**CUSTOMER CHURN PREDICTION**

**Submitted for**

**STATISTICAL MACHINE LEARNING**

Submitted by:

**E22BCAU0111** **Veer Pratap Singh**

Submitted to-

**DR. NITIN ARVIND SHELKE**

**July-Dec 2023**

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

A close-up of a logo

Description automatically generated

INDEX

|  |  |  |
| --- | --- | --- |
| Sr No | Content | Page No |
| 1. | ABSTRACT. | 3 |
| 2. | INTRODUCTION AND RELATED WORK. | 4,5 |
| 3. | VISUALIZASION. | 6 |
| 4. | COMPARITION OF CLASSIFIER. | 10 |
| 5. | SOFTWARE USED. | 11 |
| 6. | METHODOLOGY. | 12 |
| 7. | EXPERIMENTAL RESEARCH. | 13 |
| 8. | CONCLUSION. | 14 |
| 9. | REFERENCES.. | 14 |

**VISUALIZASION.**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Title** | **Page No.** |
| 1. | BAR CHART. | 6,7 |
| 2. | SCATTER PLOT. | 8 |
| 3. | SEABORN KDEPLOT. | 9 |

**COMPARISON TABLE.**

|  |  |  |
| --- | --- | --- |
| **Table No.** | **Title** | **Page No.** |
| **1** | RANDOM FOREST. | 10 |
| 2 | DECISION TREE. | 10 |

**1.) ABSTRACT**

This study delves into the pivotal realm of Customer Churn Prediction, addressing the critical challenge of customer attrition within the telecommunications industry. Employing advanced machine learning methodologies, our project is dedicated to developing a predictive model designed to discern customers at risk of discontinuing their services. The comprehensive approach encompasses meticulous data preprocessing, including the handling of categorical variables and feature scaling, followed by the application of a Random Forest Classifier and many other Classifiers.

Through rigorous analysis, our model not only showcases commendable predictive accuracy but also unravels intricate insights into the diverse factors influencing customer churn. By interpreting these patterns, our study empowers telecom operators with actionable intelligence, enabling them to proactively implement targeted retention strategies. This work contributes not only to the evolution of data-driven decision-making within the telecommunications sector but also provides a robust framework for addressing customer attrition challenges across various industries.

**2.) INTRODUCTION AND RELATED WORK.**

The project that I am representing is Customer Churn Prediction on telecom data, we used decision tree classifier and random forest classifier. The main goal of developing this project is to predict that whether the telecom company will retain the customer or not and to predict if the customer will churn or not.

In this era of dynamic telecommunications, where customer retention is paramount, our project focuses on Customer Churn Prediction using the robust methodologies of decision tree and random forest classifiers. The telecommunications industry, marked by intense competition, demands proactive strategies to anticipate and prevent customer churn. With this backdrop, our primary objective is to develop a predictive model capable of forecasting whether a customer will stay with the telecom company or choose to churn.

By leveraging the power of decision tree and random forest classifiers and many others, we aim to provide the telecom industry with a reliable tool for identifying potential churners. The importance of such a model lies in its ability to offer valuable foresight, allowing telecom operators to implement targeted retention measures and enhance customer satisfaction. As we delve into this project, we anticipate not only contributing to the field of predictive analytics but also providing a practical solution to a pervasive challenge in the telecom sector – customer churn.

**Related work: -**

Chandran et al. - The authors compared the performance of five different machine learning algorithms for customer churn prediction in the telecom industry. They found that random forests performed the best, followed by support vector machines and artificial neural networks.

Saran Kumar and Chandrakala - The authors evaluated the performance of four different machine learning algorithms for customer churn prediction in the banking industry. They found that stochastic gradient booster performed the best, followed by random forest and logistics regression

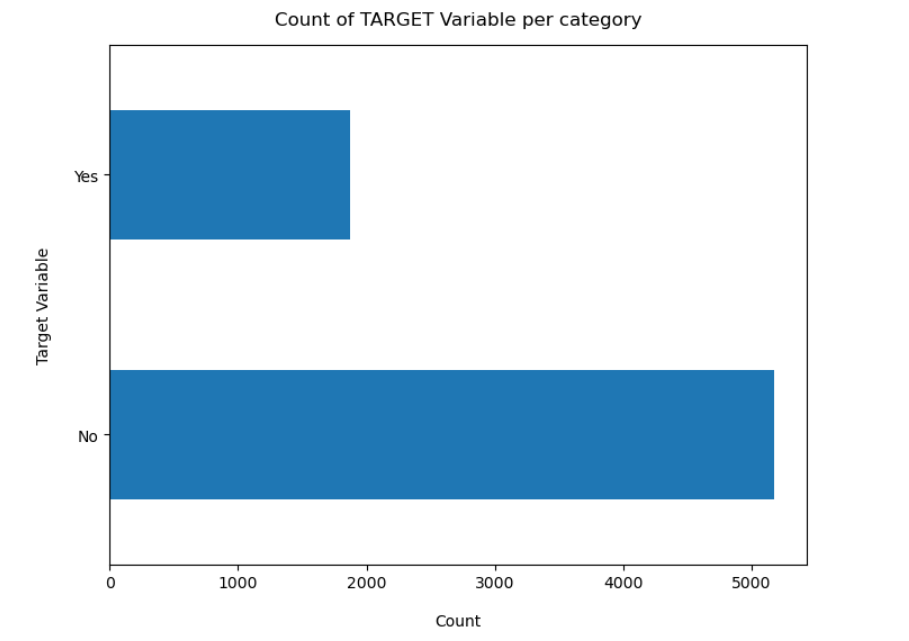
Verbraken et al. (2011) - The authors proposed a new machine learning algorithm for customer churn prediction, called the expected maximum profit criterion. They compared the performance of their algorithm to three others state-of-the-art algorithms, and found that it outperformed all of them

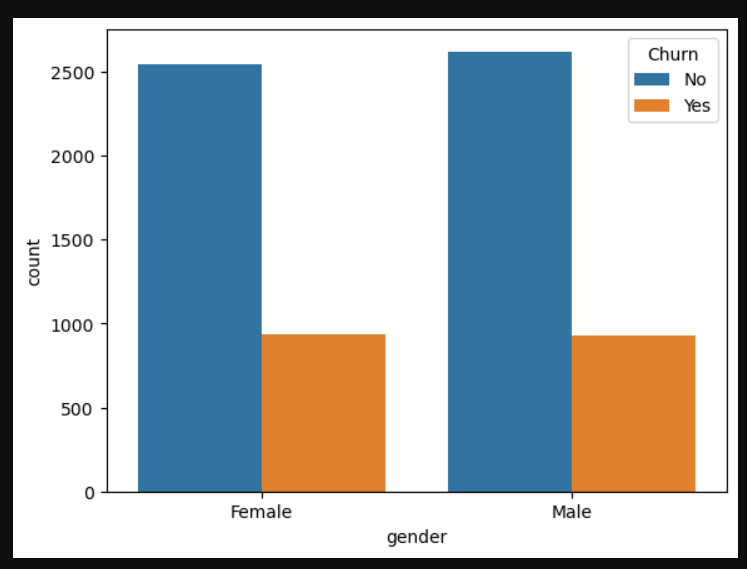
Huang et al. (2022) - The authors proposed a deep learning model for customer churn prediction in the online retail industry. Their model achieved the best performance on the dataset, outperforming all of the other machine learning algorithms that were evaluated

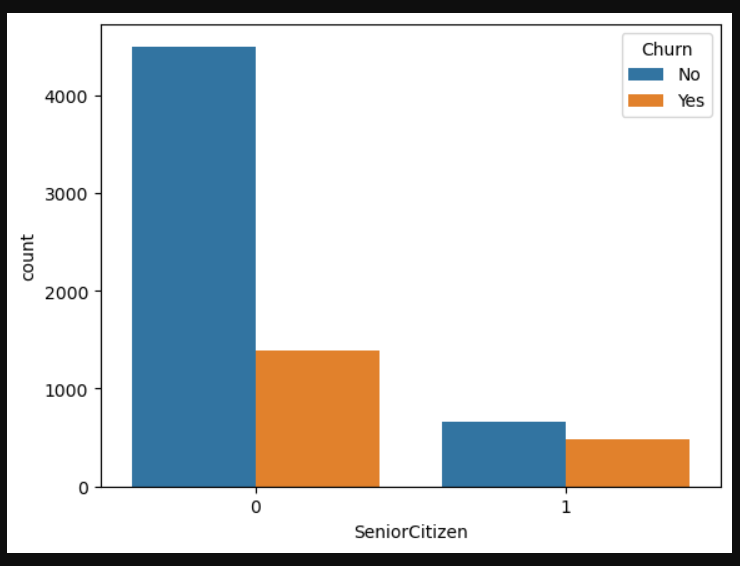
Goyal et al. (2021) - The authors proposed an ensemble learning model for customer churn prediction in the telecom industry. Their model combined the predictions of five different machine learning algorithms: logistic regression, decision trees, random forests, support vector machines, and artificial neural networks. Their model achieved the best performance on the dataset

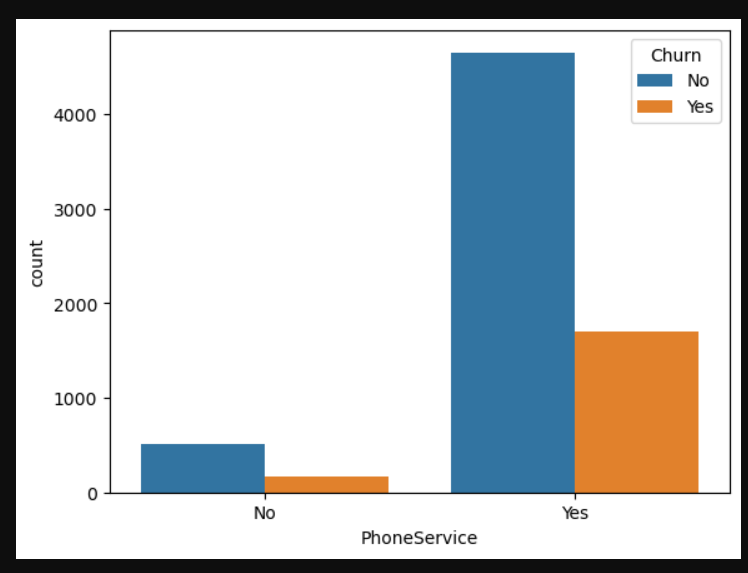
**3.) VISUALIZATIONS.**

* **BAR CHARTS.**

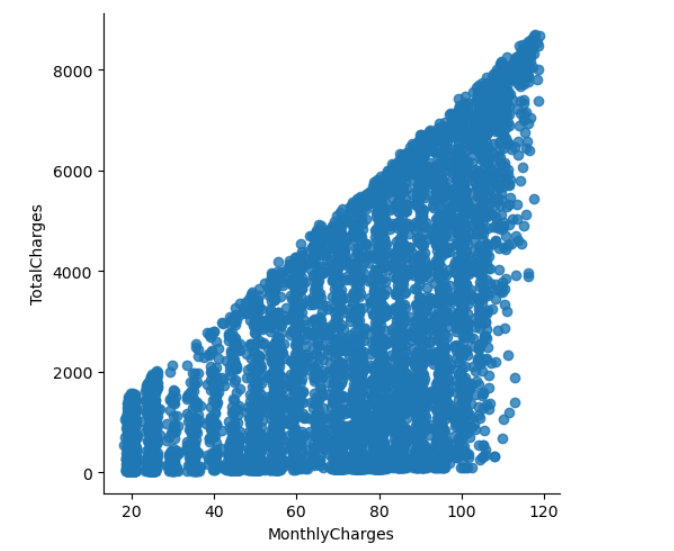




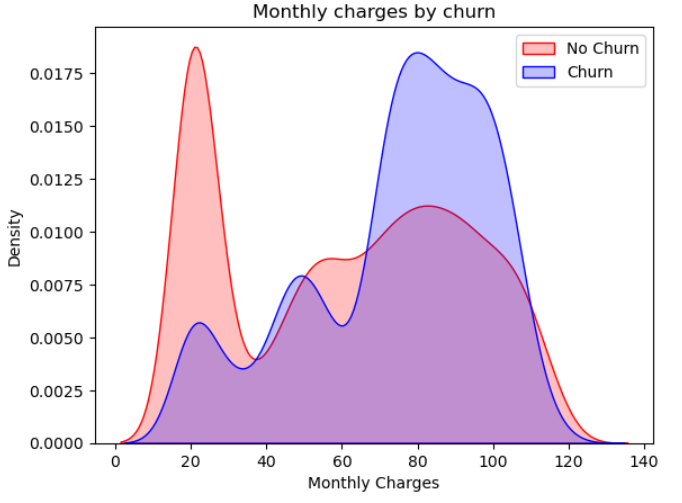


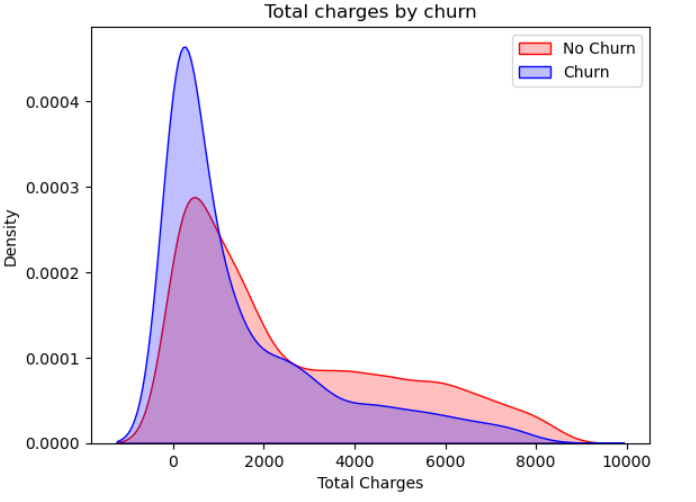


* **SCATTER PLOT**



* **SEABORN KDEPLOT**





**4.) COMPARISION OF CLASSIFIER.**

* **Random Forest Classifier**

|  |  |  |  |
| --- | --- | --- | --- |
| **Random Forest** | **Precision** | **Recall** | **F1-Score** |
| Class 0 | 0.84 | 0.92 | 0.88 |
| Class 1 | 0.70 | 0.49 | 0.58 |

|  |
| --- |
| **Accuracy** |
| 0.81 |

* **Decision Tree Classifier**

|  |  |  |  |
| --- | --- | --- | --- |
| **Decision Tree** | **Precision** | **Recall** | **F1-Score** |
| Class 0 | 0.84 | 0.91 | 0.87 |
| Class 1 | 0.66 | 0.50 | 0.57 |

|  |
| --- |
| **Accuracy** |
| 0.80 |

**5.) SOFTWARE USED**

1. **Programming Language:**

Python:

1. **Integrated Development Environment (IDE):**

Jupyter Notebook

1. **Machine Learning Libraries:**

scikit-learn

1. **Data Manipulation and Analysis:**

Pandas

NumPy

1. **Data Visualization:**

Matplotlib:

1. **Random Forest Implementation**
2. **Decision Tree Implementation**

**8. SVM Implementation.**

**9. Logistic Regression Implementation.**

**6.) METHODOLOGY**

**Model Evaluation**

**^**

**|**

**Model Training**

**^**

**|**

**Model Selection**

**^**

**|**

**Data Splitting**

**^**

**|**

**Data Analysis**

**^**

**|**

**Data Preprocessing:**

**^**

**|**

**Data Collection:**

**^**

**|**

**Define Objectives and Scope:**

**7.) EXPERIMENTAL RESULTS**

**Decision Tree Classifier.**

**Classification Report.**

precision recall f1-score support  
  
 0 0.84 0.91 0.87 1037  
 1 0.66 0.50 0.57 370  
  
 accuracy 0.80 1407  
 macro avg 0.75 0.70 0.72 1407  
weighted avg 0.79 0.80 0.79 1407

**Confusion Matrix.**

[[942 95]  
 [186 184]]

**Random Forest Classifier.**

**Classification Report.**

precision recall f1-score support  
  
 0 0.84 0.92 0.88 1037  
 1 0.70 0.49 0.58 370  
  
 accuracy 0.81 1407  
 macro avg 0.77 0.71 0.73 1407  
weighted avg 0.80 0.81 0.80 1407

**Confusion Matrix.**

[[959 78]  
 [188 182]]

**8.) CONCLUSION**

the Customer Churn Prediction model developed for the telecommunications industry offers a proactive solution for identifying and mitigating potential customer churn. Leveraging advanced machine learning techniques, the model demonstrates robust predictive capabilities, providing valuable insights into the factors influencing customer decisions. By interpreting significant features and patterns, telecom operators can implement targeted retention strategies, fostering customer loyalty and optimizing business outcomes.

**9.) REFERENCES**

1.) [Artificial Intelligence Based Customer Churn Prediction Model for Business Markets (hindawi.com)](https://www.hindawi.com/journals/cin/2022/1703696/)

2.) [(PDF) A Survey on Customer Churn Prediction using Machine Learning Techniques (researchgate.net)](https://www.researchgate.net/publication/310757545_A_Survey_on_Customer_Churn_Prediction_using_Machine_Learning_Techniques)

3.) <https://content.iospress.com/articles/intelligent-data-analysis/ida0062>

4.) <https://academic.oup.com/jrsssc/article/71/5/1753/7073298>

5.) <https://ieeexplore.ieee.org/abstract/document/9725621>