DATA REPORT

PREDICTING CUSTOMER CHURN

BUSINESS UNDERSTANDING

Capturing customer churn is important for several reasons. First and foremost, churn can have a significant financial impact on a business, as losing customers means losing revenue. By tracking churn, companies can identify the financial impact and take steps to reduce it. Secondly, understanding why customers are leaving can provide valuable insights into the company's products, services, and customer experience. This information can help improve customer satisfaction and reduce churn in the future. Additionally, by tracking customer churn, companies can identify at-risk customers and take steps to intervene before they leave. This can include offering incentives, addressing customer concerns, or making changes to improve the customer experience. Finally, having accurate and up-to-date information on customer churn enables companies to make informed decisions about resource allocation and business strategy. This dataset contains various. Data used to predict contains numerous columns about customers such as This dataset contains 3333 instances and 21 columns. Columns contain different information on customers such as state, account length, area code, phone number,'international plan', 'voice mail plan', 'number vmail messages','total day minutes', 'total day calls', 'total day charge','total eve minutes', 'total eve calls', 'total eve charge','total night minutes', 'total night calls', 'total night charge','total intl minutes', 'total intl calls', 'total intl charge','customer service calls', 'churn (target)'

DATA REPORT CUSTOMER CHURN

OBJECTIVES

The main objective of this project is to proactively address customer churn, which can have a significant financial impact on a business. **By developing a predictive model, the project aims to identify customers who are at risk of churning, so that targeted actions can be taken to prevent them from leaving**. This is in line with the importance of preserving existing customers, as keeping existing customers is often more cost-effective than acquiring new ones. By flagging customers who are likely to churn, the company can take steps to improve their customer experience, address any concerns they may have, or offer incentives to retain them as customers. The predictive model will be based on historical data and machine learning algorithms, and will be regularly updated to ensure its accuracy. The successful implementation of this project will result in a more stable and growing customer base for the company. **Additionally, the project seeks to uncover important determinants that can affect the decision of a customer of either to churn or not churn.**

DATA UNDERSTANDING

Data used for this project contains 20 columns where 5 are categorical and 15 are numerical.

Data for this project was obtained from **Kaggle** a popular data science platform.

DATA CLEANING AND WRANGLING

Data was in form of a comma separated file (csv) and was loaded into python using python's data wrangling library 'pandas'

Data was checked for any missing values, outliers and duplicates and appropriate action taken where necessary. Outliers in the dataset where kept as they would form a crucial part of the training data. Removing them would mean that the model would miss on the opportunities of learning from such instances in order to predict future records well.

DATA REPORT CUSTOMER CHURN

EXTERNAL DATA VALIDATION

Data was validated against a outside source to ensure that it was fit for use for coming up with a predictive model. According to this report, the values in the dataset were well within what was expected.

EXPLORATORY DATA ANALYSIS

Data was visualized using different charts such bar charts, and heatmaps to unveil hidden patterns that were present in the data before modelling.

Questions asked in the exploratory stage included:

- Distribution of continuous variables
- Churn count
- Which region had the highest number of customers
- Which region had the highest churn rate
- Does international plan has some sort of relationship with churning
- Does voice mail plan has some sort of relationship with churning
- Average customer charge for churning and non-churning customers
- Total day charge for churning and non-churning customers
- Total evening charge for churning and non-churning customers
- Today day calls churning and non-churning customers
- Total night calls for churning and non-churning customers

MODELLING

DATA PREPROCESSING

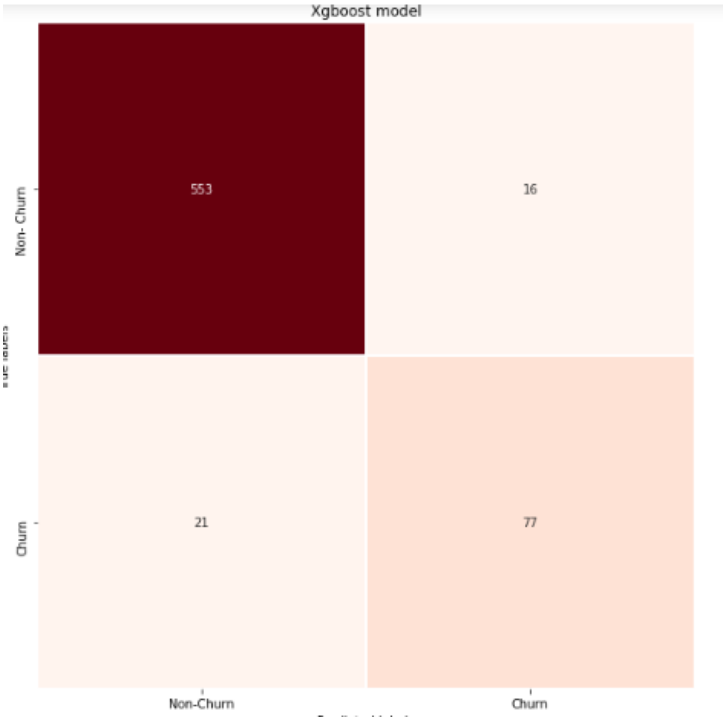
The first preprocessing step was to convert categorical data into dummy variable. This is in line with machine learning principles that require categorical data to be converted into numeric data

that can be used for prediction. Secondly, data was also normalized to make sure that all variables were on the same scale. Finally, due to the class imbalance data was resampled using Synthetic Minority Oversampling Technique (SMOTE) to ensure that there were enough instances of the minority class for the prediction model to train on.

ALGORITHMS USED

For an initial baseline model, a logistic regression model was fitted on the training data and tested on the testing data. Subsequent models such *random forest, support vector machine, Xgboost, and Adaboost* were fitted on the training and evaluated with the test set.

After fitting the models, the two best models (*random forest and tuned Xgboost model*) were compared against each other to determine which solved best the problem of customer churn. The Xgboost model was chosen as the best model as it had an overall accuracy of **94%** and a recall of **79%** which were pretty impressive.



CONCLUSION

It is always important for businesses to take care of their customers to prevent customer churn for several reasons:

1. Cost-effectiveness: Retaining existing customers is often more cost-effective than acquiring new ones. It takes less resources and money to retain a current customer than to acquire a new one.
2. Customer loyalty: Satisfied customers are more likely to remain loyal to a company, leading to long-term, recurring business.
3. Positive word of mouth: Happy customers are more likely to recommend a business to others, leading to organic growth and new customer acquisition.
4. Increased revenue: Retaining customers can lead to increased revenue, as satisfied customers are more likely to purchase additional products or services from a company they trust.
5. Improved customer experience: By taking care of customers, companies can improve their overall customer experience, leading to increased customer satisfaction and reduced churn.