# Telecom Customer Churn Prediction Using IBM Watson

# Introduction:

**Overview:**

Customer churn has become highly important for companies because of increasing competition among companies, increased importance of marketing strategies and conscious behaviour of customers in recent years. Customers can easily trend toward alternative services. Companies must develop various strategies to prevent these possible trends, depending on the services they provide. During the estimation of possible churns, data from the previous churns might be used. An efficient churn predictive model benefits companies in many ways. Early identification of customers likely to leave may help to build cost effective ways in marketing strategies.

Telecommunication industry always suffers from a very high churn rates when one industry offers a better plan than the previous there is a high possibility of the customer churning from the present due to a better plan in such a scenario it is very difficult to avoid losses but through prediction we can keep it to a minimal level. A machine learning model is built and this helps to identify the probable churn customers and then makes the necessary business decisions.

purpose

The role and function of telecommunication is to provide an exchange of communication or information at a distance between people, satellites or computers. With the advent of the Internet on a global scale in the last two and half decades, humankind has entered the Information Age, which makes information available at the touch of a finger to anyone with the appropriate equipment. Begun in the imagination of those who went before, such as Arthur C. Clarke, a science fiction author who first envisioned satellites in orbit around the earth in 1945, the unknown forms of telecommunications reside with those who follow in the footsteps of the world's leading physicists, scientists and engineers.

# **Literature Survey:**

Existing problem:

Customer churn is a major problem and one of the most important concerns for large companies. Due to the direct effect on the revenues of the companies, especially in the telecom field, companies are seeking to develop means to predict potential customer to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn. The main contribution of our work is to develop a churn prediction model which assists telecom operators to predict customers who are most likely subject to churn. The model developed in this work uses machine learning techniques on big data platform and builds a new way of features’ engineering and selection. In order to measure the performance of the model, the Area Under Curve (AUC) standard measure is adopted, and the AUC value obtained is 93.3%. Another main contribution is to use customer social network in the prediction model by extracting Social Network Analysis (SNA) features. The use of SNA enhanced the performance of the model from 84 to 93.3% against AUC standard. The model was prepared and tested through Spark environment by working on a large dataset created by transforming big raw data provided by SyriaTel telecom company. The dataset contained all customers’ information over 9 months, and was used to train, test, and evaluate the system at SyriaTel. The model experimented four algorithms: Decision Tree, Random Forest, Gradient Boosted Machine Tree “GBM” and Extreme Gradient Boosting “XGBOOST”. However, the best results were obtained by applying XGBOOST algorithm. This algorithm was used for classification in this churn predictive model.

Proposed solution:

Based on best practices from our work with the leading communication service providers, we’ve developed solution accelerators for common analytics and machine learning use cases to save weeks or months of development time for your data engineers and data scientists.

This solution accelerator complements our work doing customer lifetime value, attrition for subscription services, and profitable customer retention, but with a telco-specific lens.

Using sample telco datasets from IBM, and the Lifelines library, this solution accelerator will:

Introduce survival analysis, a collection of statistical methods used to examine and predict the time until an event of interest occurs.

Review three methods that are commonly used for survival analysis: Kaplan-Meier, Cox Proportional Hazards, Accelerated Failure Time.

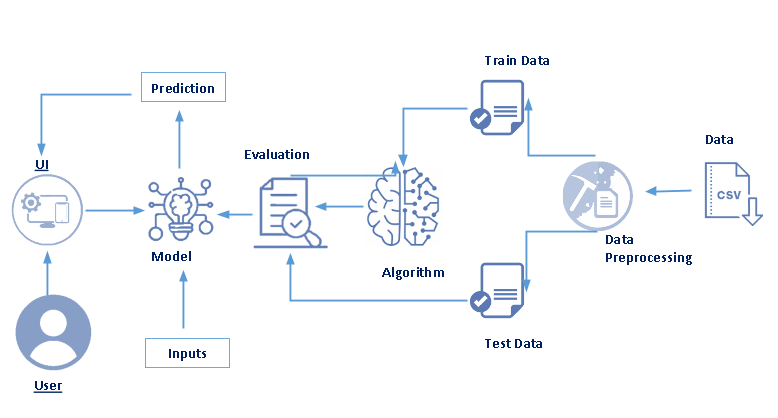
Build a churn prediction model and use the model output as an input for calculating lifetime value.

Build an interactive dashboard for calculating the net present value of a given cohort of subscribers over a three-year time horizon.

The contents of this solution accelerator are contained in Databricks notebooks that are linked to at the end of this post.

**Theoretical Analysis:**

Block diagram:



Hardware/Software designing:

**Skills Required:**

Python,Python Web Frame Works,Machine Learning,Classification Algorithms,Python-Flask

To build Machine learning models you must require the following packages

**Software Installations**

In order to develop this project we need to install following software/packages

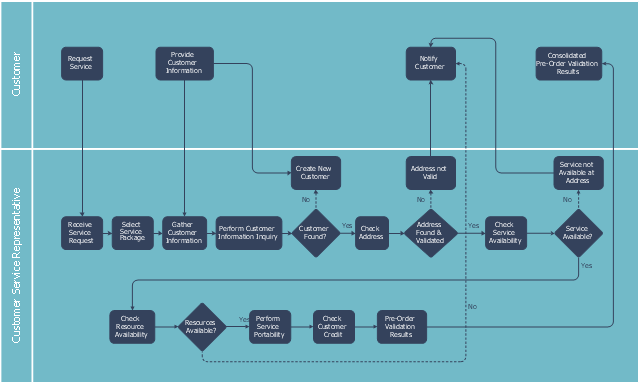
**Anaconda Navigator:**

Anaconda Navigator is a free and open-source distribution of the Python and R programming languages for data science and machine learning related applications. It can be installed on Windows, Linux, and macOS.Conda is an open-source, cross-platform, package management system. Anaconda comes with so very nice tools like JupyterLab, Jupyter Notebook,

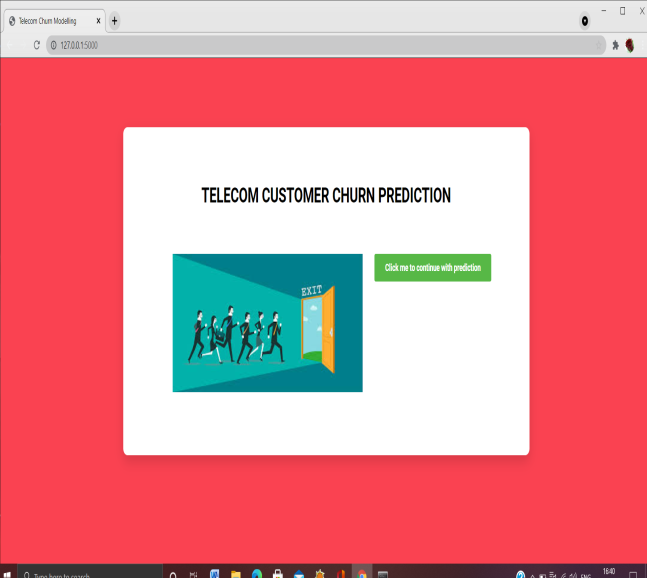
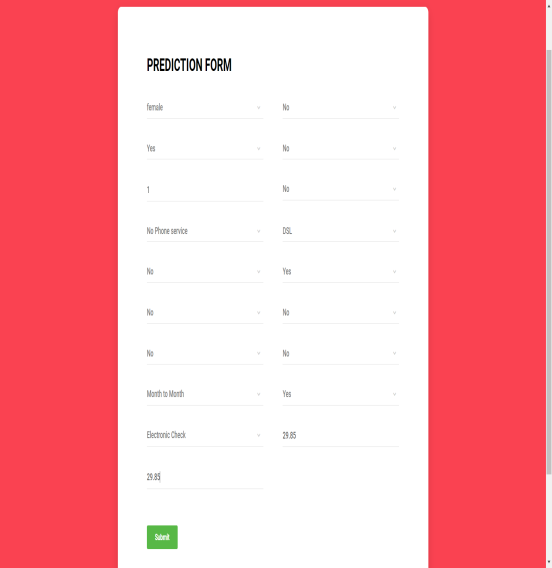
QtConsole, Spyder, Glueviz, Orange, Rstudio, Visual Studio Code. For this project, we will be using Jupyter notebook and Spyder

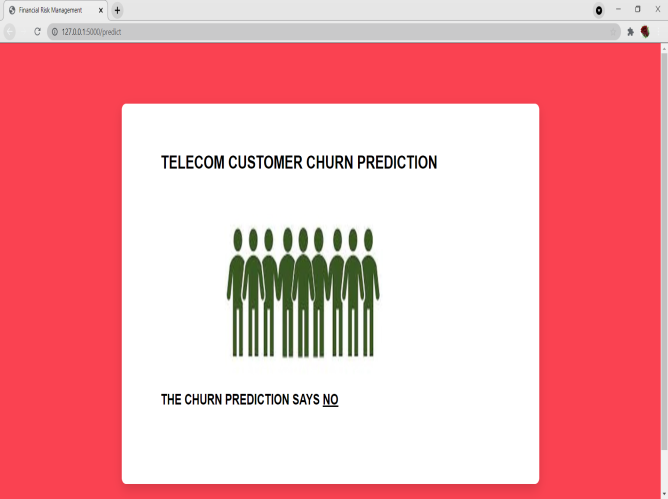
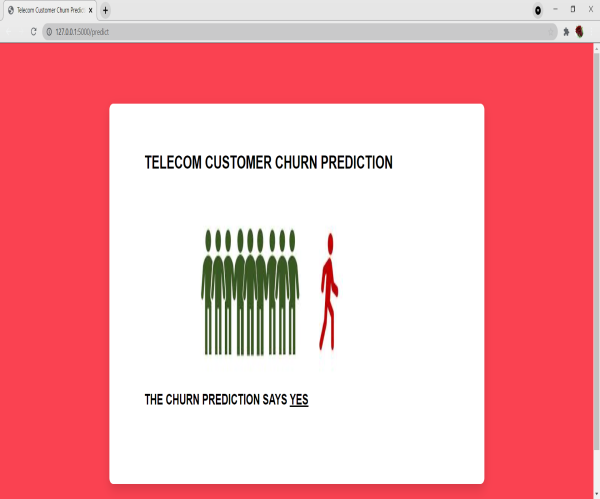
* **Numpy:**
* It is an open-source numerical Python library. It contains a multidimensional array and matrix data structures and can be used to perform mathematical operations
* **Scikit-learn:**
  + It is a free machine learning library for Python. It features various algorithms like support vector machine, random forests, and k-neighbours, and it also supports Python numerical and scientific libraries like NumPy and SciPy
* **Matplotlib and Seaborn:**
  + Matplotlib is mainly deployed for basic plotting. Visualization using Matplotlib generally consists of bars, pies, lines, scatter plots and so on. Seaborn: Seaborn, on the other hand, provides a variety of visualization patterns. It uses fewer syntax and has easily interesting default themes.
* **Flask:**
  + Web framework used for building Web applications

# Flow chart:



# Result:

Advantages&Disadvantage:

Advantages of Telecommunication :

* Quick and accessible communication
* Lack of time period
* Saves time
* Saves gasoline (do not need to drive distance)
* More than two people can communicate with at least one another at an equivalent time
* Next “best thing” to being there
* Easy to exchange ideas and knowledge via phone and/or fax
* Worldwide access
* Easy access to the people you would like to contact.
* Less effort in using transportation just to satisfy a private personally.
* You can just occupy your home and use a telephone or a cellphone if you would like to speak to someone.
* Enable end-users to speak electronically and share hardware, software, and data resources.
* This make corporation to do the transaction at the point only and in a very fast way from many remote locations, exchange business documents electronically with customers and suppliers, or remotely monitor and control production processes.
* Interconnect the pc systems of a business so their computing power is often shared by end-users throughout an enterprise.
* Make the organization work with collaboration and communication among the staff inside and out of doors a corporation
* Speed
* Develops new products and inventions

Disadvantages of Telecommunication :

* Cultural Barrier
* Misunderstanding
* Prank calls
* Sometimes expensive
* High electric bills
* Remote areas don’t have access
* Remote areas might not be ready to afford the necessary equipment
* Cannot see whom you’re speaking with
* Cannot see facial expressions, therefore results in misunderstandings
* Cultural barriers
* Poor connections or downed power lines during/after storms

# **Applications:**

The range of telecommunications applications is broad and includes telephony and video conferencing, facsimile, broadcast and interactive television, instant messaging, e-mail, distributed collaboration, a host of Web- and Internet-based communication, and data transmission.

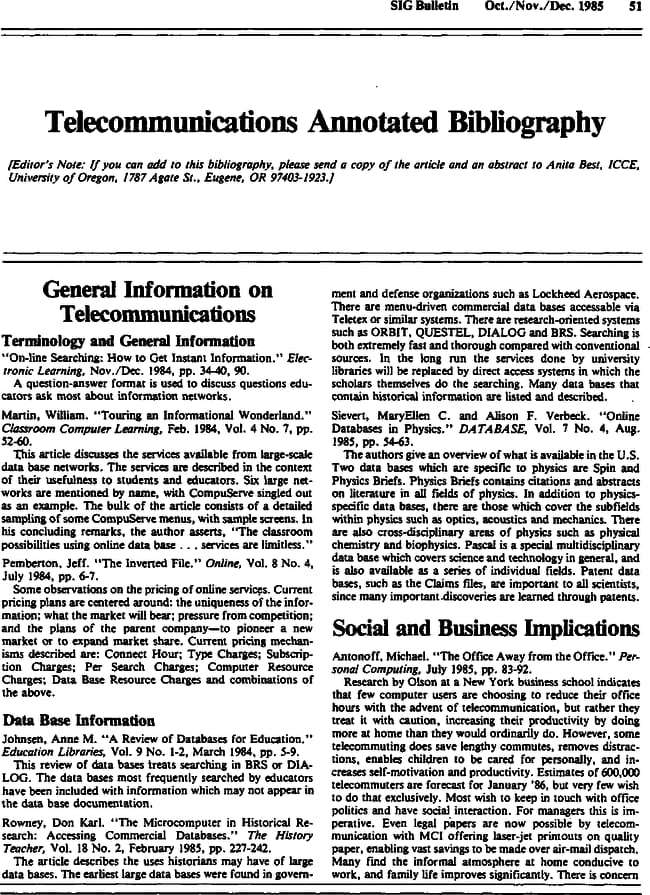
**Conclusion:**

The importance of this type of research in the telecom market is to help companies make more profit. It has become known that predicting churn is one of the most important sources of income to telecom companies. Hence, this research aimed to build a system that predicts the churn of customers in SyriaTel telecom company. These prediction models need to achieve high AUC values. To test and train the model, the sample data is divided into 70% for training and 30% for testing. We chose to perform cross-validation with 10-folds for validation and hyperparameter optimization. We have applied feature engineering, effective feature transformation and selection approach to make the features ready for machine learning algorithms. In addition, we encountered another problem: the data was not balanced. Only about 5% of the entries represent customers’ churn. This problem was solved by undersampling or using trees algorithms not affected by this problemThese algorithms are Decision Tree, Random Forest, GBM tree algorithm, and XGBOOST algorithm. The method of preparation and selection of features and entering the mobile social network features had the biggest impact on the success of this model, since the value of AUC in SyriaTel reached 93.301%. XGBOOST tree model achieved the best results in all measurements. The AUC value was 93.301%. The GBM algorithm comes in the second place and the random forest and Decision Tree came third and fourth regarding AUC values. We have evaluated the models by fitting a new dataset related to different periods and without any proactive action from marketing, XGBOOST also gave the best result with 89% AUC. The decrease in result could be due to the non-stationary data model phenomenon, so the model needs training each period of time

# Future Scope:

India's telephone subscriber base expanded at a compound annual growth rate of 19.16 percent during the last year. Due to stiff competition, some consolidation is going on in the industry. Telecom companies are also in the process of optimizing and rationalizing their operations.

Bibliography:

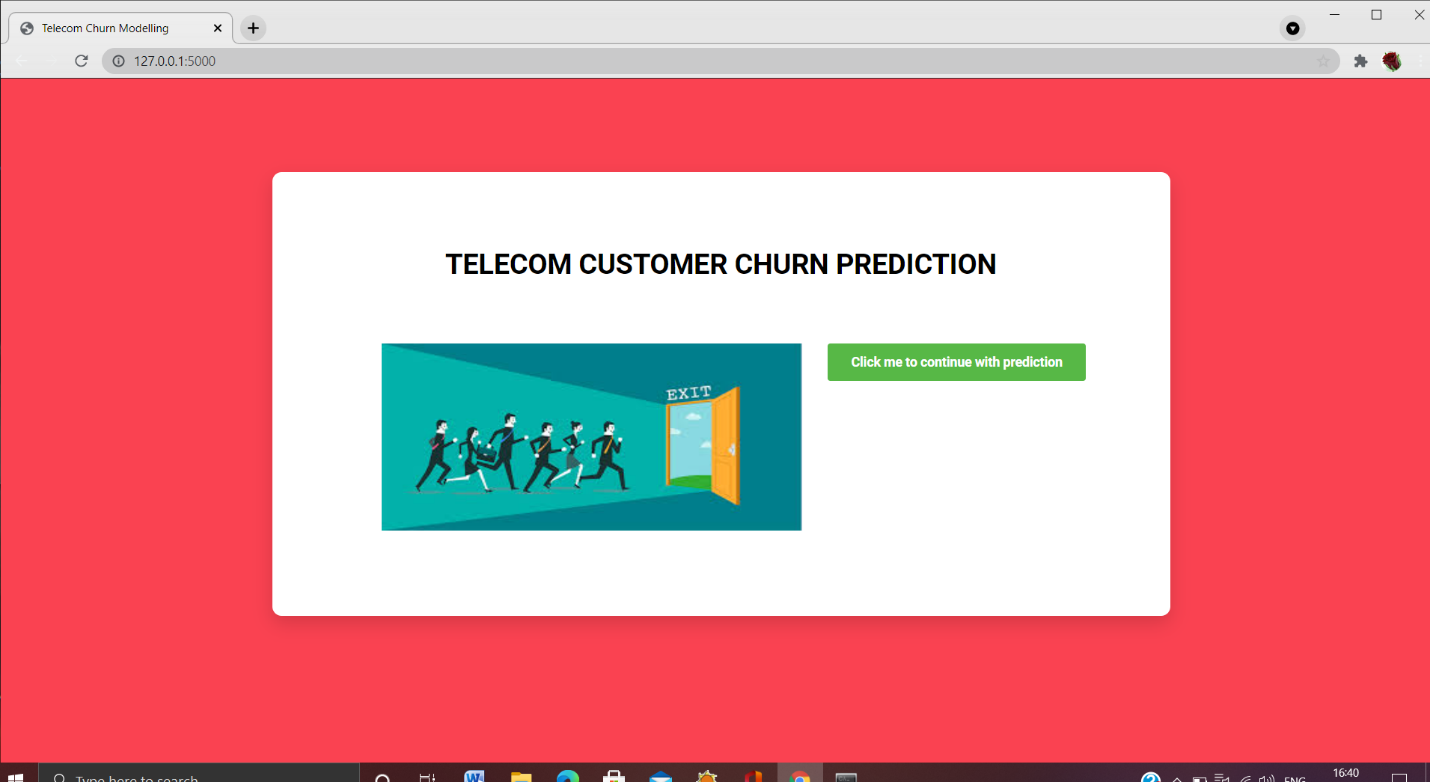


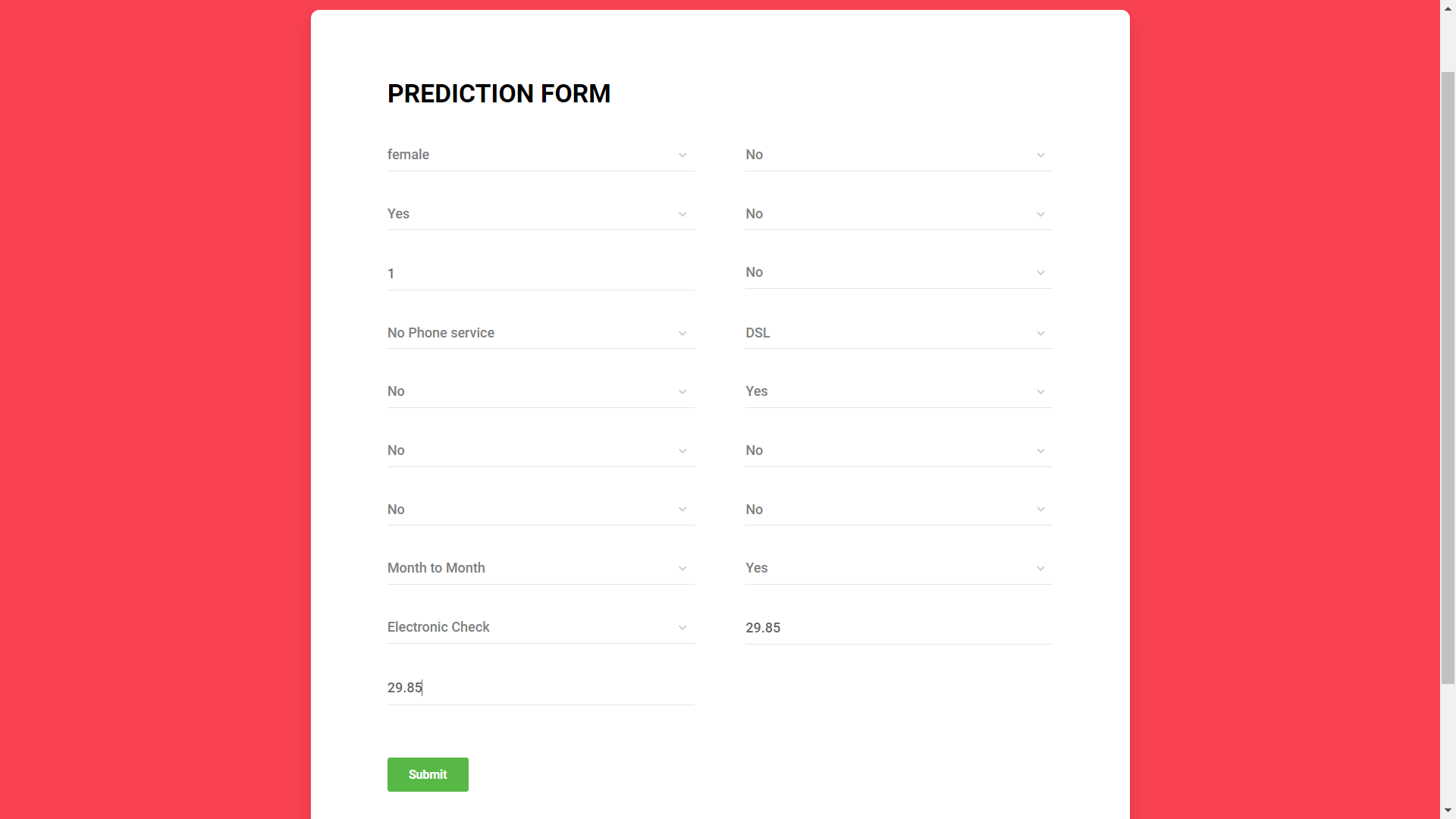
Appendix:

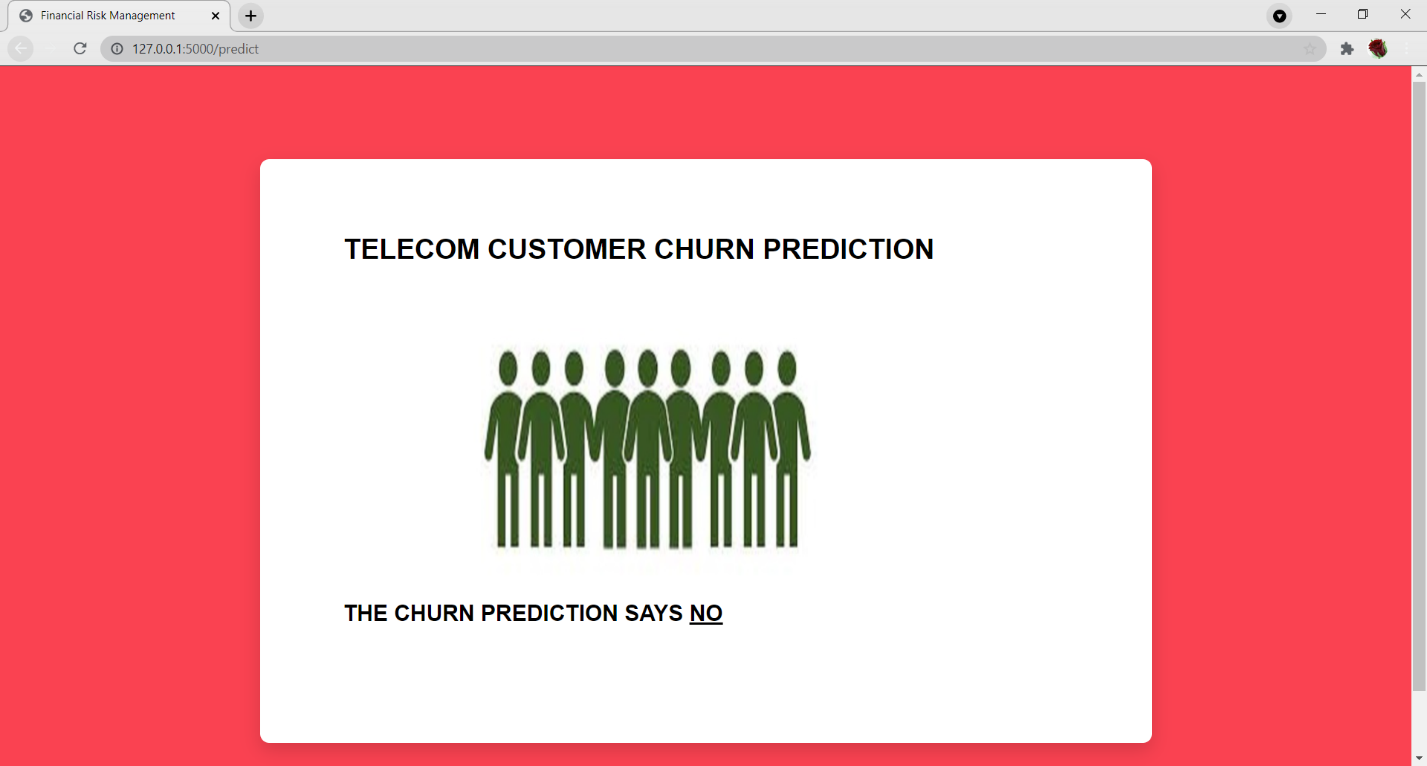
Source code:

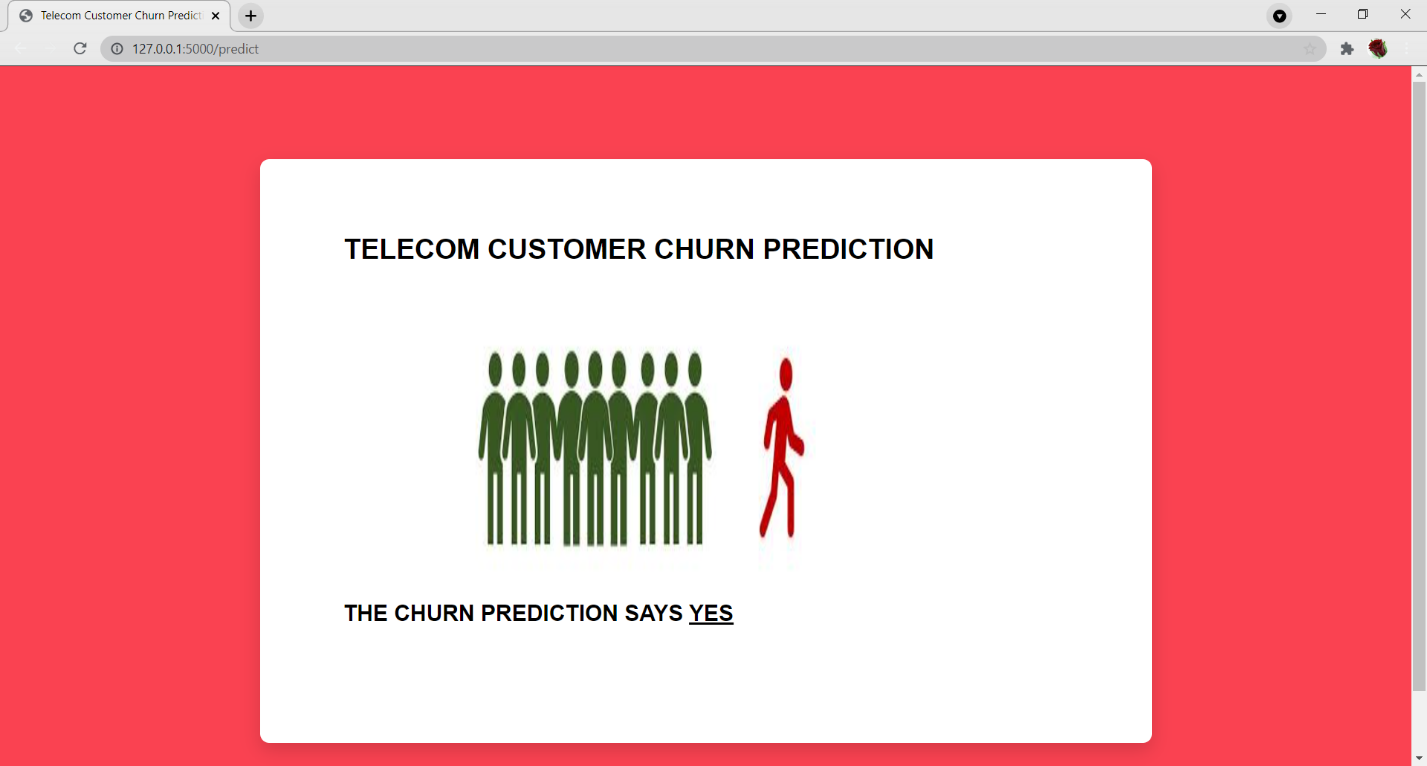
<https://github.com/smartinternz02/SI-GuidedProject-3377-> 1624078172/blob/f339872389542840834ee9857b4ff0c8b47da635/telecomcustomerchurnpred.ipynb

UI Ouput ScreenShot:









**Aiswarya Menda(18481A1269) Sai Alekhya Tadivaka(18481A1292)**

**Naradala Vijaya Durga (19481A1211)**

**Sanagapalli pavan sai kumar(18481A1287)**