CUSTOMER CHURN

CAPSTONE PROJECT

PGDSBA

SANJAM PREET SINGH BHULLAR

Business problem understanding

Defining problem statement

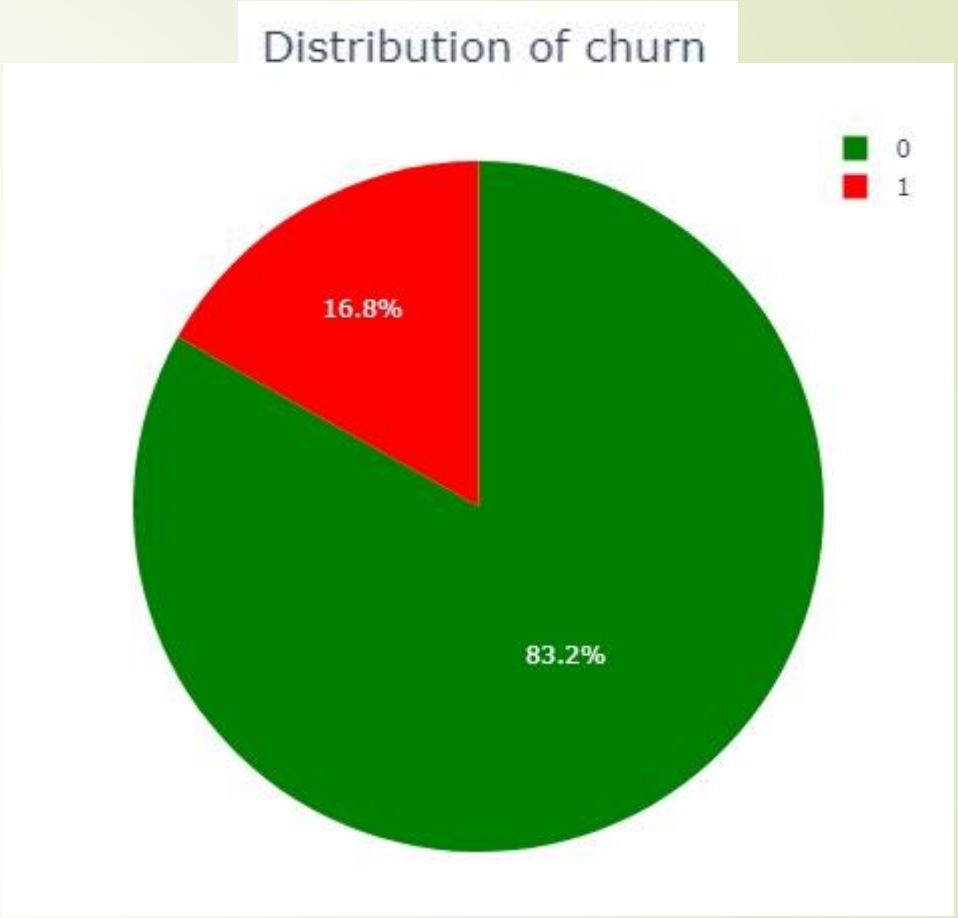
- The dataset is about an e-commerce company that has been hit by a customer churn rate.
- The company wants to build a model that predicts the churn rate and helps to draw invaluable insights.
- It is important to identify customers who might churn as the cost of acquiring new users is more than retaining the existing ones.

Objective of the study

- Propose an actionable roadmap aimed at retaining customers who are on the verge of churning.
- Develop prediction models and identify different variables related to customer churn.

Constraints to churn prediction

- The limitation to churn rate prediction is the imbalanced data. It means the number of churn customer constitutes a small portion of the data.
- There is ‘noise’ in the data. In other words, there are outliers in the dataset.
- The dataset has bad data. As a result, certain numerical variables have been typecast as ‘object’.
- Null values comprise 1.8% of the data points.



Methodology

Missing value treatment

- Numerical values have been imputed with median as the data is skewed.
- Categorical values have been imputed with mode.

Outlier treatment

- To ensure maximum number of data points are included for model-building, numerical variables are log transformed.
- Another reason to transform the attributes is to reduce the variation in the dataset.

Modelling approach used & why

The following four approaches will be undertaken to build several models.

- **Model-building with imbalanced data:** Different models such as Logistic Regression, LDA, SVM, Decision Tree and Random Forest have been built with the imbalanced data.
- **Model-building with balanced data:** The Sampling Minority Oversampling Technique (SMOTE) have been employed to deal with the class imbalance.
- **Ensemble modelling:** Different ensemble techniques such bagging and boosting have been used to build machine learning models.
- **Model-building with hypertuning parameters:** Hypertuning parameters have been used to improve the models' performance.

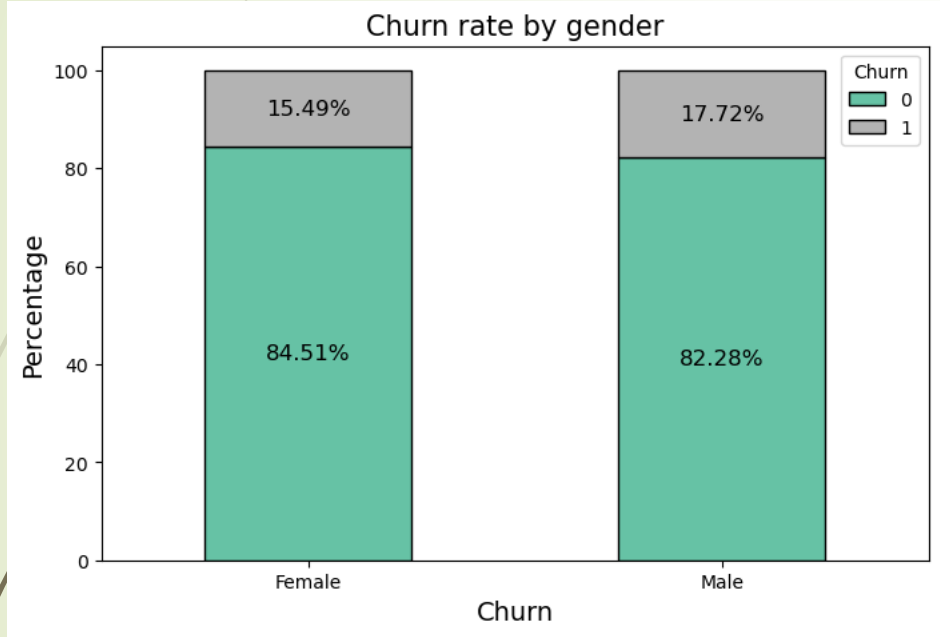
Best model

- In all, 26 models have been built for the customer churn prediction.
- Some of the models such as KNN, Decision Tree and Random Forest are overfitting.
- The best model is Gradient Boost. Its performance metrics are:

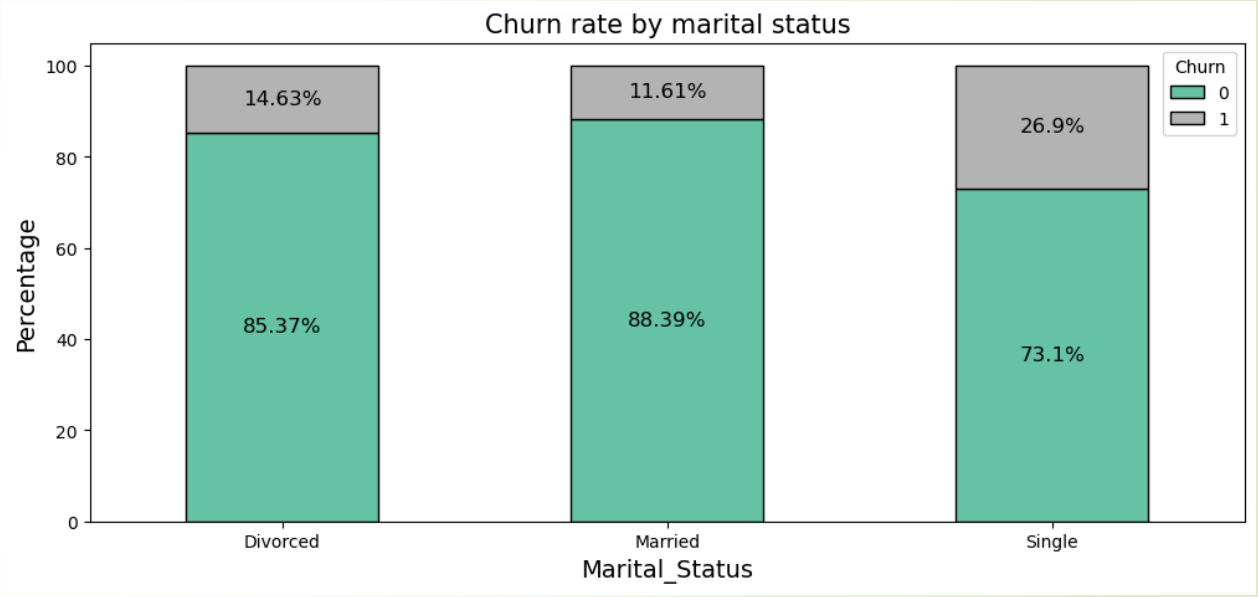
Accuracy	**Recall**	**Precision**	**F1 score**	**AUC score**
Train: 0.92 Test: 0.91	Train: 0.64 Test: 0.60	Train: 0.85 Test: 0.83	Train: 0.73 Test: 0.70	Train: 0.95 Test: 0.93

- The model's performance is consistent over the training and unseen data.

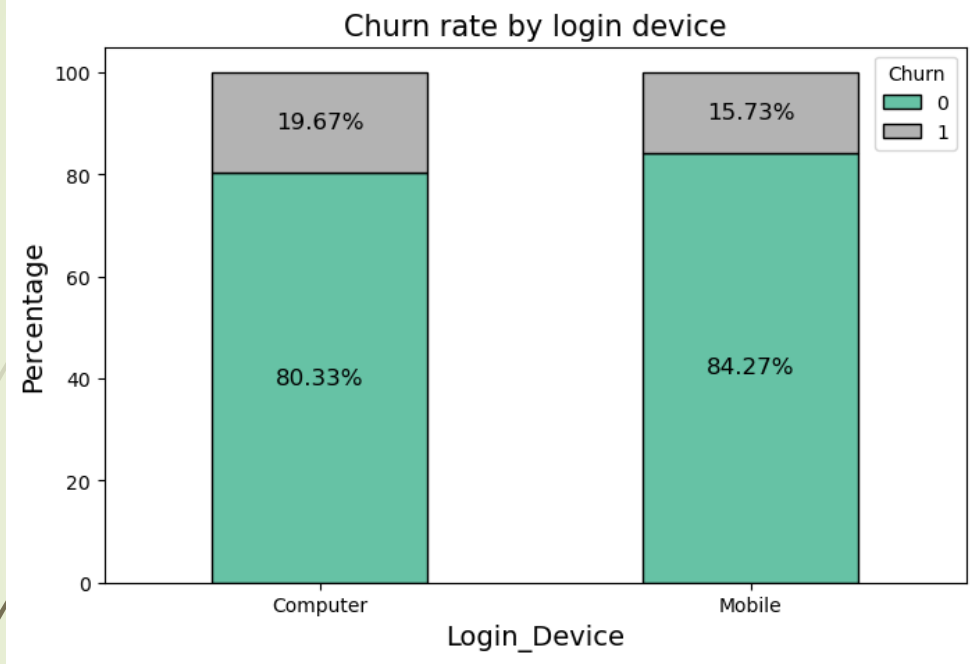
Business insights

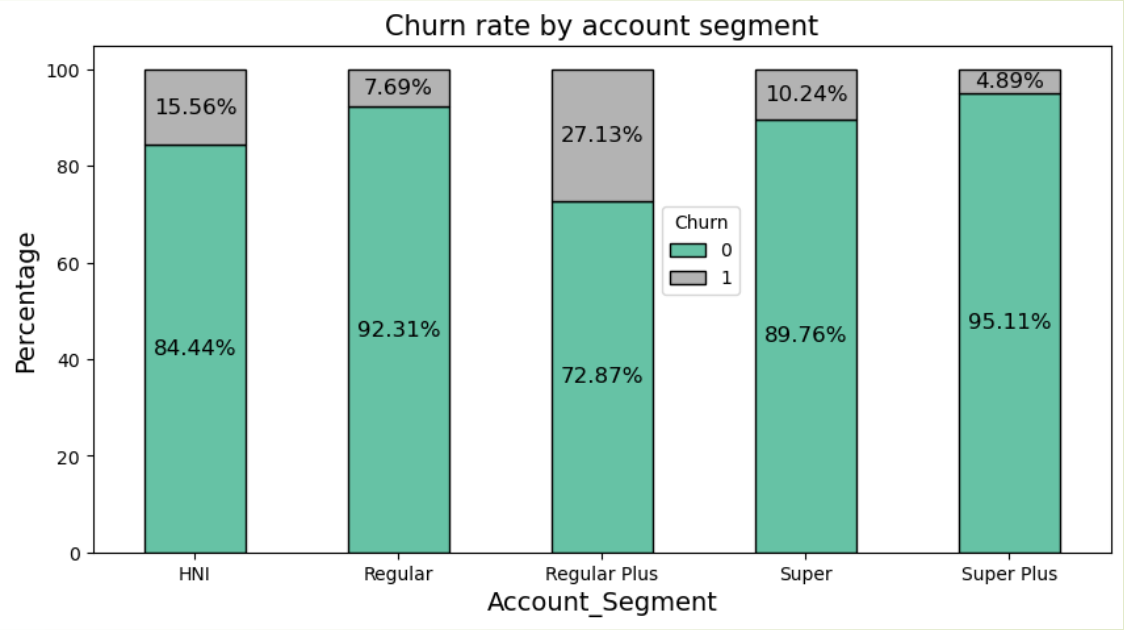


Males have a higher tendency to churn.



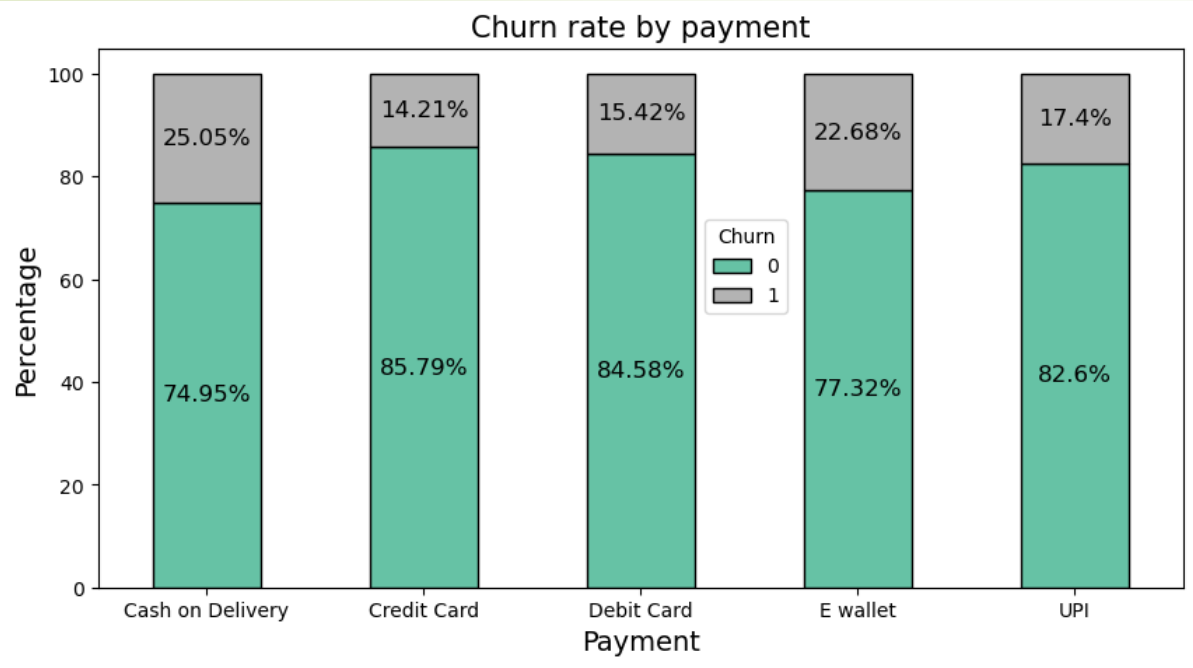
Single persons are more likely to churn.

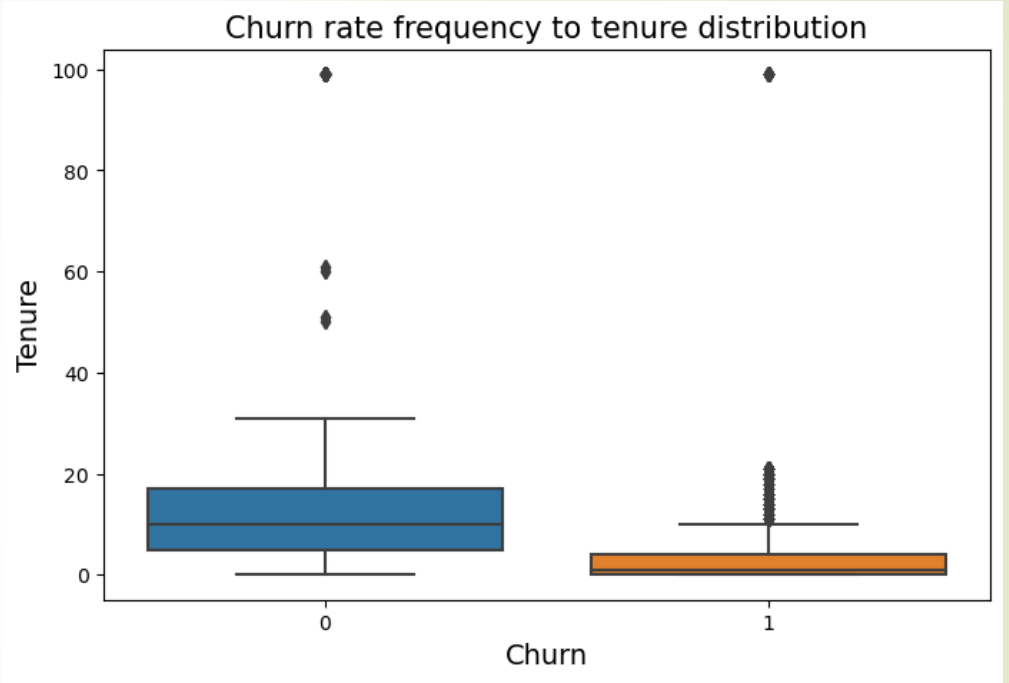




Computer users have a higher probability to churn than those using mobile.

Regular Plus users are more likely to churn followed by HNI subscribers.

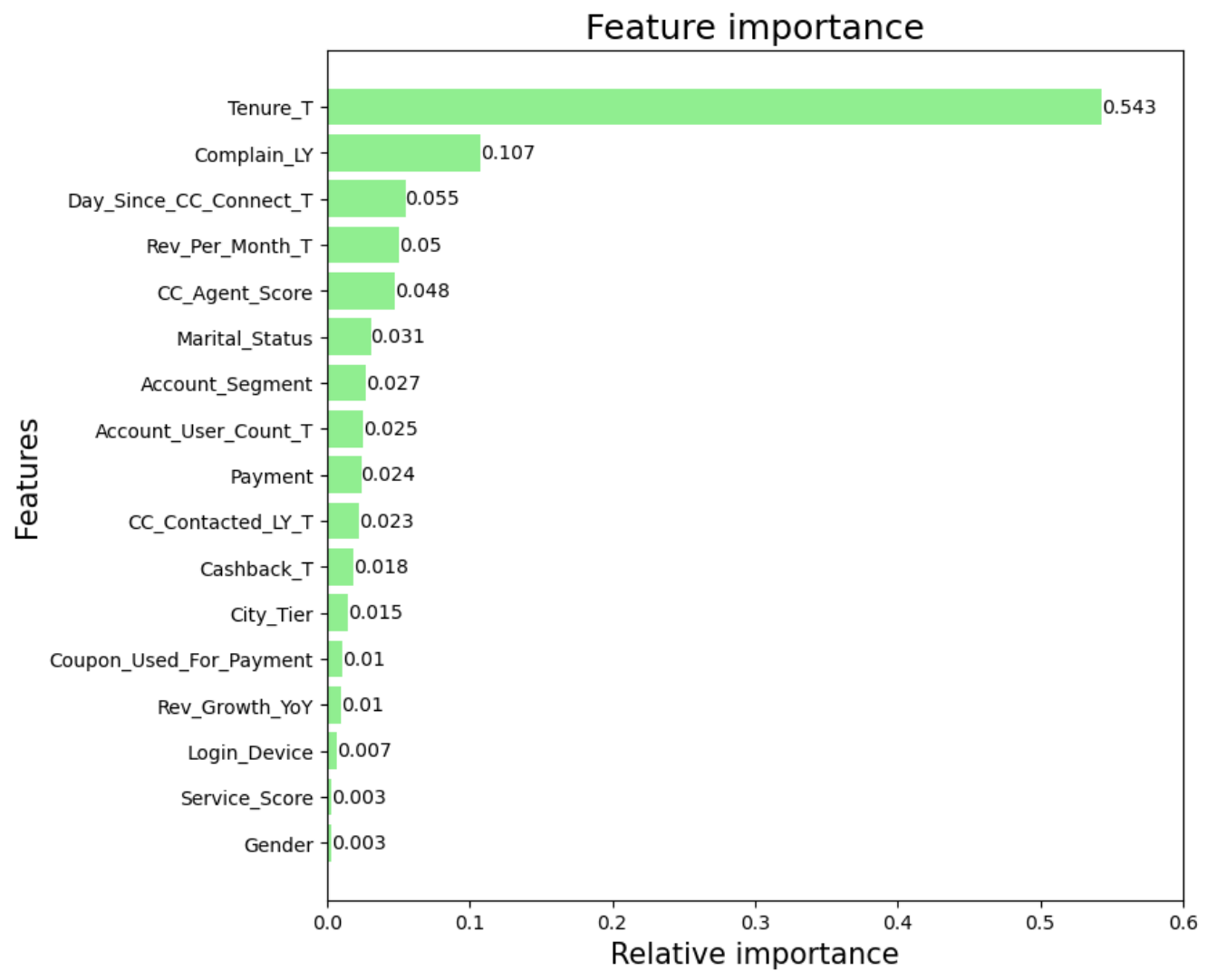




Customers who prefer cash on delivery are more likely to churn followed by e-wallet users.

Tenure has a significant impact on the churn rate. Churned customers have a shorter tenure.

Recommendations



The top five features based on the Gradient Boost model are:

1. Tenure
2. Complain last year
3. Days since CC connect
4. Revenue per month
5. CC agent score

Recommendations based on the important features are as follows:

- **Tenure:** The business team must focus on increasing the tenure of customers by offering them special pricing and long-term plans.
- **Complain last year:** The team must ensure that complaints are handled promptly. The e-commerce company must ensure that customer care service is proficient enough to deal with complaints.
- **Days since CC connect:** Some users churn silently. Even if users have not contacted the customer care for days, the company must conduct telephone surveys to get the clients' feedback.
- **Revenue per month:** Churned customers are generating slightly more revenue. Therefore, the focus should be on finding the root cause of their complaints so that such users can be targeted in a better manner.
- **CC agent score:** The majority of the customers gives a rating of 3 to customer care agents. The agents must ask for feedback so that new policies can be devised that will enhance the customers' experience.