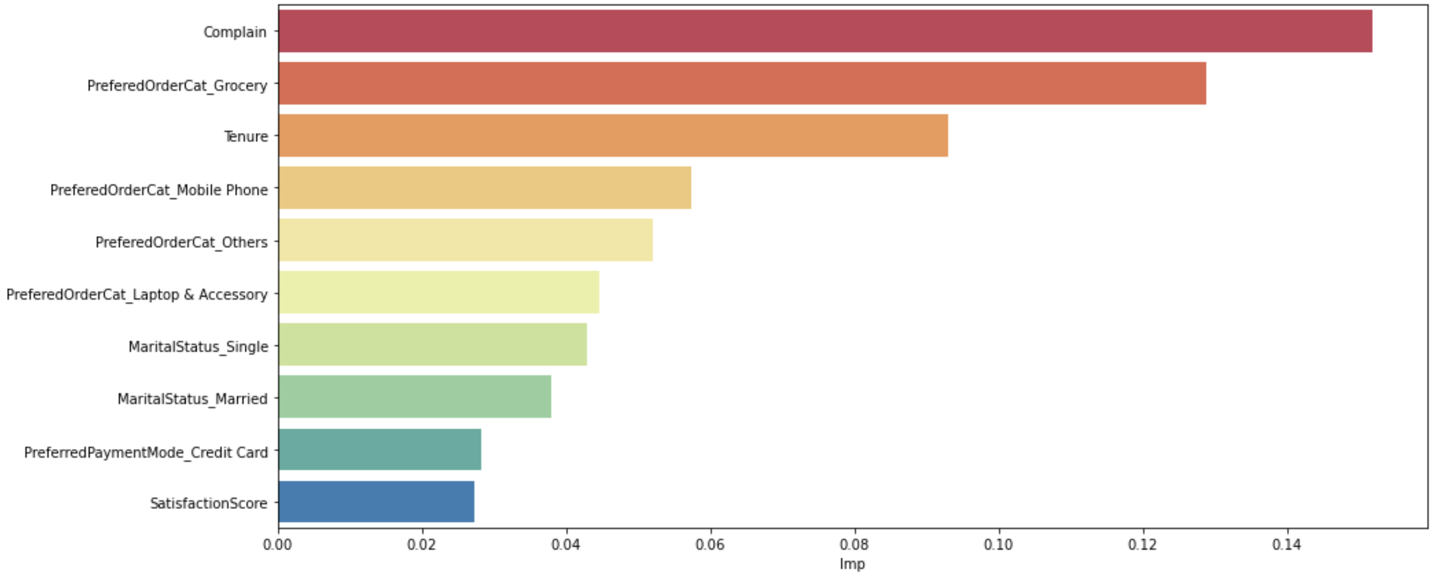
Evaluation Metrics on Test Data for Churn

	Recall (Churn)	Precision (Churn)	Accuracy	AUC
SVM (Tuned)	0.91	0.8	0.94	0.93
Logit (Tuned)	0.8	0.47	0.81	0.81
Artificial Neural Network (Tuned)	0.85	0.93	0.96	0.92
Bagging: Random Forest	0.81	0.52	0.84	0.83
Boosting: XGBoost	0.89	0.88	0.96	0.93

Insights from Analysis

Some of the **segments that are more likely to churn** than others are:

- **People who have made a Complain in the last month**: 37% of customers that complained in the last month, churn.
- **Singles**: 27% of customers that are Single, churn.
- **Customers that prefer buying Mobile Phones**: 27% of those that prefer Mobile Phone, churn.
- **Customers that prefer Cash on Delivery and E-wallets**: ~25% of customers who prefer COD and E-wallets, churn.
- **Customers with low Tenure**, typically 7 days or less.



- The chart above lists the top 10 features, as determined by the XGBoost model.
- This corroborates largely how our Linear models also weighted their important features, and what we inferred in the course of our EDA.
- **Complain, Tenure, Preferred Order Category, Marital Status, and Preferred Payment mode are by and large important discriminators.**

Recommendations

COMPLAINT REDRESSAL as an opportunity:

- Complain appears to be the biggest signifier of Churn. **Close to 40% of customers that placed a complaint in the last month, churned**.
- So the immediate opportunity would be to ensure **resolving the customer's complaint to their satisfaction and preferably, delight**, giving them a reason to continue on the platform.
- Also, feedback from the complaints need to be actioned to **improve service**

MOBILE PHONE CATEGORY as an area of focus:

- Customers that prefer buying mobile phones are the **largest segment** from a preferred category standpoint.
- This is also the segment that has a high probability of churn (~ 30%). Mobile phone buyers are possibly loyal to whichever platform offers them the best deals.
- Given that they form a significant portion of customers, special focus to be given to the category itself, to induce them to prefer this platform over others.

SINGLES SEGMENT as an area of focus:

- Singles are at twice as likely to churn than married and divorced customers.
- Perhaps the offering on the platform caters more to non-singles, and it would help to explore if **offerings can be tailored keeping Singles in mind.**

TENURE EXTENSION as a strategy:

- Ultimately if the above measures, and the service in general, succeed in **keeping customers engaged on the platform for longer than 8 months**, the possibility of churn is significantly lowered.
- Hence a customer needs to be particularly attended to and engaged more in the initial months on the platform, to effect their continued association.

Thank you.

Shubhadeep Bhattacharya
Program: PGP DSBA
Batch: October 2019