Supervised Machine Learning for predicting impact of faults on RAN KPIs.

Adetomiwa Adedeji

adetomiwa.adedeji@yahoo.com

Abstract:

Various machine learning algorithms exist for classification tasks and selection of algorithm depends on the dataset, the size of data and the number of features present in the dataset. In this project our task was a time series problem of predicting average data rate change at first hour when fault occurred in a network. In addition to the data available at hour when the fault occurred (prediction time), most NE also had lag/historical features from hourly data recorded before the fault occurred which improved predictability of average data rate change.

Keywords: NE, Machine Learning, Classification

I. Introduction:

Fault management involves all processes from the detection, isolation, diagnosis and resolution of faults in information technology or telecommunications network to ensure its continuous uninterrupted operation. Other core areas of fault management include fault reporting, proactive management, automation (using machine learning, artificial intelligence and other network monitoring tools) and continuous improvement.

Machine learning can be classified into supervised and unsupervised learning. In supervised machine learning, a model is developed based on what the algorithm can learn from previous (historical) events. Labeled data is used to train the train and validate the model and can be applied to both regression and classification tasks. In regression tasks the aim is to predict a continuous variable while in classification tasks we are predicting discrete values for example, whether a fault occurred or did not occur, or in our case whether the value of data rate is less than the value right before the fault. Examples of algorithms for supervised machine learning include Decision Trees, Logistic Regression, Support Vector Machines, Neural Networks and Naïve Bayes [1].

On the other hand, unsupervised learning involves finding patterns or identifying attributes in the data that can be used to classify the data into *n number of clusters*. K-means and Agglomerative clustering are common unsupervised learning algorithm used to stratify data into clusters [2].

Aim: To develop a machine learning-based model to predict how each NE's average data rate changes when a fault occurs based on network topology and historical data.

II. Method

Hourly recordings of access success rate, resource utilization rate, time advanced (TA), block error rate (bler), channel quality indicator (cqi), modulation and coding scheme (mcs), data rate, fault duration and relation before fault occurred were provided for all NE IDs . For the training data, fault duration and relation were provided at the hour when the fault occurred.

Hour	Access success rate	Resource utilization rate	TA	bler	cqi	mcs	Data rate	Fault duration	Relation
0	99.714558	8.678	3.71541	7.653169	5.964988	5.582824	5.66777	0	0
1	99.927484	24.264	3.14540	10.015796	6.141206	6.544645	6.77711	0	0
2	99.357688	84.004	2.92336	14.209819	5.582824	5.667775	1.17528	0	0
3	-	-	-	-	-	-	-	301.0	0.65416

Figure I: Sample Hourly data recording for an NE

Fault duration and relation were always zero before a fault occurred, once a fault occurred, the duration of the fault in each hour was recorded with the relation which indicates the adjacency between the NE and neighbouring NE where the fault occured. If the value is 1, it means that the fault occurred exactly at the NE.

The number of hours of data available for each NE was extracted. Lag/historical KPI values were extracted for NE with more than one hour of data before fault occurred and trends/ratios were calculated. Fault duration, relation and the product (fault duration multiplied by relation) at present and past hour were also extracted. All hourly data from all available NE whether fault occurred or not were used for training and validation. Fault duration and Relation when at hours when fault did not occur was left as 0. Time features (hour of day, timestamp, day, month) were extracted for each hour. LGBMClassifier was used train and cross-validate with target (0 or 1) if data rate increased or not at the next hour. The positive class probabilities were extracted and used to select appropriate threshold for classification from validation sets.

III. Results & Analysis:

Lightgbm and CNN models have been proven to be effective in the past for fault prediction for distribution network [3].

Our training data (908,922 rows) was split using train_test_split from sklearn.model_selection stratified by NE and test_size = 0.33.

Two lightgbm classifiers (n_estimators=1000) were trained with and used to predict data NE with historical and NE without historical KPI. The most important features across both models were fault_duration, relation, fault_duration multiplied by relation, NE suffix, NE suffix count encoded and day.

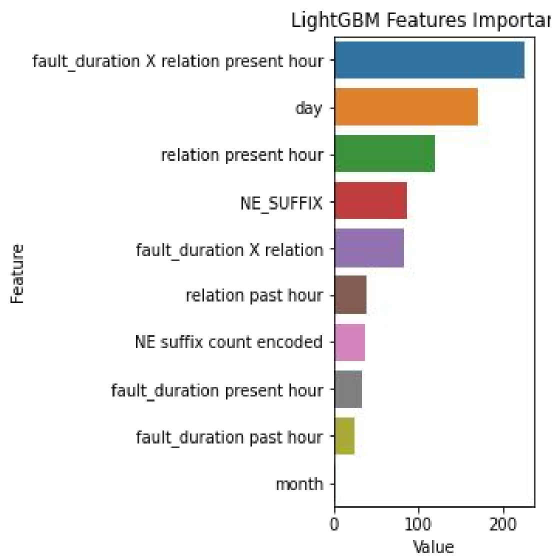

Figure II: Feature importance

The choice of classification threshold significantly affected the accuracy of the model's predictions. Manual selection or tuning of classification threshold has been proven to be effective in anomaly detection using RandomForestClassifier [4].

F1_score peaked at threshold greater than or equal to 0.4

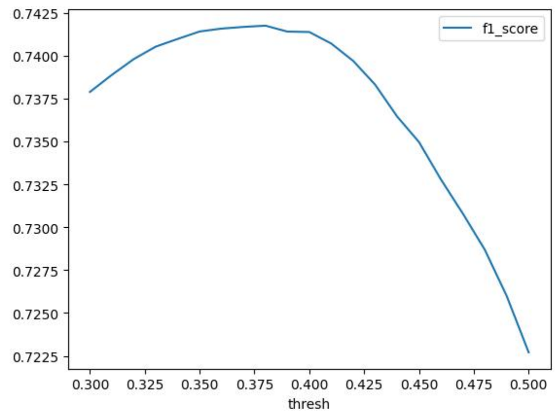

Figure III: Plot of f1_score vs threshold

IV. Conclusion:

We were able to train a machine learning classifier using Lightgbm to predict average data rate changes when at first hour when a fault occurs with mean f1_score of 0.74 using historiacal KPI values, trends and manual threshold selection of >= 0.4.

References:

1. Osisanwo, F. Y., et al. "Supervised machine learning algorithms: classification and comparison." *International Journal of Computer Trends and Technology (IJCTT)* 48.3 (2017): 128-138..
2. Mahesh, Batta. "Machine learning algorithms-a review." *International Journal of Science and Research (IJSR).[Internet]* 9.1 (2020): 381-386.
3. Min, Fan, et al. "Fault prediction for distribution network based on CNN and LightGBM algorithm." *2019 14th IEEE International Conference on*

Electronic Measurement & Instruments (ICEMI). IEEE, 2019.

4. Liu, Dapeng, et al. "Opprentice: Towards practical and automatic anomaly detection through machine learning." *Proceedings of the 2015 internet measurement conference*. 2015.