iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: X Month of publication: October 2023

DOI: https://doi.org/10.22214/ijraset.2023.55720

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

Customer Churn Analysis

Tahseen Begum[1], E. Pravallika[2], M. Rushitha[3], K. Kavya[4]
Computer Science Engineering, Koneru Lakshmaiah University, Hyderabad, India

Abstract: Customer churn analysis is a critical process for businesses to understand the reasons why customers stop using their products or services. It involves examining customer data and identifying patterns that can indicate potential churn. By analyzing this data, businesses can take proactive measures to retaincustomers, such as improving the quality of their offerings or targeting specific customer segments. Ultimately, effective customer churn analysis can help businesses reduce customer churn and increase profitability.
Keywords: Customer satisfaction, Customer feedbackCustomer engagement, Customer loyalty.

I. INTRODUCTION

Customer churn analysis is the process of examining customer data to understand why customers stop using a product or service. It involves analyzing customer behavior, identifying patterns, and developing strategies to reduce customer churn. The goal of customer churn analysis is to retain customers and increase customer loyalty. Some common metrics used in customer churn analysis include customer retention rate, churn rate, and customer lifetime value. By analyzing these metrics, businesses can identify the factors that contribute to customer churn and take steps to address them. Customer churn analysis often involves using predictive models to identify customers who are at risk of churning. These models use historical data to predict which customers are most likely to stop using a product or service. Once at-risk customers have been identified, businesses can develop targeted retention strategies to keep them engaged. These strategies might include personalized offers, improved customer service, or product improvements. In addition to reducing customer churn, customer churn analysis can also help businesses identify new growth opportunities. By analyzing customer behavior, businesses can identify new markets or products that might be of interest to their customers. Overall, customer churn analysis is a critical process for any business that wants to retain customers and grow its customer base. By understanding why customers leave and taking steps to keep them engaged, businesses can improve customer loyalty anddrive long-term growth.

II. LITERATURE SURVEY

1) A Survey on Data Mining Techniques In CustomerChurn Analysis For Telecom Industry - Amal M. Almana, (2020).

This survey focuses on data mining techniques utilized for customer churn analysis in the telecom industry. The survey investigates the most commonly used data mining techniques for customer churn analysis, such as decision trees, logistic regression, and artificial neural networks.

The survey also examines the effectiveness of these techniques in predicting customer churn and retaining customers. The study sheds light on the importance of customer churn analysis and the role of data mining in enabling telecom companies to understand their customers better and develop effective retention strategies. The survey concludes with recommendations for future research in this area. Customer churn analysis refers to the process of identifying customers who are likely to switch to another service provider or terminate their service altogether.

Telecom companies use customer churn analysis to identify the factors that contribute to customer churn and develop strategies to retain customers. Data mining techniques are often used in customer churn analysis to extract valuable insights from large datasets. These techniques include classification, clustering, association analysis, and predictive modeling. Classification involves categorizing customers into groups based on their behavior or characteristics. Clustering involves grouping customers based on similarities in their behavior or characteristics.

Association analysis involves identifying relationships between different variables, such as the products or services that customers are using. Predictive modeling involves using historical data to make predictions about future customer behavior. Overall, data mining techniques can help telecom companies to identify the factors that contribute to customer churn, predict which customers are most likely to churn, and develop strategies to retain customers.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

2) *Mobile Phone Data: A Survey of Techniques, Features,and Applications- Mohammed Okmi, Lip Yee Por, Tan Fong Ang, and Chin Soon Ku, (2020).*

A comprehensive review paper that provides an overview ofthe various techniques, features, and applications of mobile phone data. The paper discusses how mobile phone data can be used for a wide range of applications such as mobile network optimization, urban planning, transportationanalysis, and health monitoring.

The paper also explores various techniques for processing and analyzing mobile phone data, such as data pre-processing, data fusion, clustering, classification, and regression. It describes the advantages and disadvantages of each technique and provides examples of their applications in different contexts.Additionally, the paper discusses various features of mobile phone data, such as location data, call and text message data, and app usage data. It explains how these features can be used to extract useful information and insights about user behavior, preferences, and activities. The results also show that classification and clustering approaches have been widely used where algorithms such as SVM and k-means are used to classify or cluster human activities based on theircalling or mobility features.

The results show that the spatiotemporal calling feature has been widely used to depict human behaviors. This feature allows for the estimation of hourly dynamic population, the classification of land use, the estimation of ambient population, the investigation of the relationship between human dynamics and crime spatial–temporal patterns, defining the actual populations at risk, and others. Overall, the survey provides a comprehensive overview of the potential uses of mobile phone data and the various techniques that can be applied to analyze it. It highlights the importance of mobile phone data in providing insights into user behavior and improving various services and systems.

3) *Analysis of customer churn prediction in the telecomindustry using decision trees and logistic regression:lonu Brându oiu, Gavril Toderean, Horia, (2021).*

Customer churn is a major challenge for telecom companiesas it leads to a loss of revenue and market share. Customer churn prediction using decision trees and logistic regressionis a common approach to identify factors that contribute tocustomer churn and develop strategies to retain customers. Decision trees are a powerful tool for analyzing customerchurn prediction as they can capture non-linear relationshipsbetween variables and help identify key drivers of customerchurn. Decision trees split the data into subsets based on thevalues of predictor variables, creating a tree-like structurethat shows how different variables interact to influencecustomer churn.

Logistic regression is another commonlyused approach for customer churn prediction. It models therelationship between customer churn and predictor variablesby estimating the probability of churn based on the values ofthese variables. Logistic regression allows for theidentification of significant predictors of churn and thecalculation of their impact on the likelihood of churn. The telecom industry generates large amounts of data that can beused for customer churn prediction.

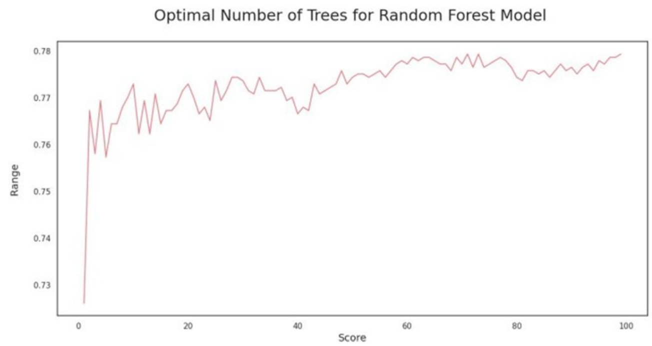
Variables that can beused for churn prediction include demographic information,usage patterns, and customer satisfaction ratings. By analyzing these variables using decision trees and logistic regression, telecom companies can develop targeted retention strategies that focus on the key drivers of customer churn. Overall, customer churn prediction using decision trees and logistic regression is an effective method fortelecom companies to identify factors that contribute tocustomer churn and develop targeted retention strategies.

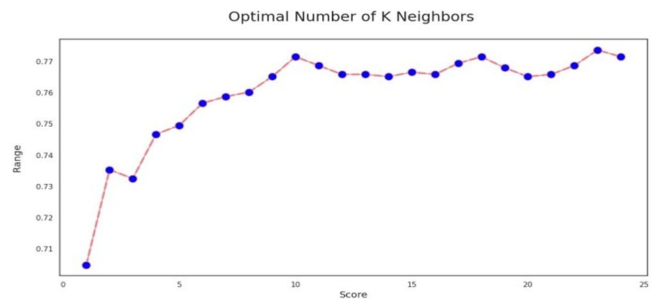
III. PROPOSED WORK

Customer churn analysis is a process of identifying customers who are likely to discontinue using a product or service, and understanding the reasons behind their decision. The analysis involves collecting and analyzing data on customer behavior, demographics, and purchase history, among other factors, to identify patterns and potential predictors of churn. The insights gained from churn analysis can help businesses develop strategies to retain customers, improve customer experience, and ultimately increase revenue. Some common techniques used in churn analysis include statistical modeling, machine learning, and data visualization.

A. Graph Analysis

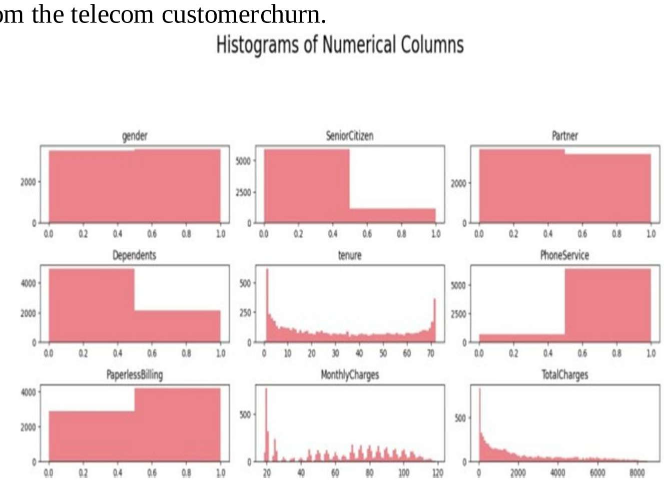
Graph analysis is the process of examining and interpreting data represented in graphical form. It involves identifying patterns, trends, and relationships within the data to gain insights and make informed decisions.

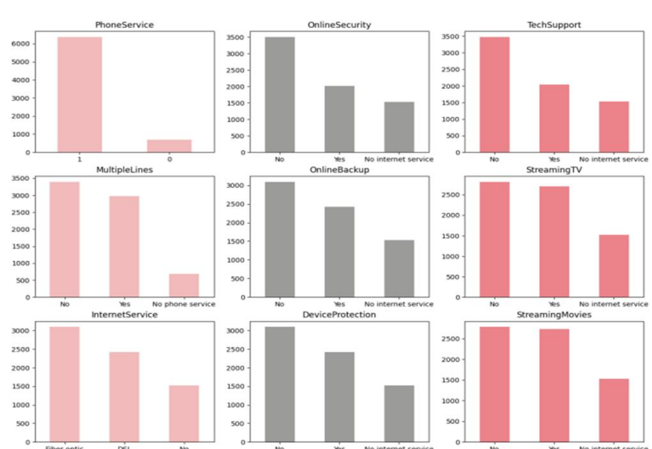




B. Data Analysis

We analyzed the collected data from the telecom customerchurn.





C. Big Data Analysis

We did an analysis of data using python here we have alarge dataset so we used google collab to analyze it.

```
#Recheck Column Datatypes and Missing Values:

dataset.columns.to_series().groupby(dataset.dtypes).groups

{int64: ['SeniorCitizen', 'tenure'], float64: ['MonthlyCharges'], object: ['customerID', 'gender', 'Partner', 'Dependents', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'TotalCharges', 'Churn']}

[10] dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   7043 non-null   object
 15  Contract          7043 non-null   object
 16  PaperlessBilling  7043 non-null   object
 17  PaymentMethod     7043 non-null   object
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   object
 20  Churn             7043 non-null   object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
```

Step 6: Clean the Dataset

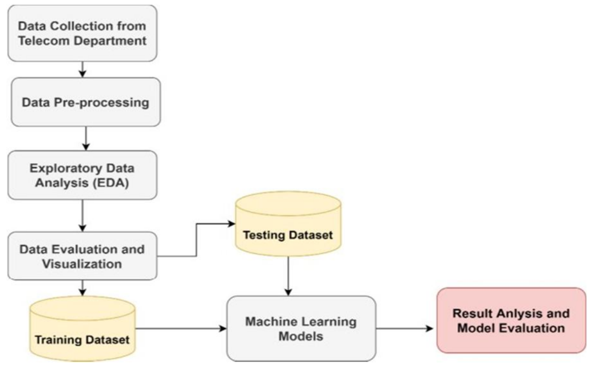
```
[18] dataset['TotalCharges'] = pd.to_numeric(dataset['TotalCharges'],errors='coerce')

     dataset['TotalCharges'] = dataset['TotalCharges'].astype("float")
```

Step 7: Take care of missing data

```
[19] dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   7043 non-null   object
 15  Contract          7043 non-null   object
 16  PaperlessBilling  7043 non-null   object
 17  PaymentMethod     7043 non-null   object
```

```
[11] dataset.isna().any()

customerID          False
gender              False
SeniorCitizen       False
Partner             False
Dependents          False
tenure              False
PhoneService        False
MultipleLines       False
InternetService     False
OnlineSecurity      False
OnlineBackup        False
DeviceProtection    False
TechSupport         False
StreamingTV         False
StreamingMovies     False
Contract            False
PaperlessBilling    False
PaymentMethod       False
MonthlyCharges      False
TotalCharges        False
Churn               False
dtype: bool

[12] #Unique values in each categorical variable:

     dataset["PaymentMethod"].nunique()

4

[13] dataset["PaymentMethod"].unique()

array(['Electronic check', 'Mailed check', 'Bank transfer (automatic)',
       'Credit card (automatic)'], dtype=object)
```

Flowchart



IV. RESULTS

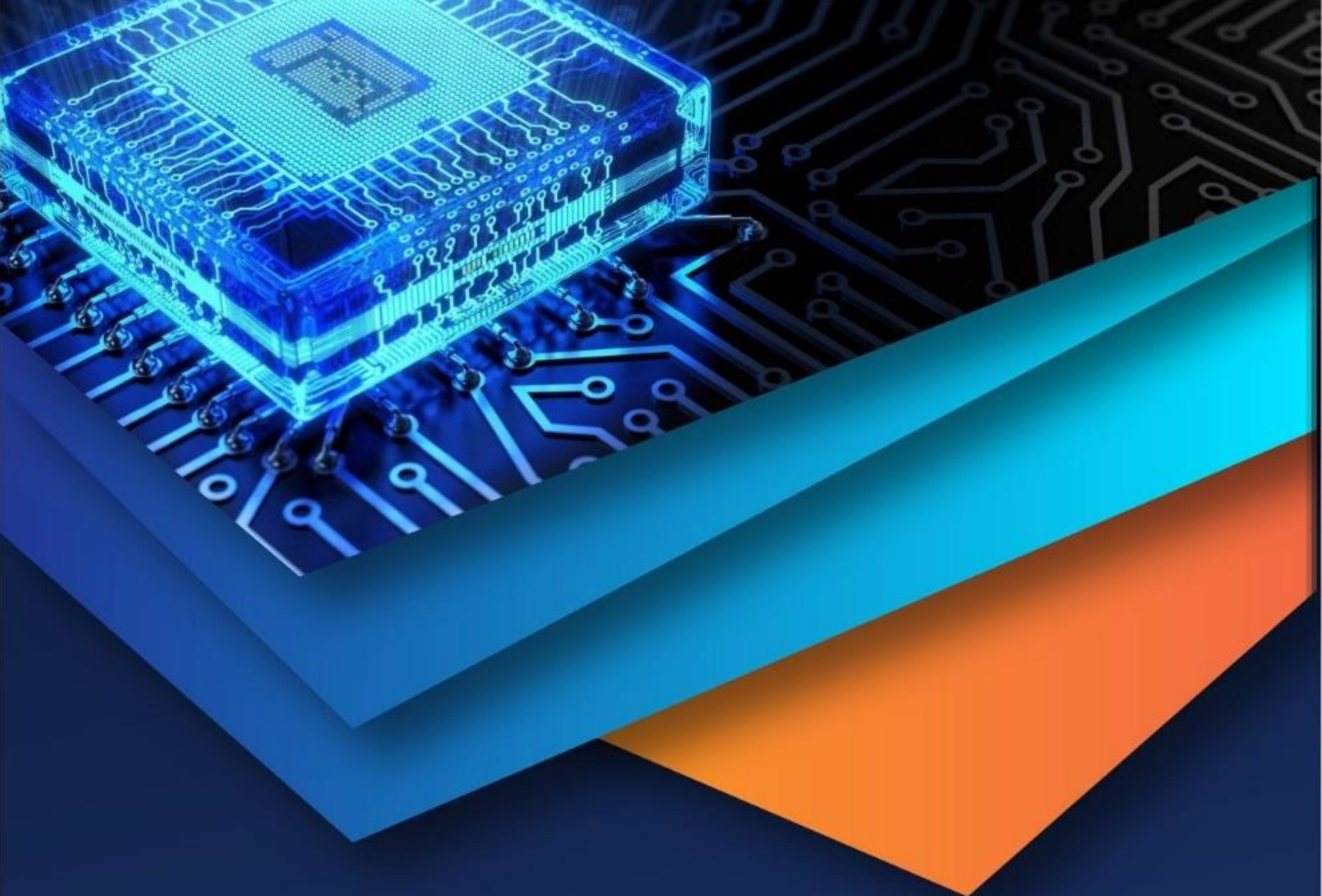
The application is an online application, that has customer churn data analysis.

V. CONCLUSION

Collect data from telecom to conduct customer churn analysis here we will use data analytics tools to extractinsights from the data. Create visual representations of the data through graphs and charts. Develop strategies to address the underlying issues and reduce churn. Present the findings and recommendations through a website.

REFERENCES

[1] https://www.sciencedirect.com/science/article/pii/S24054526 18304472
[2] https://towardsdatascience.com/customer-churn-analysis-predicting-customer-retention-and-loyalty-9bc9e1f0066d
[3] https://machinelearningmastery.com/reduce-customer-churn-with-data-analysis/
[4] https://www.analyticsvidhya.com/blog/2020/10/customer-churn-analysis-predictive-modeling-python/
[5] https://towardsdatascience.com/customer-churn-prediction-using-machine-learning-techniques-5c5e5c56c1c1



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY