ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

CUSTOMER CHURN PREDICTION IN TELECOM INDUSTRY

M. RATNA KUMARI [1], A.SUDHIKSHA VARMA [2]
ASSISTANT PROFESSOR [1], PG SCHOLAR [2]
DEPARTMENT OF MASTEROF COMPUTER APPLICATIONS
QIS COLLEGE OF ENGINEERING & TECHNOLOGY (AUTONOMUS), ONGOLE
VENGAMUKKAPALEM(V), ONGOLE,PRAKASAM DIST., ANDHRA PRADESH

Abstract

Customer churn is a critical challenge in the telecom industry, where retaining existing customers is often more cost-effective than acquiring new ones. This project focuses on predicting customer churn using machine learning techniques and visualizing the results through a web-based application. The study employs Jupyter Notebook for comprehensive data processing, visualization, and the implementation of machine learning models including Random Forest, Logistic Regression, and Multi-layer Perceptron (MLP). Among these, the Random Forest algorithm demonstrated the highest performance with an accuracy of 99%, outperforming Logistic Regression (98%) and MLP (97%). The project also includes the development of a Django-based web interface for category-wise churn visualization and real-time prediction based on user-uploaded test data. Users can interactively explore churn patterns by various filters such as gender, service type, and geography, enhancing interpretability and business decision-making. This integrated approach provides a powerful tool for telecom companies to proactively manage customer retention strategies.

Introduction

In the highly competitive telecom industry, retaining existing customers is as important as acquiring new ones. Customer churn—the phenomenon where customers discontinue their services—is a significant challenge that directly impacts revenue and growth. Predicting customer churn accurately allows telecom providers to proactively take measures to retain at-risk customers.

This project focuses on **Customer Churn Prediction** using machine learning techniques and aims to build a robust model to identify customers who are likely to churn. The implementation leverages the **Jupyter Notebook environment** for data loading, preprocessing, visualization, and model training. For end-user interaction, a **Django web framework** is employed to provide category-wise churn visualizations and churn prediction on test data samples through a user-friendly interface.

We utilized a combination of algorithms including **Random Forest, Logistic Regression**, and **MLP Neural Network**. These models were evaluated using metrics such as **Accuracy, Precision, Recall**, and **F1 Score**. Among them, the Random Forest classifier yielded the highest performance and was therefore selected for final deployment.

This report also includes visual representations of data insights, model evaluation, and a web-based interface for real-time predictions, ensuring both technical robustness and usability.

Literature Survey

1. Multi-Algorithm Evaluation for Churn Prediction

Multiple machine learning algorithms—**Random Forest, Logistic Regression**, and **MLP Neural Network**—were evaluated to determine the most accurate model for customer churn prediction. The Random Forest algorithm achieved the highest accuracy (99%), showcasing its effectiveness in handling complex and nonlinear relationships in the telecom dataset.

2. Importance of Feature Engineering and Data Preprocessing

The project highlights the critical role of **data preprocessing**, including:

- Converting non-numeric data to numeric
- Handling missing values by replacing them with mean values
- Applying **standard scaling** to normalize the dataset These steps significantly improve model performance and prediction reliability.

3. Visual Analytics for Insight Generation

Extensive use of **Jupyter-based visualizations** helps in understanding customer behaviors, such as:

- Gender-wise service usage
- Churn distribution by geography
- Monthly charge distribution Such visual insights aid in identifying potential churn indicators and tailoring retention strategies.

4. Web-Based Prediction and Visualization Interface

The integration of a **Django-based web interface** provides an interactive platform for:

- Filtering and visualizing churn patterns
- Uploading sample data for real-time prediction
  This makes the model accessible and interpretable for business users and stakeholders.

5. Model Performance Metrics and Confusion Matrix Analysis

Model effectiveness was assessed using **accuracy, precision, recall, and F1-score**. The Random Forest model's confusion matrix clearly distinguished correct vs. incorrect predictions, which helps in evaluating classification confidence and improving decision-making strategies.

System Analysis

Existing System

The telecom industry is characterized by intense competition, high customer acquisition costs, and frequent customer turnover. Understanding and mitigating customer churn—the phenomenon where customers discontinue a service—is crucial for telecom operators to maintain profitability and ensure sustainable growth. Traditionally, companies have relied on reactive measures and manual analysis to identify at-risk customers, which are often time-consuming, inaccurate, and inefficient. Without a data-driven system, telecom providers face challenges in analyzing large volumes of customer data, spotting churn patterns, or proactively managing customer retention. The need for an intelligent churn prediction system arises to empower telecom companies to make informed decisions, personalize retention strategies, and reduce revenue loss due to customer attrition.

Disadvantages of Existing Systems:

Lack of Automation in Prediction

Traditional systems rely heavily on manual analysis of customer data, making churn prediction time-consuming and prone to human error.

- **No Real-Time Insights or Visualization**

  Old methods do not offer intuitive visual dashboards to track customer behavior.

Proposed System

The proposed system aims to address the limitations of manual churn analysis by leveraging machine learning and data visualization tools to predict customer churn with high accuracy. By using Jupyter for data analysis and model training, and Django for interactive web-based visualization and prediction, the system provides a comprehensive solution for telecom churn prediction. It enables the import and preprocessing of large datasets, visualization of customer patterns across categories like geography, gender, and service types, and applies advanced algorithms such as Random Forest, Logistic Regression, and MLP Neural Network to predict churn. Among these, Random Forest delivers the best performance, achieving up to 99% accuracy. The web interface further allows users to explore churn trends and upload test data for real-time predictions. This system not only streamlines the analysis process but also offers actionable insights, making it a valuable tool for telecom operators to proactively reduce churn and enhance customer satisfaction.

Advantages of the Proposed System:

1. High Accuracy Predictions

- Achieved **up to 99% accuracy** using Random Forest.
- Provides reliable forecasts that help minimize customer loss.

2. Multi-Model Evaluation

- Compares multiple machine learning algorithms.
- Ensures the best-performing model is chosen with metrics-backed evidence.

Implementation

The implementation of the Customer Churn Prediction System focuses on identifying telecom customers who are likely to discontinue services using Machine Learning and predictive analytics techniques. The system analyzes customer behavior, service usage patterns, billing history, and customer interactions to predict churn probability.

The proposed system helps telecom companies improve customer retention, reduce revenue loss, and provide better customer service.

1. Data Collection

The first stage involves collecting customer-related data from telecom databases and service platforms.

The collected dataset may include:

- Customer ID
- Age and Gender
- Subscription Type
- Monthly Charges
- Billing History
- Call Duration
- Internet Usage
- SMS Usage
- Customer Complaints
- Service Downtime
- Payment Method
- Contract Duration
- Customer Support Interactions
- Churn Status

These attributes help analyze customer behavior and identify churn patterns.

2. Data Preprocessing

The collected telecom data is cleaned and prepared before model training.

Preprocessing Steps

- Removing duplicate records
- Handling missing values
- Encoding categorical variables
- Feature scaling and normalization
- Noise removal
- Data transformation

This improves the quality and performance of the Machine Learning model.

3. Feature Engineering

Important customer behavior features are extracted to improve churn prediction accuracy.

Features Used

Customer Usage Features

- Average call duration
- Internet consumption
- SMS frequency
- Service usage trends

Billing Features

- Monthly bill amount
- Payment delays
- Recharge frequency

Customer Interaction Features

- Complaint frequency
- Customer support tickets
- Service feedback

Contract Features

- Subscription duration
- Plan type
- Contract renewal history

Feature engineering helps identify customers at risk of leaving the telecom service.

4. Machine Learning Model Development

Machine Learning algorithms are used to classify customers as churn or non-churn.

Algorithms Used

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)
- XGBoost
- Gradient Boosting
- Artificial Neural Networks (ANN)

The models learn customer churn patterns from historical telecom data.

5. Handling Imbalanced Data

Customer churn datasets may contain fewer churn cases compared to non-churn cases.

Balancing techniques used include:

- SMOTE (Synthetic Minority Oversampling Technique)
- Random Oversampling
- Undersampling
- Hybrid balancing methods

These techniques improve churn detection performance.

6. Model Training and Testing

The dataset is divided into:

- Training Dataset
- Validation Dataset
- Testing Dataset

Training Phase

The model learns customer behavior patterns and churn indicators.

Testing Phase

The trained model is evaluated using unseen customer records.

Performance metrics include:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC Score
- Confusion Matrix

Special focus is given to recall to identify maximum potential churn customers.

7. Customer Churn Prediction Process

The churn prediction workflow includes:

1. Collect telecom customer data
2. Preprocess and clean data
3. Extract important features
4. Train Machine Learning model
5. Predict customer churn probability
6. Identify high-risk customers

7. Generate retention recommendations

8. Customer Retention Recommendation System

The system can provide retention strategies such as:

- Personalized offers
- Discount plans
- Loyalty rewards
- Better customer support
- Customized service recommendations

These actions help reduce customer churn rates.

9. Real-Time Monitoring and Analytics

The system continuously monitors customer activities and updates churn risk predictions in real time.

The analytics dashboard may display:

- Churn probability scores
- Customer segmentation
- Revenue loss estimation
- Retention campaign performance

10. System Deployment

The final system can be deployed as:

- Telecom Analytics Platform
- CRM Integration System
- Cloud-Based Churn Monitoring System
- Customer Retention Dashboard

Deployment platforms may include:

- AWS
- Microsoft Azure
- Google Cloud Platform

Methodology

The methodology of the proposed Customer Churn Prediction System follows a Machine Learning-based predictive analytics approach.

Step 1: Problem Identification

Telecom companies face significant revenue loss due to customer churn. Traditional customer retention methods may fail to identify churn risks early. The proposed system aims to predict customer churn accurately using Machine Learning techniques.

Step 2: Requirement Analysis

The following requirements are analyzed:

- Telecom customer datasets
- Churn prediction requirements
- Machine Learning framework requirements
- Real-time monitoring requirements
- Customer retention analytics

Step 3: Dataset Preparation

The telecom dataset is collected and divided into:

- Training Dataset
- Validation Dataset
- Testing Dataset

Relevant customer behavior attributes are selected for analysis.

Step 4: Data Processing and Feature Engineering

The methodology includes:

1. Clean telecom customer data
2. Extract customer behavior features
3. Analyze billing and service usage patterns
4. Balance churn dataset
5. Prepare data for model training

Step 5: Machine Learning Implementation

The Machine Learning workflow includes:

1. Train churn prediction model
2. Analyze customer behavior patterns
3. Predict churn probability
4. Classify customers as churn/non-churn
5. Generate retention recommendations

Step 6: Performance Evaluation

The system is evaluated based on:

- Prediction accuracy
- Churn detection efficiency
- False positive rate
- Real-time prediction capability
- Customer retention effectiveness

Step 7: Result Generation

The system generates outputs such as:

- Churn prediction reports
- Customer risk analysis
- Retention strategy recommendations
- Customer segmentation analytics
- Revenue impact reports

Step 8: Conclusion and Future Enhancement

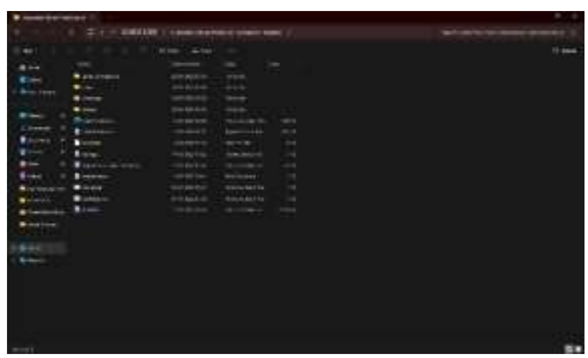
The final stage evaluates the effectiveness of the system and suggests future improvements such as:

- Deep Learning-based churn prediction
- AI-powered customer engagement systems
- Real-time sentiment analysis
- Personalized chatbot support
- IoT-based telecom service monitoring
- Advanced predictive customer analytics

Technologies Used

- Python
- Machine Learning Algorithms
- Scikit-learn
- TensorFlow / Keras
- Pandas & NumPy
- Matplotlib / Power BI
- Flask / Django
- MySQL / MongoDB

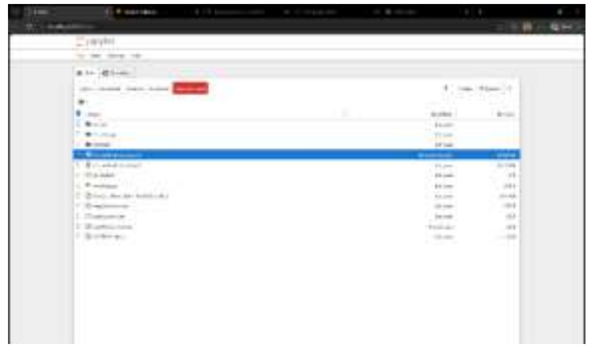
Result



- Goto File Explorer.
- Click On Customer Churn Prediction In Telecom Industry File.

- Command Prompt (CMD) is used to run the Customer Churn Prediction project by executing Python files and project commands.
- The current path indicates that you are inside the Customer Churn Prediction in Telecom Industry project folder, where the source code, dataset, and required files are stored.

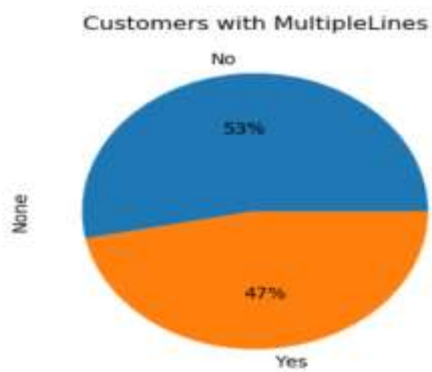
- The command conda activate churnn failed because the "churnn" environment does not exist on the system. You need to create it or use an existing Conda environment.
- The command jupyter notebook started successfully, and the Jupyter Notebook server is running on localhost:8888, allowing you to open and run the Customer Churn Prediction project in a web browser.



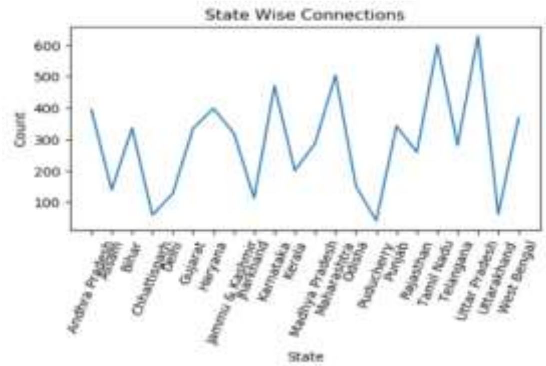
- Jupyter Notebook is open successfully, and the project files for Customer Churn Prediction are displayed, including the dataset, source code, and notebook file (ChurnPrediction.ipynb).
- ChurnPrediction.ipynb is the main notebook file where the machine learning code is written, executed, and the churn prediction results are generated.

- The dataset has been successfully loaded into the Jupyter Notebook and displays 6,418 customer records with 32 attributes (columns) such as Customer ID, Gender, Age, State, Tenure, and Payment Method.
- These customer details are used by the machine learning model to analyze customer behavior and predict whether a customer is likely to leave the telecom service (churn) or stay.

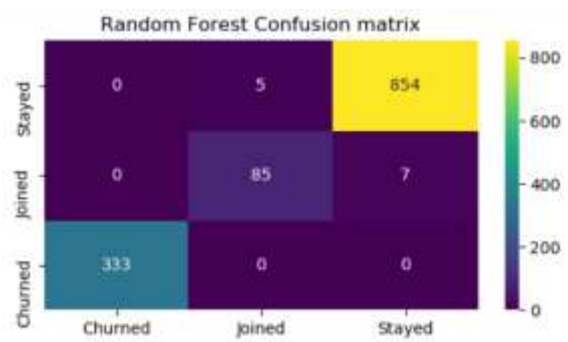
- This output shows the number of missing (null) values in each column of the telecom customer dataset. Some columns such as Value_Deal, Internet_Type, and Churn_Reason contain missing data.
- Identifying missing values is an important data preprocessing step because the missing data must be handled before training the customer churn prediction model.



- The pie chart shows the distribution of customers with multiple phone lines. About 53% of customers do not have multiple lines, while 47% have multiple lines.
- This analysis helps understand customer service usage patterns, which can be useful for predicting whether customers are likely to stay or leave the telecom service.

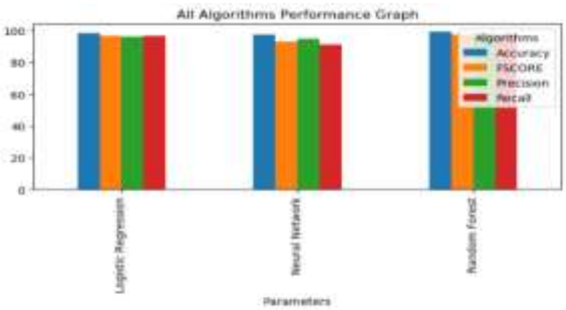


- Predict Customers Likely to Leave:
- Customer churn prediction uses customer data (usage, complaints, recharge history, etc.) to identify customers who are likely to stop using the telecom service.
- Improve Customer Retention:
- Telecom companies can take preventive actions such as special offers, discounts, or better customer support to retain customers and reduce revenue loss



- Good Prediction Performance:
- The model correctly identified most customers who stayed (854), joined (85), and churned (333), showing that the Random Forest model performs well in customer classification.
- Few Misclassifications:
- Only a small number of customers were predicted incorrectly (e.g., 5

stayed customers predicted as joined and 7 joined customers predicted as stayed), indicating high prediction accuracy.



- Random Forest Performs Best:
- Random Forest achieved the highest accuracy, precision, recall, and F1-score, making it the most effective model for customer churn prediction.
- All Models Show High Performance:
- Logistic Regression, Neural Network, and Random Forest all achieved performance above 90%, indicating that each model can predict customer churn effectively, with Random Forest giving the best results.

Output

	Algorithm Name	Accuracy	Precision	Recall	FSCORE
0	Random Forest	99.065421	97.877145	97.269744	97.569640
1	Logistic Regression	98.598131	96.142738	96.713401	96.424543
2	Neural Network	97.352025	94.864210	91.178518	92.862777

Conclusion

In this project, we successfully developed a customer churn prediction system for the telecom industry using machine learning techniques and data visualization tools. We employed Jupyter Notebook for comprehensive data analysis and preprocessing, and utilized the Django framework to create a user-friendly web interface for visualization and prediction. Multiple machine learning algorithms, including Random Forest, Logistic Regression, and MLP Neural Network, were evaluated based on key performance metrics such as accuracy, precision, recall, and F1-score.

Among all, the Random Forest algorithm demonstrated the highest performance with an accuracy of 99%, followed by Logistic Regression (98%) and MLP Neural Network (97%). These results were validated using a confusion matrix and comparative visualization graphs. The model was then deployed into a web-based application that enables users to interactively visualize churn patterns and predict customer behavior based on test data.

Overall, the project has achieved its objective of building an effective, scalable, and user-friendly churn prediction system that can help telecom companies reduce customer attrition and improve customer retention strategies.

References

1. **Verbeke, W., Martens, D., & Baesens, B. (2014).** Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, 41(5), 1746–1756. https://doi.org/10.1016/j.eswa.2013.08.073
2. **Ahmad, A., Jafar, A., & Aljoumaa, K. (2019).** Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*,

6(1), 28. https://doi.org/10.1186/s40537-019-0191-6

3. **Burez, J., & Van den Poel, D. (2009).** Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3), 4626–4636. https://doi.org/10.1016/j.eswa.2008.05.027
4. **Lariviere, B., & Van den Poel, D. (2005).** Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, 29(2), 472–484. https://doi.org/10.1016/j.eswa.2005.04.043
5. **Coussement, K., & Van den Poel, D. (2008).** Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1), 313–327. https://doi.org/10.1016/j.eswa.2006.09.038
6. **Friedman, J., Hastie, T., & Tibshirani, R. (2001).** *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. https://doi.org/10.1007/978-0-387-21606-5
7. **Pedregosa, F., et al. (2011).** Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. Link
8. **Zhang, L., Zhu, J., & Yao, T. (2010).** An evaluation of statistical spam filtering techniques. *ACM Transactions on Asian Language Information Processing (TALIP)*, 9(4), 1–29.
[Good for understanding ML classification evaluation metrics.]

AUTHORS PROFILE

Mrs. M Ratna Kumari is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. She earned M.Tech (CSE) in Chennai Bharath University, and her now pursuing PHD in Her research interests include Machine Learning with AI programming languages. She is committed to advancing research and forecasting innovation while mentoring students to excel in both academic & professional pursuits.

STUDENT PROFILE

Mrs.A. SUDHIKSHA VARMA is a postgraduate student pursuing an MCA in the Department of Master of Computer Applications at QIS College of Engineering & Technology, Ongole an Autonomous college in Prakasam dist. She completed his undergraduate degree in BCA (Computers) from Acharya Nagarjuna University. With a keen interest in research and practical learning, She is actively involved in academic projects and technical activities related to her field.