Broadband Network Fault Prediction. Case Study in Palestinian Telecommunication Company

Rabee Qasem, Loai Bitawi, Mohannad Hendi

In today's world, Internet has become a major part of our daily life. Firms in this industry spends millions on network development and enhancement in order to deliver best service quality. However, network faults prevent firms from providing the best performance and cause firms to allocate large budgets annually spent on faults maintenance in order to prevent customer churn. Predicting faults using machine learning techniques will help firms to reduce maintenance cost, enhance customer experience, increase quality of service, and eventually reducing churn. In this paper, multiple ML algorithms are studied to find which of them will provide the best prediction performance. Using AUC and other evaluation metrics proofed that Random Forest and Gradient Boosting were the best algorithms to adopt. Gradient Boosting was able to identify the most important features that contributes to fault prediction.

Data Science | Telecommunication | Predictive Maintenance | Machine Learning

1. Introduction

The internet has become a life necessity like water and electricity, and with the investments in IOT and cloud computing it is becoming essential. Telecommunication as a business sector suffers from fierce competition, thus, telecommunication firms facilitate all their resources on maintaining the best service quality in addition to competitive prices.

One of the main challenges in telecommunication is to maintain the quality of service for customers and decrease the number of faults tickets by continuous maintenance of the network elements. Although periodic maintenance ensures the quality of service and the stability of the network, it suffers from high cost, especially with network expansion, and by the time it becomes infeasible. Telecommunication companies aim to minimize network faults by estimating the lifespan of network elements and periodically maintain or replace the elements. This process has its financial drawbacks since it is like replacing the whole network each time a lifespan occurs. However, after the revolution of automation and the invasion of data science in different business aspects, network maintenance transferred from reactive maintenance, where maintenance is performed when a specific network element experiences a fault, to proactively predict when a fault might occur and raise the flag to maintain the element before it becomes faulty.

Customers can experience network faults in many forms like noise, interruption on the service, or it might be worse like full disconnection of the service. Other reasons might be soft reasons like modem configurations or AAA authentication. All these reasons will result in a service that doesn't meet customer expectations and satisfaction, and this might result in customer churn and loss in companies' revenues.

This research aims to accurately help decision makers to know the network elements that need maintenance before they experience faults using a prediction model for network faults that might happen with prediction on the type of the fault that might occur in order to let them take proactive actions to achieve better service quality and customer satisfaction. Moreover, the research focuses on finding the highest possible accuracy for the prediction.

In this research, a prediction model that studies several technical line parameters and predicts within a time period if this line needs maintenance before it encounters network fault. The faults that are studied in this research are the network-related faults which require maintenance on the network elements. The parameters considered in this study are derived from literature discussion, and interviews with experienced technical engineers in the telecommunication sector in Palestine.

1.1 Service Terminology

There are two types of DSL technologies that are commonly used in Palestine: Asymmetric Digital Subscriber Line 2 (ADSL2), and Very-high-bit-rate Digital Subscriber Line 2 (VDSL2). After a connection is established, measurements of signal quality are collected like ping and signal to noise ratio (SNR) [1]. Ping is the process of sending and receiving a data packet and measuring the time needed for the packet to travel, while SNR is the comparison of the signal level with the background noise in any communication channel, and it is measured in dB [2]. The service is provided through two combined parties, the access and the internet. The internet access is given through copper lines to customers, and the internet service is provided through Internet Service Provider (ISP) through Broadband Remote Access Server (BRAS), which routes the traffic between ISP and customers. After closing the session, information is provided regarding the session duration, upload and download bytes, timestamps and other information [1].

1.2 Service Faults

Telecommunication companies receive customer complaints about the service through trouble tickets. The customer contacts the company to submit a ticket describing the issue he's experiencing. Date and time of the ticket is taken with reasoning of the ticket type through a predefined list of values. The technical team receives the ticket, analyzes the symptoms, reviews the service readings and takes actions. In other words, service problems are analyzed through two sets of information: the trouble tickets and network parameters and system logs.

However, due to rapid changes in technology and the evolution of data science, companies are moving from reactively solving the trouble tickets to proactively predict network faults and do the needed repairments. This would help companies to provide better services, which is translated into customer satisfaction and service reliability.

Many researches that predict network faults either by aggregating trouble tickets, or by time series reading of system logs, or both. They are discussed in this research and a proposed model that uses both data sources to predict the faulty network element within a prediction window, with a prediction of network fault type that might happen within the window.

In this research, several algorithms are discussed and compared through different evaluation metrics. The algorithms are applied to a dataset collected from the telecom company involved in the study. Data combines features from different systems including customer profile, network readings, system logs and others. The best two models are selected and optimized through Random search to provide the best accuracy and cross-validation scores.

2. Literature Review

There are many researches that adopt machine learning algorithms to predict network faults. A research conducted on behalf of Indonesia's Telco that used a statistical time series analysis method which is Autoregressive Integrated Moving Average (ARIMA) on network trouble tickets to predict the amount of network faults [3]. However, the research results in many equations that might be practically inapplicable since it is computationally heavy and requires huge data sets that might not be applicable at real time application, in addition to the fact that its prediction is on the short term. Also, it is an unstable method with respect to change in data[4].

On the other hand, other researchers used an aggregated 7-day window of sessions data, customer trouble tickets and system logs [1]. The proposed model used Random Forest and C5.0 algorithms to classify and predict network faults within a 7-day prediction window. The model results in high AUC values (0.82 on RF and 0.65 on C5.0) for the VDSL service. The mentioned methods provide better accuracy and can classify the predicted faults per type. The researchers used features from customer trouble tickets, customer internet usage, and signal quality measurements.

Other researchers used Naïve Bayes algorithms on features from customer trouble tickets, network trouble tickets, and system logs to classify the input features into groups, then used the groups to predict different network failures. The model accuracy was tested for different prediction windows, and achieved an overall accuracy above 72% for 1-day prediction window [5]. Although the accuracy is relatively high, it drops significantly when the prediction window is enlarged. In paper [6], the researchers used a linear regression model with hypothesis testing, and then applied a rule-based analysis model on the same features from the same sources as other mentioned researches [1], [5]. The model hypothesis that the pattern per each warning obtained prior to network fault is similar to all network equipment, and the second hypothesis is that when a cumulative number of warnings obtained, an equipment fault happens on the next day. The model achieved R2 approximately 0.99 when the prediction is 4 days prior the actual fault, while R2 drops to 0.66 when it is 9 days prior the actual fault. The mentioned hypotheses are valid.

3. Proposed Methodology

Fig. 1. proposed Methodology Workflow.

Figure 1 shows the proposed methodology that is being used in this research. At first, the data is collected from several systems within the telecom company. The data is aggregated over the historical period taken as an analysis period. Secondly, the data is labeled and a stratified sample is taken. Then, data is split into 80% for training and 20% for testing. Then, different types of algorithms are used and validated to select the best approach. Finally, the best model is selected and analyzed to understand the features that significantly affect fault prediction.

3.1 Data Collection

One week of customer's data was taken and analyzed. The customers involved in the study are the ones who have ADSL and VDSL technologies only. The data were collected from four main datasets, (I) Customer Profile, (II) System Logs, (III) Network Alarms, (IV) Line Technical Readings. Table 1 shows the features selected for the model.

Customer Profile	System Logs	Network Alarms (Count)	Line Technical Readings
Customer ID	SESSION_COUNTS	F-38/1	Attainable Downstream (Mbps)
Bandwidth (Mbps)	TERMINATE_SESSION_MINT	F-38/6	Attainable Upstream (Mbps)
Trouble Ticket Flag	NUMBER_FAULTS	F-38/15	Port Type (ADSL, VDSL)
	TIME_T_M	F-38/10	
	Download (MB)	F-38/4	
	Upload (MB)	F-38/12	

Table 1. Proposed Model Features

3.2. Data Aggregation and Sampling

In order to perform the prediction, and since data is generated in different time windows, for example, alarms are event-related, while attainable downstream can be collected every 10 minutes, so data aggregation is needed. The customers were labeled by whether they have submitted a fault or not within a week period, then a 7-day analysis window was taken before the collection week. Some features were aggregated by one or more measurements like sum, average, max, min, count, and standard deviation and inserted as new features. The resultant dataset is 33 features with 213K records. After removing highly-correlated features, the resultant is 18 features with 211.6K unique independent records.

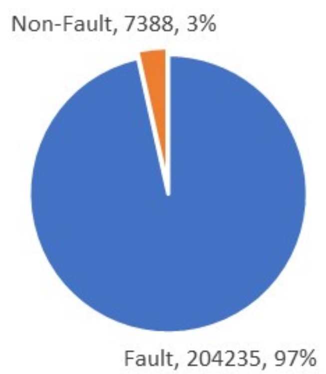

Fig. 2. Data Balance (Before downsampling)

Figure 2 shows the data labels resulted from the data exploration phase. It shows that data is imbalanced, and the classification algorithm would rather to classify the entry as non-fault. Due to that, we have to down sample the dataset to have 50%/50%. The final data is 14776 samples. The data is then split to train and test data. The ratio is 80% for training and 20% for testing. After this step, the data is ready for predictions.

3.3. Model Testing

In this section, 8 algorithms were applied to data with cross validation with 10 folds. the target is to predict whether a customer would experience a technical fault or not for the next 7 days, given the input parameters of his line through the past 7 days. The algorithms were (I) Logistic Regression, (II) K-Nearest Neighbor, (III) Gaussian Naïve Bayes, (IV) Support Vector Machine, (V) Linear SVC, (VI) Multi-layer Perceptron Classifier, (VII) Random Forest Classifier, and (VIII) Gradient Boosting Classifier. Several evaluation metrics were assessed in this research in order to choose the best algorithm that fits the data. The metrics involved confusion matrix, or Discriminator Metrics are a type of metrics that is used in binary classifiers, which compares the correctly classified entries with miss-classified ones, resulting in a set of measures like accuracy, precision, recall and F-1 measure. Another metric used is Area Under the Curve (AUC), which is popular for comparing classifiers since it measures the overall performance of the classifier. It is known that AUC is a better metric for classifiers than accuracy measurements [7].

4. Result and Analysis

4.1. Model Evaluation and selection

Figure 3 shows the Accuracy score for each algorithm. Random Forest Classifier and Gradient Boosting Classifiers provide the best cross-validation scores of 0.59 and 0.61 respectively.

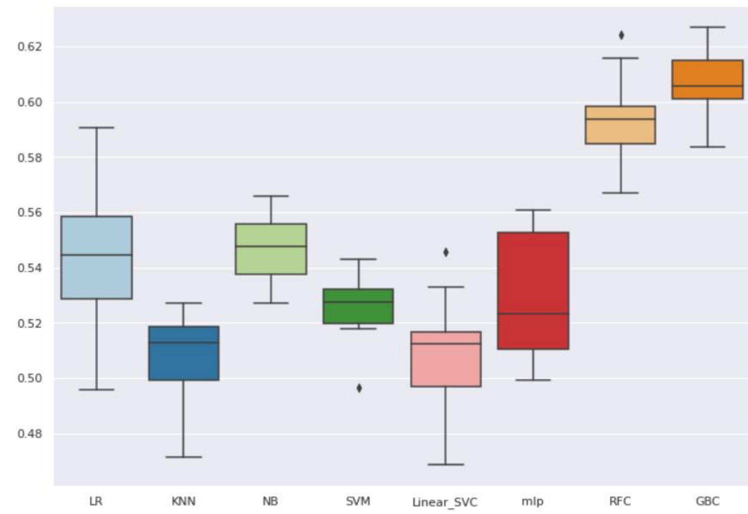

Fig. 3. Accuracy Scores Box Plot

Table 2 shows the average accuracy, Precision, recall, F-1 measure, and AUC scores for all tested models.

Algorithm	Class	Precision	Recall	F1-Score	Accuracy	AUC
Logistic Regression	No faults	0.67	0.01*	0.01*	0.49	0.5
	Faults	0.49	1	0.66		
KNN	No faults	0.53	0.52	0.52	0.51	0.51
	Faults	0.5	0.51	0.5		
Gaussian Naïve Bayes	No faults	0.55	0.86	0.67	0.57	0.56
	Faults	0.64	0.26*	0.37*		
SVM	No faults	0.53	0.52	0.53	0.52	0.52
	Faults	0.51	0.52	0.52		
Linear SVC	No faults	0.51	0.93	0.66	0.51	0.51
	Faults	0.5	0.07*	0.13*		
MLP	No faults	0.57	0.5	0.54	0.55	0.55
	Faults	0.54	0.6	0.57		
Random Forest Classifier	No faults	0.61	0.64	0.62	0.6	0.6
	Faults	0.6	0.57	0.58		
Gradient Boosting Classifier	No faults	0.6	0.68	0.64	0.6	0.6
	Faults	0.61	0.53	0.57		

Table 2. Evaluation Metrics (* Indicates very low scores that indicates low model performance)

From Table 2, it is clearly seen that the RFC and GBC provide the best scores. Since both algorithms have scores that are very close to each other, they were both selected for model tuning to choose the best one.

4.2. Model optimization

In this section, the aim is to maximize the scores as much as possible. Since RFC and GBC performance relies on many hyperparameters, optimization is not an easy task. Thus, Random search is used on both algorithms to search for the best hyperparameters that maximizes the performance score. In this research, the performance measure used is AUC. The hyperparameters involve the number of trees, depth of the tree, minimum sample per leaf, and others. Also, the random search was set with 5-fold cross-validation with 100 parameter sets. The results are listed in Table 3.

Algorithm	Class	Precision	Recall	F1-Score	Accuracy	AUC
Random Forest Classifier	No faults	0.61	0.64	0.62	0.6	0.6
	Faults	0.6	0.56	0.58		
Gradient Boosting Classifier	No faults	0.61	0.65	0.63	0.6	0.6
	Faults	0.6	0.55	0.58		

Table 3. T Evaluation Metrics (Optimized Models)

From Table 3, it appears that optimization didn't improve the performance of the selected algorithms significantly. To investigate this, a revision of the performance measurements should be studied, and the parameters extracted from the random search. The precision and recall values are nearly the same, with a better performance for no fault label. Since the dataset is balanced to 50/50 ratio, this indicates that both algorithms work well in predicting both labels. Thus, this leaves the reasons to the hyperparameters, the features, and the size of the data. The hyperparameters listed in Table 4 show that the two algorithms have near values in terms of number of trees. however, the depth in RFC is higher, and this is due to the algorithm mechanism itself. This leaves the features used in prediction and the size of the data. The features are studied in the next section. The data size would be tested by increasing the amount of data used in training the model and testing the outcomes. To do this, we've changed the data balance to be 40% no fault, to 60% fault and tested the models. Table 5 shows that the performance for the fault label has increased, while for no fault it has decreased. This indicates that the data size should be enlarged and the analysis period should be extended.

Hyperparameter	RFC	GBC
max_depth	14	7
min_samples_leaf	6	13
min_samples_split	6	6
n_estimators	179	165
Bootstrap	True	

Table 4. Optimized Hyperparameters

4.3. Feature importance

The last step in analyzing the results is knowing what are the features that have the most contribution to the model. Figure

Algorithm	Class	Precision	Recall	F1-Score	Accuracy	Cross-Validation	AUC
Random Forest Classifier	No faults	0.51	0.39	0.44	0.63	0.6	.58
	Faults	0.68	0.78	0.72			
Gradient Boosting Classifier	No faults	0.53	0.3	0.39	0.63	0.61	.57
	Faults	0.66	0.83	0.74			

Table 5. Evaluation Metrics of Selected Models (After Data imbalancing)

4 and Figure 5 shows the importance percentage per feature for both algorithms.

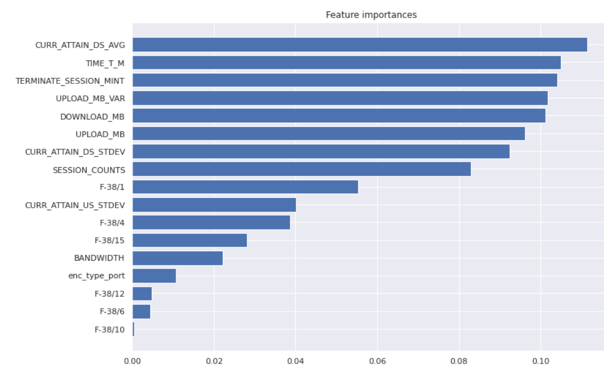

Fig. 4. Feature Importance for Random Forest

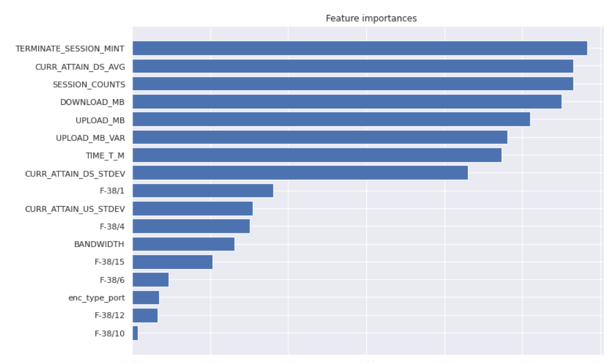

Fig. 5. Feature Importance for Gradient Boosting

From the figures, the features that contributed the most in predicting the fault is system logs related like number of sessions terminated, duration of the sessions, and upload and download amounts. Also, other features like average attainable downstream have contributed significantly from other features. This can be explained since fluctuations and continuous link interruptions are signs of bad network components, and since it is a physical wired connection, this fluctuation indicates signal loss due to high impedance on the line, or an open circuit, which are both considered technical faults. Moreover, the customer usage info would be affected by the bad connection, which results in change in the behavior of customer usage.

However, the maximum contribution per feature is up to 12%, which leaves a question whether extra features are needed or not. The answer is yes; Following the literature which included other variables related to line technical health like SNR and attenuation, these features have contributed the most in predicting the faults according to [1]. Unfortunately, acquiring data regarding these features needed development inside the company since this data is used for live troubleshooting and not stored in any of the internal systems. Also, the low contribution per feature means that they describe the data almost the same amount, which emphasizes the fact the performance can be improved by adding more features.

4.4. Conclusion

In this research, the data is used to predict network faults. The data included information from several systems within the telecom network which indicates the line technical health, and other customer related variables like usage and customer profile. By using a 7-day analysis period and comparing the results from 8 different algorithms, the results show that Random Forest and Gradient Boosting were able to predict the network faults better than other algorithms. Following the results analysis, it appears that the model needs extra features to be added, and to increase the analysis period in order to provide better predictions. This model is a practical one that is important for telecom companies that seek to optimize its performance in different ways. At first, this model can predict the network faults, and management would be able to respond proactively and solve the issue before it happens, which helps increase customer satisfaction, reduce maintenance cost, reduce failures and downtimes, and optimize the field force management in terms of time and effort.

5.Future Work

This research involves many improvement opportunities. At first, more features and data can be added to increase the performance of the model. Secondly, the model can be upgraded to predict the type of fault that might happen in addition to fault prediction. Finally, it can be improved to include other broadband technologies like FTTH and even for mobile technologies.

6. Limitations

This model was developed using a dataset given from a telecommunication company in Palestine and it might introduce some limitations to this research. The mentioned features were extracted from the network performance equipment, which might differ from a vendor to another. Moreover, the study involved network faults on ADSL and VDSL Access technology only, regardless of the faults introduced by service providers, which is more configuration-related not network health related. Finally, this model is applicable for DSL technology, but might not give the same performance for other technologies like FTTH.

Acknowledgement

This research is supported by Palestinian Telecommunication Company (PALTEL), and it was embedded as an internal project to enhance the customer's technical experience and proactively maintain network health.

References

[1] J. S. Tan, C. K. Ho, A. H. L. Lim, and M. R. bin M. Ramly, "Predicting network faults using random forest and C5.0," Int. J. Eng. Technol., vol. 7, no. 2, pp. 93–96, 2018.

[2] K. Ermolov, "Detecting Patterns in Telecommunication Data to Improve Fault Management," Aalto University, 2015.

[3] S. Yuhensky, R. Munadi, and Hafiddudin, "Forecasting formulation model for amount of fault of the CPE segment on broadband network PT. Telkom using ARIMA method," ICCEREC 2016 - Int.Conf. Control. Electron. Renew. Energy, Commun. 2016, Conf. Proc., pp. 185–191, 2017.

[4] A. MEYLER, G. KENNY, and T. QUINN, "Forecasting irish inflation using ARIMA models," Cent. Bank Financ. Serv. Auth. Irel. Tech. Pap. Ser., vol. 1998, no. 11359, pp. 1–48, 1998.

[5] T. Y. Fei, L. J. Yan, L. H. Shuan, G. Xiaoning, and S. W. King, "Detection on network equipment failure using Naïve bayes classification," Proc. 2017 IEEE 2nd Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2017, vol. 2018-Janua, pp. 286–290, 2018.

[6] L. H. Shuan, T. Y. Fei, S. W. King, G. Xiaoning, and L. Z. Mein, "Network equipment failure prediction with big data analytics," Int. J. Adv. Soft Comput. its Appl., vol. 8, no. 3, pp. 59–69, 2016.

[7] H. M and S. M.N, "A Review on Evaluation Metrics for Data Classification Evaluations," Int. J. Data Min. Knowl. Manag. Process, vol. 5, no. 2, pp. 01–11, 2015.