DATA MINING AND ANALYTICS

MINI PROJECT REPORT

**CUSTOMER CHURN ANALYSIS**

**Team Members:**

Aditi Mittal(RA1911003010226)

Anweasha Saha (RA1911003010235)

R. Vijay (RA1911003010239)

CONTENTS

Contents

[Introduction 3](file:///C:\Users\Aditi%20Mittal\Downloads\DataMiningAndAnalyticsProjectReport.docx#_Toc87541008)

[Motivation 4](file:///C:\Users\Aditi%20Mittal\Downloads\DataMiningAndAnalyticsProjectReport.docx#_Toc87541009)

[Novelty 4](file:///C:\Users\Aditi%20Mittal\Downloads\DataMiningAndAnalyticsProjectReport.docx#_Toc87541010)

[Modules 5](file:///C:\Users\Aditi%20Mittal\Downloads\DataMiningAndAnalyticsProjectReport.docx#_Toc87541011)

[**Architecture Diagram** 5](file:///C:\Users\Aditi%20Mittal\Downloads\DataMiningAndAnalyticsProjectReport.docx#_Toc87541012)

[**Dataset** 5](file:///C:\Users\Aditi%20Mittal\Downloads\DataMiningAndAnalyticsProjectReport.docx#_Toc87541013)

[Implementation 7](file:///C:\Users\Aditi%20Mittal\Downloads\DataMiningAndAnalyticsProjectReport.docx#_Toc87541014)

[Output 9](file:///C:\Users\Aditi%20Mittal\Downloads\DataMiningAndAnalyticsProjectReport.docx#_Toc87541015)

[Conclusion 11](file:///C:\Users\Aditi%20Mittal\Downloads\DataMiningAndAnalyticsProjectReport.docx#_Toc87541016)

# Introduction

**Data Analysis** is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data. Data analysis is important in business to understand problems facing an organisation, and to explore data in meaningful ways. Data in itself is merely facts and figures. Data analysis organises, interprets, structures and presents the data into useful information that provides context for the data. Data analysis tools are used to extract useful information from business data, and help make the data analysis process easier.

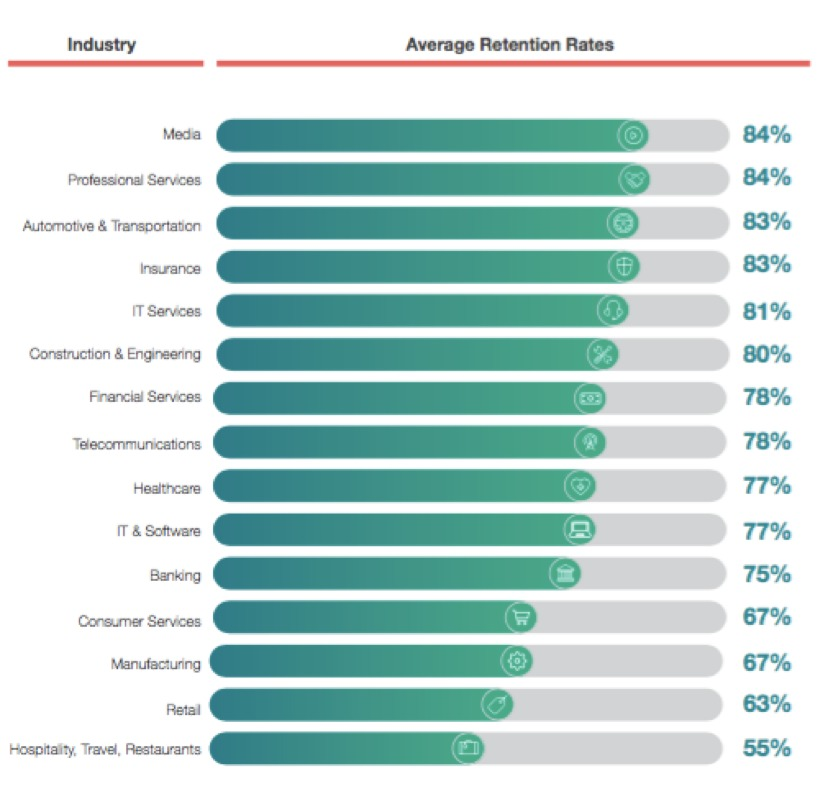
Customer Churn is one of the most important and challenging problems for businesses such as Credit Card companies, cable service providers etc. Customer churn metrics can help businesses improve customer retention. We can classify customer churn (also known as customer attrition) by grouping them into different categories. Contractual Churn, which is applicable to businesses such as cable companies and SAAS service providers, is when customers decide not to continue with their expired contracts. Voluntary Churn, on the other hand, is when a customer decides to cancel their existing service, which can be applicable for companies such as prepaid cellphones and streaming subscription providers.

# Motivation

All the SAAS companies nowadays should measure customer churn rate and need a solution to predict customers which are on verge of leaving the organization. This is where the machine learning algorithms comes into the act as the savior for SAAS organizations.

Machine learning can help predict customer churn. Customer churn is a critical metric because it is much less expensive to retain existing customers than it is to acquire new customers. Earning business from new customers means working leads all the way through the sales funnel, utilizing your marketing and sales resources throughout the process. Customer retention, on the other hand, is generally more cost-effective as you’ve already earned the trust and loyalty of existing customers.

For example: - If you lost 20% of your customers yearly, it would take 25% increase in new customer just to get you even again. What if you could reduce that churn rate to 10%? It would only take 11% increase in new customers to get you even again!



# Novelty

Increasing churn, or attrition, could be a nightmare for any marketer, business analyst, Head of Sales, or CEO. Obviously, when customers don't extend contracts or stop regular purchases, it affects not only revenue but also reputation. Since the cost of client acquisition is usually much higher than retention, it brings difficulties to businesses, especially when the cause and the remedy are unclear. Different lose clients due to different reasons as there are a lot of details which not everyone takes into account that the clients do. The model we have built in our project helps to narrow down the probable causes and forecast customer churn. Hereafter, companies can reconsider and rebuild their products and change their business strategy accordingly. Since our model generates a small prioritised list of potential defectors, it can prove to be effective at focusing customer retention marketing programs on the subset of the customer base who are most vulnerable to churn.

# Modules

* Data Collection: - Collecting customer data from credit card company
* Data Visualization and Exploratory Data Analysis: - Gaining insights of the dataset by applying univariate and bivariate analysis.
* Data Processing: - Encoding categorical datausing LabelEncoder
* Model Creation: - Two machine learning models were used to predict a customer churn
* Model Evaluation: - Both models were evaluated using classification report and confusion matrix

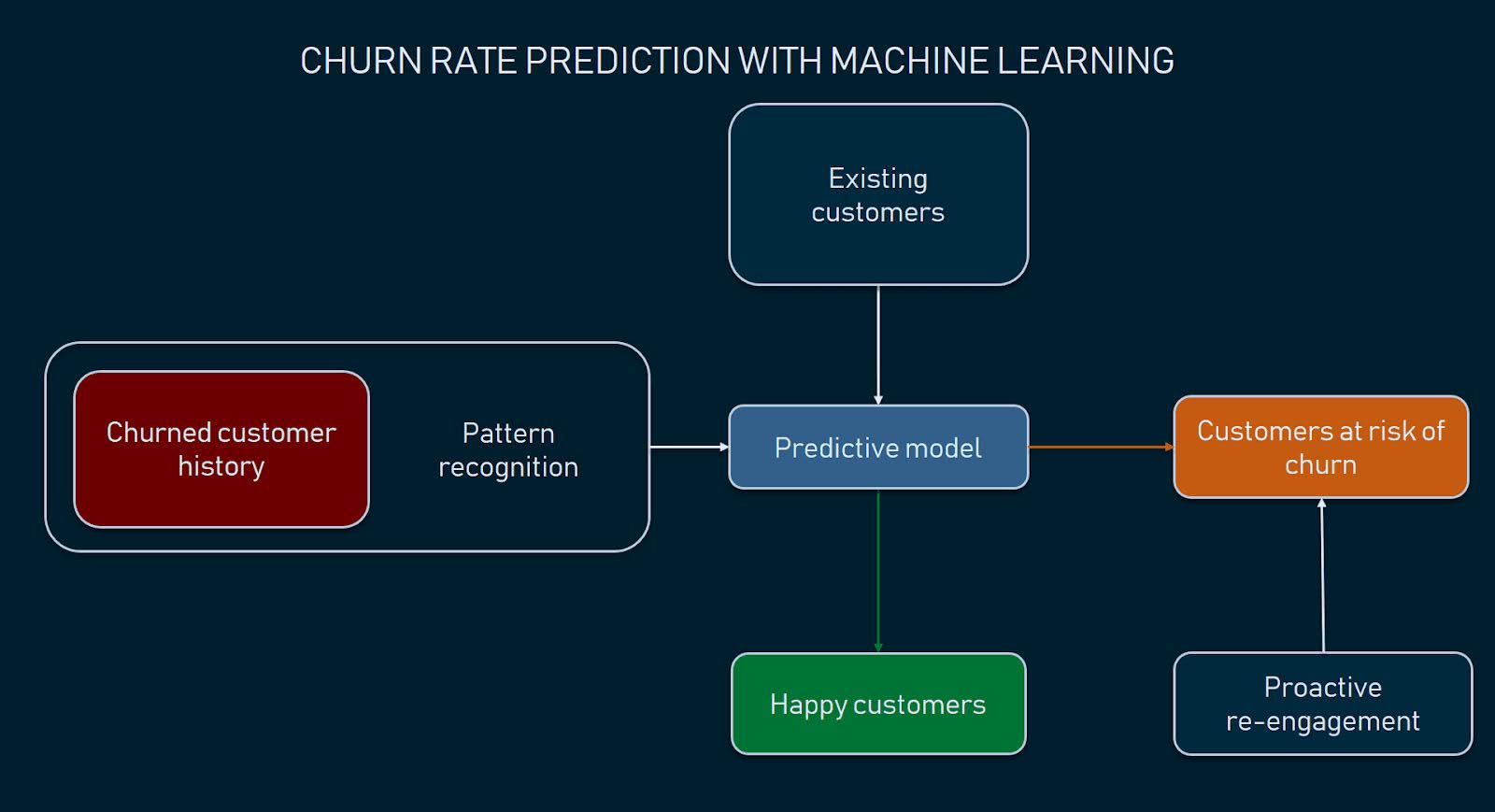
ALGORITHM/TECHNIQUE USED:

1. **Decision Tree Classifier**

Decision Tree algorithm belongs to the family of supervised learning algorithms. It can be used for solving **regression and classification problems**.The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by **learning simple decisionrules**inferred from prior data(training data).In Decision Trees, for predicting a class label for a record we start from the **root** of the tree. We compare the values of the root attribute with the record’s attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

1. **Random Forest Algorithm**

Random forest, supervised learning algorithm is a flexible, easy to use [machine learning algorithm](https://builtin.com/data-science/introduction-to-machine-learning) that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because of its simplicity and diversity (it can be used for both classification and regression tasks). It has nearly the same hyperparameters as a decision tree or a bagging classifier. Random forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model.

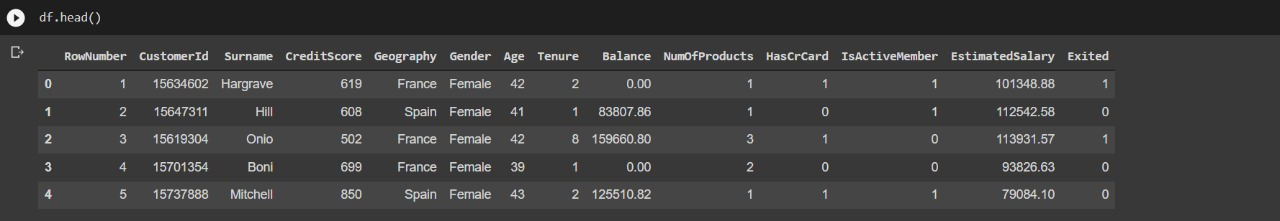
****

**Architectural Diagram**

DATASET:

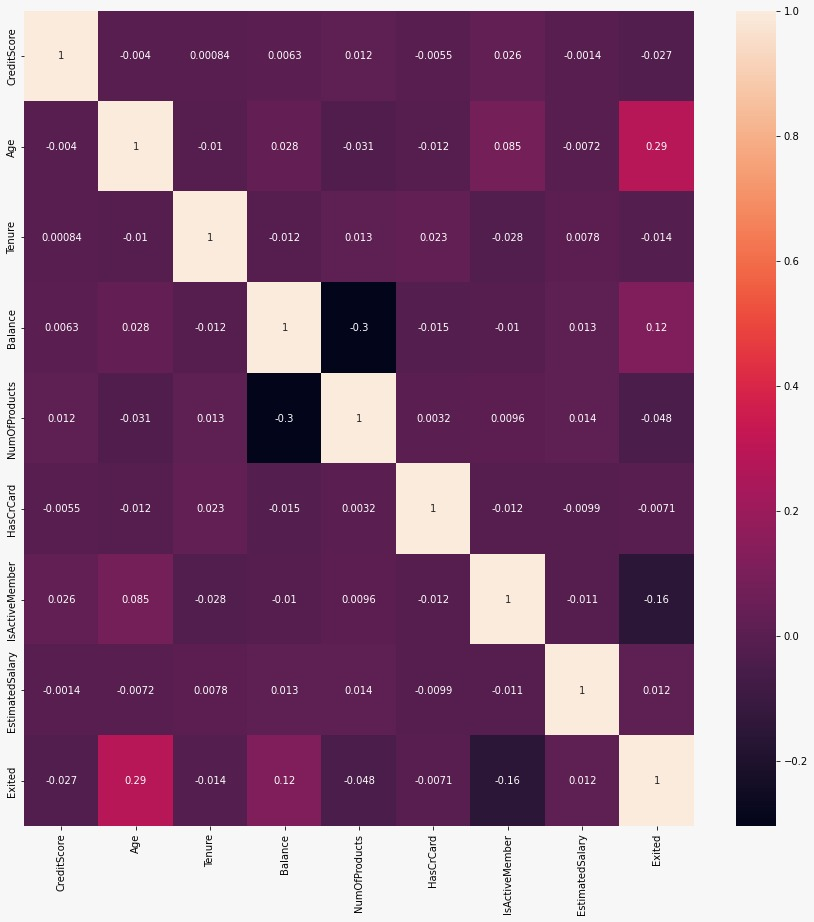
**Our customer data set** is from a credit card company, where we are able review customer attributes such as gender, age, tenure, balance, number of products they are subscribed to, their estimated salary and if they stopped the subscription or not.

Link to the dataset: ["https://raw.githubusercontent.com/anilak1978/customer\_churn/master/Churn\_Modeling.csv"](https://raw.githubusercontent.com/anilak1978/customer_churn/master/Churn_Modeling.csv)



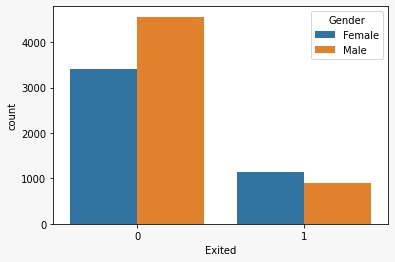
ANALYSING THE DATASET:

Heatmap



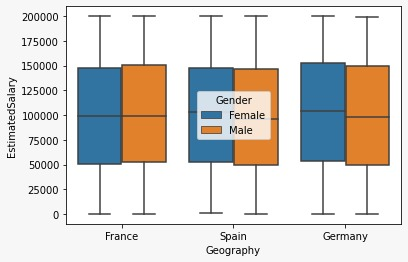
* Age is the most correlated variable to Exited feature.

Count Plot:



* The given count plot shows the count of exited employees spread across the two genders, i.e., Male and Female.

Box Plot:

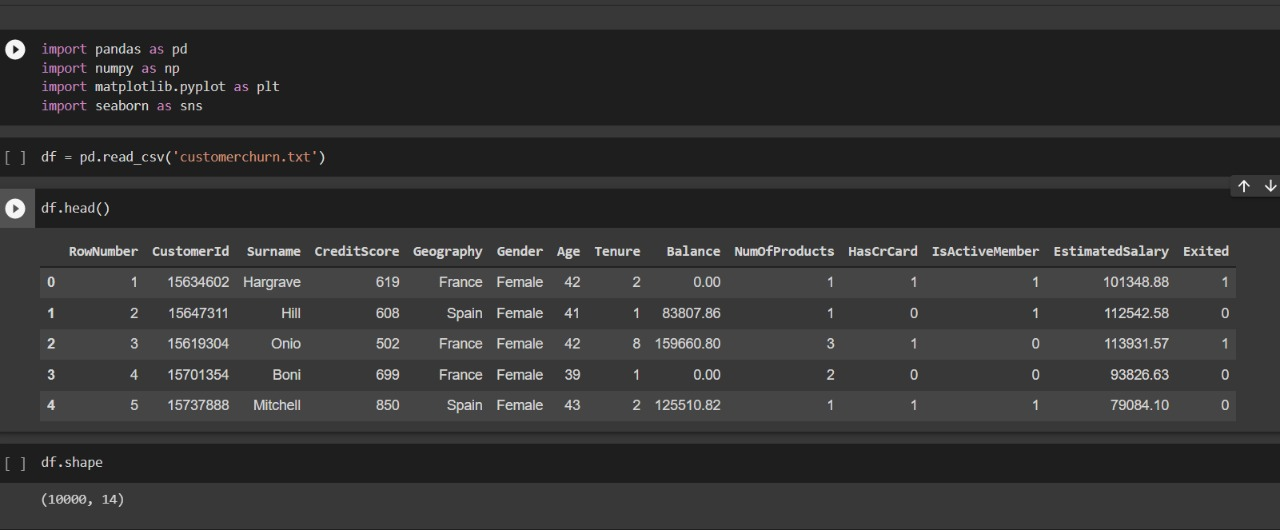


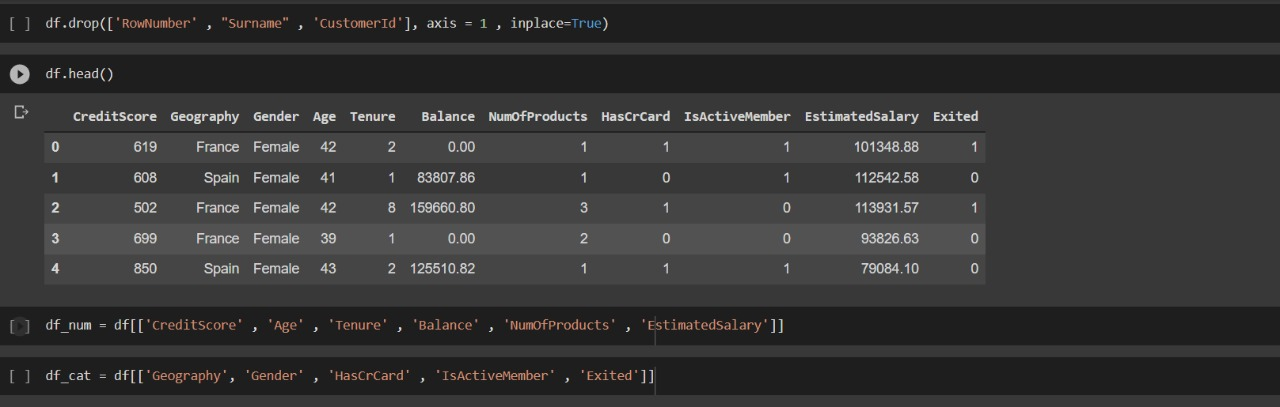
* The given box plot shows the estimated salary of employees across different countries and the two genders.

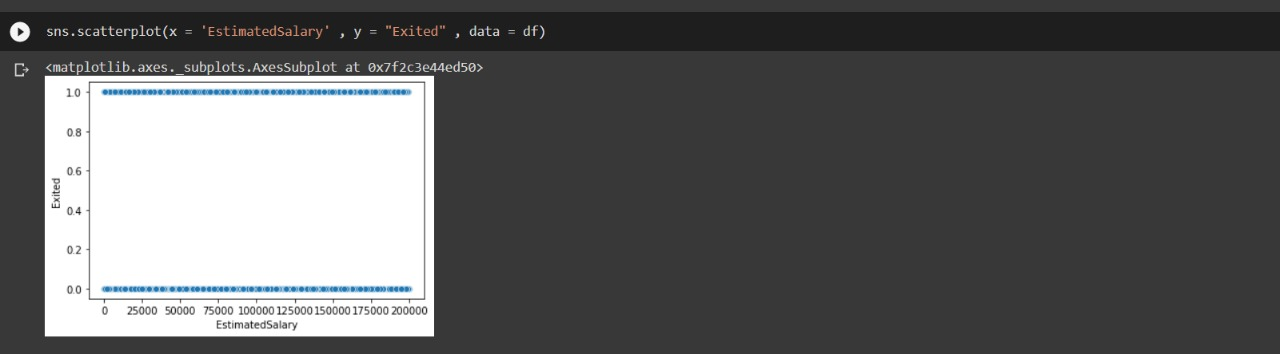
# Implementation

Tools used for our analysis are:

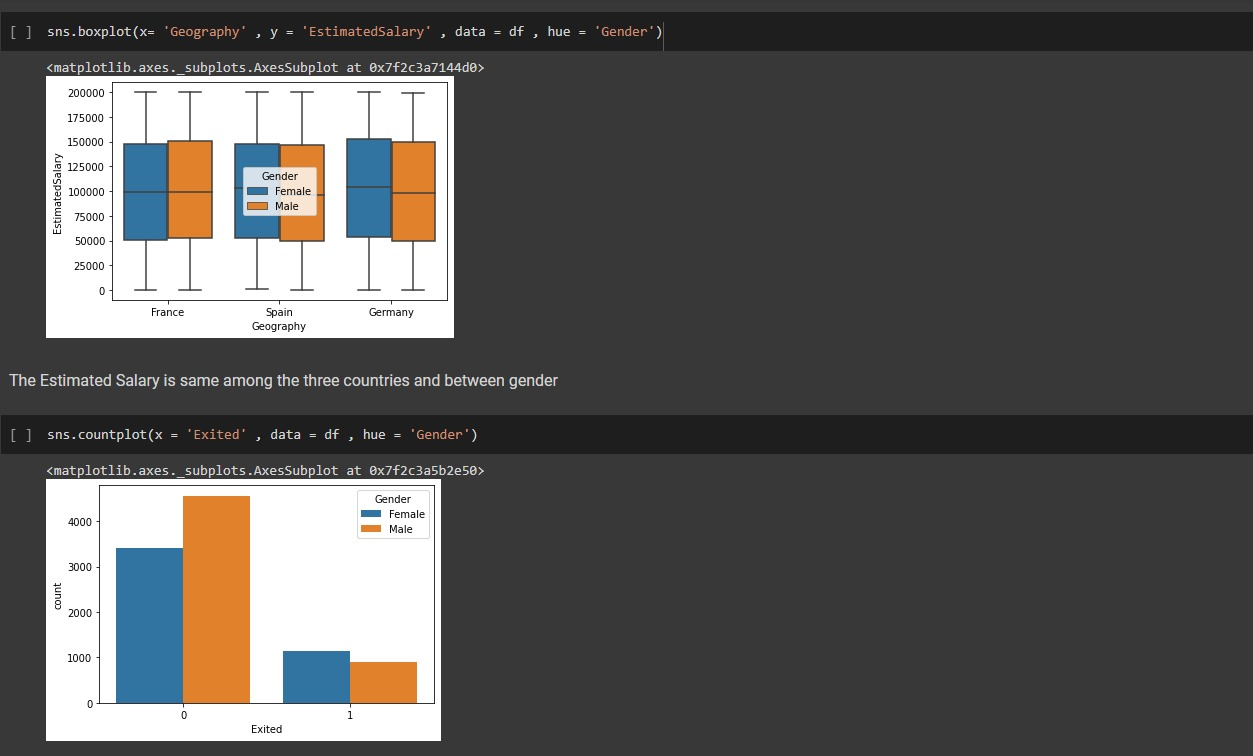
* Google Co-Lab
* Matplotlib
* Pandas
* Seaborn
* NumPy
* Sci-kit learn

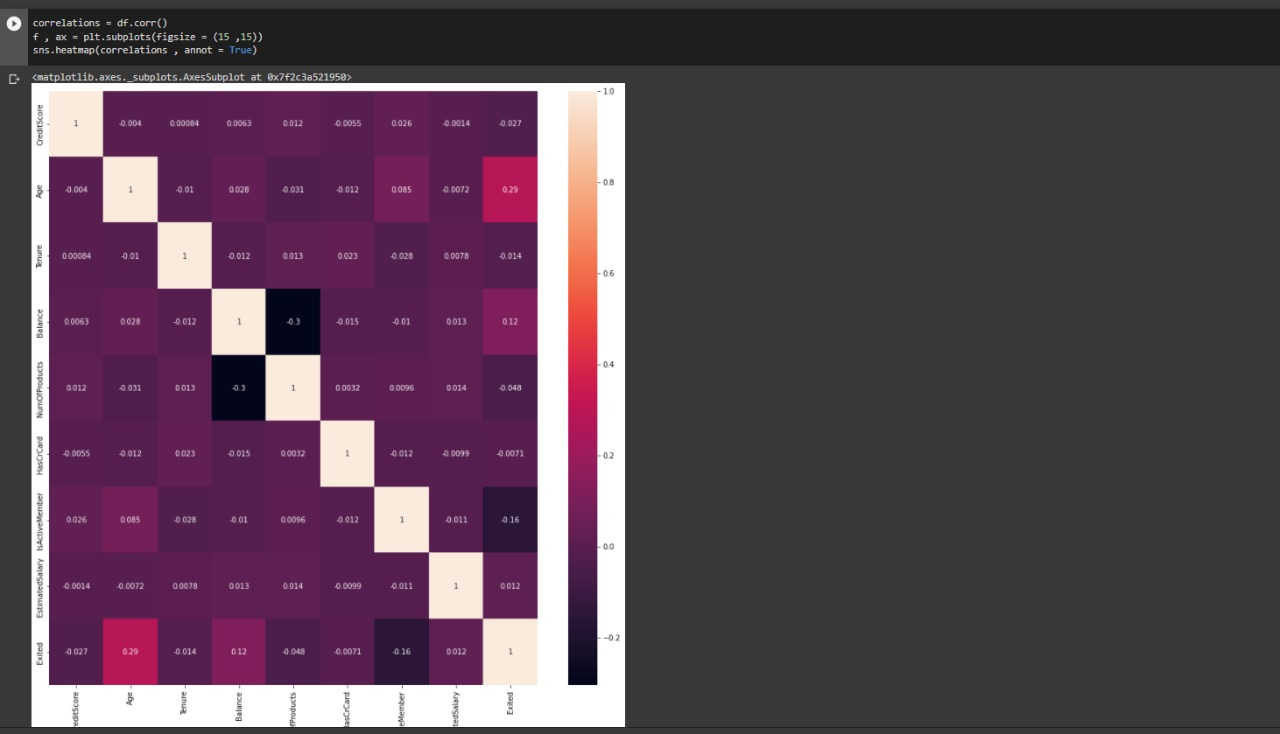


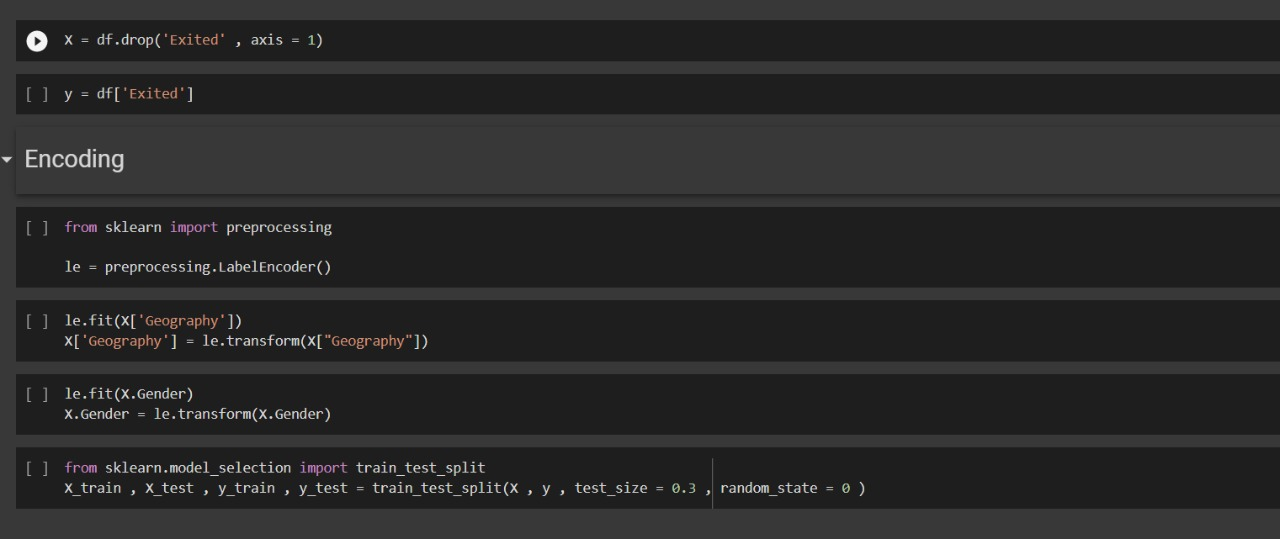


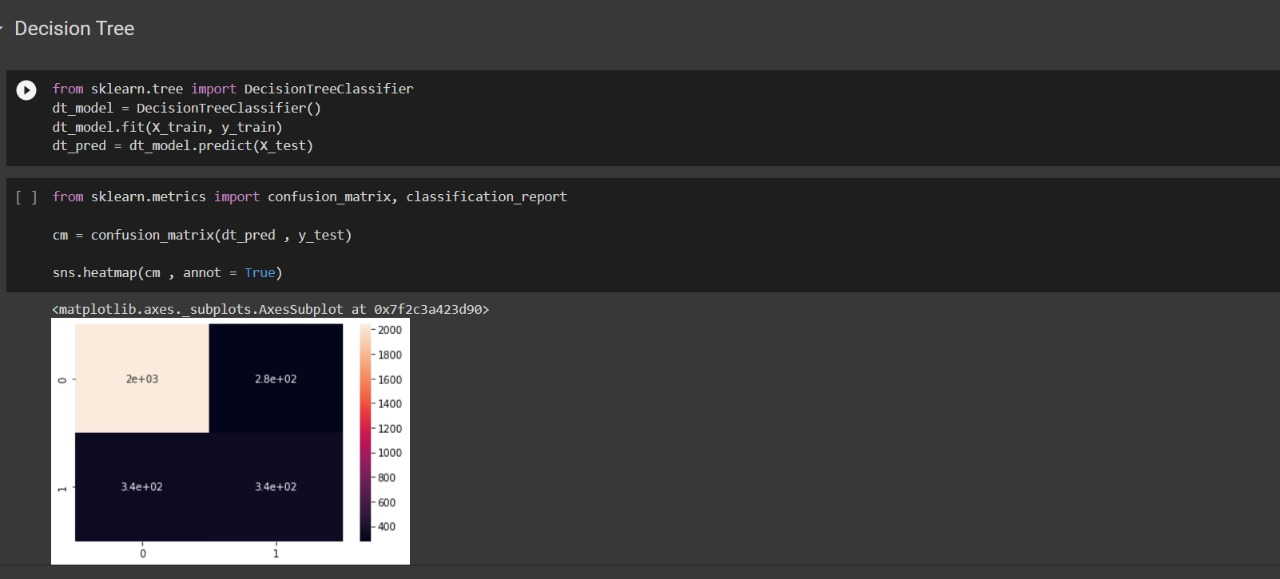


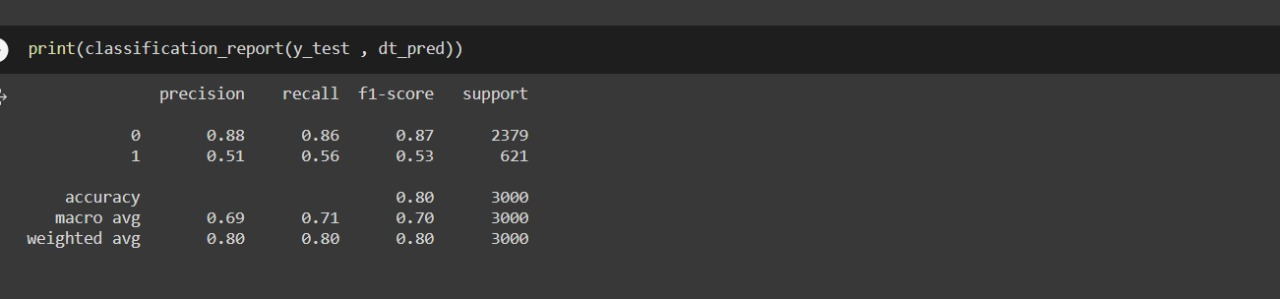


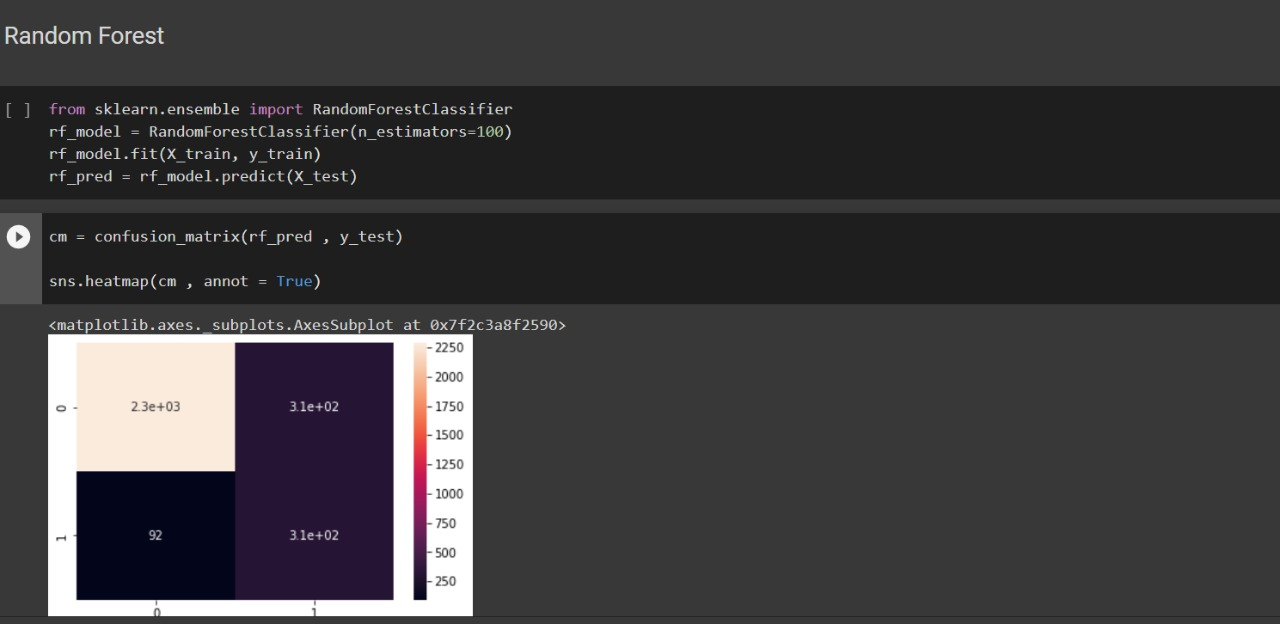


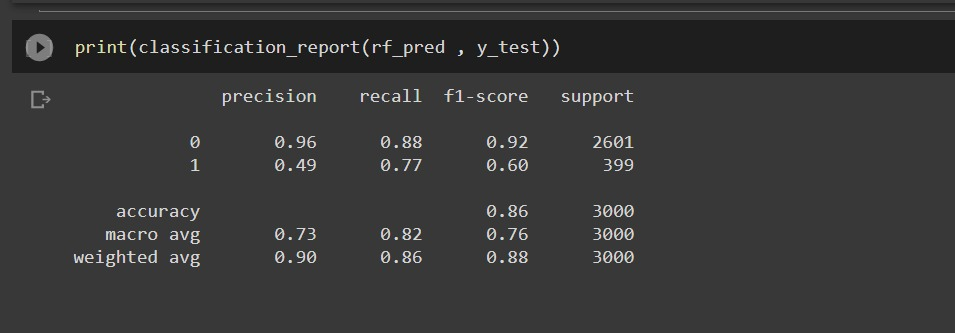








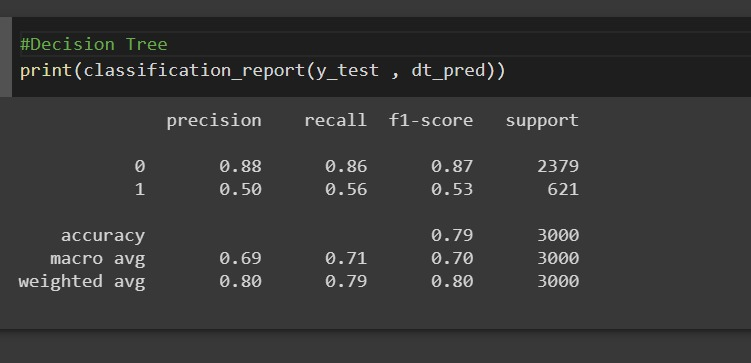




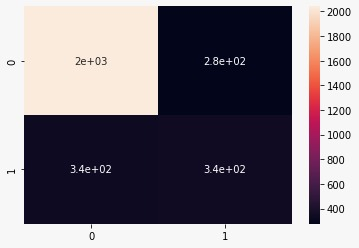
Output

**Decision Tree Classifier Model:**

Classification report

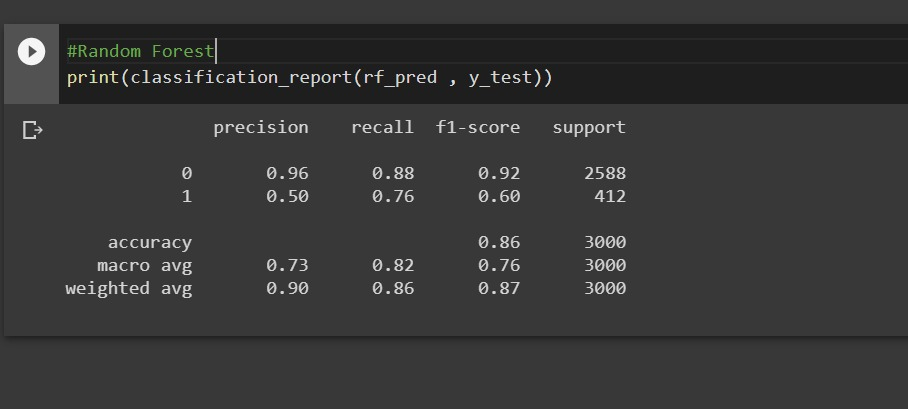


Confusion Matrix

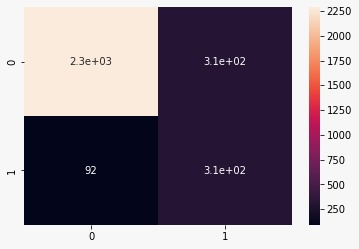


**Random Forest Model:**

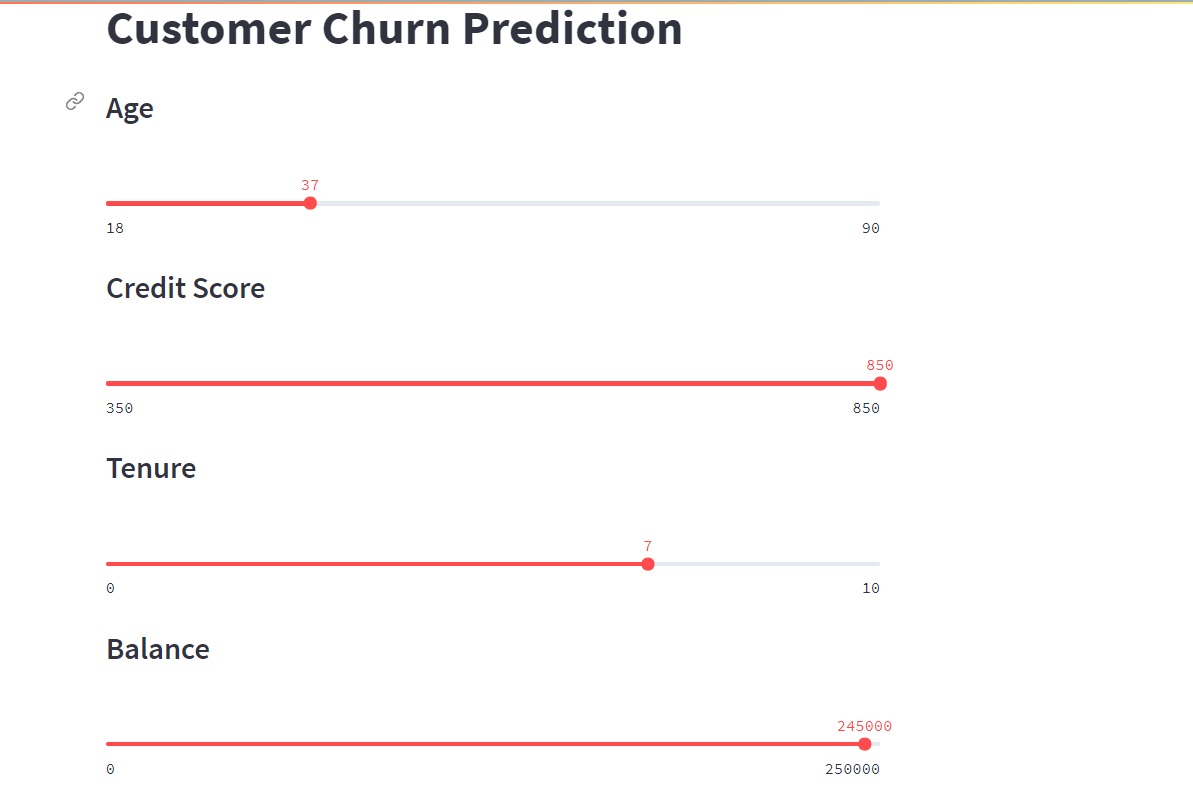
Classification report

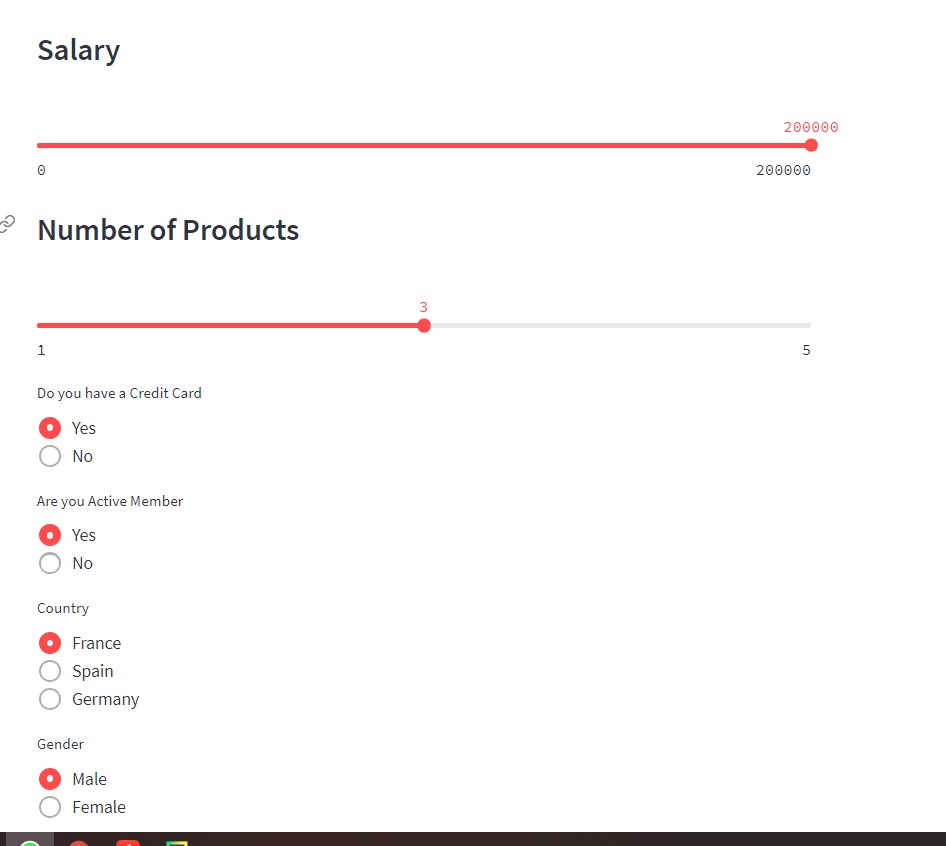


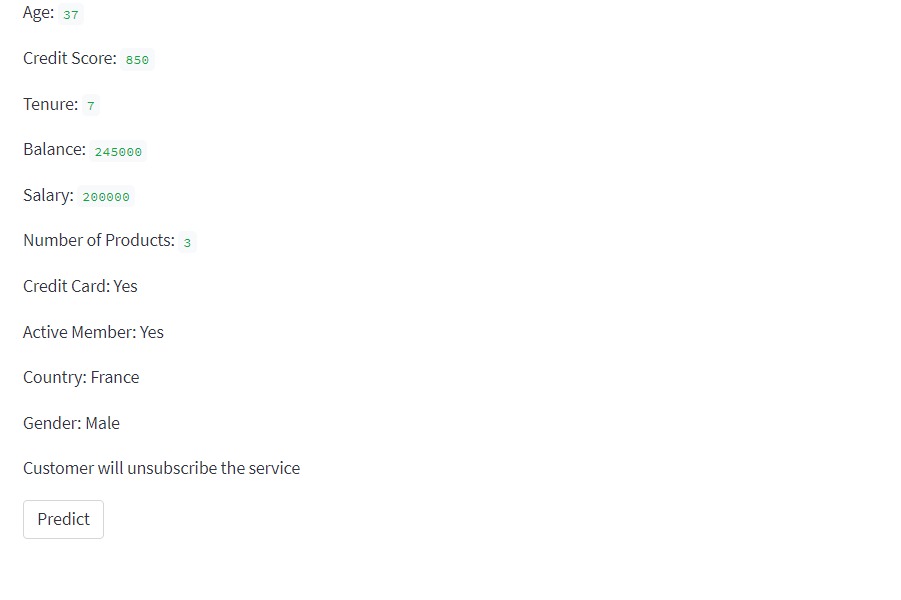
Confusion Matrix



**Deployment on Streamlit:**







Conclusion

* Based on the two predictive models we created, the one we created with Random Forest Classifier would be a better choice as it has better performance.
* With the existing consumer insights through data, companies can predict customers’ possible needs and issues, define proper strategies and solutions against them, meet their expectations and retain their business.
* Based on the predictive analysis and modelling, businesses can focus their attention with targeted approach by segmenting and offering them customized solutions.
* Analysing how and when the churn is happening in customer’s lifecycle with the services will allow the company to come up with more pre-emptive measures.