Analysis of Factors Affecting the Use of the 64QAM Modulation on the Long-Term Evolution Network by Using Random Forest Method

Mochammad Jainul[1], Raden Mohamad Atok[2]
1Department of Technology Management, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
2 Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
e-mail: mochammadjainul.19092@mhs.its.ac.id

Abstract—**Nowadays internet traffic using cellular telecommunication network is increasing very rapidly. Good LTE (Long-Term Evolution) cellular network performance is very important for any telecommunication operator to maintain customer satisfaction. Poor network performance can also cause customers to switch to other operators. One of the indicator variables in observing the radio quality of the LTE cellular network is Penetration using 64QAM Modulation. 64QAM modulation can transmit higher bitrates with lower power usage. 64QAM modulation will be used if the Channel Quality Index (CQI) condition is very good. Network quality improvement can be done by adding new BTS or optimizing existing BTS. The addition of new BTS will increase coverage, quality, and capacity but cost is high, and the time required to build BTS is also long, while improving network quality by optimizing BTS can be done by purchasing LTE features and costs incurred still relatively low. In increasing the penetration of using 64QAM modulation, it is necessary to analyze the other variables. The traditional method to improve this Key Performance Indicator (KPI) requires an expert and professional but is often inaccurate and spends a lot of time finding the factors that cause it. To solve this problem, Random forest method is proposed. By knowing the variables that have a significant effect on network quality, the capital costs incurred by cellular operators for improving network quality will be more effective and efficient because the capital costs invested only focus on influencing variables such as purchasing LTE network features only done for those related to these variables. The results of this study, we make CQI improvement flow based on the classification of the random forest method that produces feature/variable importance.**

Keywords—**LTE, Channel Quality Index, CQI, Random Forest, Base Transceiver Station, BTS, QPSK, 16QAM, 64QAM, AUC.**

I. INTRODUCTION

IN this digital era, the development of technology is very rapid, and the internet is one of the most important needs for society. The industrial revolution 4.0 has grown in recent years, such as the existence of the internet of things (IoT), block chains, and others which have resulted in the emergence of various new business models and managed in new ways. According to the results of a survey conducted by the Indonesian Internet Service Providers Association (APJII), internet users in Indonesia in 2018 reached 171.17 million people.

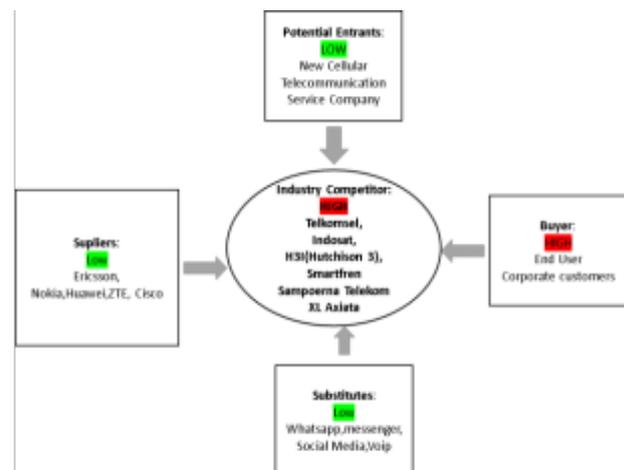LTE is an evolution of cellular technology that can

Figure 1. Porter's Five Forces Analysis on Cellular Operators.

provide an increase in internet access speeds that are far more than previous technologies, namely 3G (HSDPA) and 2G. In theory, 4G technology can reach data access speeds / throughput of 1 Gbps. This technology is a solution to the increasing need for data communication. In LTE network, the modulation systems used are QPSK, 16QAM, and 64QAM in the downlink and uplink directions. 64QAM modulation can transmit higher bitrates with lower power usage. The modulation will change dynamically depending on the network quality conditions Channel Quality Index (CQI) condition. 64QAM modulation will be used if the network quality is very good (CQI≥10).

The quality of the LTE cellular network will greatly affect retaining customers or getting new customers for telecommunications operators. Poor network quality such as difficult internet access, low internet speed will trigger complaints from customers and when this problem is left too long it will cause customers to switch to another telecommunication operator. In a study conducted by [1]. Porter Five Forces Analysis on cellular operators, industry competitors and buyers have a high category as illustrated in Figure 1. This means that these two factors greatly influence the company to achieve success.

In increasing the penetration of using 64QAM modulation, it is necessary to analyze the other variables. The procedures

IPTEK Journal of Proceedings Series No. (7) (2020), ISSN (2354-6026)
The 2nd International Conference on Global Development - ICODEV
December 5th, 2020, Online Conference

32

Table 1.
CQI 4-bit CQI Table (Source: (3GPP, 2019))

CQI index	modulation	code rate x 1024	efficiency
0		out of range	
1	QPSK	78	0.1523
2	QPSK	120	0.2344
3	QPSK	193	0.3770
4	QPSK	308	0.6016
5	QPSK	449	0.8770
6	QPSK	602	1.1758
7	16QAM	378	1.4766
8	16QAM	490	1.9141
9	16QAM	616	2.4063
10	64QAM	466	2.7305
11	64QAM	567	3.3223
12	64QAM	666	3.9023
13	64QAM	772	4.5234
14	64QAM	873	5.1152
15	64QAM	948	5.5547

Table 2.
Results of the Stratified Sampling Performance Test

Fold-to	Accuracy	AUC (micro)
1	92,29%	0,981
2	92,18%	0,981
3	91,91%	0,979
4	92,03%	0,98
5	87,97%	0,957
Average	91,28%	0,9756

Table 3.
Top 20 Importance Features Related to Modulation

No	Variabel	Gini Importance
1	MIMO	0.1831
2	PDCCH_CCE_Aggregation_8_	0.1106
3	SE_3	0.0775
4	sinr_pusch	0.0629
5	sinr_pucch	0.0375
6	_125__PL__130_	0.0265
7	_130__PL__135_	0.0250
8	DL_PRB_Utilization_	0.0200
9	_135__PL__140_	0.0132
10	lte_active_user_dl	0.0107
11	pmrasucccfra	0.0106
12	Average_Active_User	0.0105
13	dl_uu_latency	0.0104
14	pmrrcconnestabattreatt	0.0103
15	_145__PL_	0.0101
16	pmraattcfra	0.0098
17	ul_packet_loss_rate	0.0097
18	pmprbuseduldtch	0.0092
19	pmactiveueulsum	0.0091
20	dl_packet_loss_rate	0.0089

used today are still less efficient in terms of time, energy and require an expert and professional in their field but are often inaccurate in finding factors that cause the quality of LTE networks to deteriorate. To solve this problem, in this study the researchers tried to analyze the network quality data in this case using 64QAM modulation indicators and other influencing variables using the random forest classification method.

By knowing the variables that have a significant effect on network quality, the capital costs incurred by cellular operators to improve network quality will be more effective and efficient because the capital costs invested only focus on influencing variables such as purchasing features, network optimization is only done for those associated with these variables.

II. LITERATURE REVIEW

Modulation and Channel Quality Index (CQI)

Modulation is the process of laying on information signals on the carrier signal. In the LTE network, the modulation used is QPSK, 16QAM, and 64QAM. 64QAM modulation can transmit higher bitrates with lower power usage. In 64-Quadrature Amplitude Modulation (64-QAM) it is included in the high order modulation category because 64QAM consists of 64 symbols where each symbol consists of 6 bits. 64QAM modulation is used when the channel conditions between the sender and receiver are very good. From [3]. Table 1 to get 64QAM modulation, CQI must be greater than 10.

Random Forest

The Random Forest algorithm [4]. is one of the supervised machine learning methods. This algorithm was developed and first introduced by Leo Breiman in 2001. This method is a development of the classification and Regression Tree method and applies the Bagging (Bootstrap Aggregating) method and random feature selection. Basically, the Random Forest algorithm is a set of decision trees whose results will be averaged into a prediction result.

In terms of classification, the result of a classification formed from several trees is then selected based on Majority-Voting.

Performance Evaluation of Classification Methods

The actual and predictive data from the classification model are presented using cross tabulation (Confusion matrix), which contains information about the actual data class represented in the matrix row and the predicted data class in the column [4].

Classification evaluation includes accuracy, sensitivity, specificity, and precision.

$$Accuracy = \frac{TP}{TP+TN+FP+FN} \qquad (1)$$

$$Precision = \frac{TP}{TP+FP} \qquad (2)$$

$$Recall/Sensitivity = \frac{TP}{TP+FN} \qquad (3)$$

$$Specificity = \frac{TN}{TN+FP} \qquad (4)$$

In the case of imbalance, classification accuracy is not sufficient as a standard criterion measure. Area Under Curve
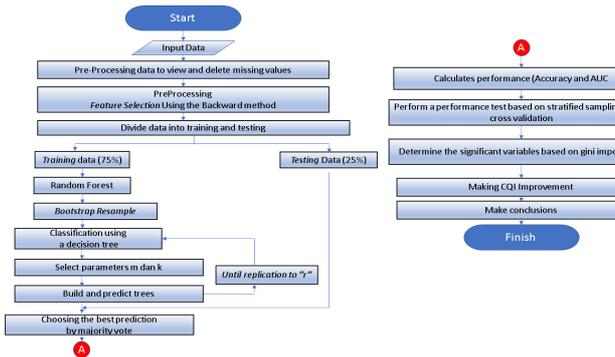
Figure 2. Flowchart of research methodology.

Table 4.
Correlation Coefficient of 8 Significant Feature Importance

Variables	*Polychoric Correlation (y=Modulation)*	Pearson Correlation (y=CQI)
MIMO	0.663	0.78
PDCCH_CCE_Aggregation_8_	-0.551	-0.429
SE_3	0.61	0.148
sinr_pusch	0.604	0.562
sinr_pucch	0.52	0.339
_125__PL__130_	-0.501	-0.404
_130__PL__135_	-0.536	-0.446
DL_PRB_Utilization_	-0.31	-0.322

(AUC) and metrics such as precision, recall have been used to understand the performance of learning algorithms in minority classes.

Area Under Curve (AUC) provides a single measure of classifier performance for evaluating which model is better on average. The size of the AUC is obtained by calculating the true positive rate (TPR), which is the number of objects in the positive class that are classified correctly and the false positive rate (FPR), which is the number of objects in the positive class that are misclassified.

$$TPR = \frac{TP}{TP+FN} \qquad (5)$$

$$FPR = 1 - \frac{TN}{TN+FP} \qquad (6)$$

The area under the ROC curve is called the Area Under Curve (AUC) measure. The AUC measure is used to summarize the ROC curve into a value, the greater the AUC value, the better the model can be [6].

Relevant research

There is some research used as study material in the preparation of this study which discusses the use of machine learning methods in the field of cellular telecommunications. The first journal that was used as a reference was a research [7] This study uses the SWP (sliding window partitioning) and Random Forest machine learning methods to analyze the KPI and KQI relationships of the 5G cellular network. The research [8] applies the Neural Network machine learning
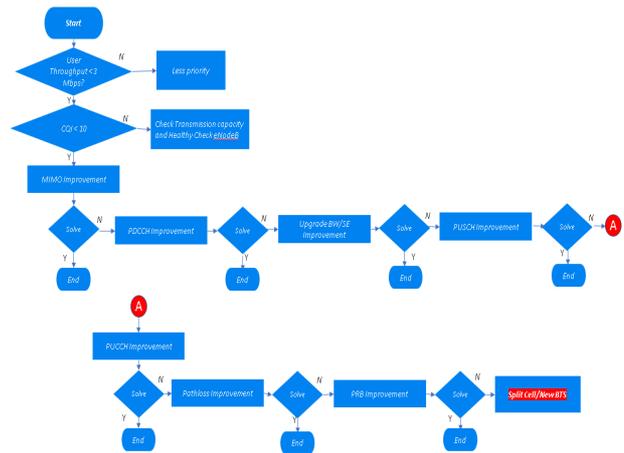

Figure 3. CQI improvement flow.

method to optimize fiber optic network operating costs in telecommunications operators. Research [9] use ANN and Kmeans method. This research aims to predict SIR (signal interference ratio) and produce a coverage map. This model can be considered a promising candidate for studying coverage maps and can be used for efficient spectrum management in the framework of a 5G mobile network. This research [10] uses machine learning to predict the occurrence of base station damage using alarm data from BTS. The aim of this research is that immediate treatment is possible if there is an indication of damage to the BTS. In this study [11], the authors predict the churn rate of subscribers to American Orange operator customers by using logistic regression method.

III. METHOD

The flowchart of research methodology is presented in Figure 2. The dataset is obtained from the one of the telecom operators in Indonesia.

In data mining techniques, the first thing to do before processing data is pre-processing data. The first step is to deal with the missing value. In this study, variables with missing values will be removed from the study. The number of predictor variables in this data is 106, therefore the next step is to carry out feature selection to select the predictor variable (x) to be used. Feature selection is carried out to determine whether multicollinearity occurs between the independent variables. The method used in selecting predictor variables (Feature Selection) is using the Backward Elimination. After feature selection, the number of predictor variables X was reduced from 106 to 77.

The steps after pre-processing are as follows:

1. Apply the Random Forest algorithm to the research data. There are three classifications used in this study, namely 1 = QPSK, 2 = 16QAM, 3 = 64QAM with the total proportion used is 75% for training data, 25% for testing data. The number of k is the number of trees used in this study is 100, while the value of m is the value of the important predictor variable used is √p, p is the number of predictor variables used in this study (p=77).

2. calculate the performance of the classification method

based on accuracy and AUC.

3. Perform a performance test based on stratified sampling 5-fold cross validation.
4. Determine the significant variables based on gini importance.
5. Creating a CQI Improvement Flow based on the variable results obtained from the Random Forest classification
6. Make conclusions

IV. RESULTS

A. Classification Results

In this study, it involved performance testing based on stratified sampling, namely sampling by constructing random subsets and ensuring that the distribution of classes in the subset was the same as the entire dataset. So that each subset contains approximately the same proportion of the two class label values. According to the number of k-folds inputted in this study, the 5-fold test was used. Here are the results that have been done as seen in Table 2 below.

In this study, it is necessary to determine the variables that significantly influence the penetration of using 64QAM modulation (CQI≥10). The following features the importance of the random forest classification results. The following features the importance obtained based on the results of the random forest classification as depicted in Table 3.

Based on Table 3, there are eight variables that have a very significant effect on modulation (quality) as seen in Table 4 below.

B. CQI Improvement Flow

To increase the penetration of using 64QAM modulation, it is necessary to increase the value of CQI $\geq$ 10. Recommendations for increasing CQI are obtained from the results of feature importance using random forest as illustrated in Figure 3.

The priority of LTE features purchased by operator X for both the old BTS software and the latest BTS software should be based on the priorities described in Figure 4. The implementation of these features requires resources (engineer for performance monitoring and engineer for feature implementation so it needs to be done in stages in its implementation)

V. CONCLUSION

The conclusions obtained from this study are:

1. The Random Forest classification model generated and tested using Cross Validation is very good because it has an average accuracy value of 91.28% and an average AUC value of 0.9756.
2. From the classification results on operator X, there are eight significant variables, which have the highest Gini Importance values. **MIMO** is KPI showing MIMO usage, **PDCCH_CCE_Agg_8** is KPI that shows the use of the PDCCH channel AGG 8, **SE_3** is related to spectral efficiency, **Sinr_pusch** is KPI which shows the level of network quality in the Uplink Channel. PUSCH is
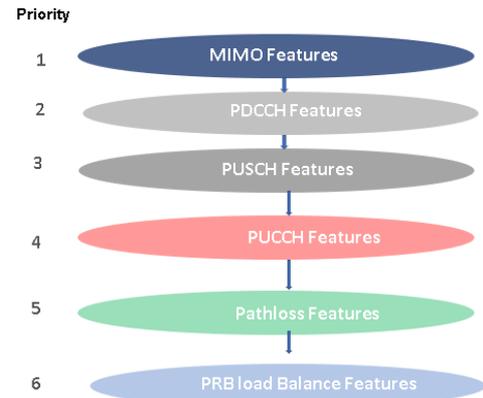

Figure 4. Priority Purchasing LTE Features.

an uplink channel that will carry CQI information if the UE does download and upload activities simultaneously, **Sinr_puuch** is KPI which shows the level of network quality in the Uplink Channel. PUSCH is an uplink channel that will carry CQI information if the UE only performs downloading activities, **125__PL__130** is KPI shows the Pathloss range between 125 dB to 130 dB measured from BTS to UE, **130__PL__135** is KPI shows the Pathloss range between 130 dB to 135 dB measured from BTS to UE, **DL_PRB_Utilization** is KPI that shows the Physical Resource Block capacity of the BTS. The higher the value indicates that the use of PRB in BTS is very high.

3. The addition of a new BTS / Split Cell to improve the quality of the LTE network is carried out after the network optimization process steps in the CQI improvement diagram have been carried out and there is no improvement in the quality of the cellular network

REFERENCES

[1] E. Arvian, "How Can The Indonesian Telecommunication Giant Persist Amid The Competition?," p. 24, 2015.
[2] D. Cahyanto, "Analisa potensi keunggulan kompetitif re-farming," EEICT, vol. 1, p. 7, 2018.
[3] 3GPP, "UE procedure for reporting Channel State Information (CSI)," in 3GPP TS36.213 V15.9.0 (2019-12), 2019, p. 214.
[4] L. Breiman, Random Forests, California: University of California, 2001.
[5] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd penyunt., USA: Kaufman, 2006.
[6] Gorunescu, " Data mining: concepts and techniques," 2011.
[7] G. Fan, "Machine learning based quantitative association," IEEE Access, vol. 7, p. 6, 2019.
[8] K. Rahouma, "Applying machine learning technology to optimize the operational cost of the egyptian optical network," IEEE Access, p. 12, 2019.

The 2nd International Conference on Global Development - ICODEV
December 5th, 2020, Online Conference

[9] O. Rozenblit, "Machine learning methods for sir prediction in cellular networks," IEEE Access, p. 5, 2018.

[10] Y. Hsin Liu, "Predicting malfunction of mobile network base station using machine learning approach," IEEE Access, p. 3, 2019.

[11] H. Jain, "Churn prediction in telecommunication using logistic regression and logit boost," IEEE Access, p. 10, 2019.