

MACHINE LEARNING - IT SEM 6 GRP 13

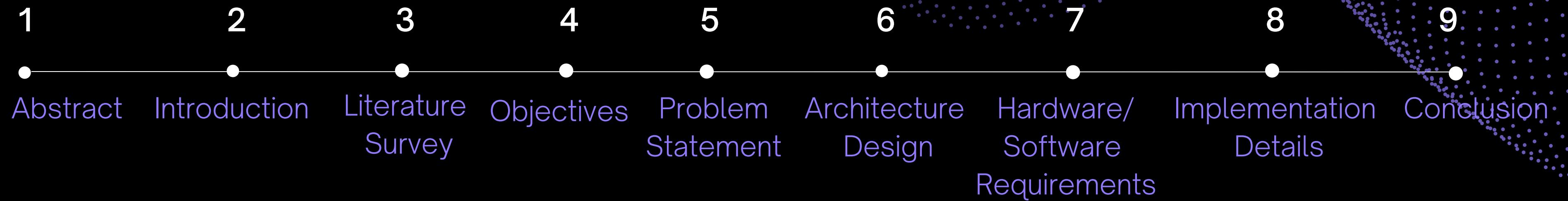
# Customer Attrition Prediction System

5019106 - Pratik Bhosale

5019146 - Saurabh Phadke

5019164 - Sam Thomas

# Topics



# Abstract

Customer churn, is the rate at which customers stop doing business with an entity. It is most commonly expressed as the percentage of service subscribers who discontinue their subscriptions within a given time period.

## Need for Customer churn

1. Decrease value flowing from customers to enterprises.
2. Loss of competitive advantage
3. Identify plausible reasons for attrition.

## Features of our dataset

## How customer churn prediction will benefit the company

1. Identify high risk customer
2. Enhance customer loyalty
3. Reduce customer maintenance cost

## Modes of Prediction

# Introduction

---

Globalization and  
advancements in  
telecommunication  
industry

Importance of Customer  
Attrition.

---

Database - The Most  
valuable asset

# Problem Statement

- To analyze attributes of dataset.
- To predict Costumer Attrition rate.
- Thus help companies retain their customer base and survive in the market.
- Deploy the model to ease usability.

# Scope

- To predict Customer Churn.
- Highlighting the main variables/factors influencing Customer Churn.
- Use various ML algorithms to build prediction models, evaluate the accuracy and performance of these models.
- Finding out the best model for our business case & providing executive summary.  
subheading

# Literature Survey

Sr. No.	Title of Paper	Year of Publication	Authors	Features
1	Customer churn prediction system: a machine learning approach	January 2021	Praveen Lalwani, Manas Kumar Mishra, Jasroop Singh Chadha, Pratyush Sethi	<ul style="list-style-type: none"> <li>1. Tested using multiple algorithms.</li> <li>2. Uses K-fold cross validation.</li> <li>3. Uses gravitational search algorithm to perform feature selection.</li> <li>4. Evaluation on test set using confusion matrix and AUC Curve.</li> </ul>
2	Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping	May 2021	Tianpei Xu, Ying Ma, Kangchul Kim	<ul style="list-style-type: none"> <li>1. System consists of feature construction, stacking model and soft voting.</li> <li>2. The stacking model consists of two levels with four algorithms: Xgboost (XGB), Logistic regression (LR), Decision tree (DT) and Naive Bayes classifier (NBC) to achieve better prediction accuracy.</li> <li>3. The third step consists of a soft voting.</li> </ul>
3	Analysis of Customer Churn Prediction in Telecom Industry Using Logistic Regression	July 2021	K. Sandhya Rani, Shaik Thaslima, N.G.L. Prasanna, R.Vindhya, P. Srilakshmi	<ul style="list-style-type: none"> <li>1. Uses R programming language.</li> <li>2. Uses Logistic Regression and Decision Tree Machine Learning algorithms.</li> <li>3. Confusion matrix is used for evaluation.</li> </ul>
4.	Customer churn prediction in telecom using machine learning in big data platform	March 2019	Abdelrahim Kasem Ahmad, Assef Jafar , Kadan Aljoumaa	<ul style="list-style-type: none"> <li>1. Uses Big Data.</li> <li>2. Makes use on Social Network analysis (SNA) for improving accuracy.</li> </ul>

# Architecture Design

---

Data Preprocessing



Exploratory Data Analysis



Model Creation



Model Improvement



Model Evaluation and Selection



Comparing Predictions against Test Set



Deploying Model to Application

# Hardware Requirements

- RAM : 4 GB or more.
- Processor : Intel i5 or equivalent.
- Hard disk: Minimum 10 GB.

# Software Requirements

- Python 3
- IDE (Jupyter notebook)
- Streamlit

# Implementation Details

Our Prediction Model for Customer Attrition in Telecom Industry uses a dataset from Kaggle which consists of 7043 tuples and 21 attributes consisting of both Categorical and Numerical values.

## Preprocessing

- The dataset is checked for any missing values and the same are handled using methods like replacing with mean.
- For categorical values containing 2 or less unique values we perform Label Encoding.

## Exploratory Data Analysis

- Establish relations between the attributes to extract meaning from the data.
- For this process, libraries like matplotlib and scipy are used.
- Plotted different types of Univariate and Multivariate graphs like Bar, Stacked graph, Scatter plot, Heatmap, etc.

## Model Selection

- We have used multiple models like Logistic Regression, Support Vector Machine, Naïve Bayes, K-Nearest Neighbours, Decision Tree and Random Forest Classifier and then compared these models based on ROC AUC and Accuracy values.
- AUC (Area Under The Curve) - ROC (Receiver Operating Characteristics) curve is a performance measurement for the classification problems.
- The Precision, Recall, F1 score and F2 score is also calculated for all the algorithms.

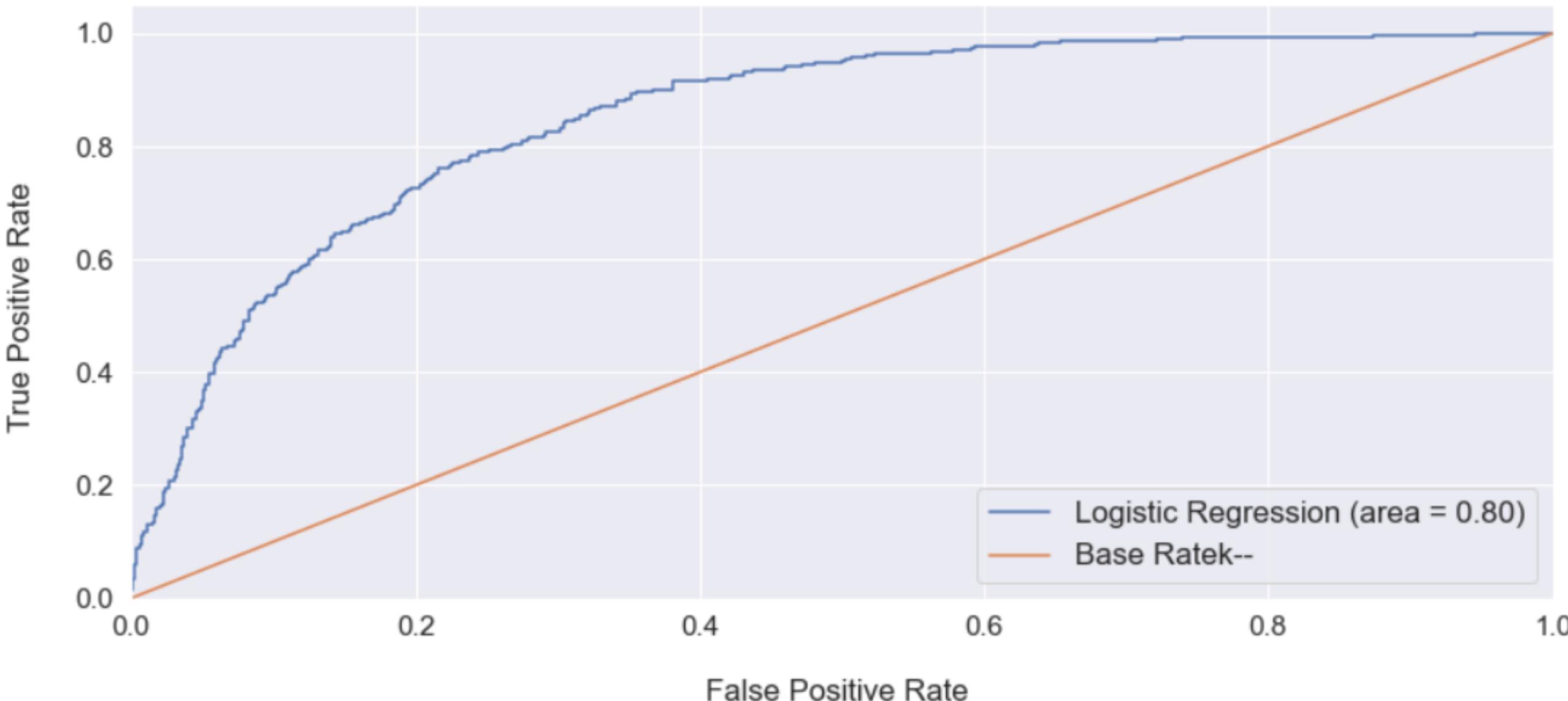
- From Comparison on these values for all the algorithms we found that Logistic Regression performs the best.
- Rechecked the accuracy of Logistic Regression using K-Folds Cross Validation and got results as 0.80 (+/- 0.04).
- Finally we will deploy this model to an app.

## Logistic Regression

---

Logistic regression is a process of modeling the probability of a discrete outcome given an input variable

ROC Graph



# Result and Conclusion

- Successfully performed EDA and created a model using Logistic Regression Algorithm to accurately predict the Customer Attrition.
- Successfully deployed the model to a Web App created using Streamlit which is intuitive and user-friendly so as to enable users to make use of the model effortlessly and even without much knowledge about the background working of the model.
- The App provides two modes :
  1. Online Mode
  2. Batch Mode



# Future Scope

- Improve Accuracy and Performance
- Improve User Interface to make it attractive, intuitive and user-friendly.

# References

- [1] Customer churn prediction system: a machine learning approach.  
<https://link.springer.com/article/10.1007/s00607-021-00908-y>
- [2] Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping.  
<https://www.mdpi.com/2076-3417/11/11/4742>
- [3] Analysis of customer churn prediction in telecom industry using decision trees and logistic regression.  
<https://ieeexplore.ieee.org/document/7570883>
- [4] Customer churn prediction in telecom using machine learning in big data platform.  
<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0191-6>
- [5] <https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

THANK YOU