ChurnShield

Proactive Churn Management for B2B Subscriptions

Github Repository : https://github.com/namritaansh02/ME781-Churn-Prediction-Estimation
Gradio Hosting : https://huggingface.co/spaces/namritaansh020/me781-churn-prediction

Group 1 - Team Members

- Amritaansh Narain - 200100022
- Jayesh Ahire - 22B2101
- Pawan Kumar Meena- 200100113
- Korimi Vennela - 20d100013
- Sahil Barbade - 210040131

Churn? What is Churn?

Churn refers to the number of subscribers that leave a provider or the number of employees that leave a firm in a given period

Objective

(1) Use various modelling functions to predict customer churn from absolute customer data

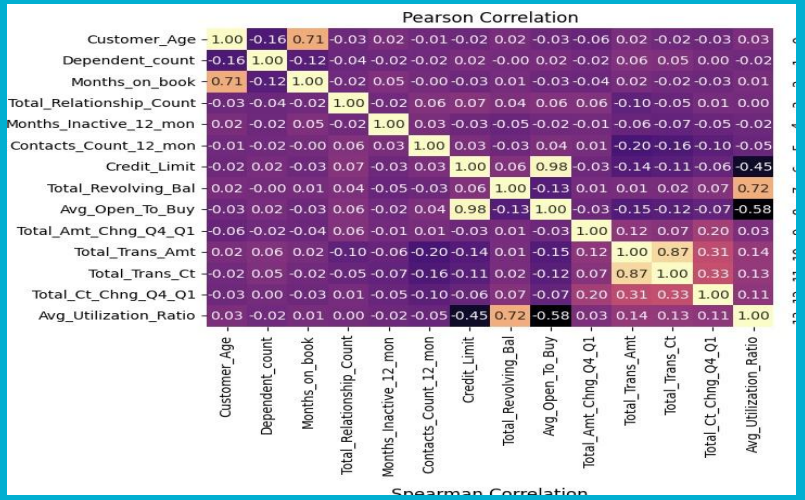
(2) Observe the effect of incorporating customer feedback in churn prediction as a dynamic element since the probability of churning a customer changes over time

(3) Allow the functionality of giving input as possible attributes and their values and use their semantic meaning for churn estimation

Data Insights

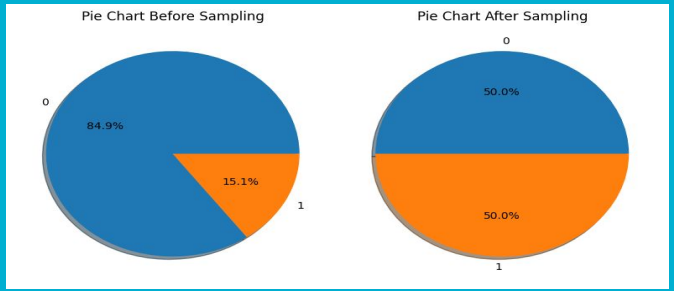
Numerical Features :

- Data Cleaning and Outlier Removal
- Presence of skewness/ kurtosis
- Presence of Multicollinearity



Categorical Features :

- Data Cleaning and One-hot encoding
- Nominal features are one-hot encoded and Ordered Features are label encoded
- Resampling of target features is done using SMOTE Analysis



Model Details - Logistic Regression

Dataset - Customers Exited Data
Train Test Split - 0.75 : 0.25
Scoring Metric - Recall

Confusion Matrix - shows that majority Actual Positive cases are predicted positively.

	Actual Positive	Actual Negative
Predicted Positive	1948	50
Predicted Negative	474	28

	Precision	Recall	f1-score	Support
Macro Avg	0.58	0.52	0.49	2500
Weighted Avg	0.71	0.79	0.72	2500

Model Details - Support Vector Classifier

Kernel - Linear
Train Test Split - 70 : 30
Metric - Accuracy

Every column is converted to its one hot representation before moving onto the classification phase

Identify outliers using Z- score analysis for data quality assessment

Metric	Result
Accuracy	0.79
Recall	0.05
Precision	0.32
F1 Score	0.087

Model Details - Customer Feedback

Model - Linear SVC
Dataset - Customer Feedback Data of 1 Year
Train-Test Split - 80:20
Loss Function - Squared hinge Loss

Review body was converted to a matrix of TF-IDF features using TF-IDF Vectorizer before fitting and training the model

Metric	Result
Accuracy	0.92
Recall	0.77
Precision	0.63
F1 Score	0.70

Model Details - Neural Network

Optimizer - Adam
Train Test Split - 0.75 : 0.25
Loss Function - Binary Cross Entropy
Metric - Accuracy - Fraction of correct classifications
Epochs - 100

Every column is converted to its one hot representation before moving onto the classification phase. Neural network hyper parameters are tuned via grid search

Metric	Result
Accuracy	0.77
Recall	0.45
Precision	0.6
F1 Score	0.51

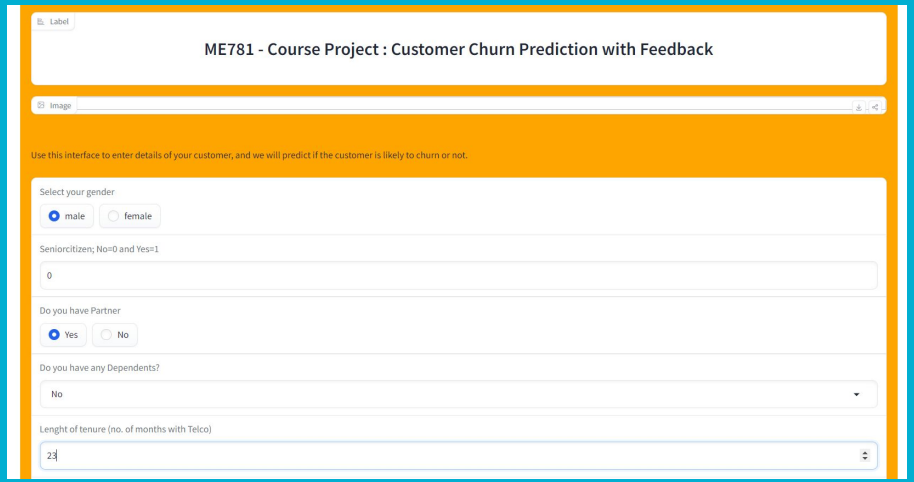
Results and Testing Report

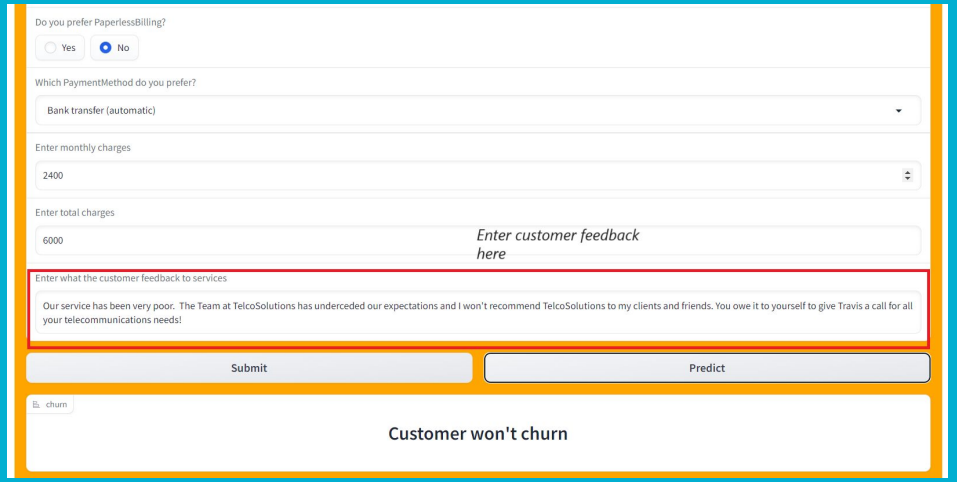
Using customer feedback **helps** in predicting if the customer will churn. However prolonged feedback is required for better prediction.

Service allows user to input attribute details and a sample customer feedback and model uses the information provided to predict if the customer will churn

Test Condition	Accuracy
Without Feedback	0.79 (best using SVC)
With 1 sample feedback	0.83 (using the SVC over NN encodings)

Screenshots of User Interface and Output





Try the model out yourself :
https://huggingface.co/spaces/namritaansh020/me781-churn-prediction

References

1. Gradio Documentation - *https://www.gradio.app/docs/interface*
2. Hosting Application on Gradio - *https://ismailouahbi.medium.com/gradio-build-deploy-and-share-your-machine-learning-models-7b38baba659c*
3. Scikit-learn Documentation - *https://scikit-learn.org/*
4. Tensorflow Documentation - *https://www.tensorflow.org/api_docs*

Thank You