Hybrid Boosted Model with an Approach Inspired by Mixture of Experts for 5G Energy Consumption

Rafael Zimmermann

ITU AI/ML in 5G Challenge

25 October 2023

Abstract — The accelerated deployment of 5G networks has brought forth concerns regarding the energy consumption of such infrastructures. This study introduces a Hybrid Boosted Ensemble Model tailored for predicting energy utilization in 5G base stations. The methodology merges Ridge Regression for linear trend analysis, XGBoost to tackle non-linear intricacies, and a final refinement strategy built upon Mixture of Experts. Preliminary results demonstrate the model's capability to adjust to new and diverse data scenarios. This research constitutes a pivotal initiative in confronting adaptability hurdles in 5G energy predictions, setting a foundation for subsequent inquiries in this essential domain.

1. INTRODUCTION

The advancement of 5G networks has brought about significant innovations in terms of services and technologies but has also imposed challenges related to energy consumption. According to estimates, 5G networks are about 4 times more energy-efficient than their 4G predecessors, yet their energy consumption is approximately 3 times higher due to the need for a greater number of cells and the additional processing required for broader bandwidths [1]. Network Operational Expenditure (OPEX) already accounts for about 25% of the total cost for network operators, and 90% of that value is allocated to high energy bills [2]. In light of these concerns, this paper focuses on the development of a model for accurate prediction of energy consumption of base stations in 5G networks.

To achieve this, a Hybrid Boosted Ensemble Model [3] is introduced that combines the feature transformation power of Ridge Regression with the target transformation flexibility of XGBoost. Inspired by the Mixture of Experts approach, a final refinement model is applied to enhance prediction accuracy [4]. The objective of this study is to optimize the parameters of base stations and energy-saving methods, providing a deep understanding of how these elements influence energy consumption. The model was tested in scenarios with significant variations in data characteristics, demonstrating great adaptability and accuracy in estimates.

2. DATASET

The data used in this study encompass over a thousand base stations. The dataset is categorized by three main groups:

Data Types and Scale:

• Base Station Information: Predominantly categorical data such as 'RUType' and 'Mode', as well as continuous numerical data like 'Frequency' and 'TXpower'.

• Cell-Level Traffic Statistics: Continuous numerical data normalized for 'load' and energy-saving modes ('ESMode1' to 'ESMode6').

• Energy Consumption Statistics: Continuous numerical data represented in relative units of energy. [5]

Relevance for Modeling:

• Temporal Variability: The inclusion of temporal information allows for the modeling of hourly seasonal patterns, which are vital for understanding the behavior of 5G users in the region.

• Granularity: Data collection at the cell level enables granular analysis that can reveal insights into the impact of different configurations on overall energy efficiency.

• Inter-variable Correlations: The presence of multiple numerical and categorical variables offers the opportunity to explore complex correlations, which can be modeled to predict energy consumption.

Gaining an in-depth understanding of the interactions between these variables and their intrinsic characteristics is imperative not only for the precise calibration of machine learning models but also to ensure a robust and applicable interpretation of the results in real-world scenarios.

3. PROPOSED MODEL

1. Introduction

In this study, three central objectives are addressed, which have significantly influenced both the conceptualization and the adopted methodology. A hybrid boosted model was employed, combining residual fitting techniques similar to gradient boosting methods with an ensemble approach influenced by Mixture of Experts. This algorithm was specifically designed to meet the three main objectives of this study, identified as Objectives A, B, and C.

Objective A: Estimating Energy Consumption in Specific Base Station Products:

To meet this objective, a data sample comprising 816 Base Stations (BS) was meticulously analyzed. Each base station in the sample provided a complete seven-day history of energy consumption, along with statistics related to traffic and the activation of energy-saving systems. Figure 1 illustrates a potential seasonality in both "Energy" and "Load" variables, while also implying a possible correlation between them.

Objective B: Generalization Across Different Base Station Products:

In this objective, the issue arises of the absence of base station (BS) information in the training dataset, which could potentially limit the consideration of seasonality and historical traffic of that particular 5G network region in the predictions. However, we have at our disposal the specific configurations of these stations in the dataset in question. This presents the possibility of prioritizing these configurations in an attempt to generalize energy consumption, without the need to anchor it to the historical traffic of the 5G bases. This approach will be explored and elucidated in subsequent sections.

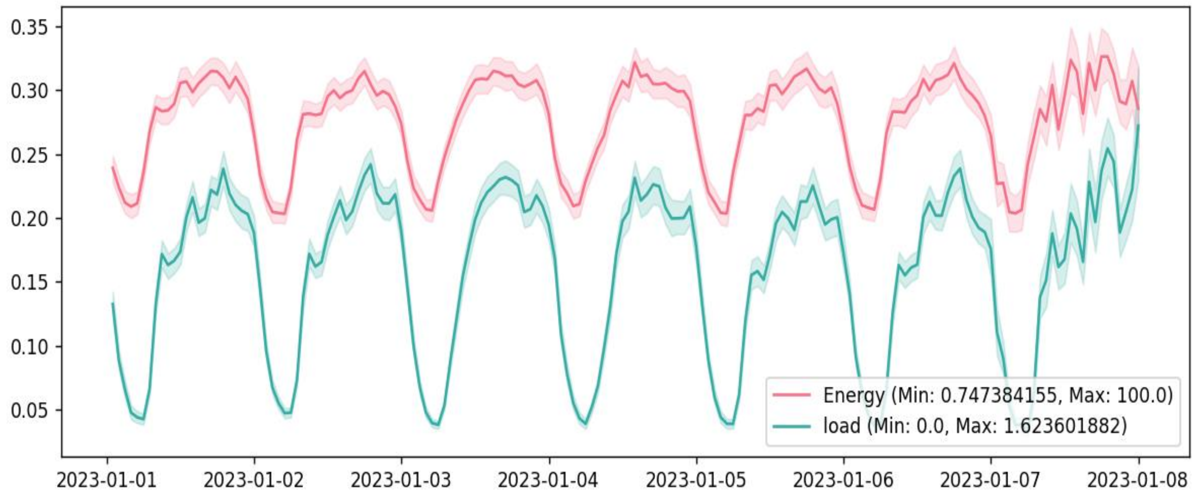

Figure 1 depicts a normalized hourly graph of "Energy" and "Load" over a 7-day span. The red line signifies "Energy" and the green, "Load", with their respective value ranges indicated in the legend.

Objective C: Generalization Across Different Base Station Configurations:

In Objective C, we face an exacerbated challenge, as, in addition to the Base Stations (BS), the associated configurations are also absent in the training dataset. This scenario implies a high degree of generalization required by the algorithm, given the need to operate under conditions considerably different from those observed during training. It is also noteworthy that this dataset has a specific limitation: the available data is restricted to day 02.

2. Inputs of the model

In the implementation of this approach, a structure based on a "mixture of experts" is employed. This strategic decision allows for the creation of three specialized experts: Expert A, Expert B, and Expert C, each designated to specifically address one of the proposed objectives.

In this section, specific inputs for each of these experts are outlined, emphasizing the importance of feature engineering and selection to optimize the performance of the algorithms and achieve the desired outcomes.

Expert A - Emphasis is placed on temporal feature engineering. Attributes such as lags, differences, and purpose-specific attributes like load, EMS, and time have been meticulously developed. A notable addition is the target encoding on the hour variable, a critical feature for capturing the historical hourly seasonality of the Base Station (BS). It is imperative to highlight that these temporal features are customized for each base station, ensuring accuracy while preventing data leakage.

Expert B - The focus has been on the implementation of advanced feature engineering and meticulous selection. Robust polynomial features were developed on attributes such as 'load,' 'ESMode6,' 'Antennas,' 'TXpower,' and 'Frequency.' Additionally, k-means clustering was applied to the Base Station (BS) configurations. The primary intent was to enhance the algorithm's ability to generalize over new stations, minimizing reliance on historical BS data and accentuating the significance of station settings and load. This method provides a hybrid approach, merging time-series features with robust BS configurations.

Expert C - The approach was notably streamlined. Despite initial efforts to implement specialized features, such attempts proved ineffective. Optimized analyses consistently indicated that simpler approaches outperformed complex ones. Therefore, the judicious selection of features emerged as the crucial component in this context, eliminating attributes that led to overfitting or that could introduce contamination from the history of the Base Station (BS) or its robust settings.

In the engineering process, all features were formulated in a singular stage. However, for each Expert, identified as A, B, or C, specific pre-processing was carried out, along with a judicious selection of attributes based on theoretical expertise and algorithmic optimization. This level of detail became essential in defining the input attributes for each specialized model.

3. Outputs of the model

The primary output of our Hybrid Boosted Ensemble Model is a quantitative projection of energy consumption. This projection stems from the synthesis of insights derived from Ridge regression and XGBoost, complemented by the appropriate model selection via approach inspired by the Mixture of Experts. Through this composite methodology, the aim is to provide a rigorous and robust estimate of energy consumption, considering the intrinsic particularities of the provided data.

4. Model architecture

In the field of energy consumption forecasting, we face unique challenges in terms of generalization with respect to the key objectives of energy prediction. To address these challenges, we have developed a Boosted Hybrid Ensemble model that combines the feature transformation capabilities of Ridge Regression with the target transformation flexibility of XGBoost, culminating in the application inspired by the Mixture of Experts for final refinement. [Figure 2]

Model Overview

Ridge Regression: This approach is selected for its ability to handle multicollinearity and extrapolate trends based on the characteristics of the input data.

XGBoost: It operates on the nuances and complexities not captured by Ridge Regression. It relies on the decision tree technique, clustering similar target values.

Mixture of Experts: Conceptualized from the Mixture of Experts paradigm, this refined approach segments the dataset and trains specialists, or individual models, for specific portions of the data. The gating mechanism, which in this case is a conditional model based on hot or cold features of the data relative to the training set, determines which specialist is most appropriate to consult, considering the specificities of the input data to be forecasted.

Boosted Hybrid Model

The essence of our approach is based on an ensemble model that combines Ridge Regression with XGBoost. Let's detail its workings:

Ridge Regression: Base model of the ensemble, particularly effective in identifying and mapping linear trends, belongs to the set of algorithms for feature transformation, which also encompasses linear regression and neural networks. This type of algorithm is notable for its ability to perform extrapolative inferences beyond the scope of the initial training set.

XGBoost on Residuals: After obtaining the predictions from Ridge Regression, we calculate the residuals and train XGBoost on these residuals. This allows XGBoost to understand and capture non-linear nuances and trends that the Ridge model may have missed.

XGBoost is a target transformation-focused algorithm that employs feature clustering to optimize outputs based on the mean of these clusters. Its predictive capability is confined to the range of the training set, lacking the ability for extrapolation. Algorithms such as decision trees and k-NN also fall into this category.

Prediction: Initially, we use estimates from Ridge Regression, which are subsequently adjusted by XGBoost predictions on the residuals. The sum of these predictions constitutes the final estimate.

Mixture of Experts:

Drawing inspiration from the Mixture of Experts (MoE) method, the test data is segmented based on specific features, determining whether they are "hot" (similar to the training data) or "cold" (different from the training data).

Model Adaptability: The adaptive ability of this strategy excels in accurately differentiating features in the data, allowing for the selection of the most suitable expert for each context. As new patterns emerge, the technique has the potential to recognize and proactively adjust its forecasting approaches.

Subsampling: In addition to segmentation, subsampling through Adversarial Validation is implemented in experts B and C. The primary objective here is to adjust the distribution of the training data to align as closely as possible to what is observed in the test data.

Model Workflow

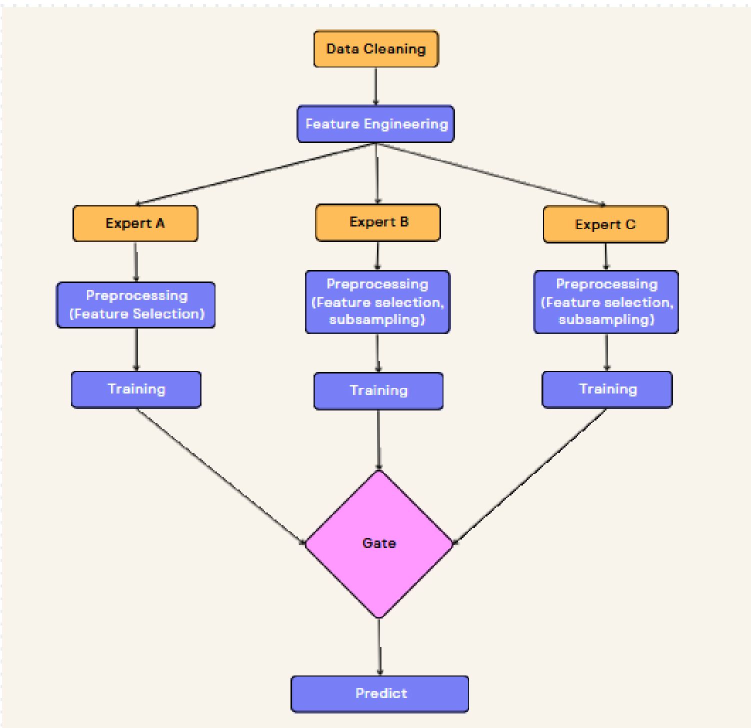

Figure 2: Schematic representation of the model's workflow, detailing the steps from data cleaning to forecasting. The diagram illustrates the differentiation in preprocessing approaches for experts A, B, and C, culminating in the decision-making stage through the "Gate" for the generation of the final forecast.

Training of the model

The model was trained with the aim of optimizing three specific Experts: Expert A, Expert B, and Expert C. Each of these Experts has its own characteristics and nuances, making a robust as well as meticulously optimized validation strategy essential.

Data Management: The technique of MultilabelStratifiedKFold with 10 folds was utilized for the crucial task of validation, selected after demonstrating the best results in multiple optimizations. The subsequent challenge was to identify the columns that offered the best stratification of the data for each Expert. After several rounds of optimization, it was decided to utilize the same columns that were employed in the adversarial subsampling process for each Expert.

Implementation: The training process of the model involved the use of 10 algorithms derived from Stratified Kfold. The complexity of the model and the diversity of metrics to be optimized resulted in variable training durations. However, on average, each fold iteration required approximately 110 seconds for its complete execution.

Model performance evaluation

To assess the performance of the proposed models, meticulous comparisons were conducted between the forecasted results and the actual metrics obtained during the testing phase. The metric employed was WMAPE (Weighted Mean Absolute Percentage Error).

To illustrate the model's efficacy, a comprehensive **Table 1** was prepared with complete results. This table details how the model behaves under different conditions and configurations of the 5G network, offering an extensive view of its performance.

Table 1: Performance of Simple and Hybrid Models

Strategy	Cv Exp A	Cv Exp B	Cv Exp C	Cv Final	Public LB	Private LB
Only XGB	0.04256	0.11232	0.04229	0.05730	0.0978	0.0968
Only Ridge	0.06154	0.11986	0.11712	0.08373	0.0958	0.0955
Boosted Hybrid	0.04083	0.11743	0.04165	0.05721	0.0690	0.0699

Analyzing the Cv Final and Private LB of the three models - Only XGB, Only Ridge, and Boosted Hybrid - we can observe some peculiarities.

- **Only XGB:** This model presents a Cv Final of 0.05730 and a Private LB of 0.0968. It is notable that XGB performs better in terms of Cv Final when compared to Ridge. However, it doesn't fare as well on the Private LB. This may indicate that although XGB is efficient on the training data, it may not generalize and extrapolate as well to unknown data.

- **Only Ridge:** This model has a Cv Final of 0.08373 and a Private LB of 0.0955. Even with a worse Cv Final than XGB, Ridge surpasses the latter on the Private LB. This highlights the importance of feature transformation, which appears to be a crucial factor in improving the model's generalization on test data.

- **Boosted Hybrid:** This is clearly the most robust model, with a Cv Final of 0.05721 and a Private LB of 0.0699. It not only excels on the training data but also significantly outperforms on the actual test data. This suggests that the combination of feature transformations with target adjustments yields a powerful hybrid model that delivers superior performance across all aspects.

In conclusion, it is evident that feature transformation is vital for the model's efficacy in a real-world environment. Furthermore, the boosted hybrid approach has proven to be the most effective, surpassing other models in both training and test metrics.

Table 2: Performance with and without Mixture of Experts (MOE)

Strategy	Cv Exp A	Cv Exp B	Cv Exp C	Cv Final	Public LB	Private LB
Without MOE	0.04083	-	-	0.04083	0.1026	0.1027
With MOE	0.04083	0.11743	0.04165	0.05721	0.0690	0.0699

In the **Table 2**, the efficacy of models with and without the implementation of the Mixture of Experts (MOE) is evaluated. The key point of this analysis lies in the comparison between the generalization power of the two strategies, as assessed through the Cv Final and Private LB metrics.

It was observed that the "With MOE" model displayed a Cv Final of 0.05721, which is higher than the "Without MOE" model that recorded a Cv Final of 0.04083. However, when evaluating the Private LB scores, the "With MOE" model demonstrated significantly better performance, with a score of 0.0699, compared to the score of 0.1027 for the "Without MOE" model.

This observation is especially relevant as it indicates that, despite a slightly higher Cv Final, the "With MOE" model exhibited a considerably greater generalization power on the test set. It is noted that the Cv Final and Private LB scores for the "With MOE" model are closer to each other, corroborating its higher efficacy in adapting to new conditions and unseen variables during the training phase.

5. DISCUSSION AND FUTURE WORK

One of the most crucial findings of this study lies in the need for algorithms with high generalization power to tackle the specificities of the energy consumption scenario in 5G networks. Particularly, it was observed that the test dataset includes what we term as "cold data," which are substantially different from the "hot data" used during the training phase.

This discrepancy between the two sets of data necessitates that the employed models be capable of interpreting and adapting to new information effectively. This study highlighted that the hybrid Boosted Model with an Approach Inspired by Mixture of Experts exhibits substantially more robust generalization power in test environments. It was able to interpret and extrapolate the "cold data" in an organized and intelligent manner, resulting in more precise and efficient generalization.

This result not only reaffirms the efficacy of the hybrid model with MOE in practical situations but also signals a promising path for future research. The next logical step would be to delve deeper into how replacing the Ridge algorithm with neural networks [1] could optimize the precision of predictions in the boosted hybrid model. As for future enhancements inspired by the Mixture of Experts paradigm (MOE), the focus could be on [6]:

1. Adaptive Masks: for dynamic adjustment based on the reliability of predictions.
2. Iterative Feedback: for continuous refinement of the masks.

In summary, an algorithm's ability to adapt to new data while maintaining accuracy and efficiency is of paramount importance in the practical application for predicting energy consumption in 5G networks. This study serves as an initial milestone for future investigations aiming to develop even more robust and adaptable models.

6. REFERENCES

[1] N. Piovesan, D. López-Pérez, A. De Domenico, X. Geng, and H. Bao, “Power Consumption Modeling of 5G Multi-Carrier Base Stations: A Machine Learning Approach,” in Proceedings of the IEEE International Conference on Communications (ICC), Rome, Italy, 28 May-1 June 2023. Accessed on: October 2023.

[2] GSMA, "5G Energy Efficiencies: Green is the New Black," Technical Report, November 2020. [Online]. Available: https://data.gsmaintelligence.com/api-web/v2/research-file-download?id=54165956&file=241120-5G-energy.pdf. Accessed on: October 2023.

[3] R. Holbrook, "Hybrid Models," Kaggle, October 2023. [Online]. Available: https://www.kaggle.com/code/ryanholbrook/hybrid-models.

[4] S. E. Yuksel, J. N. Wilson, and P. D. Gader, "Twenty Years of Mixture of Experts," IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 8, pp. 1177-1190, August 2012.

[5] Zindi Africa, "AI/ML for 5G Energy Consumption Modelling," 2023. [Online]. Available: https://zindi.africa/competitions/aiml-for-5g-energy-consumption-modelling.

[6] T. Zadouri, A. Üstün, A. Ahmadian, B. Ermiş, A. Locatelli, and S. Hooker, "Pushing Mixture of Experts to the Limit: Extremely Parameter Efficient MoE for Instruction Tuning," Cohere for AI, September 2023. [Online]. Available: https://github.com/for-ai/parameter-efficient-moe.