Customer churn prediction in telecommunication industry using machine learning models

Laiba Nadeem
School of Computing
Asia Pacific University of Technology & Innovation (APU)
Kuala Lumpur, Malaysia
TP060221@mail.apu.edu.my

Chandra Reka Ramachandran
School of Computing
Asia Pacific University of Technology & Innovation (APU)
Kuala Lumpur, Malaysia
Chandra.reka@staffmail.apu.edu.my

Mandava Rajeswari
School of Computing
Asia Pacific University of Technology & Innovation (APU)
Kuala Lumpur, Malaysia
prof.dr.mandava@staffemail.apu.edu.my

***Abstract*—In recent times the mobile computing is becoming one of the most apparent means of communication. The increase in the telecommunication service providers has resulted in the fierce competition making the market saturated to an extent where companies are struggling to retain their customers. This in turn has shifted the focus of the companies from building a large customer base into retaining customers. Customer churn refers to the tendency of customers to cancel their service or subscription and switch to competitors. Since the cost of attaining a new customer is greater than retaining existing customers companies need to adopt strategies to effectively predict customer churn. The companies are storing a vast amount of resources related to customers (big data) but fail to realize their potential in solving business problems. To this date, few companies have adopted machine learning techniques to accurately predict the customer churn in the telecommunication sector thereby more work is required to fill the gap. This research aims to analyze the application of machine learning in the Telecommunication Industry, application of machine learning models for customer churn prediction in the telecommunication sector and churn prediction challenges. Based on the findings the most common machine learning models adopted in the telecommunication sector include Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and K-nearest neighbor (KNN). The researchers mostly overlook the class imbalance problem. The issue of high dimensionality data can be solved by incorporating Principal Component Analysis (PCA) in machine learning models and the adoption of data transformation techniques improves the overall performance of the model. The hybrid models and models incorporated with genetic algorithms provide better performance compared to generic machine learning models.**

Keywords—Customer churn, big data, random forest, support vector machine, gradient boosting-nearest neighbor.

I. INTRODUCTION

RFID Telecommunication (telecom) refers to the exchange of information over varying distances through electronic means which comprises voice, video, and data transmission. This broad term encompasses the wide range of communication infrastructures and information transmitting technologies which includes mobile devices, wired phones, satellites, fibre optics, television broadcasting, and radio. The telecommunication sector comprises companies that allow global communication to take place. The majority of companies in this sector consist of telephone operators (wired, wireless), satellite, cable, and internet service providers. The company caters to two categories of customers Business to Consumer (B2C) where the transactions take place between Telecom Company and individual consumer and Business to Business (B2B) where transactions take place between Telecom Company and businesses [1]. Artificial intelligence (AI) is changing the dynamics of service sectors. In simple terminology, artificial intelligence refers to the ability of machines to carry out tasks that require some form of human intelligence. AI is an umbrella term encompassing Machine Learning and Deep Learning. Machine Learning refers to the ability of machines to learn from the data without being explicitly programmed and Deep Learning is a subset of Machine Learning and incorporates algorithms and models that imitate the structure of neural networks present in the brain to learn from the information fed and derive appropriate solutions [2, 3]. Any program or service that incorporates machine learning or deep learning is termed to have adopted artificial intelligence. The machine learning models used in the telecommunication sector are summarized in the Table. 1 [8, 13,18].

TABLE I. MACHINE LEARNING MODELS

Machine Learning Models	Model Description
Decision Tree	It is used to solve the classification problem and works for both continuous dependant variables and categorical data and works by splitting the data into 2 or more homogenous sets.
Random Forest	It is an ensemble method used for both regression and classification problems. It works by creating multiple decision trees and considers the average of all the decision tree results to propose a final solution.
Extreme Gradient Boosting	This is an ensemble method created by the implementation of various gradient boosting decision trees and considers the average of multiple trees to propose a final solution.
Naïve Bayes	This method is used to solve the classification problem based on the probabilistic approach
Support Vector Machine	This method is used to solve both classification and regression problems and works by dividing the data into two groups separated by a hyperplane.
Adaboost	It is an ensemble method used for classification problems and operates by converting weak classifiers into strong ones based on weighted classification errors.
Logistic Regression	It is used to solve classification problems (binary values yes/no, true/False) and predicts the probability of occurrence of the event.
Density-Based Spatial Clustering of Application with Noise	This is a clustering technique based on data density.

K-means Clustering	This method can solve the classification problem and works by starting with several clusters and pieces of data from the centroid and based on the distance of data from each other and centroid the data are grouped into clusters.
Neural Network	This method can solve the classification problem by using neurons similar to biological neural networks found in the human brain.
K-nearest neighbor	It is used for classification problems and works by classifying new cases based on similarity measures such as distance and the case is classified based on the majority of neighboring data points.

According to MarketandMarkets [4], the application of Artificial Intelligence in the telecommunication sector is experiencing rapid growth and is expected to rise from $235.7 million in 2016 to $2,497.8 million by the year 2022 resulting in the Compound Annual Growth Rate (CAGR) of 46.8%.

According to Statista in the year 2019, the adoption of Artificial Intelligence in the telecommunication sector is divided into multiple areas. AI in service operations accounted for 74%, product/service development accounted for 48%, marketing and sales accounted for 28%, manufacturing accounted for 21 %, supply chain management accounted for 27% and risk assessment accounted for 30%. This shows there is a gap in the adoption of AI techniques in the telecommunication sector [5].

The purpose of this research is to review the existing works of literature on the application of machine learning in the telecommunication sector, application of machine learning models for customer churn prediction in telecommunications and highlights the churn prediction challenges.

II. MACHINE LEARNING USE CASES IN THE TELECOMMUNICATION SECTOR

A. Fraud Detection Call Detail Record Using Machine Learning

Fraud in telecommunication results in huge revenue loss and degrades the customer relationships with the company especially when losses expedite over time [6]. With the advances in telecommunications, the problem of fraud has also accelerated in recent years [7]. The research carried out by [8] was focused on a Telecommunication Company which lost up to Rp 28million due to fraudulent calls in the year 2018 and the purpose was to classify potential calls as fraud. The machine learning techniques were adopted and the data of call detail records containing calls data and fraud patterns identified by the telecom company data was utilized to evaluate the accuracy of the model and for model training. The pre-processing step was carried out to remove anomalies from the data and make it more normalized and reliable.

Two machine learning algorithms were adopted namely K-means clustering Algorithm and Density-Based Spatial Clustering of Application with Noise (DBSCAN) and the accuracy was compared using confusion matrix Evaluation. The accuracy score provided by K-means clustering accounted for 0.97 whereas the DBSCAN accounted for 0.90 resulting in K-means providing a higher accuracy in anomaly detection in telecommunications [8].

Subsequently, a similar study was carried out using different machine learning models to classify the calls as fraudulent and benign and the machine learning models utilized for the research included Random Forest, XGBoost, Neural Networks, Recurrent Neural Network, Support Vector Machine, and Logistic Regression, and the models were trained on one province and tested on other provinces and utilized similar research framework as proposed by [8].

B. Denial of Service Attack Detection In Telecommunication Network Using Machine Learning

The telecommunication network has allowed seamless connection and capacity to store and communicate an abundant scale of information in the form of voice and text which are sensitive. This, in turn, has resulted in telecommunication networks becoming prey to multiple cyber threats one of the most common of which includes distributed denial of service attacks (DDOS). The DDOS is initiated by overrunning the target with traffic and halt the correct functioning of resources in the network making it more prone to hacking and this has become a major security concern for the telecommunication sector storing numerous identifying information of their customers [10]. This section discusses the application of machine learning techniques to elude the hacking of sensitive information in the form of text and voice. The distributed denial of service detection involves steps taken to identify and segregate attacks from normal traffic. The distributed denial of service attack has three types Volume Based, Protocol based, and Application-Based.

The data was fed and the model was trained to identify DDOS attacks. The information related to major resources that are prone to be attacked was utilized to segregate the normal and the abnormal traffic flow, packet size, IP spoofing, and train the model. The machine learning models used in the research included Neural Network, Support Vector Machine, and a hybrid model combining neural network and support vector machine (CNSVM).

In Comparison to accuracy scores provided by the Support Vector Machine and Neural Network the hybrid approach combining neural network with support vector machine resulted in 40% enhanced detection of denial of service attacks [11].

A similar study carried out by [12] and the machine learning model proposed in the research was the support vector machine (SVM) for detection of denial of service attack.

The pre-processing techniques were carried out to improve the reliability of data and the model was trained. The model was successful in making predictions attaining an accuracy score of 98.9%.

C. Predicting Customer Complaints in Telecom Industry

The telecommunication industry faces fierce competition and thereby is struggling to retain the most valuable asset of their business the customers. Complaints are considered as a key factor for customer dissatisfaction and in other cases customer loss. The Korean mobile telecom company data was used by Choi [13] and based on the subscribers, gender, age, device manufacturer, service quality, and complaint status were taken into account for customer complaint prediction. The purpose was to predict the complaint before they are made by the users depending on the level of service provided and historical data of complaints based on multiple customers.

The data related to the customer, network equipment utilized, and complaint history were used and necessary pre-processing steps were carried out. The models were then trained to predict customer complaints in the telecommunication industry.

The research was carried out using four machine learning models Artificial Neural Network (ANN), Support Vector Machine (SVM), K-nearest neighbor, and Decision Tree. The models were tuned to enhance accuracy. The number of hidden layers, hidden layers width, activation function, object function optimizer, dropout, learning rate, and learning steps were the parameters tuned for Artificial Neural Network (ANN). The kernel parameter was tuned for Support Vector Machine (SVM), constant k parameter was tuned for K-nearest neighbor (KNN), and for Decision Tree (DT) the split criteria and max_depth were the parameters that underwent tuning. The accuracy score provided by ANN, SVM, KNN, DT accounted for 81.04%, 79.02%, 80.56%, and 69.83% respectively and the highest accuracy for customer complaints prediction in the mobile telecom industry [13].

D. Customer Churn Prediction Using Machine Learning Models in Telecom Sector

In recent times the telecommunication companies are generating a huge amount of data on daily basis. There are multiple players in the market offering telecommunication services making the market fierce and competitive. This trend has resulted in customers having more flexibility to choose

their service providers due to multiple options available at their disposal to obtain better services. The main objective of the telecommunication sector companies is to maximize profit generation and stay competitive in the market and reduce the churn rate of their existing customers. A customer churn usually takes place either when the customer manages to find a better alternative or, as in most cases, the customer was dissatisfied with the services provided by the particular telecom company.

The general model for accurate prediction of customer churn is provided in Fig. 1:

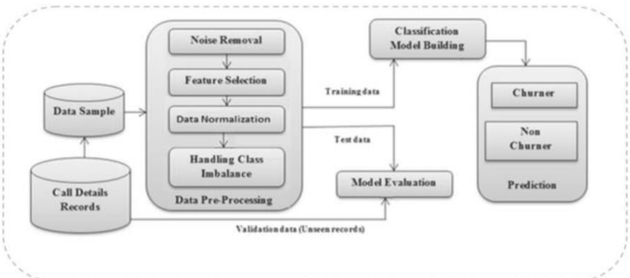

Fig. 1 Proposed Model for Accurate Customer Churn Prediction [14].

Ahmad et al [15] developed a system for customer churn prediction using SyriaTel Telecom company dataset using machine learning models Decision Tree (DT), Random Forest (RF), Global System for mobile communications tree algorithm (GSFMC), and extreme gradient boosting (EGB). His research focused on data preparation methods and feature engineering, class balancing as well as hyperparameter tuning using K-folds and cross-validation to improve the performance of the models and also incorporated the features of mobile social network and based on his research the extreme gradient boosting outperformed other machine learning models providing the accuracy score of 93.301.

III. APPLICATION OF CUSTOMER CHURN PREDICTION IN THE TELECOM INDUSTRY

The major problem faced by Telecom Company is to reduce the customer churn rate. The customer churn affects the overall reputation of the company which can often result in brand loss. A loyal high revenue-generating customer is the most valuable asset and is mostly unaffected by the competitor companies and are bound to refer the company services to friends, family members, and colleges. The telecommunication companies however have to take drastic measures such as policy shifts if the number of customers continuously drops and this in turn results in revenue loss.

The primary reasons for customer churn in the telecommunication industry arise from the level of quality provided by service providers which include service quality, network coverage, load errors, billing cost, and technology. The prediction rate of normal churn in the telecom industry is estimated to be 2% which accounts for the total annual loss of approximately 100 billion dollars [16]. The prediction of customer churn is considered 16 times cheaper than attracting new customers and the cost of inviting new customers through advertisements, workforce, and concession can easily scale up to six times higher than retaining existing customers [15, 17,18]. According to [17] the decrease in churn rate by 5% can result in profit gain from 25% to 85%.

Consequently, there is a need to predict the existence of customer churn to aid the telecom industry to create timely customer retention strategies. Due to technological advancement and big data, multiple data mining and machine learning solutions are available which can be utilized for accurate prediction of customer churn in the telecommunication sector by analyzing the trends and patterns to develop insights on customer behaviour based on historical data.

This section of the research highlights multiple approaches to the adoption of machine learning extracted from works of literature to address the common problem of customer churn which is the act of customers withdrawing from the service due to multiple factors influencing their decision.

Amin et al [18] researched about applying data transformation techniques which improves the quality of data to determine its impact on the performance of the machine learning models for Cross Company Churn Prediction which is a domain of research where the target company lacks historical data for training prediction models and utilizes the data from another company to train the models to successfully carry out predictions. The data transformation methods used included log, rank, z-score, and box-cox and were applied to the machine learning models to determine if data transformation techniques can influence the accuracy of models performance. The machine learning models used for the research included Naïve Bayes (NB), K-nearest neighbor (KNN), Gradient Boosting (GB), Single Induction Rule (SRI), and Neural Network. The initial accuracy score of the NB, KNN, SRI, and NN accounted for 0.51, 0.49, 0.52, 0.45, and 0.48 respectively and the best performance was obtained from Gradient Boosting (GB). Most of the data transformation techniques improved the performance of the model however the z-score data transformation technique failed to improve the accuracy of the models. Naïve Bayes model experienced a significant increase in performance compared to other machine learning models. SRI model did not experience a significant change in the performance after applying data transformation techniques.

Suguna et al [19] proposed a technique to reduce the manual time for computation and maximize early prediction of customer churn on Telecom dataset using the extraction of feature from the dataset by incorporating Principal Component Analysis (PCA) technique which converted high dimensionality dataset into low dimensional data without any loss of information. The reduced dataset was applied to various machine learning classifier models. Based on his research after application of PCA on telecom dataset the kernel support vector machine outperformed in performance providing the accuracy score of 95.5% compared to other classifiers Logistic Regression (LR), K-nearest neighbor (KNN), Kernel Support Vector Machine (KSVM), Random Forest (RF), Naïve Bayes (NB) and Decision Tree (DT).

Malleswari et al [20] research focused on reducing the features from the dataset to decrease computational time and focus on only the features that contribute to customer churn prediction on Orange Telecom Dataset. Three machine learning models were used to evaluate the performance of the machine learning models which included Random Forest (RF), Decision Tree (DT), and Gradient Boosted Tree (GBT) and the RF model provided the better accuracy score of 87% compared to DT and GBT. It was also observed from the research that DT had the lowest execution time compared to the other two models.

Raja and Jeyakumar [21] research focused on customer churn prediction in the telecom industry using IBM Watson dataset using three machine learning models K-nearest neighbor (KNN), Random Forest (RF), and Extreme Gradient Boost (XGB). The research focused on simple data pre-processing techniques to remove missing values and change the data types of data to the numeric data type for customer churn classification. The univariate analysis was also carried out to normalize the features of the dataset based on the value of a subscriber's service. The correlation among the features was carried out and after applying machine learning models the XGBoost algorithm provided a higher accuracy score of 0.798 compared to the other two models.

Gaur and Dubey [22] research focused on four machine learning models Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosting (GF) using telecom dataset for customer churn prediction and the Gradient Boosting (GF) provided greater accuracy score in predicting customer churn in the telecommunication sector.

Kavitha et al [23] research focused on the application of three machine learning models Extreme Gradient Boosting (XGBoost), Logistic Regression (LR), and Random Forest (RF) for customer churn prediction in the telecommunication sector. The pre-processing to remove missing values and extract features influencing customer churn were taken into account and based on the analysis carried out the Random Forest (RF) model provided the highest accuracy score of 80% compared to XGBoost and LR.

Idris et al [24] research focused on integrating the search optimization capability for searching churners and non-churners using genetic algorithm with the capabilities of classification in AdaBoost for developing an evolved system for efficient churn prediction. The particle swarm optimization genetic programming technique was utilized. The evolution of the Genetic Programming (GP) process was exploited by the integration of Adaboost to evolve multiple programs for each class and the final prediction was dependent based on the weighted sum of the output of the GP programs. The research focused on two datasets orange and Cell2Cell and the machine learning models used for the study comprised of K-nearest neighbor (KNN), Random Forest (RF), and GP with Adaboost. Based on the research the model with GP and Adaboost outperformed in accurate churn prediction compared to other models for both the datasets acquiring the accuracy score of 63% and 89% for orange and Cell2Cell datasets respectively.

IV. CHURN PREDICTION CHALLENGES

This area of research highlights some of the common challenges associated with churn prediction models using machine learning.

The most common problem associated with churn prediction involves the class imbalance problem where the target class has more non-churn data compared to churn data resulting in the hindrance of attaining accurate predictions due to class bias problem [25, 26]. To address this issue various class balancing techniques which include under-sampling, oversampling, etc are available.

The other issues associated with the churn prediction involve scarcity of data and absence of parameter tuning for enhancing the model's performance or high dimensionality dataset and the computation time required for accurate predictions of model performance. This in turn cause hindrance in the performance of machine learning models. Many studies address these issues. Suguna [19] in his research proposed the technique of Principal Component Analysis (PCA) which converts high dimensionality dataset into low dimensional data without any loss of information. Another study of Cross Company Churn prediction also addresses the issue of scarcity of data to train the model and using the data from other companies for model training [18]. There are various hyperparameter tuning options available exclusive to each machine learning model to improve the performance of the model as showcased in the study of predicting customer complaints in the telecom industry [13].

V. CONCLUSION

In Conclusion, based on the extensive research carried out on the use cases of machine learning models in the telecommunication sector, the application of machine learning models for customer churn prediction in the telecom sector and churn prediction challenges it is evident that the most common machine learning models used in telecommunication sector includes Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting and K-nearest neighbor (KNN). Many of the researchers did not include the hyper parameter optimization technique to improve the model performance and utilized simple pre-processing techniques including missing values analysis and data type conversions. The research also concludes that different machine learning models provide different performance scores depending on the dataset used. There was also a gap in addressing the most common issue associated with customer churn prediction which is the class imbalance problem where the target class data is present in minority resulting in class bias problem which impacts the overall performance of the machine learning models. Furthermore, the issue of scarcity of data to train the model was addressed and the approach to overcome this common issue that halts or slows the adoption of machine learning models in the telecommunication sector was discussed by [18] using Cross Company Churn Prediction which is a domain of research where the target company lacks historical data for training prediction models and utilizes the data from another company to train the models to successfully carry out predictions. Multiple kinds of research were carried out to address the issues faced by machine learning models to carry out accurate predictions. The issue of increased manual time for computation for successful customer churn prediction was addressed by [19] and his research proposed the incorporation of the Principal Component Analysis (PCA) technique which converts high dimensionality dataset into low dimensional data without any loss of information and this contributed to increase in the performance speed of machine learning models in accurately predicting customer churn in telecom. Subsequently, another research was carried out by

[24] with the focus of further improving the performance of machine learning models by incorporating search optimization capability for searching churners and non-churners using genetic programming with the capabilities of classification in AdaBoost for developing an evolved system for efficient churn prediction and this improved the performance of machine learning model for accurate churn prediction. Another research was carried out by [24] to determine the influence of data transformation techniques to improve model performance and the data transformation techniques that were utilized included log, rank, z-score, and box-cox and it was observed that adoption of data transformation techniques improved the performance of the machine learning algorithms. In comparison to the existing studies, most of the researchers focused on customer churn prediction as a whole and did not consider parameters such as customer loyalty towards the company or the overall business loss with the withdrawal of service from high revenue-generating customers utilizing the company's services for a relatively long period which is crucial for companies to retain their valuable customers and carry out effective retention strategies.

VI. FUTURE RESEARCH

In the Future Research, the adoption of external environmental factors (social, economic, political) influencing customer churn along with the segregation of customers into high and low revenue-generating customers and their impact on the performance of machine learning models is the main focus of study. The research on various pre-processing techniques to enhance the quality of data and the comparison of machine learning models and deep learning models for predicting customer churn should be taken into account. The in-depth study of various dimensionality reduction techniques and incorporation of multiple genetic algorithms and their influence on machine learning models performance would also be the focus of research.

REFERENCES

[1] "What Is the Telecommunications Sector?," Investopedia, 2019. https://www.investopedia.com/ask/answers/070815/what-telecommunications-sector.asp.

[2] H.-D. Wehle, "(PDF) Machine Learning, Deep Learning, and AI: What's the Difference?," ResearchGate, 2017. https://www.researchgate.net/publication/318900216_Machine_Learning_Deep_Learning_and_AI_What. (accessed Nov. 25, 2020)

[3] D. Jakhar and I. Kaur, "Artificial intelligence, machine learning and deep learning: definitions and differences," Clinical and Experimental Dermatology, Jun. 2019, doi: 10.1111/ced.14029.

[4] "AI in Telecommunication Market Size, Share and Global Forecast to 2022 | MarketsandMarketsTM," www.marketsandmarkets.com, 2018. https://www.marketsandmarkets.com/Market-Reports/ai-in-telecommunications-market-130504876.html.

[5] S. Liu, "AI adoption by industry/function in global organizations 2019," Statista, 2019. https://www.statista.com/statistics/1112982/ai-adoption-worldwide-industry-function/.

[6] M. Arafat, A. Qusef, and G. Sammour, "Detection of Wangiri Telecommunication Fraud Using Ensemble Learning," 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), Apr. 2019, doi: 10.1109/jeeit.2019.8717528.

[7] M. Liu, J. Liao, J. Wang, and Q. Qi, "AGRM: Attention-Based Graph Representation Model for Telecom Fraud Detection," IEEE Xplore, May 01, 2019. https://ieeexplore.ieee.org/document/8761665 (accessed Nov. 25, 2020).

[8] M. A. Jabbar and S. Suharjito, "Fraud Detection Call Detail Record Using Machine Learning in Telecommunications Company," Advances in Science, Technology and Engineering Systems Journal, vol. 5, no. 4, pp. 63–69, Jul. 2020, doi: 10.25046/aj050409.

[9] H. Li et al., "A Machine Learning Approach to Prevent Malicious Calls over Telephony Networks," IEEE Xplore, May 01, 2018. https://ieeexplore.ieee.org/document/8418596.

[10] S. Newman, "Under the radar: the danger of stealthy DDoS attacks," Network Security, vol. 2019, no. 2, pp. 18–19, Feb. 2019, doi: 10.1016/s1353-4858(19)30025-x.

[11] P. A and S. S, "DDOS ATTACK DETECTION IN TELECOMMUNICATION NETWORKUSING MACHINE LEARNING," Journal of Ubiquitous Computing and Communication Technologies, vol. 01, no. 01, pp. 33–44, Sep. 2019, doi: 10.36548/jucct.2019.1.004.

[12] M. S. Hoyos Ll, G. A. Isaza E, J. I. Vélez, and L. Castillo O, "Distributed Denial of Service (DDoS) Attacks Detection Using Machine Learning Prototype," Distributed Computing and Artificial Intelligence, 13th International Conference, pp. 33–41, 2016, doi: 10.1007/978-3-319-40162-1_4.

[13] C. Choi, "Predicting Customer Complaints in Mobile Telecom Industry Using Machine Learning Algorithms - ProQuest," www.proquest.com, 2018. https://www.proquest.com/docview/2054017508.

[14] M. Azeem, M. Usman, and A. C. M. Fong, "A churn prediction model for prepaid customers in telecom using fuzzy classifiers," Telecommunication Systems, vol. 66, no. 4, pp. 603–614, Mar. 2017, doi: 10.1007/s11235-017-0310-7.

[15] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," Journal of Big Data, vol. 6, no. 1, Mar. 2019, doi: 10.1186/s40537-019-0191-6.

[16] B. Balle, B. Casas, A. Catarineu, R. Gavaldà, and D. Manzano-Macho, "The Architecture of a Churn Prediction System Based on Stream Mining," Semantic Scholar, 2013. https://www.semanticscholar.org/paper/The-Architecture-of-a-Churn-Prediction-System-Based-Balle-Casas/916db221ad60a6daa0bded1b9aa8c3b6e0604701.

[17] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," IEEE Access, vol. 7, pp. 60134–60149, 2019, doi: 10.1109/access.2019.2914999.

[18] A. Amin et al., "Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods," International Journal of Information Management, vol. 46, pp. 304–319, Jun. 2019, doi: 10.1016/j.ijinfomgt.2018.08.015.

[19] Suguna.R, Devi.S, and Mathew.R, "Customer Churn predictive analysis by component minimization using machine learning,"

International journal of innovative technology and exploring engineering, vol. 8, no. 3229–3233, pp. 3229–3233, 2019.

[20] M. Malleswari, R. J. Manira, P. Kumar, and Murugan, "Comparative Analysis of Machine Learning Techniques to Identify Churn for Telecom Data," International Journal of Engineering & Technology, vol. 7, no. 3.34, p. 291, Sep. 2018, doi: 10.14419/ijet.v7i3.34.19210.

[21] B. Raja and P. Jeyakumar, "An Effective Classifier for Predicting Churn in Telecommunication," Journal of Advanced Research in Dynamical and Control Systems, vol. 11, no. 01, pp. 221–229, 2019.

[22] A. Gaur and R. Dubey, "Predicting Customer Churn Prediction In Telecom Sector Using Various Machine Learning Techniques," IEEE Xplore, Dec. 01, 2018. https://ieeexplore.ieee.org/document/8933783.

[23] V. Kavitha, G. Kumar, M. Kumar, and M. Harish, "Churn Prediction of Customer in Telecom Industry using Machine Learning Algorithms," 2020. [Online]. Available: https://www.ijert.org/research/churn-prediction-of-customer-in-telecom-industry-using-machine-learning-algorithms-IJERTV9IS050022.pdf.

[24] A. Idris, A. Khan, and Y. S. Lee, "Genetic Programming and Adaboosting based churn prediction for Telecom," 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Oct. 2012, doi: 10.1109/icsmc.2012.6377917.

[25] B. Zhu, B. Baesens, and S. K. L. M. vanden Broucke, "An empirical comparison of techniques for the class imbalance problem in churn prediction," Information Sciences, vol. 408, pp. 84–99, Oct. 2017, doi: 10.1016/j.ins.2017.04.015.

[26] C. Gui, "Analysis of imbalanced data set problem: The case of churn prediction for telecommunication," Artificial Intelligence Research, vol. 6, no. 2, p. 93, Jun. 2017, doi: 10.5430/air.v6n2p93.