Journal of Theoretical and Applied Information Technology
30th November 2024. Vol.102. No. 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A SYSTEM FOR ANALYSING CALL DROP DYNAMICS IN THE TELECOM INDUSTRY USING MACHINE LEARNING AND FEATURE SELECTION

[1]CHEEPU BALAKRISHNA [2]DR CINDHE RAMESH [3] S MEGHANA
[4] DR. C. DASTAGIRAIAH

[1]assistant Professor,Computer Science And Engineering, Saiapurthi Institute Of Technology
[2]professor, Department Of Cse, Cvr College Of Engineering, Hyderabad, Telangana.
[3]assistant Professor, Department Of Cse, Ngit College, Uppal, Hyderabad
[4]assistant Professor, Department Of Cse, Anurag University, Hyderabad, Telangana, India-500088,
Emails: balu.cheepu@gmail.com [2]hmcr.ramesh@gmail.com [3]meghanashiva1212@gmail.com
[4]dastagiraiah.cse@anurag.edu.in

ABSTRACT

Quality of service in the telecom industry plays a vital role in the growth and economy of the country. In India, several telecom operators provide services, and there is a regulatory authority known as the Telecom Regulatory Authority of India (TRAI). In the telecom domain, call dropping is a problem that deteriorates the performance of the telecom industry in rendering services. It also causes inconvenience and waste of time for users, as well as reducing the level of user satisfaction. There is a need for a technology-driven analysis with the help of Artificial Intelligence (AI) to analyze call drop dynamics toward making well-informed decisions. The existing research revealed that Machine Learning (ML) helps analyze call drop dynamics. However, there is a need for a framework with machine learning techniques and optimizations to improve performance in analyzing call drop dynamics in the telecom industry. In this paper, we proposed an ML framework for automatic analysis of all drops in the telecom industry across all operators. The framework also supports optimizations like future engineering and dimensionality reduction to improve the performance of machine learning models. We proposed an algorithm called Learning based Call Drop Analytics (LbCDA) which exploits feature selection and training multiple classifiers towards call drop analytics. With benchmark dataset variants of the telecom industry, our empirical study has revealed that our framework's Random Forest (RF) model outperforms other models with the highest accuracy of 87.40%.

Keywords: *Telecom Industry, Call Drop Analysis, Random Forest, Machine Learning, Artificial Intelligence*

1. INTRODUCTION

The Telecom industry is one of the industries that promote the growth and economy of a country. The performance of telecom operators is paramount due to the significance of telecom services for various communications and networks. Signal strength and connectivity are significant for successful telecommunications. Satisfactory calls in telecommunications lead to improved network performance. In this regard, call dropping is a significant problem that causes significant issues for users and telecom operators. Call dropping can also lead to the deterioration of customer satisfaction. Moreover, it is a problem that can impact the growth and economy of a country. In India, several telecom operators are rendering their services. With the emergence of artificial intelligence (AI), it is now possible to analyze the telecom industry data to understand call drop dynamics across all operators. The analysis of call drops with the technology-enabled approach can help telecom operators improve their services, leading to better quality services to users. There are some existing approaches for call drop analytics, as found in the literature. To increase prediction accuracy for user numbers, this study presents the edge-controller architecture in 5G networks [4]. Wireless technologies are essential to society's growth from 5G to the forthcoming 6G networks with AI/ML integration. For 6G applications such as resource management and biometrics, machine learning techniques, such as supervised and Deep Learning, are essential [6]. From the review of the literature, it was observed that there was little research on call drop analytics in the telecom industry using machine learning models. Our contributions to this paper are as follows.

1. We proposed an ML framework for automatically analyzing all telecom industry drops across all operators. The framework also supports optimizations like future engineering and dimensionality reduction to improve the performance of machine learning models.
2. We proposed a learning-based Call Drop Analytics (LbCDA) algorithm, which exploits feature selection and trains multiple classifiers for call drop analytics.
3. We developed a prototype to evaluate our telecom industry call drop analytics methodology.

The remainder of the paper is structured as follows: Section 2 reviews prior works on analyzing the telecom industry's data using machine learning. Section 3 provides preliminaries associated with the current study. Section 4 presents experimental results linked to the research carried out in this paper. Section 5 presents conclusions drawn from the research in this paper and gives scope for future research.

2. RELATED WORK

Lalwani et al. [1], with the best accuracy of 81.71% and AUC score of 84%, the six-phase approach comprises data pre-processing, feature selection using a gravitational search algorithm, and model testing with Adaboost and XGBoost. Technological developments in machine learning enable telecom companies to estimate client attrition more accurately. Ullah et al. [2] suggested a churn prediction model utilizing classification and clustering since customer attrition is expensive for the telecom sector. By segmenting churn clients for targeted retention offers based on behavioral patterns, CRM methods are enhanced. Future research will include trend analysis using AI. Wisesa et al. [3] accurate models for B2B market analysis are necessary for sales forecasts. This paper proposes the Gradient Boost Algorithm for B2B sales trend forecasting with MSE: 0.18. It employs machine learning to improve sales forecasts. When comparing four machine learning techniques using data from 2016–2017 to 2018, it is indicated that they are effective in forecasting B2B sales patterns. Polese et al. [4], for reduced latency, 5G networks use edge clouds. This study presents the edge-controller architecture to increase prediction accuracy for user numbers. Kaur et al. [5], wireless technologies are essential to society's growth from 5G to the forthcoming 6G networks with AI/ML integration. For 6G applications such as resource management and biometrics, machine learning techniques, such as supervised and Deep Learning, are essential.

Jain et al. [6] used an orange database; this study employed Logistic Regression and Logit Boost. The accuracy of the two approaches was equivalent in the results, pointing to the need for hybrid models in the future for better results. For more research, comparisons with neural networks, SVM, and rough sets are planned. Park et al. [7] used two computational approaches to investigate a nonlinear fractional telecommunication model. Novel solutions aid in understanding wave propagation in a dispersive medium. Lee et al. [8] examined 3,901 news stories and 660 academic publications on Industry 4.0, identifying 31 research questions. It validates previous debates on Industry 4.0 by proposing a six-dimensional convergence paradigm. This is an interdisciplinary subject that machine learning helps to understand by recommending future considerations of patent data and user perspectives. Gui et al. [9], With the rise in uncrewed aerial vehicles and general aviation, effective management of air traffic flow is essential. Traffic flow prediction using ADS-B data is made possible by the high performance of LSTM models. Future research will concentrate on model combination accuracy and airport traffic statistics. Ma et al. [10] used online data analytics and machine learning for resource allocation and traffic prediction to make 5G networks more advantageous through proactive optimization. Contextualization and real-time analytics provide challenges.

Zhao et al. [11] combined machine learning (ML) and software-defined networking (SDN), and future networks hope to achieve intelligent designs. Resources, routing, and network security are all improved by ML in SDN. Offering a roadmap for multidisciplinary academics, this study explores ML algorithms in SDN networks and future possibilities. Casali et al. [12], with a focus on trends, obstacles, and potential future

Journal of Theoretical and Applied Information Technology
30th November 2024. Vol.102. No. 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

study areas, the article analyzes machine learning (ML) studies on geographical data. ML and AI, especially urban spatial analysis, hold great potential for sustainable city planning. Liu et al. [13] increased malware rapidly with Android applications. For Android malware detection, machine learning is essential. This review includes a range of machine learning (ML) methodologies, from data collection to assessment, and highlights further research areas. Hussain et al. [14] used DL and MEC to identify anomalies like outages and congestion, and 5G combined AI for network management. The outcome lowers OPEX and improves QoS by reducing false alarms and improving accuracy. Qolomany et al. [15] developed smart buildings that should be efficient and convenient, easily incorporating technology into day-to-day activities. Machine learning and extensive data analytics will make intelligent services possible, resolving present development obstacles.

Musumeci et al. [16] automated processes related to identification, detection, and prediction; machine learning (ML) holds promise for managing optical network failures. Using real-world examples, this lesson presents ML principles and applications. Future studies have to focus on flexibility and scalability to shift network circumstances. Abdulkareem et al. [17], fog computing can manage resources more accurately, securely, and with more precision with ML. This review examines the function and uses of ML, emphasizing its difficulties and potential. Fang et al. [18] state that 5G networks are more vulnerable to spoofing attempts because of the wide range of device connections. Reliable, effective, and ongoing security improvements are provided via machine learning-based authentication. Research on human-device interaction and anomaly detection is planned for the future. Cayamcela and Lee [19], though machine learning is essential for network efficiency, 5G is critical for various applications, including e-health and driverless cars. Examining problems and potential future directions, this article investigates machine learning (ML) solutions for 5G and Beyond 5G (B5G) networks. Herrera et al. [20] analyzed the effect of ICTs on socioeconomic indicators using neural networks and XGBoost using SHAP values. The findings show that while gender disparity in Brazil still exists, age and education have a favorable impact on income, necessitating focused governmental efforts. Akrami et al. [21] revealed the important study themes and trends, focusing on blockchain technology, AI-powered 5G, and IoT through a bibliometric analysis of 700 papers. Blockchain and machine learning are two powerful technologies that might be integrated meaningfully. From the review of the literature, it was observed that there was little research on call drop analytics in the telecom industry using machine learning models.

3. PRELIMINARIES AND THEORITICAL ANALYSIS

3.1 Handoffs

Handoff refers to when a call is being processed and the channel associated with the established connection changes. Handoffs frequently occur due to signal degradation in the current traffic channel or cell boundary crossing. When a mobile user moves from one cell to another during the course of a conversation, the call is transferred from one base station to a new base station. This is another definition of handoffs. Soft handoffs and hard handoffs are the two types of handoffs that occur. Hard handoff occurs when a mobile station acquires a new channel before the traffic/communication channel is relinquished, losing an ongoing call. The major cause of this hard handoff is that the base stations are either extremely far away or busy, meaning there isn't a channel open in the base station to allow the conversation to continue.

3.2 Monitoring Call Drops

Cell splitting is one of the finest methods available to reduce call dropouts caused by hand-off. Numerous other techniques exist, including dynamic channel allocation and cell sectoring.

3.2.1 Process of Cell Splitting

The process of dividing a large, overcrowded cell into several smaller cells, each with its own base station and a commensurate decrease in transmitter power and antenna height, is known as cell splitting. If each cell's coverage area is divided or reconfigured to make room for micro and picocells. However, there can be more handoffs and handovers after using this technique. However, in this case, the handover would go extremely well since the cells' coverage regions would now overlap, resulting in a satisfactory decrease in the rate of call dropouts. The cell splitting approach simply scales up the geometry of the cell's architecture, increasing the number of handoffs while maintaining the cell's design cells that, in turn, would raise the number of clusters, implying a rise of reuse channels, which would lower the likelihood of call dropouts and raise cell capacity. One of the main benefits of using the cell-splitting technique is that power transmitted in the smaller cells created by redesigning the more giant cells is significantly less than power transmitted in the more giant cells. This is because there is now much less power wasted in the larger, older cells than in the smaller, newly created cells. It is important to note that there will be a reduced transmission power, as expressed in Eq. 1.

$$\frac{P_{transmitted\ in\ small\ cell}}{P_{transmitted\ in\ large\ cell}} = \left(\frac{R_{small\ cell}}{R_{large\ cell}}\right)^n \qquad (1)$$

Where 'n' is the path loss exponent.

3.2.2 Concept of Cell Sectoring

Most operators have installed omnidirectional antennas, frequently providing signals with weak intensities. By using directional antennas to check for interference and frequency reuse of channels, sectoring entails swapping out these omnidirectional antennas for sector antennas, which generate beams of higher/stronger intensities. There are two main methods for cell sectoring: the first involves using six single-directional antennas, each covering 60 degrees of the cell; the second method uses three single-directional antennas, each covering 120 degrees of the cell. Theoretically, segmenting the cell into sectors would increase the network's ability to handle calls as the channels previously assigned to the cell are now split among several sectors, raising the handoff rate. The network's capacity has increased by lowering the quantity of co-channel cells that interfere. Instead of interfering with channels in all directions, as is the case these days, the channels only interfere with one, two, or three channels with the same transmission angle as theirs. The primary benefit of utilizing this method is a notable reduction in the signal-to-interference ratio (SIR). The signal emitted from the tower may have alien transmissions interfering with it, as shown by the SIR factor. The quality of the services rendered and the frequency of call dropouts will decrease with decreasing SIR. The formula for SIR is expressed in Eq. 2.

$$\frac{S}{I} = \frac{R^{-n}}{\sum_{i=1}^{i_0} D_i^{-n}} \qquad (2)$$

where S denotes the signal's intensity or strength, I denotes the amount of interference that occurred, "D" for the distance to the center of the closest co-channel cell, "R" for the cell's radius, and i_0 for the total number of co-channel interfering cells. The formula above illustrates how the S to I ratio falls; thus, the call drop reduces as the distance between the cell centers grows. Additionally, a specific number of traffic channels should be set aside for handover management, as suggested in [22]. The presentation thoroughly examined how effective handoff management, sectoring, and cell splitting contribute to the expansion of radio resources.

3.2.3 Probability Analysis

Assume that we must calculate the correlation between the quantity of accessible traffic channels and call losses. The Erlang B Formula may thus be used to estimate it relatively quickly [23], [24].

Let 'B' denote the probability of loss, 'A' the provided traffic intensity in Erlang, and 'N' the total number of accessible channels. Then Erlang B Formula is as expressed in Eq. 3.

$$B = \frac{\frac{A^N}{N!}}{\sum_{k=0}^{N} \frac{A^k}{k!}} \qquad (3)$$

The chance of a drop call lowers as 'N' grows. Numerous other formulas used in probability analysis have been published in various literature [23], [24]. These formulas demonstrate that the probability of a dropped call decreases as the number of channels increases, indicating that cell splitting and sectoring will significantly increase the efficiency of the handoff mechanism in GSM. Also, the call drop rate can be calculated as in Eq. 4.

$$\text{Drop call rate} = \frac{\text{Number of dropped calls}}{\text{Number of call attempts}} \qquad (4)$$

Assume that the Poisson Probability Function with a discrete variable may be used to determine the Drop call probability about the call duration in the following way:

$$P(Y = n) = \frac{(tv_d)^n}{n!} e^{-tv_d}, n \geq 0 \qquad (5)$$

Where "t" is the time interval, v_d is the call drop rate, "n" is the confirmed call dropped, and "Y" is a random variable used to tally the number of drops [23].

3.2.4 Dynamic Channel Allocation

Few service providers worldwide have used the very effective technique of Time Division Multiple Access (TDMA) based dynamic channel allocation in heavy traffic scenarios to lower the chance of call drops in their networks. The likelihood of call dropouts due to handover in situations with heavy traffic is assessed for the performance. A bandwidth window is employed under dynamic channel allocation, and its size varies based on the traffic circumstances that are in place. Due to this phenomenon, the lowest priority calls receive minimal bandwidth, which lowers the likelihood of handover call drops. The highly prioritized and real-time handover calls, often voice and multimedia calls, are given the required bandwidth. The possibility of co-channel interference is one of the main issues with this strategy. This phenomenon, which could lead to two adjacent cells using the same channel at the same time, can be prevented by ensuring that any channel that is currently in use by one cell can only be reallocated to another cell simultaneously if and only if the two cells are separated by a distance of 'd,' where 'd' is defined as in Eq. 6.

$$d = D/R \qquad (6)$$

where the physical distance between the two cell centers is denoted by "D," and the cell's radius is represented by "R."

3.2.5 Hybrid Channel Allocation

Hybrid Channel Allocation is an unusual yet highly successful strategy to reduce the number of call dropouts. This technique's name accurately indicates a blend of many distinct procedures. This may be characterized as a carefully thought-out plan to distribute traffic channels using Fixed Channel Allocation (FCA) and Dynamic Channel Allocation (DCA), two distinct techniques. The hybrid channel allocation approach successfully lowers both call blocking and call dropping by taking into account the DCA technique for managing handoff calls and the FCA method for allocating channels to incoming calls.

4. PROPOSED SYSTEM

We proposed a machine learning-based framework for call drop analytics using data from the telecom industry of several telecom operators. This section presents the problem, framework, algorithm, and evaluation methodology.

4.1 Problem Definition

Provided test data from the telecom industry, developing a machine learning framework that performs call drop analytics and prediction is a challenging problem.

4.2 Our Framework

We proposed a framework based on machine learning for analyzing telecom industry data and moving it towards call drop analytics. The framework is shown in Figure 1. It provides exploratory data analysis, dimensionality reduction, univariate and bivariate analysis, and machine learning for call drop analytics. The telecom industry data is subjected to exploratory data analysis to arrive at details pertaining to data distribution. If the data is imbalanced, it will get balanced with a tool like SMOTE. The machine learning process involves several models that follow a supervised learning process for call drop analysis.

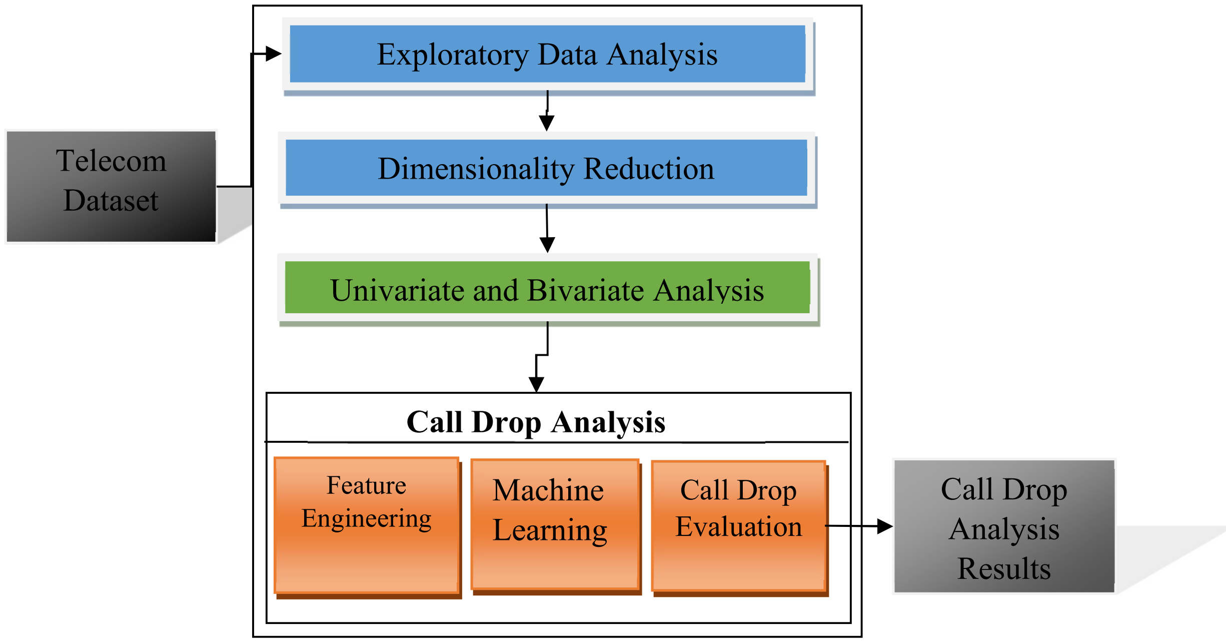

***Figure 1:** Proposed ML-based framework for call drop analysis*

The data is explored with univariate and bivariate analysis with a single feature and multiple features. Correlation analysis is preferred when the data is continuous, while the Chi-Square test is preferred for categorical data. Several machine learning models analyze call-dropping patterns in the given dataset. The dataset has several features, but we focused on the call drop category as this paper is meant to analyze call dropped dynamics. The framework takes care of feature scaling, encoding for categorical data, missing values, and outliers. After future engineering, the selected features are given to machine learning models to analyze call drop data and predict possible outcomes. The machine learning models used in this research are evaluated using confusion matrix-related performance metrics.

4.3 Feature Selection and Dimensionality Reduction

Feature selection is an integral part of this research as it could find the importance of features with the help of a correlation matrix. Feature selection helps improve the training quality of the machine learning models. Dimensionality reduction is involved in the current research and is meant to enhance data quality. Principal component analysis is the technique used to achieve dimensionality reduction. The process of dimensionality reduction exploits different numbers of clusters. The main focus of this paper is call-dropping-related data analytics, as this research is intended to analyze call-dropping exhibited by other telecom operators. Dimensionality reduction improves quality in the training process.

4.4 Machine Learning

Several machine learning models have been used in this research. These machine learning models are part of the proposal framework for call drop analysis.

K-Nearest Neighbour: Often employed in missing value imputation, KNN [25] is a straightforward supervised machine learning (ML) technique that may be used in regression or classification applications. Its premise is that we may categorize unanticipated points according to the values of the closest existing points, as the observations closest to a particular data point are the most "similar" observations in a data collection. The user can choose the number of adjacent observations to be used in the method by selecting K.

Support Vector Classifier: The Support Vector Machine algorithm is especially implemented in SVC [26], which is made for classification jobs. To put it another way, SVC is a classification-specific SVM. The hyperplane that divides the data points into distinct classes the best is what it

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

looks for. When someone refers to an "SVC," they typically mean the algorithm's classification variation; however, the words "SVC" and "SVM" are occasionally used interchangeably.

Gaussian Naive Bayes: A machine learning classification method called Gaussian Naive Bayes [27]is predicated on a probabilistic methodology and assumes that each class has a normal distribution. It assumes that every parameter can independently predict the output variable. It can forecast the likelihood that a dependent variable will fall into each group.

Decision Tree: A decision tree is an interpretable, flexible technique for predictive modelling in machine learning [28]. It may be used for both classification and regression tasks since it organizes judgments according to the data supplied.

Random Forest: An ensemble of decision trees trained using a particular kind of random noise is called a Random Forest (RF) [29]. The decision tree ensemble type that is most widely used is random forests. This paper covers several methods for constructing independent decision trees to increase the likelihood of developing a successful random forest.

4.5 Proposed Algorithm

We proposed an algorithm known as learning-based Call Drop Analytics (LbCDA), which exploits feature selection and trains multiple classifiers for call drop analytics.

Algorithm 1: Learning-based Call Drop Analytics (LbCDA)
Input: Telecom dataset D, ML models M (KNeighbors, SVC, GuassianNB, Decision Tree, Random Forest)
Output: Call drop analysis results T, performance statistics P

1. Begin
2. F←FeatureSelection(D)
3. (T1, T2)←DataSplit(D)
4. For each model m in M
5. Train m with F and T1
6. End For
7. For each model m in M
8. R←Predict(m, T2)
9. P←Evaluate(R, ground truth)
10. Print R
11. Print P
12. End For
13. End

Algorithm 1: Learning-based Call Drop Analytics (LbCDA)

As presented in Algorithm 1, it takes a telecom industry dataset and a machine learning pipeline as input. It performs call drop analytics to find the call drop patterns of each operator. It follows a multi-class classification where the call drop category has three classes: satisfactory, call dropped, and poor voice quality. The algorithm has provision for feature selection, which is essential for improving the quality of training. It has an iterative process to train all the machine learning models in the pipeline, resulting in the learned models. All the models that have been trained are used in another iterative process to analyze the test data to find patterns concerning call dropping. A.lgorithm follows a learning-based approach, which has the potential to understand the training data and then perform its functionality more efficiently. The feature selection processor enables the selection of contributing features toward leveraging performance by improving the quality of training.

4.6 Dataset Details

The empirical study uses a telecom dataset with details of the number of operators. The dataset has several features, as illustrated in Figure 2. It has details like type of network, signal strength, and call drop category Related to upload and download speed, among other demographic details.

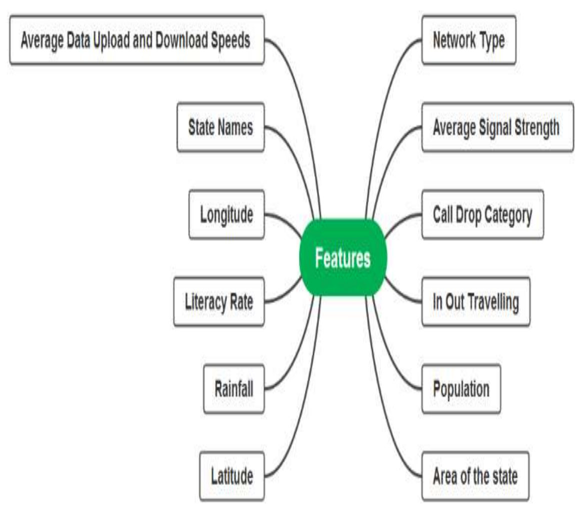

Figure 2*: Features in the dataset*

The features in the dataset are used in exploratory data analysis and machine learning to analyze call drop dynamics across different operators in the

Journal of Theoretical and Applied Information Technology
30th November 2024. Vol.102. No. 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

telecom industry. The "operator" indicates the telecom company, while "in/out traveling" indicates whether the call is inbound or outbound. "Network type" indicates the type of network, and the "rating" relates to the level of service in the company. The "caller category" contains low, high, and medium values. "Latitude" indicates the geographical latitude, and "longitude" indicates the geographical longitude, while "state name" indicates the region of the country where the data is collected. The "average data speed download" and the "average signal strength download" indicate the related details in the given area, while "population" indicates the population in the given state. "Literacy rate" indicates the literacy of the population, "rainfall" indicates rainfall statistics, and the "area of the state" indicates the geographical location of the state.

4.7. Evaluation Procedure

Since we used a learning-based approach (supervised learning), metrics derived from the confusion matrix, shown in Figure 3, are used to evaluate our methodology.

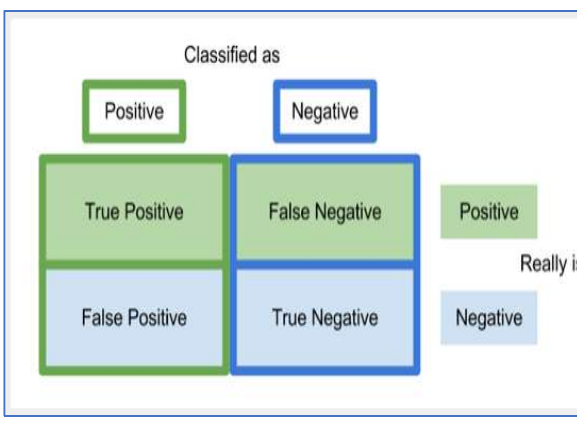

Figure 3: *Confusion matrix*

Based on the confusion matrix, the predicted labels of our method are compared with the ground truth to arrive at performance statistics. Eq. 1 to Eq. 4 express different metrics used in performance evaluation.

$$\text{Precision (p)} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall (r)} = \frac{TP}{TP+F} \quad (2)$$

$$\text{F1-score} = 2 * \frac{(p * r)}{(p+r)} \quad (3)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

The measures used for performance evaluation result in a value that lies between 0 and 1. These metrics are widely used in machine learning research.

5. EXPERIMENTAL RESULTS

This section presents experimental results regarding call drop analytics using telecom industry data covering various operators in the industry. The results are presented in terms of exploratory data analysis, feature engineering, dimensional reduction and performance of the RF model, and performance comparison among many machine learning models.

5.1 Exploratory Data Analysis

This section presents exploratory data analysis reflecting data distribution among operators, different call drop categories, and univariate and bivariate data analysis.

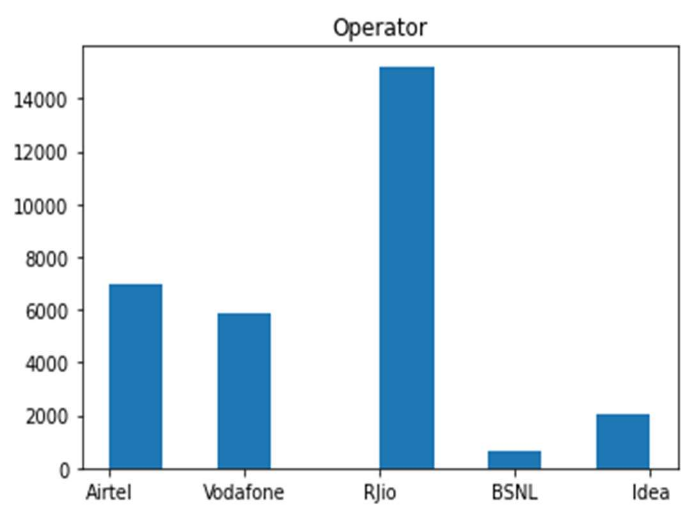

Figure 4: *Operator-wise data distribution*

As presented in Figure 4, data distribution dynamics in the given dataset are visualized for different telecom operators.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

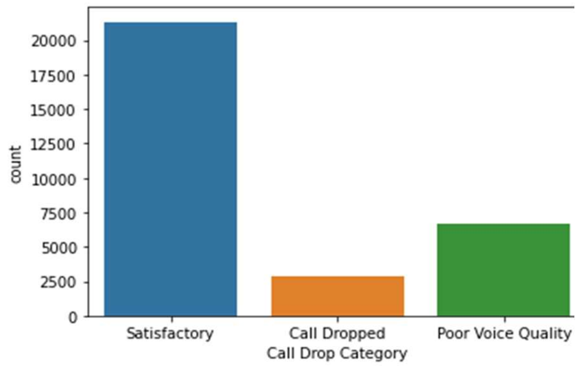

Figure 5*: Data distribution across the three call-dropping classes*

Figure 5 presents different call drops, such as satisfactory, call dropped, and poor voice quality, with the data distribution dynamics.

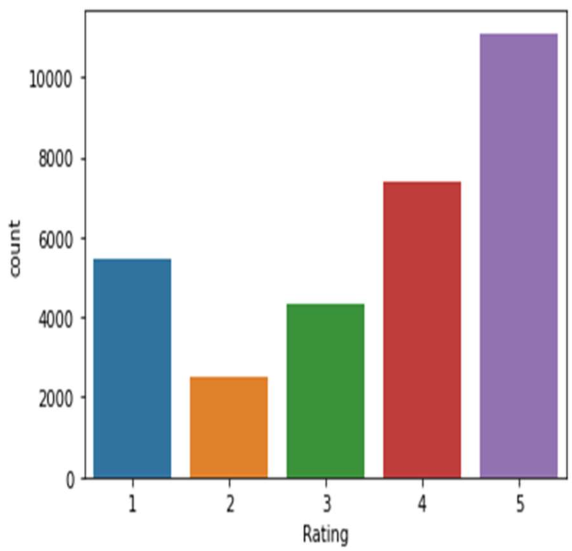

Figure 6: *Shows rating distribution of different operators*

The rating distribution across different telecom operators is provided, as shown in Figure 6.

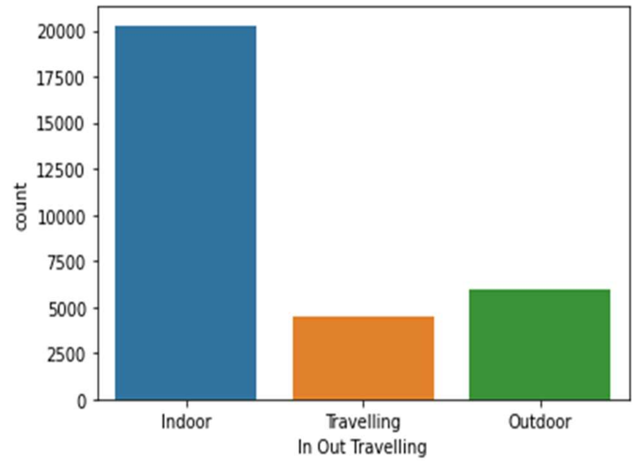

Figure 7: *Shows usage dynamics in our traveling*

The in-out traveling associated with the telecom industry is presented in terms of indoor, traveling, and outdoor, as shown in Figure 7.

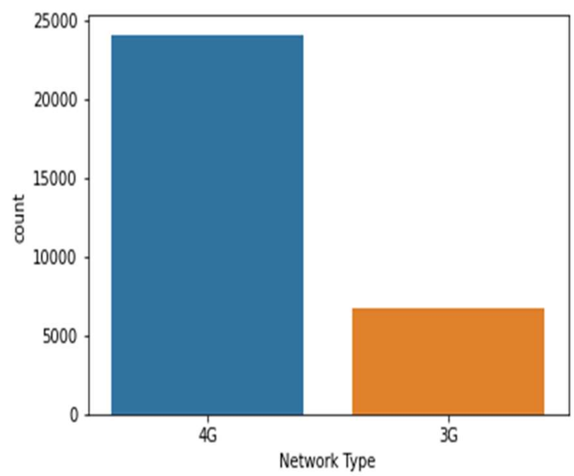

Figure 8: Shows distribution of data in terms of network type

Figure 7 shows the two categories of network type and the data distribution of these two types in the given dataset.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

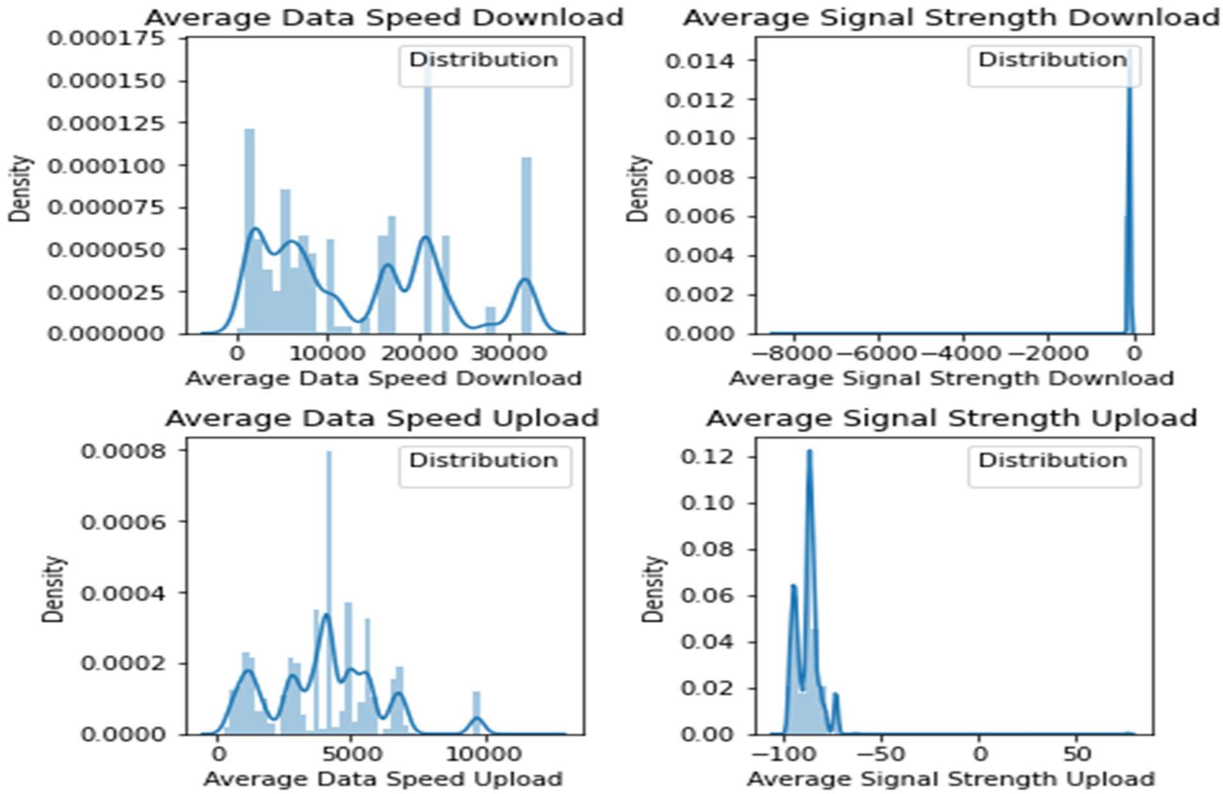

Figure 9: *Distribution of signal variables*

Figure 9 shows different signal variables and their details in terms of average data speed, download average signal strength, download average data speed upload, and average signal strength upload.

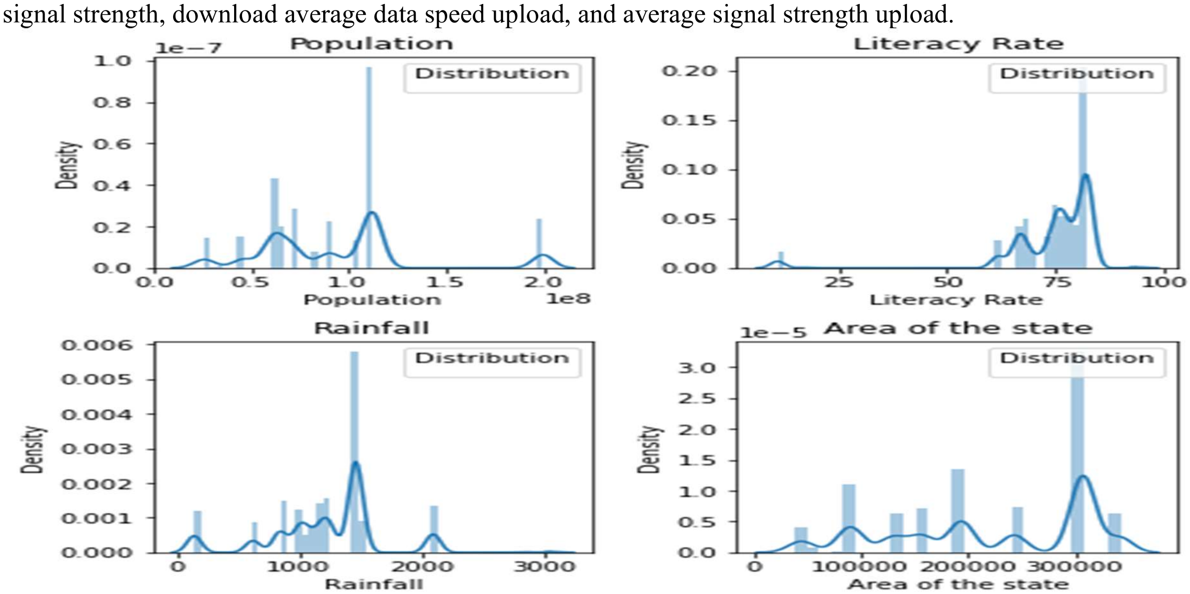

Figure 10: *Distribution of state variables*

As presented in Figure 10, different state variables are used for data analytics. The variables like population, literacy rate, rainfall, and the area of the state are provided.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

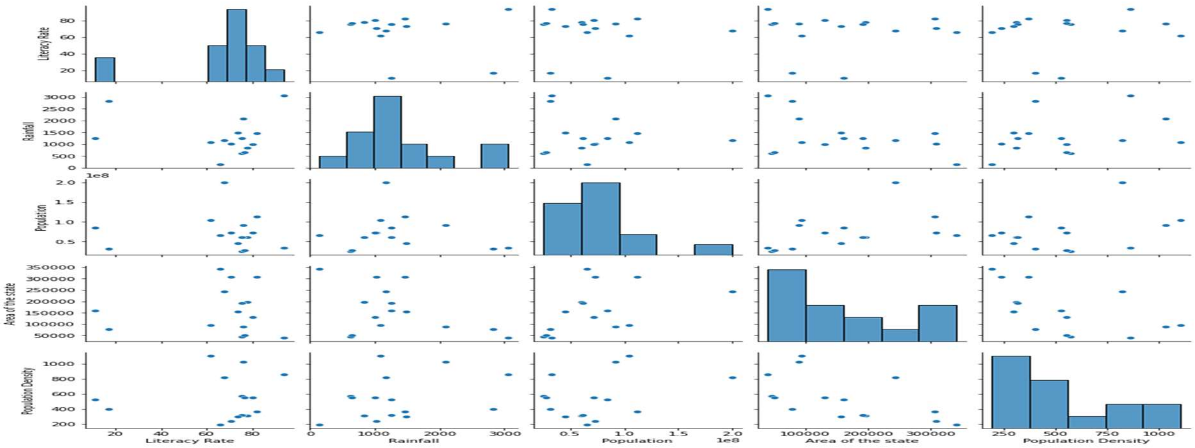

***Figure 11:** Bivariate analysis of data*

As presented in Figure 11, bivariate data analysis considers different variables, such as literacy rate, rainfall, population, area of the state, and population density.

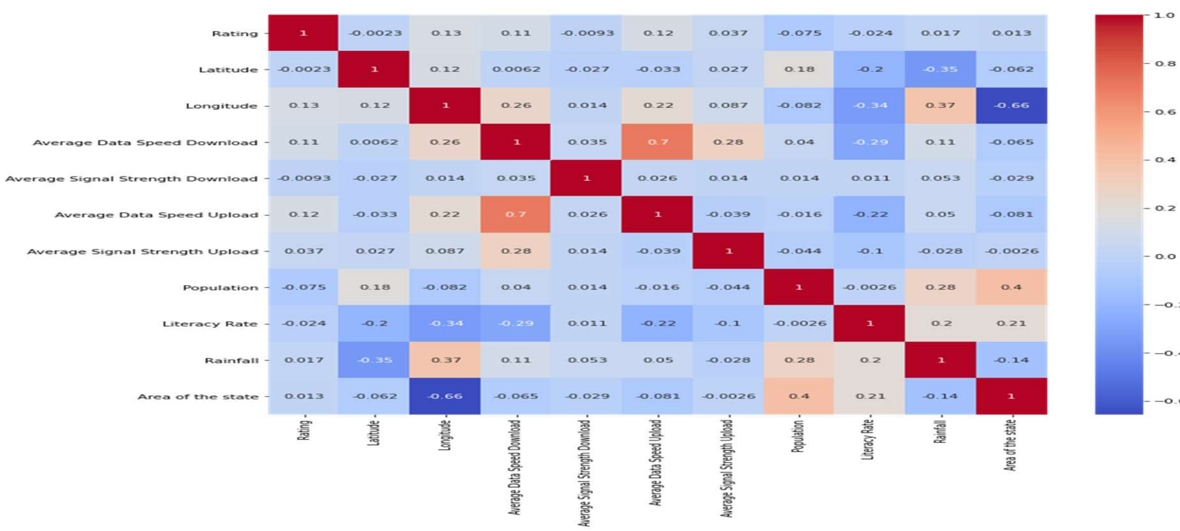

***Figure 12**: Correlation analysis of different variables*

Figure 12 presents the correlation analysis among the different variables available in the dataset. It shows the correlation details visualized in a heat map, which helps in understanding how each variable is correlated with all other variables.

5.2 Dimensionality Reduction

This section presents dimensionality reduction results associated with Principal Component Analysis (PCA).

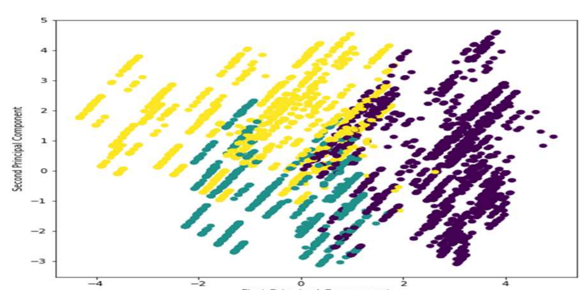

Figure 13: *Visualization of principal component analysis with three clusters*

As presented in Figure 13, the result of the principal component analysis with three clusters is visualized in the dimensional reduction process.

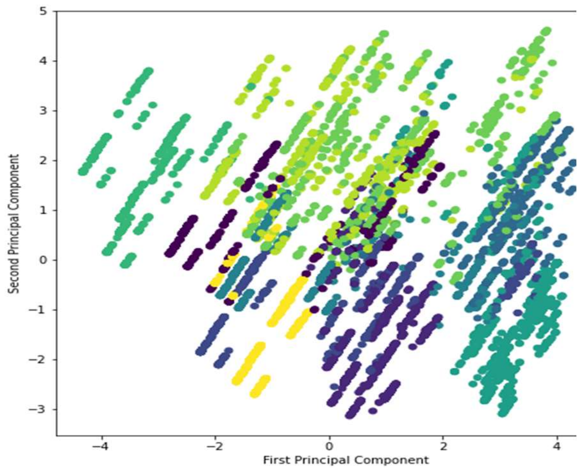

Figure 14: *Visualization of principal component analysis with 5 clusters*

As presented in Figure 14, the result of the principal component analysis with five clusters is visualized in the dimensional reduction process.

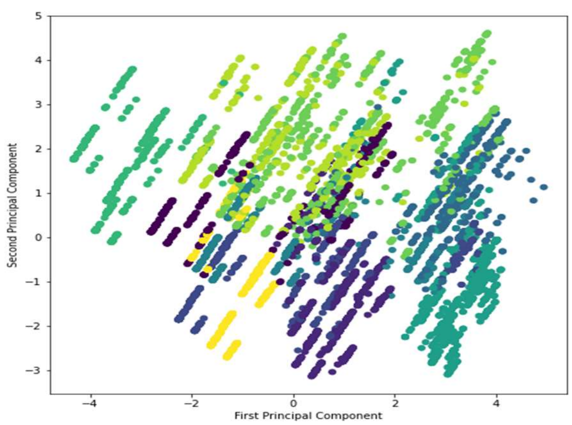

Figure 15: *Visualization of principal component analysis with 10 clusters*

As presented in Figure 15, the result of the principal component analysis with ten clusters is visualized in the dimensional reduction process.

5.3 Results of Classification Models

This section presents the results of experiments in terms of classifying call drop analytics results into three classes: satisfactory, call dropped, and poor voice quality exhibited by the RF model.

Table 1: *Different classes associated with the dropping feature*

Id	Class Drop Name
0	Satisfactory
1	Call Dropped
2	Poor Voice Quality

Table 1 shows different classes involved in the proposed multi-class classification of call drop analytics. Three classes are involved in the call drop category: satisfactory, call dropped, and poor voice quality.

Table 2 *Performance of random forest in Caldera Analytics*

Call Drop category Feature			
	Precision	**Recall**	**F1-Score**
Satisfactory	0.82	0.74	0.77
Call Dropped	0.81	0.86	0.83
Poor Voice Quality	0.88	0.91	0.89

Table 2 shows the performance of the RF model in call drop analytics for all three classes in terms of different performance metrics.

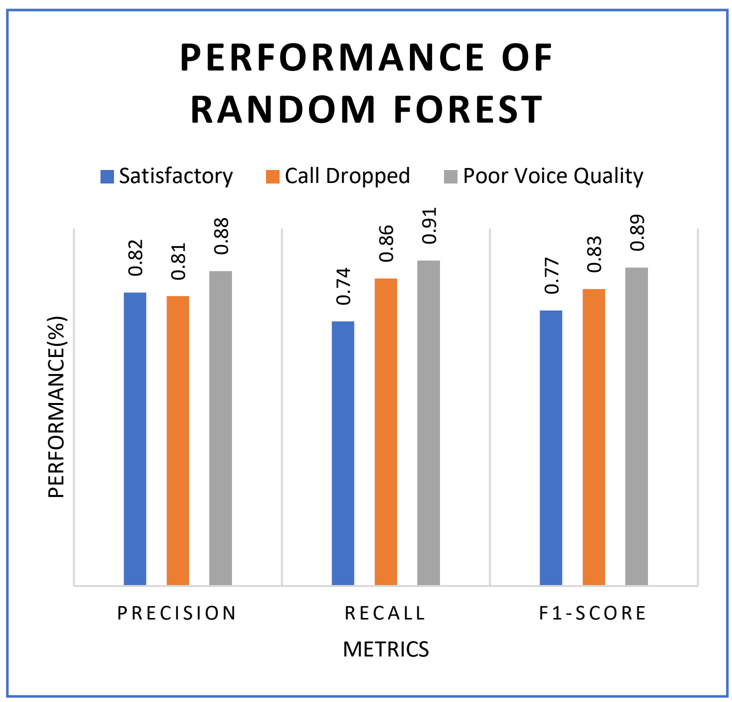

Figure 16: *Results of call drop analysis using RF model*

The performance of the RF model for all three related classes is provided, as shown in Figure 16. The results are provided for the given dataset related to telecom data. Concerning precision, the satisfactory class shows 82%, the call dropped class shows 81%, and the poor voice quality class shows 88%. Concerning recall measures, the RF model showed 74% for the satisfactory class, 86% for the call-drop class, and 91% for the poor voice quality class. Concerning F1-score measure, the RF model exhibited 77% for satisfactory class, 83% for call dropped class and 89% for poor voice quality class.

Table 3: *Performance of RF model with undersampled dataset*

Prediction through an Undersampled dataset			
	Precision	**Recall**	**F1-Score**
Satisfactory	0.82	0.8	3330.81
Call Dropped	0.85	0.86	0.85
Poor Voice Quality	0.87	0.89	0.88

The results of call drop analytics using RF for undersampled data are presented in Table 3 for all coal drop category classes.

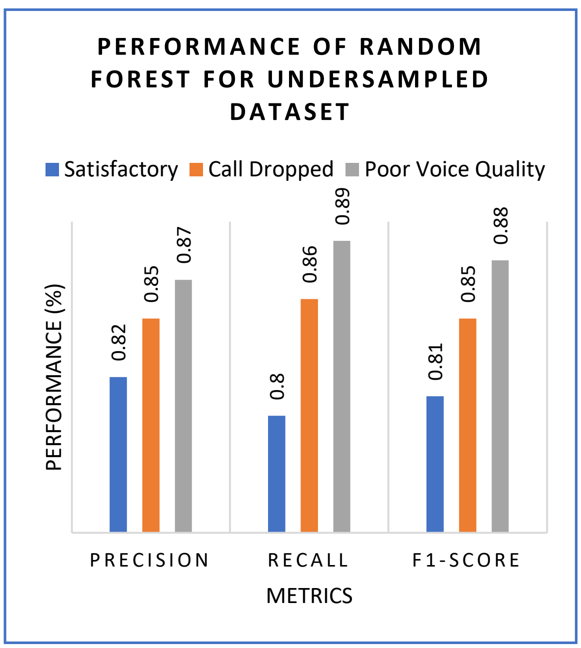

Figure 17: *Performance of RF model in call drop analysis with undersampled dataset*

The performance of RF for an undersampled dataset concerning call drop analytics and the classification of the data into three call drop categories, such as satisfactory, the call dropped, and poor voice quality, is presented in Figure 7.

Concerning the satisfactory category, the RF model achieved 82% precision, 80% recall, and 81% F1 score. Concerning the call dropped category, RF achieved 85% precision, 86% recall, and 85% F1 score. The RF model achieved 87% precision, 89% recall, and 88% F1 score regarding poor voice quality.

Table 4: *Performance of RF model with an oversampled dataset*

Prediction through the Oversampled dataset			
	Precision	**Recall**	**F1-Score**
Satisfactory	0.97	0.67	0.77
Call Dropped	0.89	0.88	0.88
Poor Voice Quality	0.83	0.98	0.93

Table 4 shows the results of the RF model in call drop analytics for all three classes in terms of precision, recall, and F1 score. These results are observed with oversampled data.

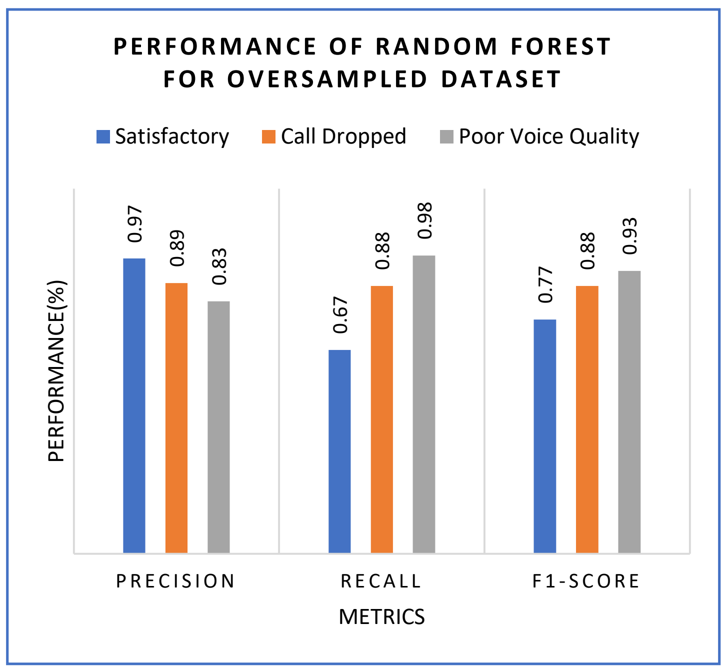

Figure 18: *Performance of RF model in call drop analytics with oversampled dataset*

The results of the RF model in call drop analytics using an oversampled dataset for all three classes

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

are provided in Figure 18. The RF model achieved 97% precision, 67% recall, and 77% F1-score for a satisfactory class. The model could achieve 89% precision, 88% recall, and 88% F1-score for call dropped class. In the same fashion, the RF model could achieve 83% precision, 98% recall, and 93% F1-score for poor voice quality class

5.4 Performance Comparison

This section compares performance among different ML models used for call drop analytics.

Table 5: *Performance of comparison among different models*

Accuracy Comparison	
Models	Accuracy
KNeighbors	0.836
SVC	0.849
GuassianNB	0.851
Decision Tree	0.832
Random Forest	0.874

Table 5 shows the accuracy of various ML models in call drop analytics.

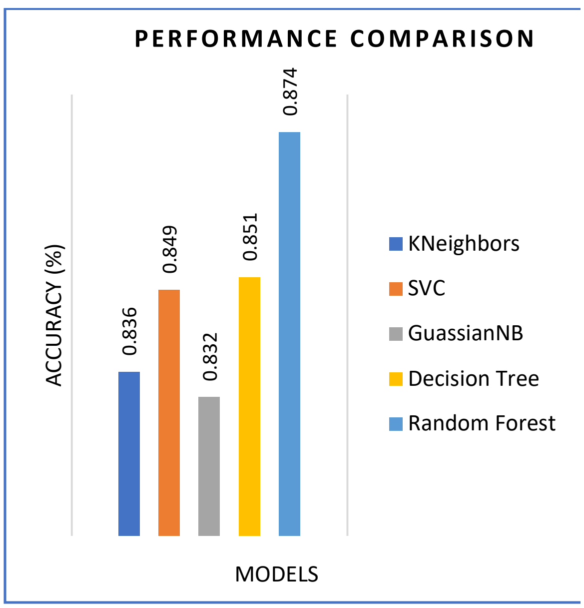

Figure 19: *Performance comparison in call drop analytics among different machine learning models*

As presented in Figure 19, different ML models are compared to analyze their performance in terms of accuracy for call drop analytics. Each model provided a different level of performance due to their architecture and mechanisms in discriminative procedures. The KNeighbors model could achieve 83.60% accuracy, SVC 84.90%, GaussianNB model 83.20%, DT model 85.10%, and RF model 87.40%. The results show that the GaussianNB model showed the lowest performance, with 83.20% accuracy. RF model achieves the highest accuracy with 87.40% accuracy in call drop analytics.

6. CONCLUSION AND FUTURE SCOPE

We proposed an ML framework for automatic analysis of all drops in the telecom industry across all operators. The framework also supports optimizations like future engineering and dimensionality reduction to improve the performance of machine learning models. Featuring engineering is finding the best-contributing features of call drop analytics in the telecom industry. It will help in improving the quality of training in supervised learning. On the other hand, dimensionality reduction is used to improve the dataset and leverage performance in training. Machine learning models are used to learn from the data and make predictions about call drops in the telecom industry by different operators. We proposed an algorithm called Learning based Call Drop Analytics (LbCDA) which exploits feature selection and training multiple classifiers towards call drop analytics. With benchmark dataset variants of the telecom industry, our empirical study has revealed that our framework's Random Forest (RF) model outperforms other models with the highest accuracy of 87.40%. The proposed system can positively impact the telecom industry as it can understand call drop problems under defy problems and serve consumers better. The proposed system has potential challenges, including a lack of sufficient data, and there is a need to improve the models further with hyperparameter tuning besides using optimization techniques. In the future, we intend to improve our framework with other optimization techniques like hyperparameter tuning and deep learning models for call drop analytics in the telecom industry.

REFERENCES

[1] Praveen Lalwani;Manas Kumar Mishra; Jasroop Singh Chadha; Pratyush Sethi; (2021). Customer churn prediction system: a machine learning approach .

Computing, pp.1 –24, . doi:10.1007/s00607-021-00908-y.

[2] Ullah, Irfan; Raza, Basit; Malik, Ahmad Kamran; Imran, Muhammad; Islam, Saif ul; Kim, Sung Won (2019). A Churn Prediction Model using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. IEEE Access, 7,pp.60134–60149. doi:10.1109/ACCESS.2019.2914999.

[3] Oryza Wisesa;Andi Adriansyah;Osamah Ibrahim Khalaf; (2020). Prediction Analysis Sales for Corporate Services Telecommunications Company using Gradient Boost Algorithm . 2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP),pp.101-106,. doi:10.1109/BCWSP50066.2020.9249397.

[4] Polese, Michele; Jana, Rittwik; Kounev, Velin; Zhang, Ke; Deb, Supratim; Zorzi, Michele (2020). Machine Learning at the Edge: A Data-Driven Architecture with Applications to 5G Cellular Networks. IEEE Transactions on Mobile Computing, pp. 1–16,. doi:10.1109/TMC.2020.2999852.

[5] Jasneet Kaur;M. Arif Khan;Mohsin Iftikhar;Muhammad Imran;Qazi Emad Ul Haq; (2021). Machine Learning Techniques for 5G and Beyond . IEEE Access, 9, pp.23472-23488, doi:10.1109/access.2021.3051557 .

[6] Jain, Hemlata; Khunteta, Ajay; Srivastava, Sumit (2020). Churn Prediction in Telecommunication using Logistic Regression and Logit Boost. Procedia Computer Science, 167, pp.101–112. doi:10.1016/j.procs.2020.03.187 .

[7] Park, Choonkil; Khater, Mostafa M.A.; Abdel-Aty, Abdel-Haleem; Attia, Raghda A.M.; Rezazadeh, Hadi; Zidan, A.M.; Mohamed, A.-B.A. (2020). Dynamical analysis of the nonlinear complex fractional emerging telecommunication model with higherâ order dispersive cubicâ quintic. Alexandria Engineering Journal, pp.1-9, S1110016820301617–. doi:10.1016/j.aej.2020.03.046

[8] Changhun Lee, Chiehyeon Lim. (2021). From technological development to social advance: A review of Industry 4.0 through machine learning. Elsevier. 167, pp.1-12. [Online]. Available at: https://doi.org/10.1016/j.techfore.2021.120653 [Accessed 23 February 2024].

[9] Gui, Guan; Zhou, Ziqi; Wang, Juan; Liu, Fan; Sun, Jinlong Sun (2020). Machine Learning Aided Air Traffic Flow Analysis Based on Aviation Big Data. IEEE Transactions on Vehicular Technology, 69(5), pp.4817–4826. doi:10.1109/tvt.2020.2981959

[10] Ma, Bo; Guo, Weisi; Zhang, Jie (2020). A Survey of Online Data-Driven Proactive 5G Network Optimisation Using Machine Learning. IEEE Access, 8, pp.35606–35637. doi:10.1109/ACCESS.2020.2975004.

[11] Zhao, Yanling; Li, Ye; Zhang, Xinchang; Geng, Guanggang; Zhang, Wei; Sun, Yanjie (2019). A Survey of Networking Applications Applying the Software Defined Networking Concept Based on Machine Learning. IEEE Access, pp.1–21. doi:10.1109/ACCESS.2019.2928564

[12] Ylenia Casali, Nazli Yonca Aydin, Tina Comes. (2022). Machine learning for spatial analyses in urban areas: a scoping review. Elsevier. 85, pp.1-18. [Online]. Available at: https://doi.org/10.1016/j.scs.2022.104050 [Accessed 23 February 2024].

[13] Liu, Kaijun; Xu, Shengwei; Xu, Guoai; Zhang, Miao; Sun, Dawei; Liu, Haifeng (2020). A Review of Android Malware Detection Approaches based on Machine Learning. IEEE Access, pp.1–30,. doi:10.1109/ACCESS.2020.3006143

[14] Hussain, Bilal; Du, Qinghe; Zhang, Sihai; Imran, Ali; Imran, Muhammad Ali (2019). Mobile Edge Computing-Based Data-Driven Deep Learning Framework for Anomaly Detection. IEEE Access, 7, pp.137656–137667. doi:10.1109/access.2019.2942485

[15] BASHEER QOLOMANY, ALA AL-FUQAHA, AJAY GUPTA. (2019). Leveraging Machine Learning and Big Data for Smart Buildings: A Comprehensive Survey. IEEE. 7, pp.90316-90356. [Online]. Available at: Digital Object Identifier 10.1109/ACCESS.2019.2926642 [Accessed 23 February 2024].

[16] Musumeci, Francesco; Rottondi, Cristina Emma Margherita; Corani, Giorgio; Shahkarami, Shahin; Cugini, Filippo; Tornatore, Massimo (2019). A Tutorial on Machine Learning for Failure Management in Optical Networks. Journal of Lightwave Technology, pp. 1–15. doi:10.1109/JLT.2019.2922586

[17] Abdulkareem, Karrar Hameed; Mohammed, Mazin Abed; Gunasekaran, Saraswathy Shamini; Al-Mhiqani, Mohammed Nasser; Mutlag, Ammar Awad; Mostafa, Salama A.; Ali, Nabeel Salih; Ibrahim, Dheyaa Ahmed (2019). A Review of Fog Computing and Machine Learning: Concepts, Applications, Challenges, and Open Issues. IEEE Access, 7, pp.153123–153140. doi:10.1109/ACCESS.2019.2947542

[18] Fang, He; Wang, Xianbin; Tomasin, Stefano (2019). Machine Learning for Intelligent Authentication in 5G and Beyond Wireless Networks. IEEE Wireless Communications, 26(5), pp.55–61. doi:10.1109/mwc.001.1900054

[19] MANUEL EUGENIO MOROCHO-CAYAMCELA, HAEYOUNG LEE, WANSU LIM. (2019). Machine Learning for 5G/B5G Mobile and Wireless Communications: Potential, Limitations, and Future Directions. IEEE. 7, pp.137184-137206. [Online]. Available at: Digital Object Identifier 10.1109/ACCESS.2019.2942390 [Accessed 23 February 2024].

[20] Gabriel Paes Herrera, Michel Constantino, Jen-Je Su, Athula Naranpana. (2023). The use of ICTs and income distribution in Brazil: A machine learning explanation using SHAP values. Elsevier. 47, pp.1-16. [Online]. Available at: https://doi.org/10.1016/j.telpol.2023.102598 [Accessed 23 February 2024].

[21] NOUHAILA EL AKRAMI, MOHAMED HANINE, EMMANUEL SORIANO FLORES. (2023). Unleashing the Potential of Blockchain and Machine Learning: Insights and Emerging Trends From Bibliometric Analysis. IEEE. 11, pp.78879-78903. [Online]. Available at: Digital Object Identifier 10.1109/ACCESS.2023.3298371 [Accessed 23 February 2024].

[22] Agustina, J., Zhang, P., & Kantola, R., Performance evaluation of GSM handover traffic in a GPRS/GSM network. Proceedings of the Eighth IEEE Symposium on Computers and Communications. ISCC 2003. doi:10.1109/iscc.2003.1214113.

[23] Boggia, G., Camarda, P., & D'alconzo, A. (2007). Modeling of Call Dropping in Well-Established Cellular Networks. EURASIP J Wirel Commun Netw EURASIP Journal on Wireless Communications and Networking, 2007(1), 017826. doi:10.1155/2007/17826.

[24] Tarkaa, N., Mom, J., & Ani, C. (2011). Drop Call Probability Factors in Cellular Networks. International Journal of Scientific and Engineering Research, 2(10). Retrieved May 24, 2016.

[25] Ali, Najat; Neagu, Daniel and Trundle, Paul (2019). Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets. SN Applied Sciences, 1(12), 1559–. http://doi:10.1007/s42452-019-1356-9

[26] Debabrata Samanta; Ahmed H. Alahmadi; M. P. Karthikeyan; Mohammad Zubair Khan; Amit Banerjee; Goutam Kumar Dalapati and Seeram Ramakrishna; (2021). Cipher Block Chaining Support Vector Machine for Secured Decentralized Cloud Enabled Intelligent IoT Architecture . IEEE Access. http://doi:10.1109/access.2021.3095297

[27] Laura Cataldi; Lara Tiberi and Giovanni Costa; (2021). Estimation of MCS intensity for Italy from high quality accelerometric data, using GMICEs and Gaussian Naïve Bayes Classifiers . Bulletin of Earthquake Engineering. http://doi:10.1007/s10518-021-01064-6

[28] Sarker, Iqbal H.; Colman, Alan; Han, Jun; Khan, Asif Irshad; Abushark, Yoosef B. and Salah, Khaled (2019). BehavDT: A Behavioral Decision Tree Learning to Build User-Centric Context-Aware Predictive Model. Mobile Networks and Applications. http://doi:10.1007/s11036-019-01443-z

[29] Lovatti, Betina P.O.; Nascimento, Márcia H.C.; Neto, Álvaro C.; Castro, Eustáquio V.R. and Filgueiras, Paulo R. (2019). Use of Random forest in the identification of important variables. Microchemical Journal, 145, 1129–1134. http://doi:10.1016/j.microc.2018.12.028