**FINAL PROJECT**

Done By

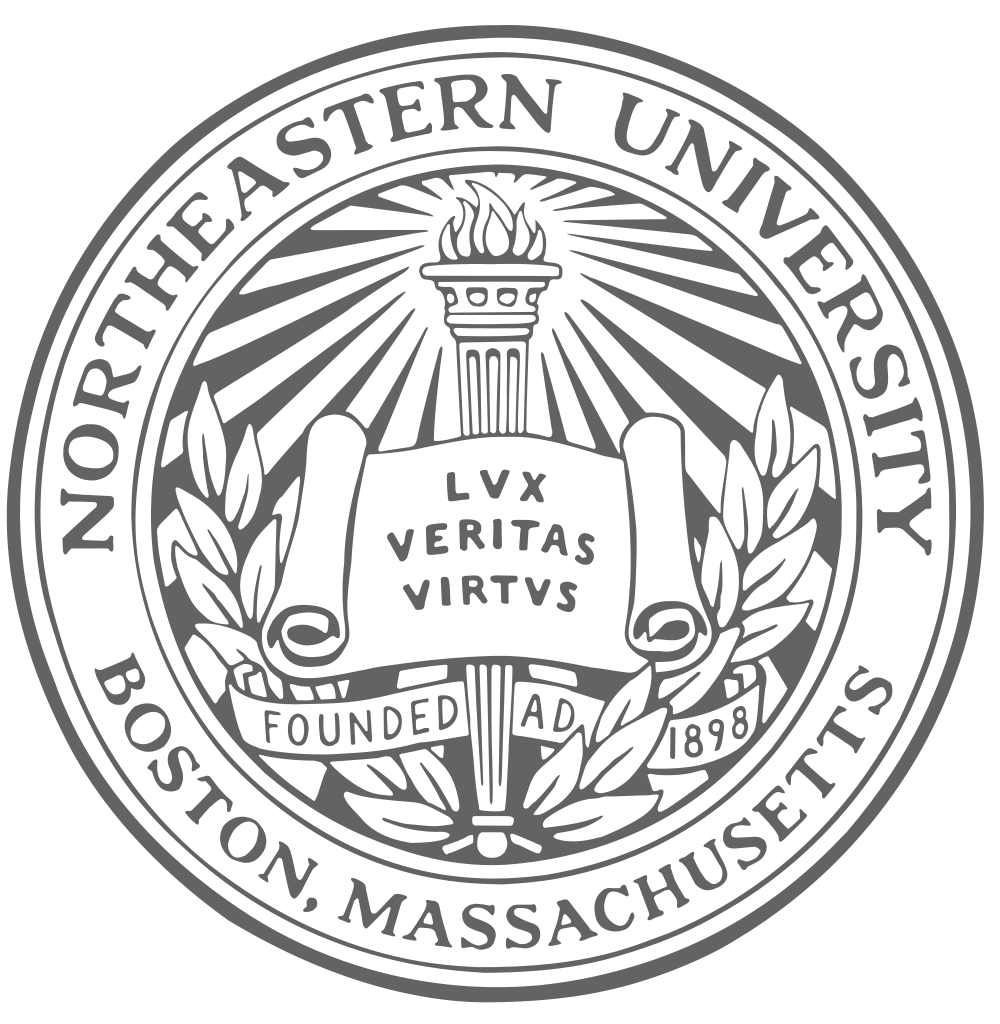
ANISH NITIN SOMAIAH

**WEEK 6: Telecom Churn Prediction using Logistic Regression (Categorical and Continuous Dataframes)**

ALY 6040: 80815 DATA MINING APPLICATIONS

Instructor: Mr. DONALD SMITH

Date: May 13, 2019



**ABSTRACT:**

The objective of the project is to analyze 2 independent datasets of Telecom churning to derive insights about the pattern of telecom churn of customers and make a prediction model to classify customers whether they churn or not. The business case of this project is to identify key parameters to identify existing patterns and retain customers despite competition in the consumer market.

**DATASET SOURCE:**

* **Categorical variable dominant dataset** Retrieved from <https://www.kaggle.com/blastchar/telco-customer-churn>
* **Continuous variable dominant dataset** Retrieved from <https://www.kaggle.com/becksddf/churn-in-telecoms-dataset>

**DATASET FEATURES:**

|  |  |
| --- | --- |
| Categorical Dominant Dataset | Continual Dominant Dataset |
| customerID | state |
| gender | account length |
| SeniorCitizen | area code |
| Partner | phone number |
| Dependents | international plan |
| tenure | voice mail plan |
| PhoneService | number vmail messages |
| MultipleLines | total day minutes |
| InternetService | total day calls |
| OnlineSecurity | total day charge |
| OnlineBackup | total eve minutes |
| DeviceProtection | total eve calls |
| TechSupport | total eve charge |
| StreamingTV | total night minutes |
| StreamingMovies | total night calls |
| Contract | total night charge |
| PaperlessBilling | total intl minutes |
| PaymentMethod | total intl calls |
| MonthlyCharges | total intl charge |
| TotalCharges | customer service calls |
| Churn | churn |

**CODE PROCESS :**

* Data Cleaning and Pre-Processing for **Exploratory Data Analysis** and Modelling to gain insights from the existing dataframe.
* **Outlier Analysis** to be done for the continuous variables to ensure accuracy of the model.
* Implement **Logistic Regression** on both models by obtaining main predictor variables using **Correlation Matrix.**
* Finally predicting test data using the training model and calculating accuracy of test model using metrics such as Specificity, Sensitivity, Accuracy and AUC using **Confusion Matrices and ROC curves**.
* **Finetuning** obtained model for **better balanced accuracy** using **variance inflation factors and proper cutoff value** estimation using graphs so that the Prediction Model that we propose to the organization facilitates the company to make necessary actions to retain customers, sustain business and compete against peer organizations in the market.
* Comparison of Models and Data Visualizations are attached in the PPT.