IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, November 2023

Telecom Churn Prediction Using Machine Learning

Ashwini Atmaram Patil[1] and Nilesh R. Wankhade[2]
Student, Department of Computer Engineering[1]
Head of Department, Department of Computer Engineering[2]
Kalyani Charitable Trust's, Late. G. N. Sapkal College of Engineering, Nashik, Maharashtra, India

Abstract: *Telecom churn prediction is a critical task for telecom companies to retain their customers. Churn refers to the phenomenon where a customer discontinues their subscription or service with a telecom company. Predicting churn helps telecom companies take proactive measures to prevent churn by identifying potential churners and offering them attractive retention strategies. This abstract presents an overview of the telecom churn prediction problem using machine learning techniques. The telecom churn prediction problem involves analyzing historical customer data, including demographic information, usage patterns, billing details, and service history, to predict whether a customer is likely to churn in the future. Machine learning algorithms are used to learn patterns and relationships from this data and make predictions based on new, unseen data. Telecom churn prediction using machine learning involves preprocessing historical customer data, feature engineering, selecting appropriate machine learning algorithms, evaluating model performance using various metrics, and deploying the best-performing model in a production environment. By implementing this process, telecom companies can reduce churn rates and improve customer satisfaction.*

Keywords: Machine Learning, Random Forest ,Decision Tree, XGBoost, Prediction, Churn

I. INTRODUCTION

In today's highly competitive telecom industry, customer retention is a critical challenge for service providers. Churn, the phenomenon where customers discontinue their subscription or service, is a significant concern for telecom companies as it leads to a loss of revenue and market share. To address this challenge, telecom companies are turning to machine learning techniques to predict churn and take proactive measures to prevent it. Churn prediction is crucial for telecom companies as it helps them retain their customers and reduce churn rates. By identifying potential churners and offering them attractive retention strategies, telecom companies can prevent customers from leaving and improve customer satisfaction. Moreover, churn prediction enables telecom companies to optimize their marketing and sales efforts by focusing on high-value customers and avoiding wastage on low-value ones. Telecom churn prediction is a crucial task for telecom companies as it helps them retain their existing customers and reduce customer attrition. Churn refers to the process of customers discontinuing their services with a telecom company. Predicting churn using machine learning algorithms can help telecom companies identify the factors that lead to customer churn and take proactive measures to retain them. Telecom churn prediction is a critical task for telecom companies as it helps them retain their existing customers and reduce customer attrition. Telecom companies can build accurate and effective churn prediction models using machine learning algorithms. The models can be deployed in a production environment to provide real-time predictions and help the company take proactive measures to retain customers.

II. OBJECTIVE & SCOPE OF PROPOSED SYSTEM

1. The objective of this proposed system is to develop a machine learning model that can accurately predict customer churn in the telecom industry.
2. Churn refers to the phenomenon where a customer discontinues their subscription or service with a telecom company. Predicting churn is crucial for telecom companies as it enables them to take proactive measures to retain customers, reduce churn rates, and improve customer satisfaction.

3. To explore the customer churn prediction in telecom using machine learning.
4. To investigate the impact of customer churn in telecom industry as a whole
5. To discuss the significance of customer churn models in telecom industry
6. To compare the algorithms that are effective in reducing churn rate in telecom companies
7. The scope of this proposed system includes collecting and preprocessing customer data, feature engineering, selecting and training appropriate machine learning algorithms, and evaluating the performance of the model. The system will also incorporate techniques such as data augmentation, ensemble learning, and hyper parameter tuning to improve the model's accuracy and robustness.
8. The proposed system will be developed using Python and the popular machine learning libraries.The system will be trained and tested on a large and diverse dataset of telecom customer data, which will be obtained from reputable sources such as Kaggle. The system will be designed to be scalable, efficient, and easy to deploy in a production environment

III. FEATURES OF PROJECT

1. Real-time monitoring
2. Fraud detection
3. Historical Customer Data
4. Customer demographics
5. Customer service interactions
6. Network performance metrics
7. Competitor information
8. Social media sentiment analysis
9. Customer satisfaction surveys/NPS scores
10. Duration since last touchpoint/activity/interaction made by user with network (login times)

IV. LITERATURE REVIEW

1. Weijie Yu, Weinan Weng proposed system aims to identify affecting customer churn and construct an efficient model, which is used to predict and analyze data with visualization results. The churn forecast consists of several phases: data prepossessing, data analysis,evaluation measure, and application of machine learning algorithms. Moreover, data pre-processing covers data cleaning, transformation, and classification. The machine learning classifiers selected are Logistic Regression, SVM, Random Forest, AdaBoost, GBDT, XGBoost, Light GBM, and CatBoost. Classifiers were evaluated using performance measures, such as accuracy, precision, recall, AUC, and F1-Score. Based on the paper, the result was shown that the Light GBM outperformed other classifiers while identifying potential churners. [1]
2. Dr. O. Rama Devi, Sai Krishna Pothini proposed a model focuses on individuals who utilize paid OTT platforms for streaming video content on any device. The study used a questionnaire to gather data from participants of all demographics. The collected data underwent various pre- processing steps to make it suitable for machine learning models.The goal of predicting subscriptions for OTT (Over-The-Top) platforms using machine learning is to devise a model which can accurately predict whether a customer will continue using this platform or not. This information is important for OTT companies to understand and optimize their marketing and retention efforts.Relevant data, such as customer demographics and viewing habits, is collected and analyzed to train the model. This process involves cleaning the data, selecting important features, and training a machine learning model. The model is then tested and validated using performance metrics.In short, this problem requires a comprehensive understanding of customer behavior and the use of machine learning to predict subscription decisions. The results can provide valuable insights for OTT companies to improve their customer understanding and retention efforts. [2]
3. QiuYing Chen, Sang-Joon Lee, proposed system used Orange3 software to construct a customer churn prediction model for delivery platforms. The most effective Gradient Boosting algorithm was chosen to study customer churn prediction on the takeaway delivery platform. The predictive models of the Gradient Boosting

algorithm show efficient and accurate results that are relatively easy to approach. In addition, unlike the results of general mechanical learning techniques, it also exhibits key characteristics that make the implementation of gradient enhancement techniques more effective. Especially as with ecommerce, it is more effective to implement incremental enhancement techniques to predict non-contractual customer churn. [3]

4. Gavril et al. presented an advanced methodology of data mining to predict churn for prepaid customers using dataset for call details of 3333 customers with 21 features, and a dependent churn parameter with two values: Yes/No. Some features include information about the number of incoming and outgoing messages and voicemail for each customer. The author applied principal component analysis algorithm "PCA" to reduce data dimensions. Three machine learning algorithms were used: Neural Networks, Support Vector Machine, and Bayes Networks to predict churn factor. The author used AUC to measure the performance of the algorithms. The AUC values were 99.10%, 99.55% and 99.70% for Bayes Networks, Neural networks and support vector machine, respectively. The dataset used in this study is small and no missing values existed. [4]
5. He et al. proposed a model for prediction based on the Neural Network algorithm in order to solve the problem of customer churn in a large Chinese telecom company which contains about 5.23 million customers. The prediction accuracy standard was the overall accuracy rate, and reached 91.1%.[5]
6. Idris proposed an approach based on genetic programming with AdaBoost to model the churn problem in telecommunications. The model was tested on two standard data sets. One by Orange Telecom and the other by cell2cell, with 89% accuracy for the cell2cell dataset and 63% for the other one. [6]
7. Huang et al. studied the problem of customer churn in the big data platform. The goal of the researchers was to prove that big data greatly enhance the process of predicting the churn depending on the volume, variety, and velocity of the data. Dealing with data from the Operation Support department and Business Support department at China's largest telecommunications company needed a big data platform to engineer the fractures. Random Forest algorithm was used and evaluated using AUC. [7]
8. This paper explains our work on subscriber churn analysis and prediction for such services. We work on data mining techniques to accurately and efficiently predict subscribers who will change-and-turn (churn) to another provider for the same or similar service. The dataset we use is a public and real dataset compiled by Orange Telecom for the KDD 2009 Competition. Number of teams achieved high scores on this dataset requiring a significant amount of computing resources. We are aiming to find alternative methods that can match or improve the recorded high scores with more efficient and practical use of resources. In this study, we focus on ensemble of meta-classifiers which have been studied individually and chosen according to their performances.[8]
9. Shin-Yuan Hung a proposed different techniques to build predictive models for telecom churn prediction. We included customer service and customer complaint log for modeling, as suggestions from prior research of Wei and Chiu (2002). We examined the impact of inadequate data on model building. Our empirical evaluation shows that data mining techniques can effectively assist telecom service providers to make more accurate churner prediction.
10. Zhang Y presents a hybrid approach for building a binary classifier. The approach is the combination of the k-nearest neighbor algorithm, handling separately m 1-dimensional data sets divided from a data set in m-dimension, and the logistic regression method. This hybrid KNN-LR classifier improves the performance of the logistic regression in classification accuracy in some situations where the predictor and target variables exhibit complex nonlinear relationships. The results of the experiment on four benchmark data sets show the proposed approach compares favorably with the well-known classification algorithms such as C4.5 and RBF. Furthermore, its effectiveness is illustrated by its application in customer churn prediction based on real-world customer data sets.[10]

V. REPRESENTATION OF THE METHODOLOGY

The basic model for predicting future customer churn is data from the past. We look at data from customers that already have churned (response) and their characteristics / behaviour (predictors) before the churn happened. The dataset contains demographic details of customers, their total charges and they type of service they receive from the company.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, November 2023

It comprises of churn data of over customers spread over 21 attributes obtained from Kaggle.By fitting statistical models that relate the predictors to the response, we will try to predict the response for existing customers. This method belongs to the supervised learning category.

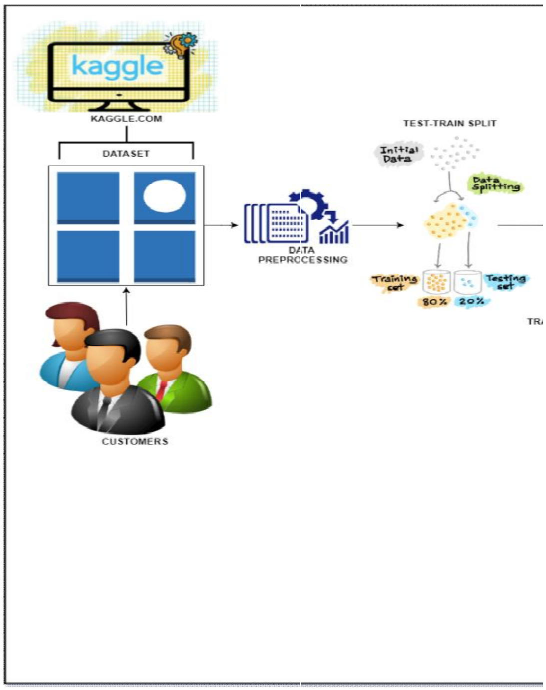

Fig : Representation Of The Methodology

VI. PROGRAMMING ARCHITECTURE

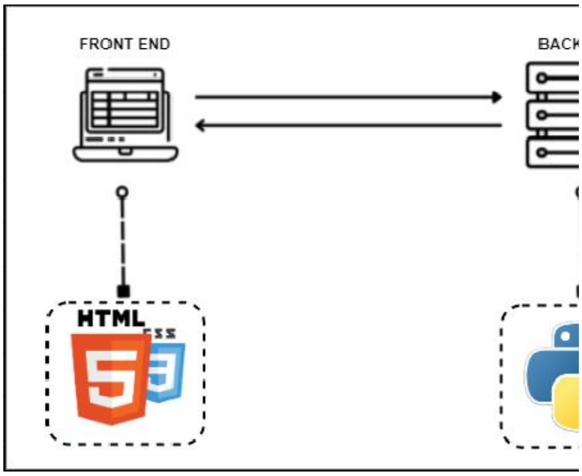

Figure: Programming Architecture

VII. ADVANTAGES

1. Creating new opportunities for cross-selling and upselling.
2. Improved Accuracy: Machine learning algorithms can analyze large volumes of customer data and identify patterns and trends that are not easily detectable by traditional methods. This helps in making more accurate predictions about customer churn, which can help telecom companies take proactive measures to retain their customers.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, November 2023

3. Real-time Predictions: Machine learning algorithms can make real-time predictions about customer churn, which allows telecom companies to take immediate action to address the issue. This can help in reducing churn rates and improving customer satisfaction.
4. Personalized Solutions: Machine learning algorithms can analyze individual customer data and provide personalized solutions to address their specific needs. This can help in improving customer satisfaction and reducing churn rates.
5. Cost-effective: Machine learning algorithms can help in reducing the cost of customer acquisition by predicting which customers are likely to churn and targeting them with retention offers. This can help in reducing the overall churn rate and improving the profitability of the telecom company.
6. Enhanced Customer Experience: Machine learning algorithms can help in improving the overall customer experience by providing personalized solutions and addressing their specific needs. This can help in improving customer satisfaction and reducing churn rates.
7. Competitive Advantage: Telecom companies that adopt machine learning algorithms for churn prediction can gain a competitive advantage over their competitors by providing better customer service and reducing churn rates. This can help in improving the overall market share and profitability of the company.
8. Continuous Learning: Machine learning algorithms can continuously learn and improve over time as they are exposed to more customer data. This helps in making more accurate predictions about customer churn and improving the overall performance of the model.
9. Scalability: Machine learning algorithms can handle large volumes of customer data and scale to accommodate the growing needs of the telecom company. This helps in improving the overall efficiency and effectiveness of the churn prediction process.

VIII. APPLICATION AREAS

1. Churn prediction can help you see which customers are about to leave your service so you can develop proper strategy to re-engage them before it is too late.
2. Customer Retention: Telecom companies can use machine learning algorithms for churn prediction to identify customers who are likely to churn and take proactive measures to retain them. This can include offering personalized solutions, discounts, and promotions to address their specific needs.
3. Network Optimization: Telecom companies can use machine learning algorithms for churn prediction to identify areas with high churn rates and optimize their network infrastructure to improve the overall customer experience. This can include improving network coverage, reducing network latency, and enhancing network reliability.
4. Pricing Strategy: Telecom companies can use machine learning algorithms for churn prediction to identify the optimal pricing strategy for their products and services. This can include offering different pricing plans based on customer usage patterns and preferences.
5. Marketing Campaigns: Telecom companies can use machine learning algorithms for churn prediction to identify the most effective marketing campaigns for retaining customers. This can include targeting customers with personalized offers and promotions based on their specific needs and preferences.
6. Sales Forecasting: Telecom companies can use machine learning algorithms for churn prediction to forecast sales and revenue based on customer churn rates. This can help in making informed business decisions and improving the overall financial performance of the company.
7. Resource Allocation: Telecom companies can use machine learning algorithms for churn prediction to allocate resources such as customer service representatives, marketing personnel, and sales teams based on customer churn rates. This can help in optimizing resource allocation and improving the overall efficiency and effectiveness of the company.
8. Network Security: Telecom companies can use machine learning algorithms for churn prediction to identify potential security threats and vulnerabilities in their network infrastructure. This can include identifying unusual network activity, detecting network intrusions, and preventing network attacks.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, November 2023

9. Network Maintenance: Telecom companies can use machine learning algorithms for churn prediction to identify potential network failures and outages before they occur. This can include predicting network faults, identifying network bottlenecks, and optimizing network performance.
10. Network Expansion: Telecom companies can use machine learning algorithms for churn prediction to identify potential areas for network expansion and growth. This can include identifying high-demand areas, predicting network usage patterns, and optimizing network capacity.
11. Network Optimization: Telecom companies can use machine learning algorithms for churn prediction to optimize their network infrastructure based on customer usage patterns and preferences. This can include improving network coverage, reducing network latency, and enhancing network reliability.

IX. HARDWARE REQUIREMENTS

1. CPU Quad Core (not counting hyper-threading) 2.4Ghz, Intel VT or AMDV (Intel i3 or better)
2. Memory 4 GB
3. The ability to install more memory is desirable. Disk 512 GB SSD or better
4. Graphics Accelerated, Gaming Support Nvidia is preferred over AMD 1920 by 1080 resolution is recommended (at least on an external port) At least 1280 by 1024 resolution
5. HDMI output recommended (perhaps with an adapter)
6. Mouse An external mouse (USB or Bluetooth) is desirable.
7. USB USB 3.0 desirable for an external disk Other USB ports may be needed for: mouse, printer, mic-in, and headphones-out, depending on how these are connected.
8. External monitor A 23” or larger HDMI monitor is recommended, with reasonable resolution.
9. Laptop or Desktop Windows 11 or macOS 12.4 or above. Linux is also acceptable if a mainstream distribution (e.g. Ubuntu).

X. SOFTWARE REQUIREMENTS

1. Operating System: Windows XP and later versions
2. Front End: HTML,CSS
3. Programming Language: Python
4. Dataset: Telecom Churn Prediction(Kaggle.com)
5. Domain: Machine Learning
6. Algorithm: Random Forest, Decision Tree, XGBoost.

XI. TEST DATA REQUIREMENTS

Unit Testing

Unit testing concentrates verification on the smallest element of the program – the module. Using the detailed design description important control paths are tested to establish errors within the bounds of the module.In this system each sub module is tested individually as per the unit testing such as campaign, lead, contact etc are tested individually. Their input field validations are test

Integration testing

Once all the individual units have been tested there is a need to test how they were put together to ensure no data is lost across interface, one module does not have an adverse impact onanother and a function is not performed correctly. After unit testing each and every sub module is tested with integrating each other.

XII. SYSTEM TESTING FOR THE CURRENT SYSTEM

In this level of testing we are testing the system as a whole after integrating all the main modules of the project. We are testing whether system is giving correct output or not. All the modules were integrated and the flow of information among different modules was checked. It was also checked that whether the flow of data is as per the requirements or not. It was also checked that whether any particular module is non-functioning or not i.e. once the integration is over each and every module is functioning in its entirety or not.

1. Functional testing: this involves testing the functionality of the system to ensure that it meets the required specifications and performs as expected. This includes testing the churn prediction accuracy, input data handling, and output interpretation.
2. Performance testing: this involves testing the system's performance under different load conditions to ensure that it can handle the expected workload and respond within acceptable time limits. This includes testing the system's scalability, resource utilization, and response time.
3. Security testing: this involves testing the system's security features to ensure that it can protect sensitive customer data from unauthorized access, theft, or misuse. This includes testing the system's authentication, authorization, and encryption mechanisms.
4. Compatibility testing: this involves testing the system's compatibility with different operating systems, databases, and hardware configurations to ensure that it can operate in a variety of environments.
5. Usability testing: this involves testing the system's user interface and user experience to ensure that it is intuitive, easy to use, and meets the needs of the end-users.
6. Regression testing: this involves testing the system's functionality after making changes or updates to ensure that the changes have not introduced any unintended side effects or regressions.
7. Acceptance testing: this involves testing the system's functionality from the perspective of the end-users to ensure that it meets their requirements and expectations. This includes testing the system's accuracy, reliability, and ease of use.
8. Recovery testing: this involves testing the system's ability to recover from failures, errors, or disasters to ensure that it can continue operating and providing service to the end-users.
9. Stress testing: this involves testing the system's performance under extreme load conditions to ensure that it can handle unexpected or catastrophic events.
10. Exploratory testing: this involves testing the system's functionality and behavior in unanticipated or unexpected scenarios to ensure that it can handle unexpected situations and provide accurate and reliable results. In this level of testing we tested the following: -

- Whether all the forms are properly working or not.
- Whether all the forms are properly linked or not.
- Whether all the images are properly displayed or not.
- Whether data retrieval is proper or not

XIII. CONCLUSION

In conclusion, churn prediction is a critical challenge for telecom companies as it affects revenue, market share, and customer satisfaction. Machine learning techniques offer a promising solution to this problem by enabling telecom companies to predict churn accurately and take proactive measures to prevent it. The challenges involved in churn prediction include dealing with large volumes of historical customer data, handling noisy, incomplete, and highly dimensional data, and addressing imbalanced datasets. Various machine learning algorithms such as Random Forest, Decision Tree, XGBoost . can be used to solve this problem depending on the nature of the data and the specific requirements of the telecom company. By implementing machine learning algorithms for churn prediction, telecom companies can reduce churn rates, improve customer satisfaction, and gain a competitive advantage in the market.

REFERENCES

[1]. Weijie Yu, Weinan Weng," Customer Churn Prediction Based on Machine Learning" 2022 4th International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), 978-1-6654-6399-7/22/$31.00 ©2022 IEEE,DOI 10.1109/AIAM57466.2022.00176

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, November 2023

[2]. Dr. O. Rama Devi, Sai Krishna Pothini," Customer Churn Prediction using Machine Learning: Subcription Renewal on OTT Platforms", IEEE Xplore Part Number: CFP23BC3-ART; ISBN: 978-1-6654-5630-2, 978-1-6654-5630-2/23/$31.00 ©2023 IEEE

[3]. QiuYing Chen, Sang-Joon Lee," A Machine Learning Approach to Predict Customer Churn of a Delivery Platform", 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) | 978-1-6654-5645-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAIIC57133.2023.10067108

[4]. Brandusoiu I, Toderean G, Ha B. Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications. 2016. p. 97–100.

[5]. He Y, He Z, Zhang D. A study on prediction of customer churn in fixed communication network based on data mining. In: Sixth international conference on fuzzy systems and knowledge discovery, vol. 1. 2009. p. 92–4.

[6]. Idris A, Khan A, Lee YS. Genetic programming and adaboosting based churn prediction for telecom. In: IEEE international conference on systems, man, and cybernetics (SMC). 2012. p. 1328–32.

[7]. Huang F, Zhu M, Yuan K, Deng EO. Telco churn prediction with big data. In: ACM SIGMOD international conference on management of data. 2015. p .607–18.

[8]. Yabas, U, Chankya, H.C. (2013). Churn prediction in subscriber management for mobile and wireless communications services. IEEE Publications.

[9]. Shin-Yuan Hung a, David C. Yen b, Hsiu-Yu Wang, "Applying data mining to telecom churn management", Expert Systems with Applications 31 (2006) 515–524,

[10]. Zhang, Y.; Qi, J.; Shu, H.; Cao, J. A hybrid KNN-LR classifier and its application in customer churn prediction. In Proceedings of the 2007 IEEE International Conference on Systems, Man and Cybernetics, Montréal, QC, Canada, 7–10 October 2007; pp. 3265–3269.

DOI: 10.48175/IJARSCT-13886