MAJOR PROJECT REPORT

Name: SATHVIK S.R

Project Name: Machine Learning Major Project

Project Deadline: 30-September-2024

Project Statement: Predicting Customer Churn for a Telecom Company.

OBJECTIVE

This project aims to forecast client attrition for a telecom firm, therefore determining which ones are most likely to discontinue using its products. This will help the business to concentrate retention initiatives on high-risk clients, hence lowering turnover and increasing profitability.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

For telecom firms, client turnover is a major issue as losing one immediately affects income. By means of prediction of turnover, businesses may spot clients most likely to depart and work to keep them. Using demographic data and past usage trends, machine learning models were applied in this research to forecast consumer attrition.

CHAPTER 2

DATA REVIEW

Customer demographic data, services consumed, and financial data including monthly and total costs comprise the dataset utilized in this study. Churn is the goal variable; it is a binary value denoting either client departure or retention.

Features in a dataset:

- **Demographics:** Gender, Senior Citizen, Partner,
- **Dependents Services:** Internet Service, Contract Type, Phone Service, Streaming Services
- **Billing Information:** Monthly Charges, Total Charges
- **Target Variable:** Churn (Yes/No)

CHAPTER 3

DATA PREPROCESSING

To prepare the data for model creation, numerous preprocessing processes were carried out:

- **Handling Missing Values:** Rows with missing values in the Total Charges column were eliminated.
- **Categorical Encoding:** Categorical data including gender, contract, and payment method were translated into numerical representation using one-hot encoding.
- **Feature Scaling**: Continuous variables such as tenure and monthly costs were scaled to improve model performance.
  After preprocessing, the data was separated into training (70%) and testing (30%) sets.

CHAPTER 4

Exploratory Data Analysis (EDA)

Exploratory data research yielded several crucial findings regarding the variables causing churn:

- **Contract Type:** Customers with month-to-month contracts churn more frequently compared to those on one or two-year contracts.
- **Monthly costs:** Higher monthly costs related to increased turnover.

- **Tenure:** Customers with shorter tenure were more likely to churn, showing that maintaining new customers is more tough.
- **Internet Service Type:** Customers with Fiber optic internet churned at a greater rate than those utilizing DSL or no internet services.

CHAPTER 5

CONSTRUCTION OF MODELS

Several classification models were tested to predict customer churn:

- **Logistic Regression:** A simple, interpretable model.
- **Decision Tree:** A model that provides easy-to-understand decision rules.
- **Random Forest:** An ensemble model that typically performs better by reducing overfitting. The models were evaluated based on accuracy, precision, recall, and F1-score. These metrics help to understand how well the models distinguish between customers who churn and those who don't.

The performance of each model is summarized below:

MODEL	ACCURACY	PRECISION	RECALL	F1-SCORE
LOGIC REGRESSION	79%	70%	65%	67%
DECISION TREE	78%	71%	61%	65%
RANDOM FOREST	81%	75%	69%	72%

Random Forest outperformed the other models with the highest accuracy, precision, recall, and F1-score, making it the most reliable model for predicting customer churn.

CHAPTER 6

IMPORTANT OBSERVATIONS

- **Contract Type**: Predicting turnover most relies on this element. Comparatively to those on longer-term contracts, customers on month-to- month contracts were considerably more likely to leave.
- **Monthly Charges:** Higher monthly prices paid by consumers increased their risk of leaving. Providing personalised discounts or competitive prices might assist to keep these clients.
- **Tenure:** Shorter term customers are more likely to leave, which emphasizes the importance of concentrating retention efforts on more recent ones.
- **Internet Service Type:** Customers utilizing Fiber optic internet left more often, maybe in response to more competition from other providers or more expenses.

CHAPTER 7

ECONOMIC CONSEQUENCES

The company depends much on this churn prediction model:

- **Targeted Retention Campaigns:** The strategy helps the business to spot high-risk consumers and start individualized retention initiatives like discounts or improved services.

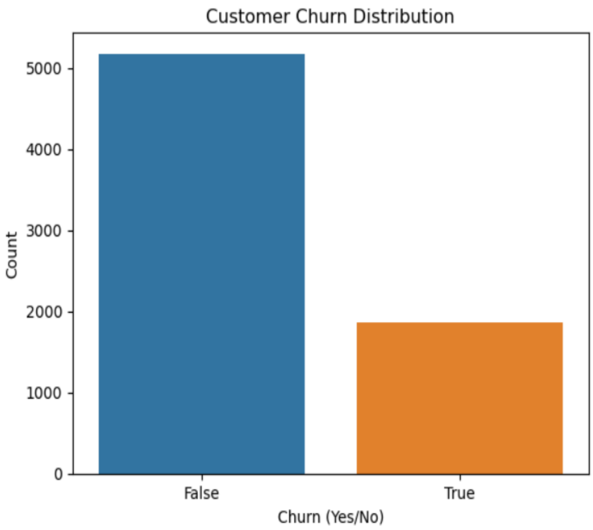
- **Cost Reduction:** By concentrating on at-risk consumers, the firm may optimize its marketing and customer service expenditures, enhancing return on investment (ROI) on retention initiatives.
- **Improved Customer Satisfaction:** Proactively addressing the requirements of consumers likely to churn might boost satisfaction and brand loyalty.

CHAPTER 8

CONCLUSION

The Random Forest model attained an accuracy of 81%, making it a viable tool for predicting telecom customer attrition. The primary variables driving turnover are contract type, monthly expenses, and client tenure. By employing this methodology, telecom businesses may focus their retention efforts on at-risk consumers, lowering churn and raising total income.

OUTPUT:



	precision	recall	f1-score	support
False	0.81	0.94	0.87	1539
True	0.71	0.40	0.51	574
accuracy			0.79	2113
macro avg	0.76	0.67	0.69	2113
weighted avg	0.78	0.79	0.77	2113

```
Accuracy: 0.79
F1-Score: 0.51
```

PROJECT PREVIEW:

https://nbviewer.org/github/sathviksr2001/Jupyter-Notebook/blob/main/Predicting%20Customer%20Churn%20for%20a%20Telecom%20Company.ipynb