JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CUSTOMER CHURN PREDICTION FOR TELECOM INDUSTRY

[1]Adesh Shelar,[2]Pratik Sasane,[3]Manish Vishe and [4]Prajkta Khaire

[1,2,,3]Scholar, [4]Professsor
Department OF Informnation Technology,
SSJCOE Dombivli, India

Abstract : This study focuses on predicting customer churn in the telecommunication industry using machine learning techniques. The research explores the use of Decision Tree, Random Forest, and XGBoost algorithms for churn prediction. Emphasis is placed on data preprocessing, feature selection, and model generation through EDA and PCA. The importance of analyzing and selecting relevant features for accurate predictions is highlighted. The study references other research on churn prediction in the telecom sector, showcasing the significance of machine learning in customer retention strategies.

Keywords: Customer Prediction, Employee Prediction, Descision Tree, Random Forest Algorithm

I. INTRODUCTION

Customer churn, the phenomenon where customers discontinue a product or service, is a significant concern for telecommunication companies. Predicting and preventing customer churn is crucial as acquiring new customers is more costly than retaining existing ones. In this context, machine learning techniques and algorithms play a pivotal role in today's commercial landscape. This project focuses on utilizing various machine learning techniques to predict customer churn in the telecom sector. Classification models such as Logistic Regression, Random Forest, and lazy learning are employed to analyze customer behavior and predict potential churners. By comparing the performance of these models, telecom companies can enhance their customer retention strategies effectively. Keywords: churn, machine learning, Logistic regression, Random Forest, K-nearest-neighbors. The study aims to address the importance of churn prediction in the telecom industry and highlight the role of machine learning in analyzing customer behavior to predict and prevent churn effectively.

I. LITERATURE SURVEY

Irfan Ullah et al., [6] identified churn factors that are essential in determining the root causes of churn. By knowing the significant churn factors from customers' data, Customer Relationship Management (CRM) can improve productivity, recommend relevant promotions to the group of likely churn customers based on similar behavior patterns, and excessively improve marketing.

Kavitha V et al., [7] used a Decision Tree, Random Forest, and XGBoost to predict the customers who are likely to cancel the subscription which can offer them better services and reduce the churn rate.By preprocessing and feature selection ,the data set for training and testing. For the above mentioned algorithm, it is necessary to do some feature engineering to have more efficient and accurate results.

Krishna Sai and Sasikala [8] implemented an EDA using Visualization, statistical tests for feature selection and Data mining methods for predicting the likely churners by utilizing a Logistic Regression Model.Here dataset has been analysed by using the data visualization techniques before entering into the modeling process. Kiran and Surbhi :- Conducted Customer Churn Analysis in Telecom at an International Conference for Reliability in India, 2015.

Trupti S. Gaikwad :- Explored the amalgamation of Mathematics, Statistics, and Electronics in Machine Learning in the International Research Journal on Advanced Science Hub, 2020.

Krishna B.N and Sasikala :- Explored Predictive Analysis and Modeling of Customer Churn in Telecom using Machine Learning Techniques at an International Conference on Trends in Electronics and Informatics in India, 2019.

Rahul J and Usharani T :-Investigated Churn Prediction in Telecommunication Using Data Mining Technology in the International Journal of Advanced Computer Science and Applications, 2013.

II. SYSTEM DESIGN

It is very crucial to make the data useful because unwanted or null values can cause unsatisfactory results or may lead to producing less accurate results. In the data set, there are a lot of incorrect values and missing values. We analyzed the whole dataset and listed out only the useful features. The listing of features can result in better accuracy.

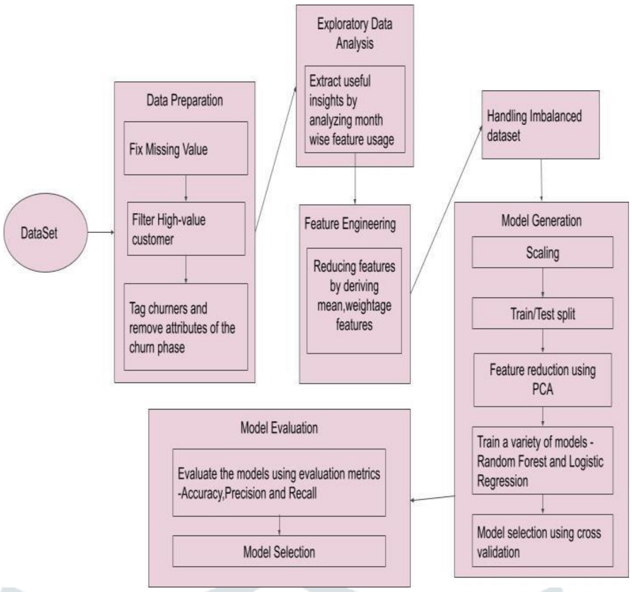


Fig 2.1 Data Flow assistant

Feature selection is a crucial step for selecting the required elements from the data set based on the knowledge. The dataset used here consists of many features out of which we chose the needed features, which enable us to improve performance measurement and are useful for decision-making purposes while remaining will have less importance. The performance of classification increases if the dataset is having only valuable variables and which are highly predictable. Thus having only significant features and reducing the number of irrelevant attributes increases the performance of classification. Many techniques have been proposed for customer churn prediction in the telecommunication industry. Here by using logistic regression, Random Forest and KNN we can predict the probability of a churn i.e., the likelihood of a customer to cancel the subscription and we can evaluate the models using performance metrics like accuracy , precision and recall score.

III. PROPOSED SYSTEM

The proposed system aims to predict customer churn in the telecommunication industry by leveraging machine learning algorithms such as Decision Tree, Random Forest, and XGBoost. Through data preprocessing, feature selection, and exploratory data analysis (EDA), the system identifies key features for accurate churn prediction. Additionally, Principal Component Analysis (PCA) is applied for dimensionality reduction and model building. The system focuses on balancing class distribution, enhancing model efficiency, and improving accuracy in predicting customer churn. By incorporating advanced machine learning techniques, the proposed system aims to provide valuable insights for customer retention strategies in the telecom sector.

VI. METHODOLOGY

6.1 Data Collection

Gather relevant data on customer behavior, usage patterns, and churn indicators in the telecom sector.Discuss the process of collecting historical customer data, including interactions, transactions, and demographics.

Determine the sources of data that are relevant to predicting customer churn. These may include:
Transactional data: Records of customer purchases, subscriptions, or usage of services.
Interaction data: Logs of customer interactions such as website visits, app usage, customer support calls, and emails.
Demographic data: Information about customer demographics such as age, gender, location, income, etc.
Feedback data: Surveys, reviews, or feedback provided by customers.

6.2 Data Preprocessing

Data preprocessing is a crucial step in the data analysis pipeline, particularly for tasks like customer churn prediction. The process involves several key steps to ensure that the data is clean, consistent, and suitable for machine learning algorithms. Initially, missing values are identified and handled using techniques such as imputation, deletion, or prediction. Categorical variables are then encoded into numerical representations, either through one-hot encoding, label encoding, or target encoding, depending on the nature of the data. Numerical features are scaled to a similar range to prevent features with larger magnitudes from dominating the model, employing techniques like standardization, min-max scaling, or robust scaling.

6.3 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a fundamental stage in understanding the data landscape and uncovering insights pertinent to predicting customer churn. By delving into the dataset, one can discern valuable patterns and relationships that inform subsequent modeling efforts. Initially, loading and structurally examining the dataset elucidates its dimensions and the composition of features. Descriptive statistics, including summary metrics like mean, median, and standard deviation, provide a foundational understanding of numerical attributes associated with customer behavior. Visualization techniques, such as histograms and scatter plots, facilitate the exploration of feature distributions and relationships, particularly between predictors and the churn outcome. Feature analysis is pivotal, as it discerns potential predictors by comparing feature distributions across churn and non-churn groups, highlighting variables with discriminatory power. Temporal analysis, if applicable, offers insights into how customer behavior evolves over time, potentially indicating precursors to churn. Correlation analysis further elucidates the relationship between features and churn status, aiding in feature selection.

6.4 Model Building

In constructing a churn prediction model, a systematic approach is paramount to ensure the model's effectiveness in identifying potential churners accurately. Commencing with data preprocessing, the dataset undergoes rigorous cleaning, encoding categorical variables, and scaling numerical features, ensuring compatibility with machine learning algorithms. Feature engineering follows suit, where novel features are crafted or existing ones transformed to encapsulate pertinent information conducive to churn prediction. Subsequently, model selection entails the careful evaluation and comparison of diverse algorithms, ranging from traditional logistic regression to more sophisticated techniques like random forests or gradient boosting machines, with consideration given to model interpretability and performance metrics. Upon selecting the optimal model, rigorous training ensues using the preprocessed data, accompanied by hyperparameter tuning to maximize predictive accuracy.

In churn prediction models, various machine learning algorithms can be employed to effectively identify customers at risk of churning. Some commonly used algorithms include:

1. **Decision Tree Classifier :-** Decision Tree Classifier is a popular algorithm used for classification tasks, including churn prediction. It operates by recursively partitioning the feature space into segments based on feature values, creating a tree-like structure of decision rules.

2. **Random Forest Classifier :-** RandomForestClassifier is an ensemble learning method based on the construction of multiple decision trees during training and combining their predictions through voting or averaging for classification tasks like churn prediction.

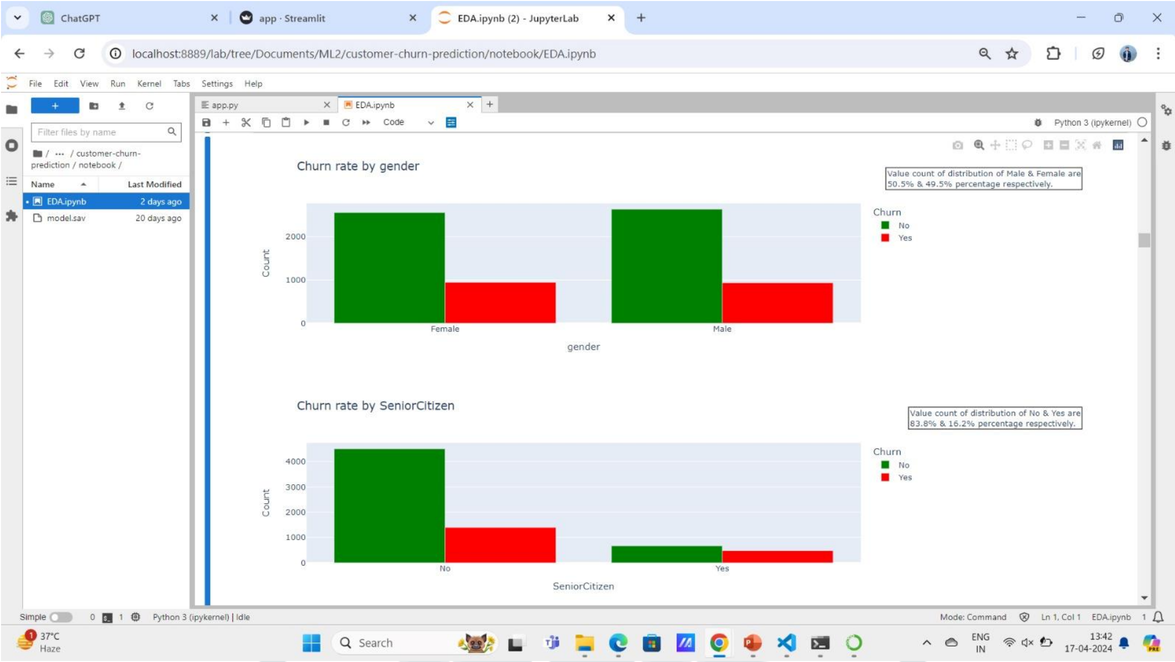
3. **Logistic Regression :-** Logistic Regression is a widely used algorithm for binary classification tasks, including customer churn prediction. Despite its name, logistic regression is a linear model that predicts the probability of a binary outcome (e.g., churn vs. non-churn) based on one or more predictor variables

6.5 Model Evaluation

Model evaluation in churn prediction involves assessing the performance of the predictive models in distinguishing between churners and non-churners. This process is critical for determining the effectiveness and reliability of the models in real-world scenarios. Various evaluation metrics and techniques are employed to gauge the predictive accuracy and generalization capability of the models. Typically, the dataset is divided into training and testing subsets, with the models trained on the training data and evaluated on the testing data.

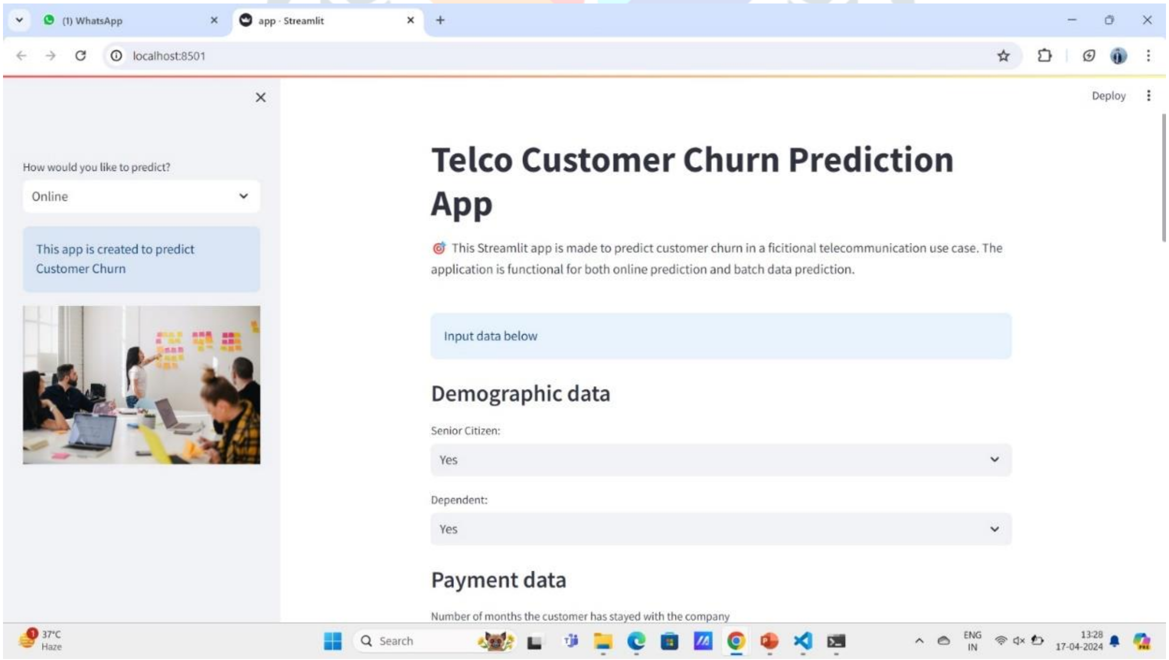
VII. RESULT

The exploratory data analysis (EDA) conducted for the project yielded insightful findings that lay the groundwork for further analysis and model development. Through meticulous examination of the dataset, several key patterns and trends emerged, shedding light on the characteristics and dynamics of customer behavior related to churn. Descriptive statistics provided a comprehensive overview of the dataset, highlighting central tendencies, variability, and distributions of key features. Visualization techniques such as histograms, box plots, and scatter plots allowed for the exploration of relationships between variables, unveiling potential correlations and outliers.



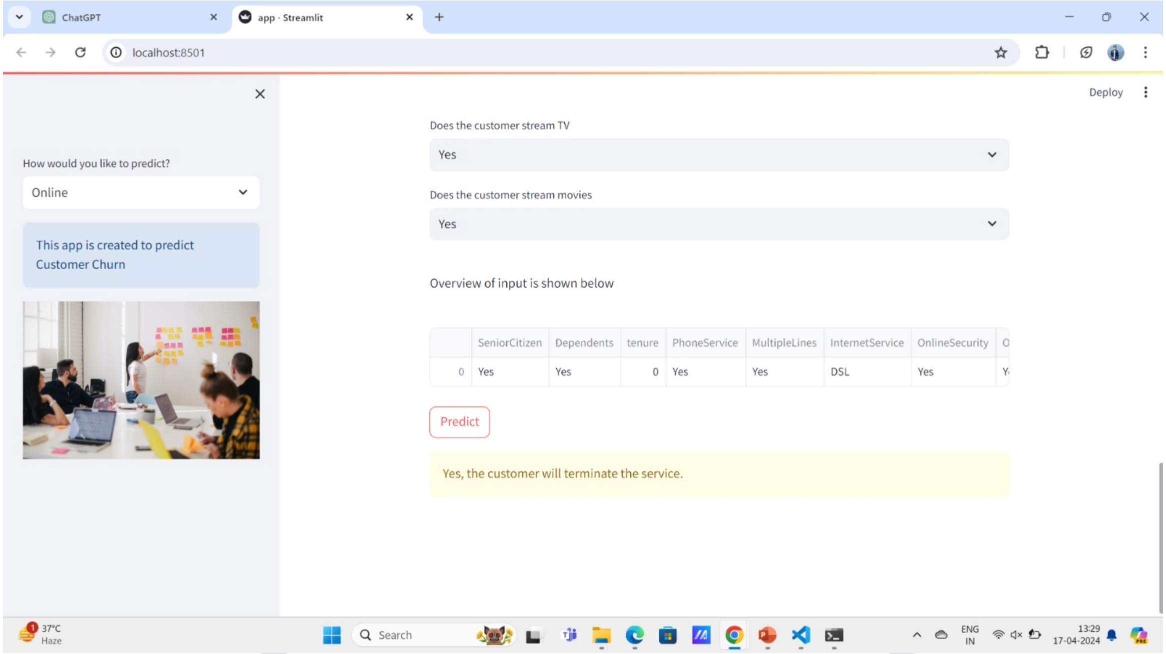
7.1 Exploratory Data Analysis

The results of churn prediction provide valuable insights into the effectiveness of the predictive models and their practical implications for customer retention strategies. After training and evaluating the churn prediction models, organizations gain a comprehensive understanding of their ability to identify customers at risk of churn.For instance, a high accuracy indicates that the model correctly predicts churn and non-churn instances, while precision reflects the proportion of predicted churners who actually churned, offering insights into the model's reliability in flagging potential churners.



7.2 Web Page

The output typically consists of individual churn predictions for each customer, along with associated probabilities or scores indicating the likelihood of churn. The output of a churn prediction model encapsulates the culmination of rigorous data analysis, model training, and evaluation processes. It provides actionable insights and predictions about which customers are likely to churn, empowering organizations to proactively engage with at-risk customers and implement targeted retention strategies.



7.3 Prediction result

VIII. CONCLUSION

Predicting customer churn is crucial for telecom companies to retain customers and reduce subscription cancellations. Machine learning techniques play a vital role in analyzing customer behavior and predicting churn accurately. The proposed churn model demonstrated better results and performance using machine learning algorithms. Continuous refinement of features and introduction of new models can further enhance accuracy and predictive capabilities. Data preprocessing, feature engineering, and exploratory data analysis are essential for improving model accuracy. Selecting valuable features and addressing incorrect or missing values contribute to better predictive outcomes. Future enhancements include exploring advanced machine learning techniques, real-time prediction systems, and personalized retention strategies.

IX.FUTURE SCOPE

Advanced Machine Learning Techniques :- Explore deep learning models like neural networks for more accurate churn prediction. Implement ensemble methods to combine multiple models for improved performance.
Real-Time Churn Prediction :- Develop real-time prediction systems to identify potential churners promptly. Implement automated alerts and interventions to retain at-risk customers.
Customer Segmentation :- Utilize clustering algorithms to segment customers based on behavior and preferences. Tailor retention strategies for different customer segments to improve effectiveness.
Continuous Learning and Adaptation :- Stay updated on emerging technologies and methodologies in churn prediction. Continuously learn from model outcomes and adapt strategies to improve customer retention.

X. ACKNOWLEDGMENT

We had a great experience working on this project and we got to learn a plethora of new skills through this project. However, it would not have been possible without the kind support and help of many individuals. We would like to extend our sincere thanks to all of them. We are highly indebted to the teachers and especially Prof. Prajkta Khaire for their guidance and constant supervision as well as providing necessary information regarding the project and also for their support in completing the project.

XI. REFERENCES

1. Abhishek and Ratnesh ,“Predicting Customer Churn Prediction in Telecom Sector Using Various Machine Learning Techniques”, In the proceedings of 2017 International Conference on Advanced Computation and Telecommunication, Bhopal, India, 2017.

2. Abinash and Srinivasulu U ,“Machine Learning techniques applied to prepaid subscribers: case study on the telecom industry of Morocco”, In the proceedings of 2017 International Conference on Inventive Computing and Informatics , Coimbatore, India, pp. 721-725, 2017.

3. Trupti S. Gaikwad; Snehal A. Jadhav; Ruta R. Vaidya; Snehal H. Kulkarni. "Machine learning amalgamation of Mathematics, Statistics and Electronics". International Research Journal on Advanced Science Hub, 2, 7, 2020, 100-108. doi: 10.47392/irjash.2020.72

4. Alae and El Hassane , “A Comparative Study of Customer Churn Prediction in Telecom Industry Using Ensemble Based Classifiers”, In the proceedings of Intelligent Systems and Computer Vision , Fez, Morocco, 2017.

5. Salini Suresh; Suneetha V; Niharika Sinha; Sabyasachi Prusty; Sriranga H.A. "Machine Learning: An Intuitive Approach In Healthcare". International Research Journal on Advanced Science Hub, 2, 7, 2020, 67-74. doi: 10.47392/irjash.2020.67

6. Anuj and Prabin ,“A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services”, International Journal of Computer Applications ,Volume 27, No.11, pp. 26-31, 2011.

7. Balasubramanian M, and Selvarani M ,“Churn Prediction in Mobile Telecom System using Data Mining Techniques ”, International Journal of Scientific and Research Publications, Volume 4, Issue 4, pp. 1-5, 2014.