

# **First Draft Submission**

## **MODEL TO PREDICT CUSTOMERS LEAVING TELECOMMUNICATION COMPANIES**

### **Team members:**

1. Rajagopal Naidu Kodavati 11596822
2. Jaya Naga Satya Pavan Ganesh Kotipalli 11594936
3. Jaya Sindhu Edara 11708783
4. Sri Mounish Seenii 11549237

### **Motivation:**

1. The motivation for developing churn prediction models in a telecommunications company is to improve customer satisfaction, reduce costs, and gain a competitive edge by making data-driven decisions and proactively addressing customer churn.
2. This project has the potential to lead to significant financial and operational benefits for the company.
3. These models are a key tool for increasing revenue and driving business growth.
4. It's a strategic initiative that combines data analysis, modeling, and business intelligence to drive tangible results and long-term success.

### **Significance:**

1. It leverages data-driven insights to inform decision-making and improve the company's competitiveness in a dynamic industry.

2. It directly impacts revenue, customer satisfaction, and overall business success.
3. Customer churn is a significant concern for telecommunications companies.
4. Losing customers can result in a loss of revenue, and it can be more costly to acquire new customers than to retain existing ones. Predicting and preventing churn can significantly impact a company's bottom line.

### **Objectives:**

1. Initially we are planning to categorize our data into numerical and categorical values and perform exploratory data analysis like (univariate, bivariate, etc.) to get a complete idea on what the dataset can do.
2. After finding out the capabilities of the dataset we are planning to perform feature engineering on the given data and develop some features which are useful for better finding out the customer churn.
3. Identifying the factors which contribute customer churn and reduce them.
4. The main objective is to improve the performance of the machine learning models by training them with more relatable features.

### **Features:**

The features which we are using are categorical and numerical

1. Categorical features: Customer\_ID, Gender, Partner, Dependents, Phone\_Service, Multiple\_lines, Internet\_Service, Online Security, Online Backup, Device Protection, Tech Support, Stream Movies, Paperless Billing, Payment Method.

2. Numerical features: Tenure, Monthly\_Charges, Totla\_Charges.
3. Also Planning on perfoming feature engineering with these features and create useful features which contribute to customer churn

## **Increment – 1:**

### **Related Work:**

Our project was on prediction of employees who leaves company or not, so for this we take data form online sources which has all the required fields to categorize the situation of employee and we make necessary changes to our model to make good features for extraction. With the required features we build a model to predict the outcome.

### **Dataset:**

<https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

We take the dataset from Kaggle platform which has required columns and categories to our model and we make changes according to best fit of our model.

### **Implementation:**

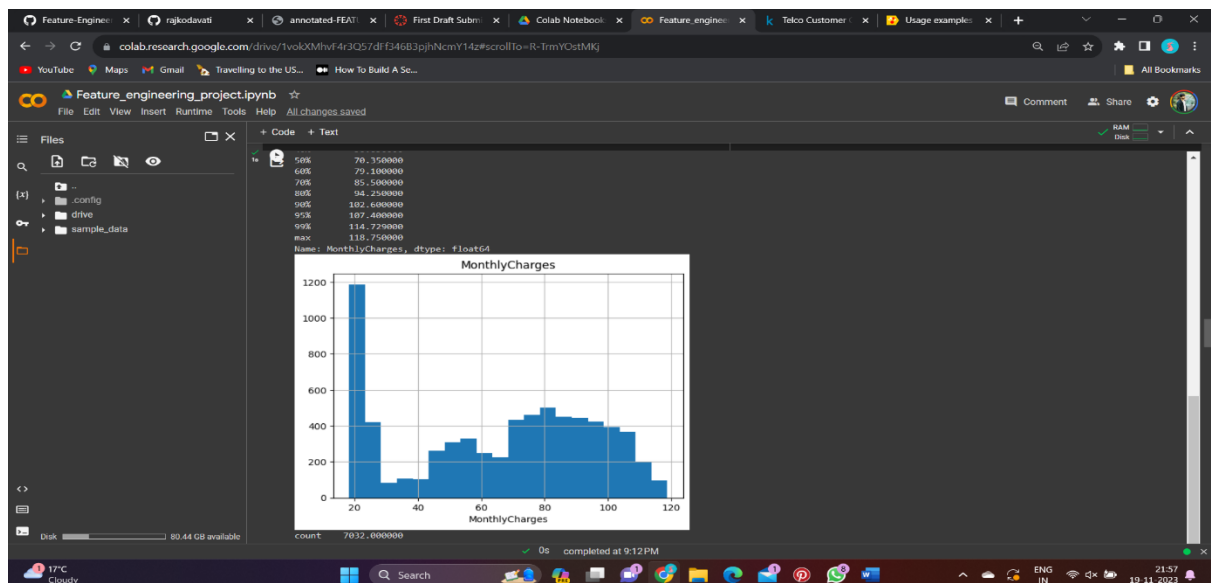
- Initially we imported our dataset from Kaggle and make necessary changes to make good fit for the model
- The changes we made in dataset is to convert some categorial data to numerical data which helps to us to make easy prediction
- The conversion is on type of 1 = True and 0 = False in many cases

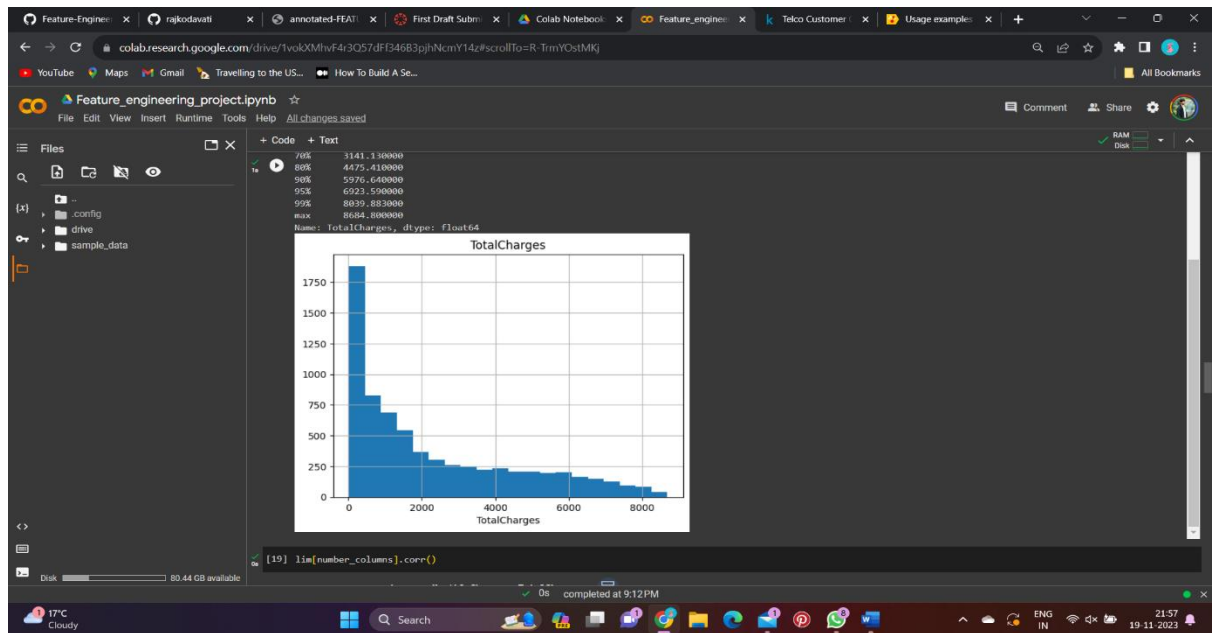
- After the data was ready we briefly summarize the data into two categories which are categorical data and numerical data for better understanding to build the model and make prediction easy
- After summarization we print corelation matrix for clear visualization to user
- Performs one-hot encoding technique which is one of the popular encoding techniques to categorical data and we split the data to test and train to find accuracy, recall, precision of the data. For this we also use catboost library.

## Analysis:

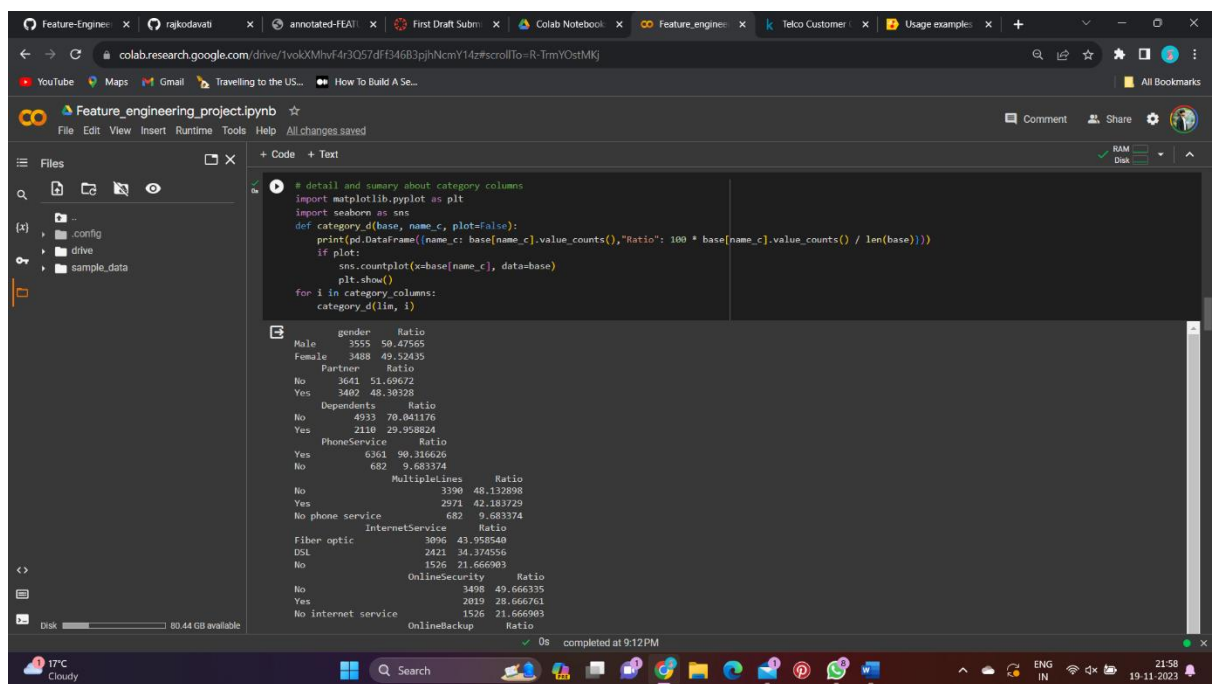
As the process mention above was done successful we analyse the data and display the data in graph formats and text formats using matplotlib libraries and displays how accurate the data which we are using.

Analysis we made on numerical data:



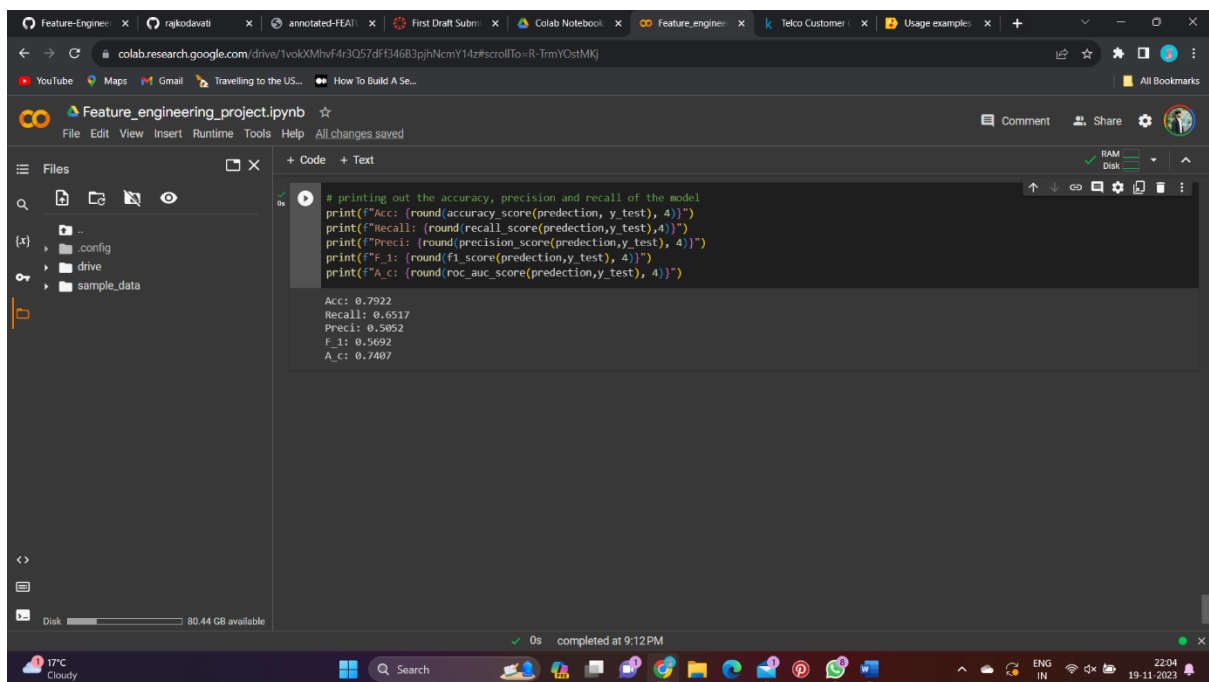
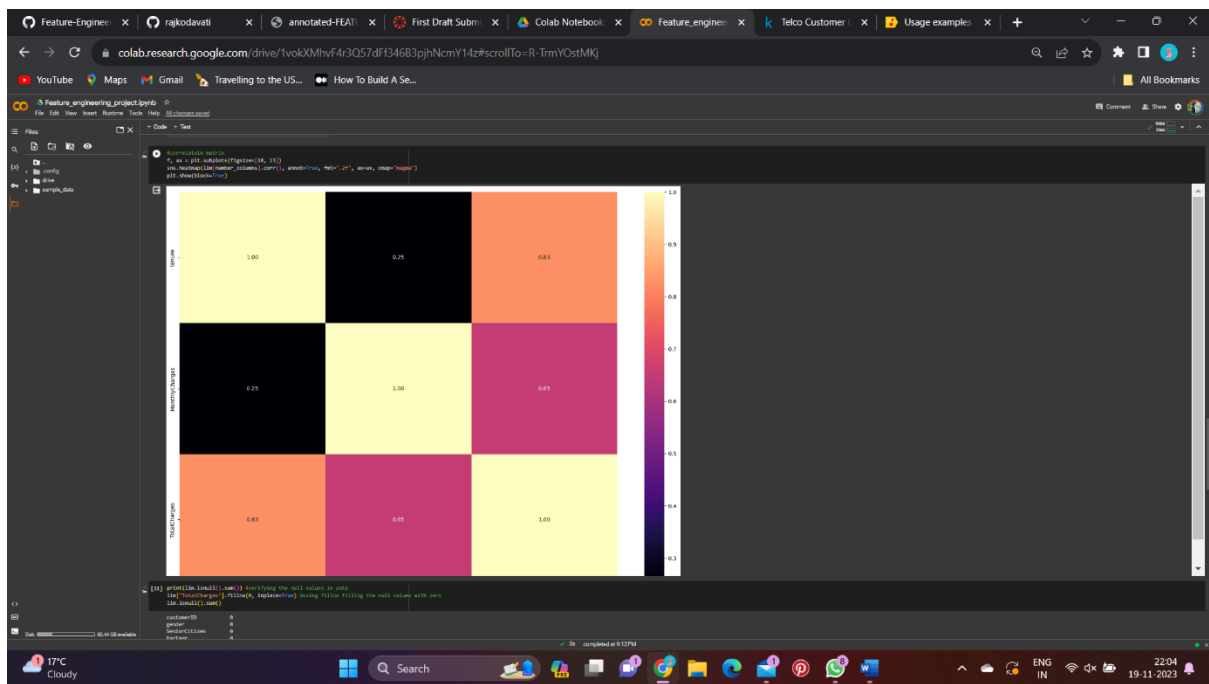


Analysis we made on categorical data:



## Preliminary results:

We print the correlation matrix of the model and we split our model to test and train dataset to find prediction and accuracy of our dataset.



## Project Management:

### Work completed:

We make our dataset perfect to our model and we perform prediction and deep analysis of our data, and we train and test the data to find how accurate is our dataset and we make features ready

to extract to build our model, for predicting the final out as employees will leave the company or not.

### **Responsibility:**

Rajagopal Naidu Kodavati make all the necessary requirements to prepare the data and perform the tasks to get best accuracy and brief analysis of data.

Jaya naga satya pavan ganesh kotipalli and Jaya sindhu Edara helped me in documentation for first draft and making proposal

Sri Mounish Seeni helps me in making proposal

### **Contribution:**

Rajagopal Naidu Kodavati – 50%

Jaya naga satya pavan ganesh – 20%

Jaya Sindhu Edara – 20%

Sri Mounish Seeni – 5%

### **Work to be completed:**

After we make our dataset and analysis perfect to our model and make features, next step we need to perform good feature selection and extraction of selected features. After the successful extraction of features, we build our model by using encoding techniques and predict the results as final output. Also displays how accurate our model.

### **Responsibility:**

Rajagopal Naidu Kodavati – Feature Extraction

Jaya naga satya pavan ganesh kotipalli – Encoding and build model

Jaya sindhu - Making in final project documentation

Sri Mounish Seeni – Displays final output and split the predicted output to perform operation to show accuracy of our model

**Contibution:**

Rajagopal Naidu Kodavati – 25%

Jaya naga satya pavan ganesh kotipalli – 25%

Jaya sindhu – 25%

Sri Mounish Seenii – 25%

**References:**

1. Makhtar M, Nafis S, Mohamed M, Awang M, Rahman M, Deris M. Churn classification model for local telecommunication company based on rough set theory. J Fundam Appl Sci. 2017;9(6):854–68
2. Suchánek P, Králová M: Customer satisfaction, loyalty, knowledge and competitiveness in the food industry. Economic Research-Ekonomska Istraživanja. 2019;32(1):1237–1255.  
10.1080/1331677x.2019.1627893 [CrossRef] [Google Scholar]
3. Pope L: How to Prevent Customer Churn with Retention Marketing. G2. 2020, 27th August. Reference Source [Google Scholar]

Github link:

<https://github.com/rajkodavati/Model-to-predict-customers-leaving-companies-using-feature-engineering>