



Amrutvahini College of Engineering, Sangamner
Department of Computer Engineering
International Conference on Recent Trends and
Advancements in Computing Technologies (ICRTACT – 2024)



Churn Prediction in Telecom Sector using NLP & ML

Paper ID: 256

Track: Computing

**Presented by: Akash Raut, Shivanjali Kadam,
Mohit Sakalkale, Ritesh Sambhus**

Amrutvahini College of Engineering, Sangamner.

Contents

- Introduction
- Motivation
- Contribution
- Literature survey
- Proposed Approach
- Methodology
- Performance metrics
- Validation & Testing
- Performance Result Analysis and Observations
- Conclusions and Future Work
- References

Introduction

- **Churn prediction:** Churn prediction means to predict earlier which customers may leave or switch to another service provider.
- The aim is to address the challenge of customer churn in the telecom sector. Here, we have developed a churn prediction model using ML & NLP to identify customers at risk of churning.
- The goal is to create an effective model that enable telecom companies to proactively retain at-risk customers and improve overall customer satisfaction.

Motivation

- Major revenue generation of telecom companies is based on number of customers. Customer churn can badly affects the revenue.
- We identify the need of strong tool for churn prediction, which will help companies to identify the areas where improvement is required and to improve overall customer satisfaction.

Contribution

- 1. Integration of NLP and ML Techniques:** Our model integrates NLP techniques with ML algorithms to predict telecom customer churn, enhancing prediction accuracy.
- 2. Improved Churn Prediction:** By analyzing structured data and unstructured feedback, our model provides deeper insights into customer behavior, leading to more accurate churn predictions.
- 3. Supporting Telecom Companies:** Our model helps telecom companies by predicting customer churn earlier, making it easy for them to retain customers and develop effective marketing strategies.

Literature Survey

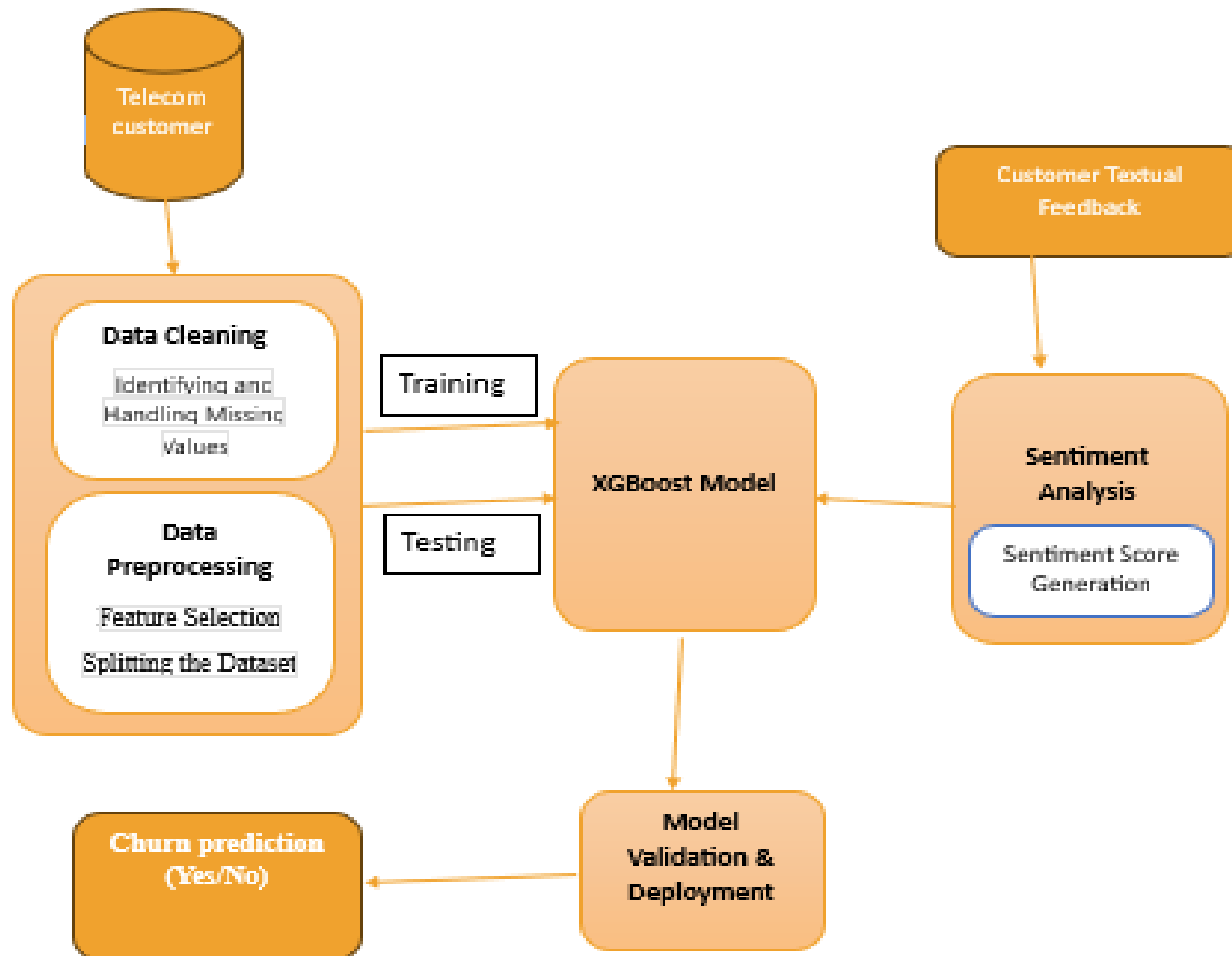
Years	Authors	Title	Methodology	Findings
2023	Chen Zhue	Prediction of Telecom Customer Churn Based on MIPCA-XGBoost Method	MIPCA-XGBoost	The MIPCA-XGBoost method for telecom customer churn prediction achieves high accuracy through mutual information feature selection.
2023	Sarkaft Saleh, Subrata Saha	Customer retention and churn prediction in the telecommunication Industry	Machine learning techniques(ML) And Algorithm	The findings indicate that age is a key predictor of churn, which highlights the need to take into account cultural variation and diversify churn analysis methods in order to improve comprehension and retention initiatives

Years	Authors	Title	Methodology	Findings
2020	pan Tang Wuhan, Hubei	Telecom Customer Churn Prediction Model Combining K-means and XGBoost Algorithm.	XGBoost and K-Means clustering	This method shows better precision and overall generalization ability while keeping high efficiency when compared to other models like decision trees.
2020	Samah Wael Fujo,Suresh Subramanian, Moaiad Ahmad Khder	Customer Churn Prediction in Telecommunication Industry Using Deep Learning.	DeepBP-ANN model ,XGBoost	The model's efficacy is enhanced by balancing approaches like ROS and Activity regularization, and performance is further improved by adjusting parameters like the number of neurons and epochs.

Proposed Approach

1. The data set taken for model training consists of 21 attributes and 7044 records which is preprocessed and Cleansed for further analysis.
2. The Dataset is split into 80% and 20% for training and testing respectively.
3. The proposed model is further trained with the XGBoost algorithm on the dataset to identify complex relationships and patterns within data.
4. The Sentiment Analysis on customer feedback gives an additional information to the ML model to predict whether the customer will churn or not.

Architecture



Methodology

1.Data Exploration and Analysis: Exploratory data analysis (EDA) is performed to understand the distribution and characteristics of the dataset and analysing the correlation between different features and their impact on churn rate.

2.Data Preprocessing and Feature Engineering: Selecting the relevant features and handling missing values, outliers, and inconsistencies in the dataset and encoding categorical variables using techniques like one-hot encoding or label encoding.

3.Model Development: Splitting the dataset into training, validation, and test sets. The XGBoost Model is developed on the training data capturing underlying patterns. The sentiment scores that are generated after sentiment analysis on text data are given as an additional parameter to the built model.

4.Model Evaluation: Evaluating the performance of trained model using relevant metrics such as accuracy, precision, recall, F1-score.

Working of Algorithm

1. The XGBoost (eXtream Gradient Boosting) Algorithm follows the boosting ensemble learning technique.
2. In this, multiple weak learners (decision trees) are sequentially trained to correct the errors of their predecessors.
3. It supports parallelization of tree construction.
4. During training, XGBoost optimizes an objective function using gradient descent by means of adjusting the weights of each tree to minimize the difference between the predicted and actual values.

5. Once all the trees are trained, XGBoost combines their predictions by summing up the weighted predictions of each tree.

6. The final prediction for each data point is the sum of the weighted predictions from all the trees in the ensemble.

Steps in XGBoost Algorithm:

1. Initialize the model.
2. Calculate the residuals.
3. Build decision trees iteratively.
4. Regularization for model.
5. Gradient Optimization.
6. Generating final predictions.

- Objective Function for XGBoost algorithm :

$$\text{Objective} = \sum(\text{Loss}(y_i, \hat{y}_i)) + \sum(\Omega(f_k))$$

Where:

- $\text{Loss}(y_i, \hat{y}_i)$: Loss function measuring the difference between predicted (\hat{y}_i) and actual (y_i) values.
- $\Omega(f_k)$: Regularization word for every tree in the group.

Role of NLP

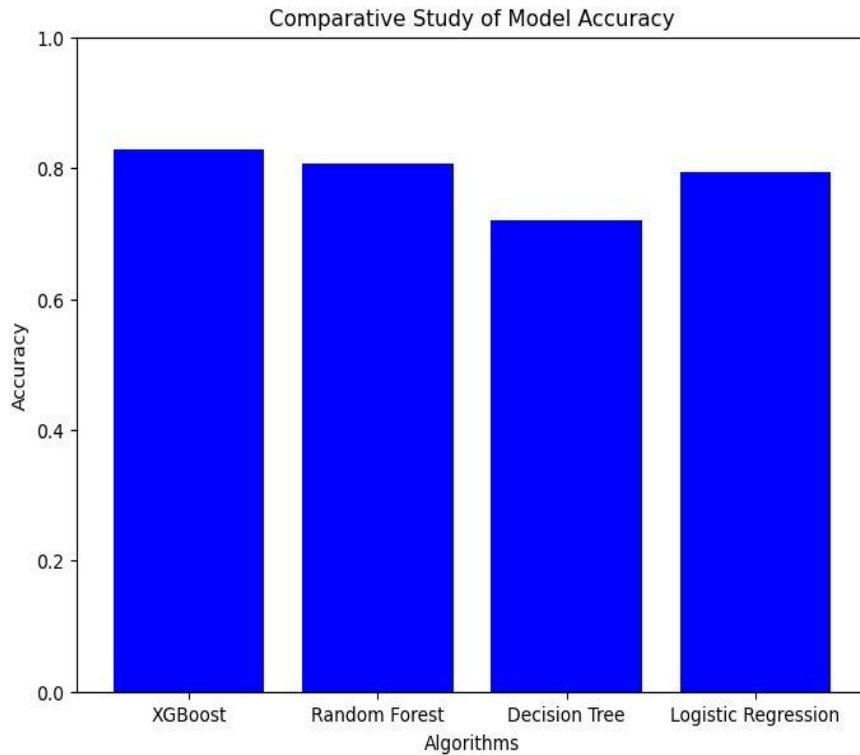
1. Used to determine the sentiment of customer feedback.
2. Sentiment Intensity Analyzer (SIA) is used for sentiment analysis.
3. SIA relies on a pre-built lexicon containing a list of words with pre-assigned sentiment scores.
4. SIA generates a compound score that reflects the general emotions of the piece of text.
5. This score normally varies from -1 (very negative) to +1 (highly positive), with zero representing neutrality.

The Compound Score Generated by SIA is Given as an additional parameter to the built XGBoost model for prediction the churning of customers

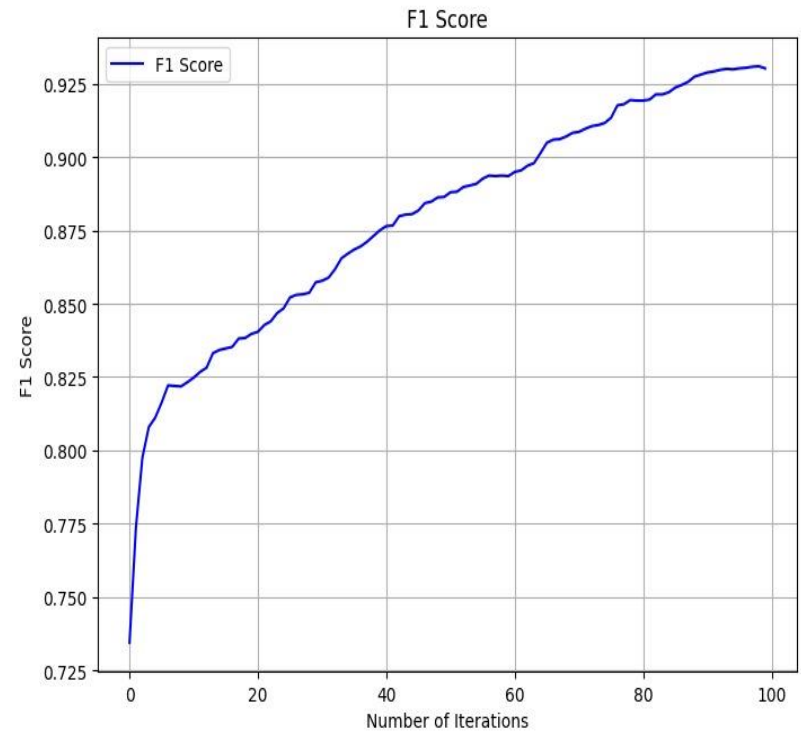
Performance metrics

- Accuracy :- 82.92 %
- Precision :- $TP/(TP + FP) = 0.91$
- Recall :- $TP/(TP + FN) = 0.85$
- F1-score = $2 \times (Precision \times Recall) / (Precision + Recall) = 0.88$

Performance Result Analysis



Accuracy of different models



F1 Curve of XGBoost Model

Validation & Testing

Test Case	Expected Result	Actual Result	Pass / Fail
Verification of registration page with valid username and password	Registration successful	Registration successful	Pass
Verification of login page with valid username and password	Login successful	Login successful	Pass
Verify the system's ability to accept structured data accurately	Accepted Successfully	Accepted Successfully	Pass
Verify whether system receives data in standard format only	Successful	Successful	Pass

Test Case	Expected Result	Actual Result	Pass / Fail
Verify the system's ability to accept Textual data accurately.	Accepted Successfully	Accepted Successfully	Pass
Verify Database Connectivity	Successful Connection	Successful connection	Pass
Verify whether the Result PDF is generate	Successfully Generated	Successfully Generated	Pass
Verify Churn Prediction Button	Predicts result	As expected	Pass
Verify Logout Button	Logout Successful	Logout Successful	Pass

Conclusions

- XGBoost emerges as the superior model for churn prediction, exhibiting higher accuracy compared to other algorithms like Random Forest and Logistic Regression.
- The integration of Natural Language Processing (NLP) provides an additional layer to Machine Learning, enabling the analysis of customer sentiments and feedback. This enriches the predictive capabilities of the model and enhances its accuracy.
- Our study demonstrates the conclusion of combining NLP and advanced ML techniques for proactive churn prediction in the telecom sector. By leveraging both structured data and unstructured feedback, our model offers valuable insights for retaining customers and optimizing business strategies.

References

- [1] Chen Zhue(2023);" Prediction of Telecom Customer Churn Based on MIPCA-XGBoost Method". ISSN: 2832-6024 | Vol. 3, No. 1, 2023.
- [2] Sarkaft Saleh, Subrata Saha(2023);" Customer retention and churn prediction in the telecommunication Industry" 27 February 2023 / Accepted: 16 May 2023.
- [3] pan Tang Wuhan, Hubei (2020);" Telecom Customer Churn Prediction Model Combing K-means and XGBoost Algorithm" Wuhan University of Technology School of Management Wuhan, Hubei, China.
- [4]Mr.Abhinav Sudhir Thorat, Dr. Vijay Ramnath Sonawane(2022);" A Random Forest Churn Prediction Model: An Investigation of Machine Learning Techniques for Churn Prediction and Factor Identification in the Telecommunications Industry" Digital Object Identifier 10.1109/ACCESS.2019.2914999.
- [5]Samah Wael Fujo,Suresh Subramanian, Moaiad Ahmad Khder;" Customer Churn Prediction in Telecommunication Industry Using Deep Learning"Inf.Sci.Lett.11.No.1185 -198.