UNIZIK JOURNAL OF ENGINEERING AND APPLIED SCIENCES

UNIZIK Journal of Engineering and Applied Sciences 2(1), June 2023, 111-126
Journal homepage: https://journals.unizik.edu.ng/index.php/ujeas
PRINT ISSN: 2992-4383 || ONLINE ISSN: 2992-4391

Machine Learning: A Panacea for Improving the Energy Efficiency of a 5G Network

Chinedu R. Okpara[1*], Ogbonna P. Ngwu[2], Victor E. Idigo[3] and Cosmas K. Agubor[4]
[1,4]Telecommunications Engineering Department, Federal University of Technology, Owerri, Imo State, Nigeria.
[2]Greenera Technologies Limited, Owerri, Imo State, Nigeria.
[3]Electronic and Computer Engineering Department, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.
[*]Corresponding Author's E-mail: chinedu.okpara@futo.edu.ng

Abstract
5G networks supports large number of devices and requires high data rates, needing much energy consumption. The need to efficiently manage this energy consumption using machine learning is a major drive for this research. A 5G production dataset was cleaned, normalized and analyzed using Python programming language. Correlation investigation results shows that the highest correlation value of 0.78 exists between the reference signal power and the reference signal received power of the neighbouring cells (nRxRSRP). Using the significance indicator, the signal to noise ratio was observed to be the most important feature of the production dataset. In determining the hyper-parameter tuning, in order to maintain good accuracy and avoid over-fitting when developing improved energy efficiency algorithms, it was found that the number of estimators should not exceed 25 and the maximum depth of gradient descent should not exceed 9. The cleaned dataset was trained, validated and various algorithms were developed. The ridge stacking regression algorithm which combines all the algorithm outputs put together, outperformed the individual algorithms with the least root mean square error (RMSE) value of 1.931 and maximum coefficient of correlation R^2 of 0.132 which measures how best the regression model fits into the data while the Xgboost was the best algorithm amongst all the individual models with RMSE value of 1.943 and R^2 value of 0.114.

Keywords: 5G, Artificial Intelligence, Machine Learning, Python, Signal to noise ratio

1. Introduction

Cellular technologies, over the years have evolved from the first generation (1G) up to the present generation (5G), and will continue to evolve in the coming years. Panwar, Sharma and Singh (2016) stated that over the past years, technologies such as the Internet of things (IoT) have resulted in billions of connected devices leading to the generation of an enormous volume of data. It was expected that the traffic volume will increase exponentially and there will be approximately 50 billion devices by 2021. Lahdekorpi, Hronec, Jolna and Moilanen (2017) asserts that in mobile networks, 80% of the total energy in cellular networks is consumed by base stations and have pivotal importance for energy efficiency improvement. According to (Yadav, 2017), cellular technologies have seen gradual evolution from the first to the fifth (5G) evolution for meeting the demands in terms of bandwidth, throughput, latency and jitter. Rajoria, Trivedi and Godfrey (2018) states that to improve the coverage and meet capacity requirements, a large number of small cells will be deployed. Small cells make the network denser, which leads to more energy consumption. Furthermore, massive MIMO also increases the power consumption due to more hardware components required for each base station. Ericsson (2020) asserts that, in 2025 the amount of user data will increase four-times compared to today's network. 5G networks demand lots of energy consumption, giving rise to the need for efficient resource management and spectral sharing for improved energy efficiency. Bjȍrnson, Kountouris and Debbah (2013) explained that to improve the cellular energy efficiency, without sacrificing quality of service (QoS) at the users, the network topology must be densified to enable higher spatial re-use.

Their goal was to minimize the total power consumption while satisfying QoS constraints at the users and power constraints at the BS and small cell access points (SCAs). Johnson (2018) tried to solve the problem of 5G energy consumption by the deployment of small cells. This makes the networks denser, leading to more energy consumption. (Fonseca, Kazman & Lago, 2019) discussed that 5G networks are being designed to provide pervasive networking, high data rates, coverage, reliability and low latency and meeting such diverse requirements has also resulted in increased ICT energy consumption; by 2025, the ICT industry itself could be responsible for 30% of power consumption globally. Mughees, *et al*, (2020) discussed that several technologies including software-defined networking (SDN), ultra-dense networks (UDN), network-function virtualization (NFV), multi-access edge computing (MAC), cloud computing and small cells are being integrated in the 5G network to realize its diverse set of services. That these comes with several challenges in terms of energy efficiency, which covers the whole network, from the radio access network, core network, data centers and technologies. Considering the energy constraints and versatile network requirements, traditional approach is not enough for network optimization, hence, machine learning techniques play important roles in assisting in the task of achieving energy efficiency in the network by learning intelligently from data and optimizing the overall operation of the network. Haidine, *et al*, (2021) states that AI and ML will unlock the power of software and algorithms that will allow for efficient deployment of assets and resources.

In Deepsig (2022), it is explained that a fully operative and efficient 5G network cannot be complete without artificial intelligence (AI), and by integrating machine learning (ML) into 5G technology, intelligent base stations will be able to make decisions for themselves, and mobile devices will be able create dynamically adaptable clusters based on learned data. According to Sharma (2022), in machine learning, various kinds of algorithms are used to allow machines to learn the relationships within the data provided and make predictions based on patterns or rules identified from the dataset. AI/ML solutions can help carriers determine where and how to deploy resources to avoid demand crunches and potential service disruptions. Regression is a machine learning statistical technique where the model predicts the output as a continuous numerical value by relating a dependent variable to one or two independent (explanatory) variables, showing whether changes observed in the dependent variable are associated with changes in one or more of the explanatory variables. This is done by essentially fitting a best-fit line and seeing how the data is dispersed around this line (Beers, 2022). Raca, *et al* (2020) collected 5G trace dataset, generated from two mobility patterns (static and car), and across two application patterns (video streaming and file download); composed of client-side cellular key performance indicators (KPIs), channel-related metrics, context-related metrics, cell-related metrics and throughput information was used. These metrics were generated from a well-known non-rooted android network monitoring application, the G-Net Track Pro. This is the first publicly available dataset that contains throughput, channel and context information for 5G networks. The research gap identified from their trace dataset is that the 5G production dataset does not contain power or energy usage. Secondly, they did not develop any machine learning model(s) with the 5G production dataset.

2.0 Material and methods

The 5G production dataset of Raca *et al*, (2020) was cleaned, normalized and analyzed using the Python programming language. Given that the production dataset is labeled, it will perform the mapping of input function to output function; hence supervised learning was used as shown in figure 1. From figure 1, we see that the input raw data (production data set) after being prepared will be trained and tested, some algorithm(s) will be chosen and processed to see which model best fits, and informed decision(s) taken. Figure 2 shows the block diagram of the steps taken to achieve the work. The dataset which came in thirty-five (35) different sets and having twenty-five (25) features were first merged together. It was then cleaned by the removal of outliers, removal of inconsistent data points, encoding of non-numeric data points and descriptively renaming of the columns (features). The statistical distribution of each feature was also explored by checking the minimum value, the maximum value, the mean, the standard deviation, the variance, the quartiles and the median points. Visual plots were used to determine the features of the dataset that needs to be normalized; the production dataset was normalized using the logarithm method in the numpy library. This is because machine learning performs well with normalized data. After the normalization of the features needing normalization, the significance indicator was used to determine the key features of the production dataset needed to improve energy efficiency of a 5G network.

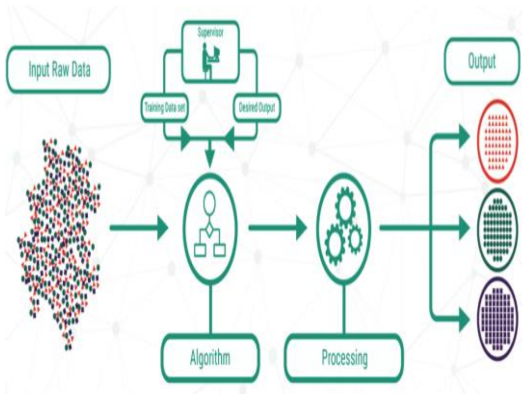


Figure 1: Supervised learning (Kaur, Khan, Iftikhar, Imran & Ul-Haq, 2021)

The density heat map, which uses colour gradient to show how concentrated a given data is, at a particular area and scatter plots, which shows how the data is distributed were used to show the relationships existing between the features. The Pearson correlation matrix was further used to investigate the features relationship pair having the highest correlation value. The cleaned dataset was then trained with 70% and tested/validated with 30% of the data. The hyper-parameters (Number of estimators and maximum depth of gradient descent) for optimal performance for improved energy efficiency were also determined for improved accuracy and in order to avoid over-fitting. After this, different regression analysis algorithms were modeled, including the random-forest algorithm, Xgboost regression algorithm, gradient boosting algorithm, ridge regression algorithm and lasso algorithm; they were all compared together with respect to the coefficient of correlation R^2, and root mean square error (RMSE).

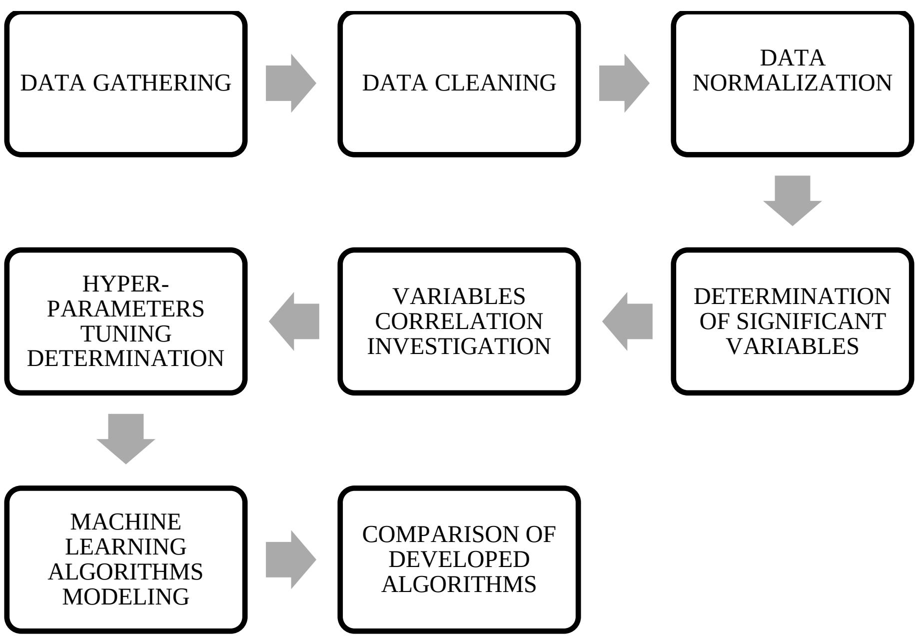
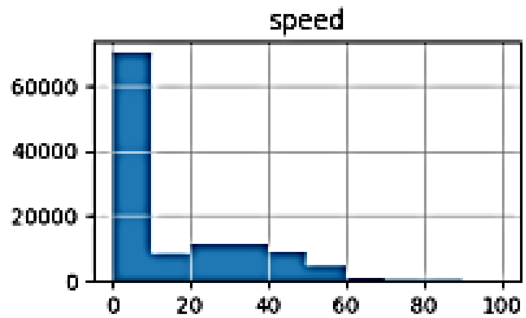


Figure 2: Steps to achieving the research.

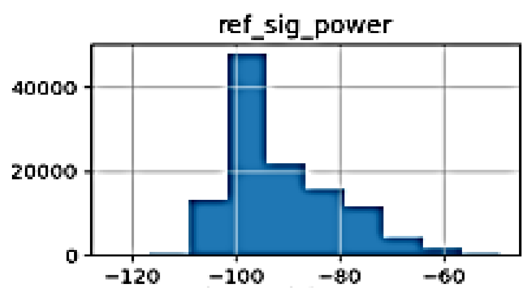
3.0 Results and Discussions

The power of visual plots was used in determining the features of the production dataset that needed to be normalized, since machine learning works better with normalized data. Figure 3(a-j) shows the resulted numerical columns with respect to determining the data transformation and normalization approach to take for each feature. From the results obtained, it is observed from Figure 3(a) that the speed data distribution is skewed heavily to the left thus needs normalization.

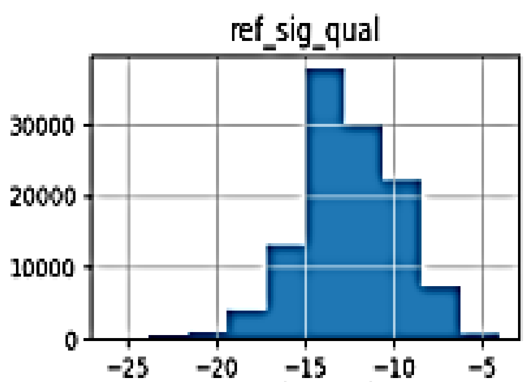
Figure3(b) shows that reference signal power data distribution is almost a Gaussian distribution, needing little or no normalization. Figure 3(c) shows that the reference signal quality data distribution is a Gaussian distribution. Figure 3(d) shows that the signal to noise ratio data distribution is a Gaussian distribution. Figure 3(e) shows that the signal quality indicator data distribution is almost a Gaussian distribution, needing a little normalization. Figure 3(f) shows that the received signal strength data distribution is a Gaussian distribution. Figure 3(g) shows that the download bit-rate data distribution needs further investigation of its values, since its values are centered between 0 and 1. Figure 3(h) shows that the up-link bit-rate data distribution needs further investigation, due to its values ranging from 0 to 1,000. Figure 3(i) shows that the nRxRSRP data distribution is slightly skewed to the right, needing normalization. Figure 3(j) shows that the nRxRSRQ data distribution needs further investigation, since its values centered between 0 and 0.3.



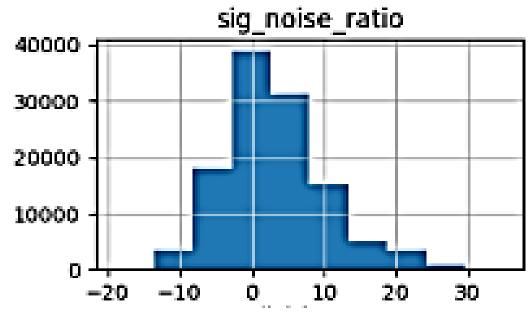
(a) Speed data distribution chart.



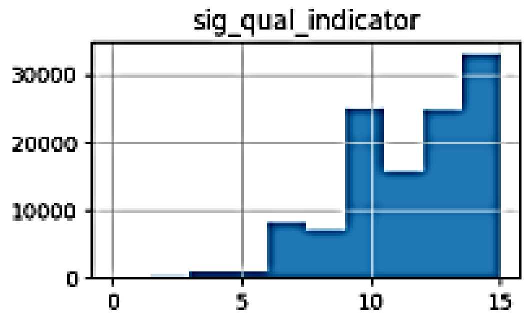
(b) Reference signal power data distribution chart.



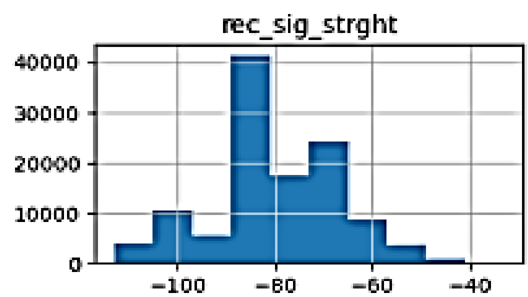
(c) Reference signal quality data distribution chart.



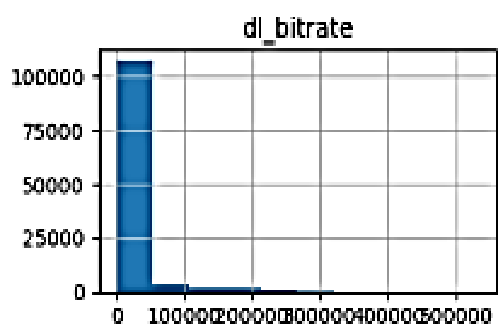
(d) Signal to noise ratio data distribution chart.



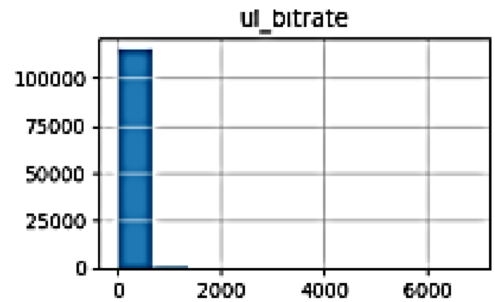
(e) Signal quality indicator data distribution chart.



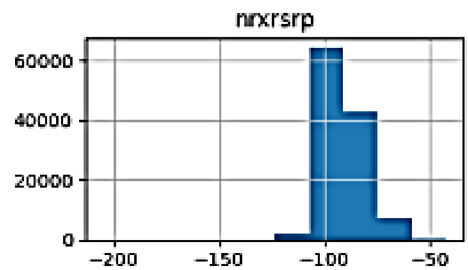
(f) Received signal strength data distribution chart.



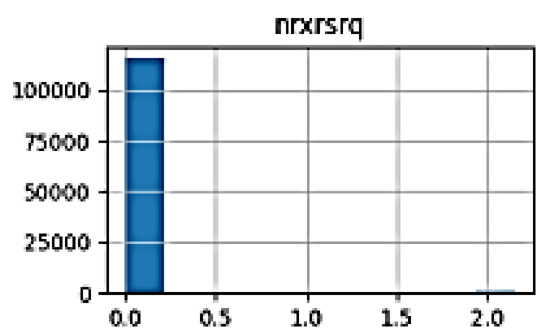
(g) Download bit-rate data distribution chart.



(h) Upload bit-rate data distribution chart.



(i) Received signal reference power of the neighbouring cells data distribution chart.



(j) Received signal reference quality of the neighbouring cells data distribution chart.

Figure 3 (a-j). Column distribution for data normalization approach.

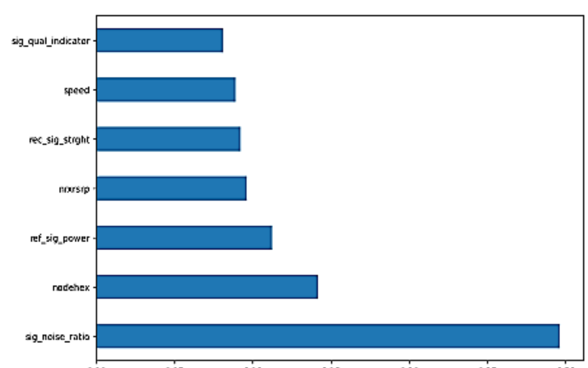


Figure 4: Key features of the production dataset for improved energy efficiency.

The significance indicator was used to determine the key important features of the production dataset of Raca *et al*, (2020) for improving the energy efficiency of a 5G network; from figure 4, the signal to noise ratio was found to be the most important. Furthermore, the density heat map which uses colour gradient to show how concentrated a given data is, at a particular area, and the scatter plot which shows how the data is distributed were used to show the relationships between the features. The density heat map and scatter plot of figure 5(a-b) shows that the linear relationship between the received signal strength and the reference signal quality is good.

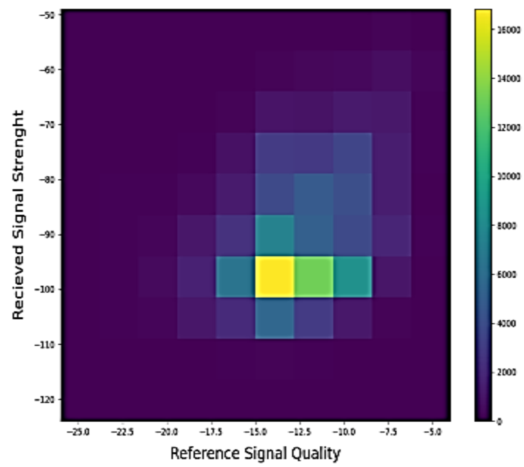


Figure 5(a): Density heat map of received signal strength vs. reference signal quality.

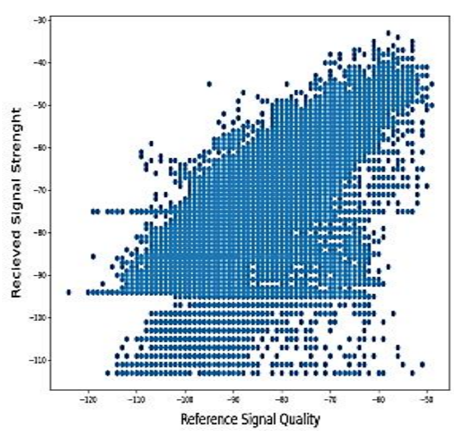


Figure 5(b): Scatter plot of received signal strength vs. reference signal quality.

The density heat map and scatter plot of figure 6(a-b) shows that the linear relationship between the reference signal quality and the reference signal power is good.

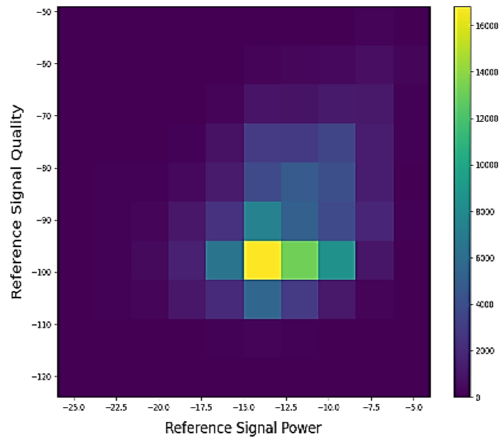


Figure 6(a): Density heat map of reference signal quality vs. reference signal power.

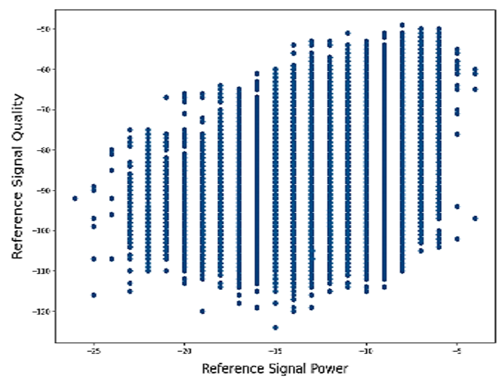


Figure 6(b): Scatter plot of reference signal quality vs. reference signal power.

The density heat map and scatter plot of figure 7(a-b) shows that very good relationship exists between the signal to noise ratio and the reference signal quality.

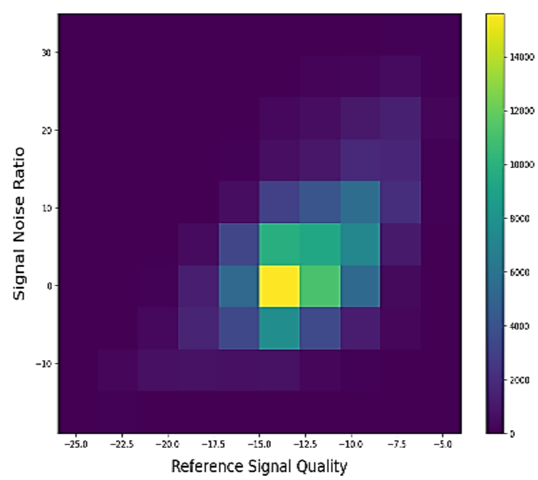


Figure 7 (a): Density heat map of signal to noise ratio vs. reference signal quality.

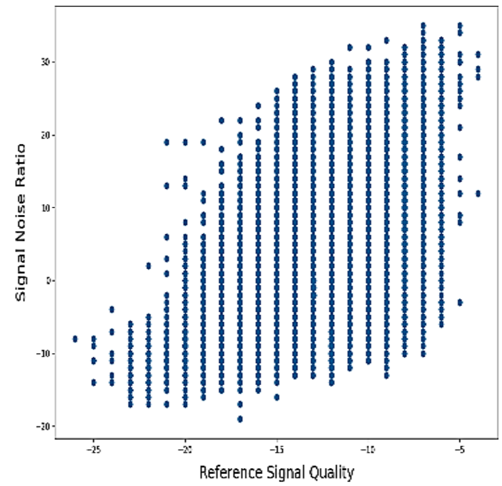


Figure 7 (b): Scatter plot of signal to noise ratio vs. reference signal quality.

The density heat map and scatter plot of figure 8(a-b) shows that good relationship exists between the download bitrate and the upload bit-rate of the data capture.

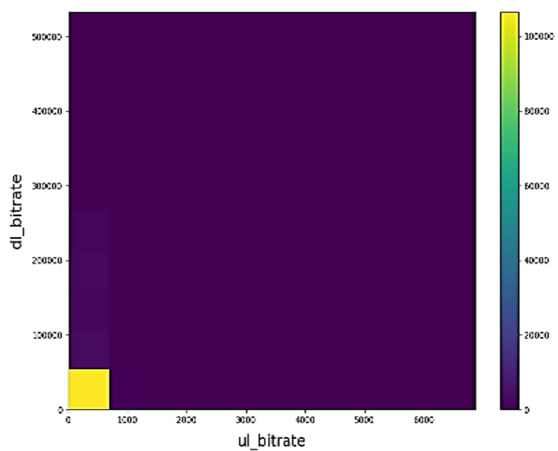


Figure 8 (a): Density heat map of download bit-rate vs. upload bit-rate of data capture.

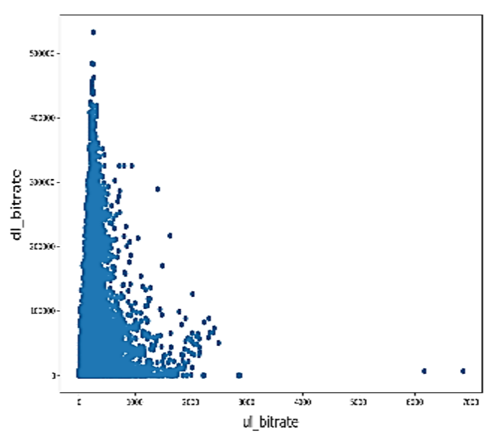


Figure 8 (b): Scatter plot of download bit-rate vs. upload bit-rate of data capture.

The density heat map and scatter plot of figure 9(a-b) shows that there is good relationship between the reference signal received power of the neighbouring cells (nRxRSRP) and the received signal power.

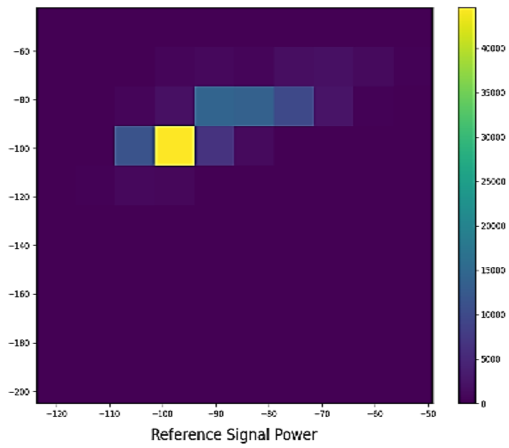


Figure 9 (a): Density heat map of nRxRSRP vs. received signal power.

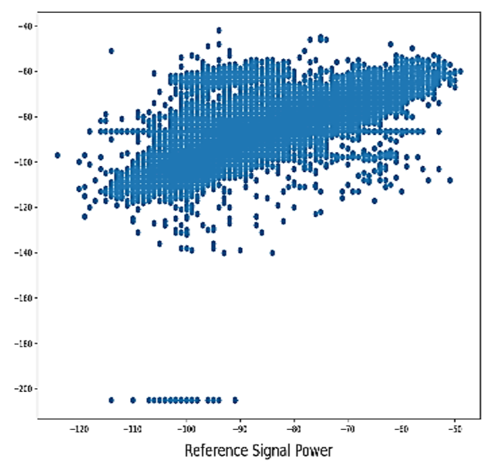


Figure 9 (b): Scatter plot of nRxRSRP vs. received signal power.

The density heat map and scatter plot of figure 10(a-b) shows that there is great relationship between the received signal power and the signal to noise ratio.

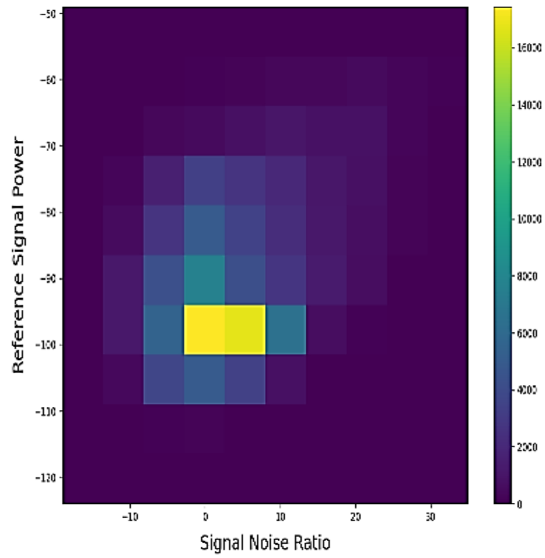


Figure 10 (a): Density heat map of received signal power vs. signal to noise ratio.

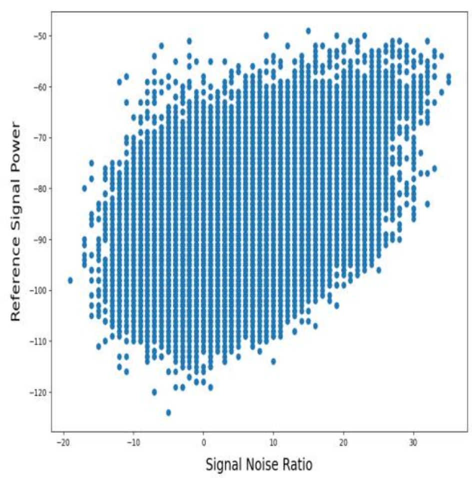


Figure 10 (b): Scatter plot of received signal power vs. signal to noise ratio.

The Pearson correlation coefficient was used to investigate how correlated the features of the production dataset are by themselves, it was found as shown in figure 11 that the highest correlation exists between the reference signal power and the received signal reference power of the neighbouring cells (nRxRSRP) with a correlation value of 0.78.

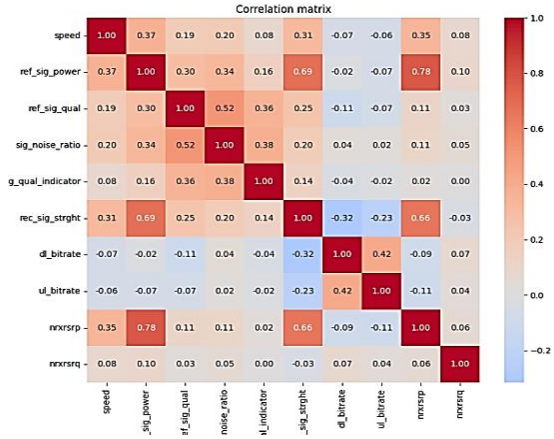


Figure 11: Correlation investigation results of the production dataset features.

In determining the hyper-parameters for optimal performance in improving the energy efficiency of a 5G network, for good accuracy and to avoid over-fitting, it was found that the number of estimators (trees) should not exceed twenty-five (25) and that the maximum depth of gradient descent should not exceed nine (9) as shown in figure 12 and figure 13 respectively.

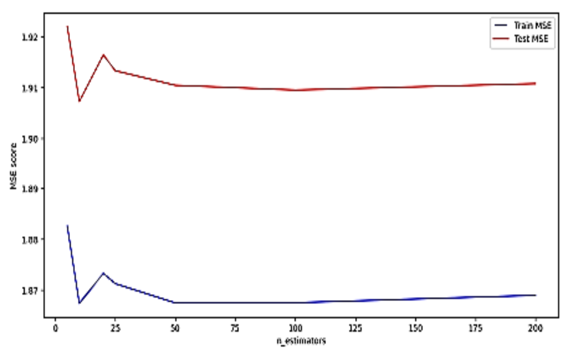


Figure 12: Determination of number of estimators for optimum performance.

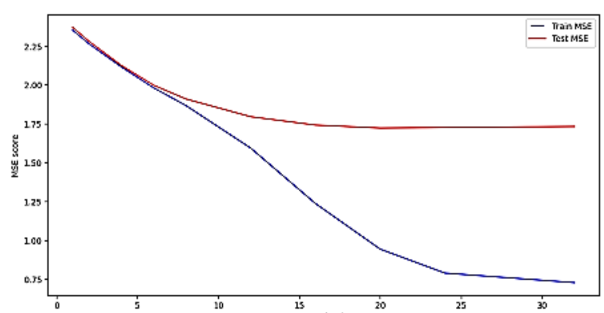


Figure 13: Determination of maximum depth of gradient descent for optimal performance.

The seven most important features determined using the significance indicator were used to train and test the models. After determining the hyper-parameter tuning, the dataset was then trained with 70% and tested/validated with 30% of the data. Several algorithms were then modeled, including the random-forest algorithm, the Xgboost regression algorithm, the gradient boosting algorithm, the ridge regression algorithm and the lasso regression algorithm, and they were analytically compared with respect to the coefficient of correlation R^2, and root mean square error (RMSE).

The results from figures 14-18 shows that given the developed algorithms: the ridge stacking regression algorithm which combined all the model outputs performed better than all the individual models with the least root mean square error (RMSE) value of 1.931 and the highest coefficient of correlation, R^2 value of 0.132, being the measure of how best a regression model fits into the data. The Xgboost regression algorithm was the best of all the individual algorithms with RMSE value of 1.943 and (R^2) value of 0.114.

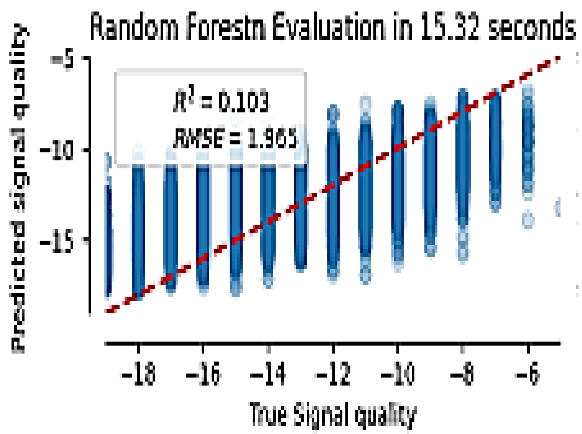


Figure 14: Predicted signal quality vs. true signal quality using random forest algorithm.

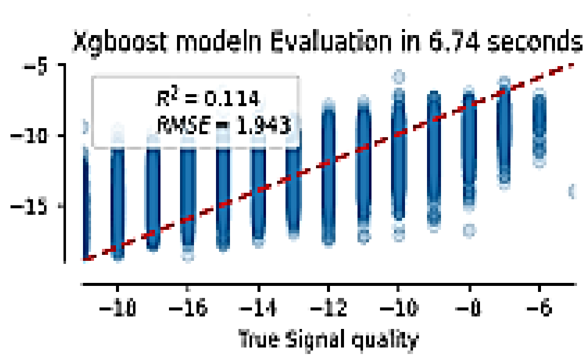


Figure 15: Predicted signal quality vs. true signal quality using Xgboost algorithm.

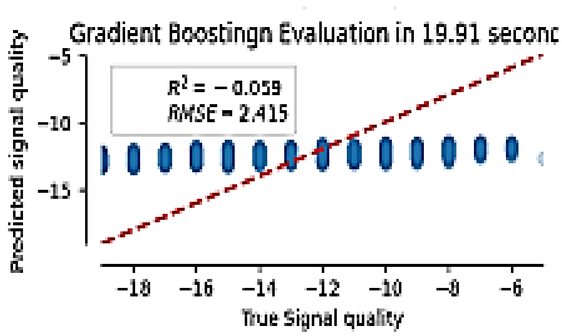


Figure 16: Predicted signal quality vs. true signal quality using gradient boosting algorithm.

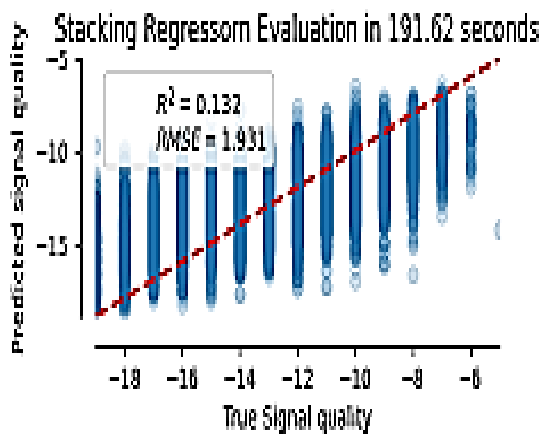


Figure 17: Predicted signal quality vs. true signal quality using ridge (stacking) regression algorithm.

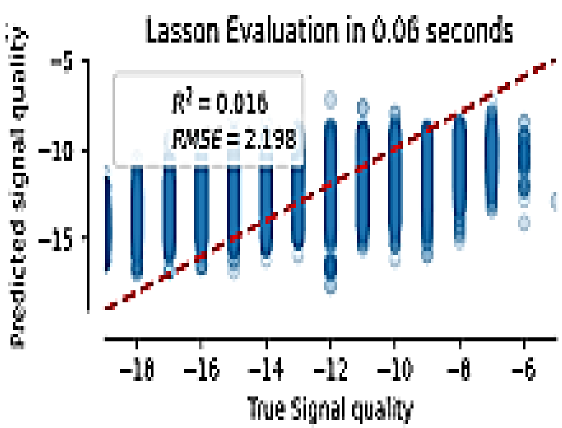


Figure 18: Predicted signal quality vs. true signal quality using lasso regression algorithm.

4.0. Conclusion

This work, machine learning: a panacea for improved energy efficiency of a 5G network is carried out to show the speed and accuracy of machine learning, given a 5G production dataset, in identifying key features of the production dataset and how it aids in improving the energy efficiency of a 5G network after training and validating. The Python programming language was the tool of use. From the results obtained after data normalization, it is observed using the significance indicator that the signal to noise ratio is the most significant feature of the given production dataset, and it was further observed that the reference signal quality, the signal to noise ratio, the reference signal power and the reference signal received power of the neighbouring cells all have key roles in improving the energy efficiency of a 5G network and that there were good relationships existing between them. These features can also be used in developing models for determining the signal strength losses in mm-wave technology. From the results obtained in choosing the hyper-parameters for optimal performance and to avoid over-fitting during modeling, the maximum number of estimators (trees) should not exceed twenty-five (25) and the maximum number of gradient descent should not exceed nine (9).

Using these key features of the production dataset to train and test various algorithms, including the random-forest algorithm, the lasso algorithm, the gradient boosting algorithm, the Xgboost algorithm and the ridge stacking regression algorithm. The results obtained showed that the stacking regression algorithm outperforms the individual algorithms, and that the Xgboost algorithm proved to be the best out of the individual algorithms.

5.0 Recommendation

From the obtained results, it can be recommended that in designing for optimal performance of a 5G network, the following features play important roles in determining the performance of a network model; they include the signal to noise ratio, the node in hexadecimal format, the reference signal received power, the RSRP values of the neighbouring cells, the received signal strength, the speed and the signal quality indicator. Amongst them, the signal to noise ratio is the most important. And in the developing of models for energy efficiency improvement and to avoid over-fitting, the maximum depth of gradient descent should not exceed 9, and the number of estimators should not exceed 25.

Acknowledgements

Special thanks goes to Jason Quinlan, Darijo Raca, Dylan Leahy and Cormac J. Sreenan for allowing the use of their conference paper on ‘Beyond throughput, the next Generation: a 5G dataset with channel and context metrics’ (MMSys ‘20) at Instanbul, Turkey for this research.

References

Beers, B. 2022. What is Regression? Definition, Calculation and Example. Available from: https:www.investopedia.com [December 12, 2022]

Bjornson E., Kountouris M. & Debbah M. 2013. Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination. *ICT 2013*. https://doi.org/10.1109/ictel.2013.6632074.

Bjornson E., Sanguinetti L., Hoydis J. & Debbah M. 2015. Optimal Design of Energy-Efficient Multi-User MIMO Systems: Is Massive MIMO the answer? *IEEE Transactions on Wireless Communications*; 14(6): 3059-3075. https://arXiv.org/pdf/1403.6150v2.

https://www.deepsig.ai/how-artificial-intelligence-improves-5g-wireless-capabilities [March 8, 2022].

Ericsson Mobility Report. 2020. Available from: https://www.ericsson.com/49da93/assets/local/mobility-report/documents/2020/june2020-ericsson-mobility-report.pdf [January 14th 2022]

Fonseca A, Kazman R, & Lago P. A. 2019. Manifesto for Energy-Aware Software. *IEEE Software*, 36(6): 79–82. https://doi.org/10.1109/ms.2019.2924498.

Haidine, A., Salman, F. Z., Aqqal, A. & Dahbi, A. 2021. Artificial Intelligence and Machine Learning in 5G and beyond: A Survey and Perspectives. Open Access peer-reviewed chapter, Moving Broadband Mobile Communication Forward Intelligent Technologies for 5G and beyond. DOI: 10.5772/intechopen.98517

https://www.deepsig.ai/how-artificial-intelligence-improves-5g-wireless-capabilities [March 8, 2022]

Johnson, D. D. 2018. The 5G dilemma: more base stations, more antennas-less energy? Available from: https://spectrum.ieee.org/energywise/telecom/wireless/will-increased-energy-consumption-be-the-achilles-heals-of-5g-networks.

Kaur, J., Khan, M. A., Iftikhar, M., Imran, M. & Ul Haq, Q. E. 2021. Machine learning techniques for 5G and beyond. DOI: 10/1109/Access.2021.3051557

Lahdekorpi, P., Hronec, M., Jolma, P. & Moilanen, J. 2017. Energy efficiency of 5G mobile networks with base station sleep modes. *In Proc. IEEE Conf. Standards for Commun. Netw.* (CSCN), pp. 163 -168.

Mughees, A., Tahir, M., Sheik, M. A. & Ahad, A. 2020. Towards Energy Efficient 5G Networks using Machine Learning: Taxonomy, Research Challenges, and Future Research Directions. *IEEE Access* vol. 8, pp. 187498 – 187522. DOI: 10.1109/ACCESS.2020.3029903

Okpara C. R. 2023. Improving the Energy Efficiency of a 5G network using machine learning. PhD Dissertation, Department of Electronic and Computer Engineering, Nnamdi Azikiwe University, Awka, Nigeria.

Okpara, C. R., Idigo V. E. & Ngwu, O. P. 2023. Improving the Energy Efficiency of a 5G network: The Machine Learning Approach. Iconic Research and Engineering (IRE) journals, ISSN: 2456-8880, IRE 1704107, Vol. 6, Issue 8, Pp. 142 – 145.

Okpara C. R., Idigo V. E. & Okafor C. S. 2023. Comparative analysis of the features of a 5G network production dataset: The machine learning approach. EJ-ENG Volume 8, Issue 2, https://www.ej-eng.org/index.php/ejeng/article/view/2994. DOI:10.24018/ejeng.2023.8.2.2994.

Panwar, N., Sharma, S. & Singh, A. K. 2016. A Survey on 5G: The next generation of mobile communication. *Phys. Commun.*, vol. 18, pp. 64-84.

Raca D., Leahy D., Sreenan C. J. & Quinlan J. J. 2020. Beyond Throughput, The Next Generation: A 5G Dataset with Channel and Context Metrics. ACM Multimedia Systems Conference (MMSys), Instanbul, Turkey.

Rajoria, S., Trivedi, A. & Godfrey, W. W. 2018. A comprehensive survey: Small cells meets massive MIMO. *Phys. Commun.* vol. 26, pp. 40-49

Yadav R. 2017. Challenges and evolution of next generations wireless communication. *In Proc. Int. MultiConf. Eng. Comput. Scientists,* 2017; 2: 619-623.