A Comprehensive Machine Learning Framework for Predicting the Energy and Economic Impact of Electric City Buses

Mailavarapu SaiLohitha[1], Dr. Bandla Srinivasa Rao[2]

[1]*PG Student, Department of Computer Science and Engineering, Teegala Krishna Reddy engineering college, India, lohithavanama@gmail.com*
[2]*Professor, Department of Computer Science and Engineering, Teegala Krishna Reddy engineering college, India, sreenibandla@gmail.com*

Abstract: The research work currently attempts to reduce the carbon emissions and make energy-efficient urban public transportation with electric buses. This research sets up a big data analytics framework with machine learning to forecast and optimize the energy consumption of electric city buses. Such system offers accurate prediction of energy economy by utilizing real-time large-scale telematics and operational information processed via batch and stream through Apache Spark. As per the objective of fast-paced transit environment subjected to continuous disturbances by traffic, weather, and vehicle load, the scalability of the framework serves as a crucial capacity for distributed computation and in-memory processing. Energy consumption can be viewed in a holistic manner charged with heterogeneous data sources. Such predictive insights enable transit agencies to undertake proactive energy strategies, model optimization on routes, and introduction of batteries that last longer into lower operational costs with reduced environmental impact. Future work will be targeted towards real-time integrated streaming tools such as Apache Kafka and Flink and deploy advanced models like LSTM and Reinforcement Learning while developing visual analytics and cloud scale. The research will explore how NLP can be subjected to use for unstructured data analysis, for instance through driver logs and maintenance reports. From intelligent transport systems, this framework is considered a great major step and indeed becomes a crucial building block towards the vision of smart energy-efficient cities.

Keywords: Electric Bus, Big Data Analytics, Apache Spark, Machine Learning, Energy Prediction, Telematics, Smart City, Real-Time Streaming, LSTM.

1. Introduction

The new trend for cities to become environmentally sustainable is changing the way urban mobility is done. In public transport systems, there is a rapid increase in the number of electric buses. Electric buses prove to be efficient replacements for old-fashioned vehicles, which emit greenhouse gases and were considered quite expensive for operation since cities are going green[1]. Managing energy consumption in electric buses is a very challenging task due to various complications theoretically ranging from traffic to meteorological changes, road gradients to even passenger grown loads. These challenges require a data-driven approach in real-time analytics generating insight and having predictive intelligence[2].
As electric buses are really good at generating data through installed sensors, conventional analytics tools become much less effective in saving time for their analysis, which would yield any actionable insight[3]. This paper proposes an intelligent framework that integrates big data analytics through Apache Spark and machine learning algorithms to forecast and optimize energy consumption for electric

buses in cities. Apache Spark's large-scale telemetry workload, in operational datasets, realized by distributed computing and in-memory processing, scalability and speed[4].
The system forecasts energy requirements accurately, thus helping route planning, efficient use of batteries. The proactive schedule for maintenance can be planned according to their needs. Heterogeneous data sources can be integrated through the use of telematics, weather feeds, and traffic data to allow a comprehensive view of the operational environment[5].
Besides this, it meets the wider desire of smart cities since the framework will develop energy-efficient, cost-effective, and environmentally friendly intelligent transport systems. Future developments will further improve the efficiency of this system in its prediction and usability areas with advanced deep learning models, real-time streaming tools, and natural language processing for unstructured data analysis. Thus, the ultimate aim of this study is to provide guidance for transit agencies in optimally operating electric fleets in a greener urban future[5].

2. Problem Statement

The transformation of public transport from conventional fuel buses to electric buses is an important mile to cover towards improving the environment and decreasing pollution in urban areas[6]. One of the major challenges in using electric buses is managing and optimizing energy consumption for these vehicles as many factors influence it, for example, veering load, driving behavior, traffic conditions, differences in terrain, and variations in weather conditions. Such variations make energy use fluctuate; thus, consumption cannot be accurately predicted to plan effective operations[7].
Conventional systems for data processing are also unable to handle the massive and varied amounts of data generated by electric bus telemetry, control unit, and external data sources. As a result, it becomes impossible for the authorities to make an efficient decision regarding route planning, battery management, maintenance scheduling, and, indeed, energy optimization[8]. Their work is limited without real-time insight due to the very low or no possibility of adjustments to changes in operations on the road.
The prototype advanced, scalable, integrated big data analytics with machine learning is needed to meet this challenge[9]. In other words, it should be able to manage, in real time or nearly real time, the very large amounts of heterogeneous data and should also provide an analytical insight to understand how the data are such to facilitate energy-efficient transit systems. In this paper, we propose a platform for such a framework by designing and implementing an open-source tool based on Apache Spark and machine learning techniques for improving decision-making around energy management of electric buses[10].

3. Related Work

Mustafa and Badr (2021) suggest a framework for real-time decision-making using big data analytics, mainly emphasizing cloud-based solutions for competitive intelligence [1]. The paper highlights the need for an integrated systemic approach toward big data analytics for timely and informed decision-making that keeps pace with fast-moving adoption environments. According to Jaiswal and Soni (2021), it elaborates further on Apache Spark's deep analytical capabilities, while providing references to competitive intelligence [2]. It discusses the way that Spark can be made effective for the large scale processing of data along with further illustrations for corporate entities in their business battle for business intelligence.
Real-time analytics is examined by Soni and Gohil (2020), incorporating Spark Streaming, which is concerned with acquiring insights from live data streams and is inevitable for timely decision-making in the context of competitive intelligence [3]. The authors, Patel and Dave (2020), also address the same subject about Apache Spark under real-time data analytics in an efficient manner processing large scale real-time data for organizations to remain permitted in changing market scenarios [4].
They have added together in Gajendran and Ramaswamy (2020) with Apache Spark as part of real-time analytic solutions by generalizing the frameworks on Scalability and Performance [5]. It is about Big Data as their work primarily focuses on the problems raised by integration of Spark and Stream Technologies to provide continuous insight from Big Data.

Finally, referring to data-driven lines of attack toward competitive intelligence, Srinivasan and Rajaraman (2020) set out how big data analytics would factor into predicting trends in the markets, with consequent strategic decisions from the insights gained [6].

Collectively, these works demonstrate how big data analytics could transform businesses and real-time processing tools such as Apache Spark into much better competitive intelligence.

4. Proposed Work

This research is basically the design of a scalable and intelligent energy prediction model for electric public transport buses integrating Apache Spark with machine learning algorithms [11]. The architecture can therefore facilitate the management of very large amounts of data generated by an electric bus, which may include telematics, sensors, and operational logs. With its distributed computing and in-memory processing features, Spark provides the construction velocity and scalability that are being required to effectively manage not huge but big data [12].

The uniqueness of the system is proved by the fact that it is able to correlate with and process various kinds of disparate types of data sources: vehicular speed, battery status, passenger load, weather, and traffic data. Thus cleaned and transformed, this data is used further to train machine learning methods including regression, decision tree, and SVM to be able to predict energy consumption per route or trip per real time [13].

Real-time prediction enables transit agencies to optimize routing and driving strategies as well as charging schedules. While it accelerates model training, real-time querying and processing of structured data is efficiently accomplished with Spark SQL and DataFrames [14].

The best of all worlds, periodic update of data and extension to incorporate real-time streams via an ecosystem of tools such as Apache Kafka or Spark Streaming: hence, it supports continuous learning and update of models, allowing the framework to evolve progressively along with dynamically changing urban mobility patterns.

The kind of optimization created through Big Data and machine learning this very solution creates will really be a game-changer when it comes to energy optimizing through proactive decision making towards smart city transport sustainability [15].

5. Implementation

Actually, on top of that, implement this model for the design of this project that will be scalable on frameworks and data-driven towards predicting the energy demand of an electric city bus through Apache Spark and machine learning techniques [16]. This entire telemetries comes from many sources: from on-board telematics units, GPS trackers and vehicle control systems to some external sources that include weather and traffic information [17]. All the datasets collected in this method are huge and heterogeneous and must undergo good treatment and preprocessing [18].

Traditionally, Spark served as the primary processing engine, owing to its capabilities for distributed computing and real-time analytics. Through SparkSQL or DataFrames, raw data would get populated into Spark through preprocessing such as missing value imputation, normalization, aligned timestamps, and feature extraction in the search of relevant features that affect energy consumption: speed, acceleration, altitude, state of charge, total onboard passengers, and environmental factors [19].

Basically, the machine learning algorithms that are going to be executed using Spark MLlib will include: Linear Regression, Decision Trees, and Gradient Boosted Trees. The data will be trained and validated on historical data, and the prediction performance would be measured by RMSE and R^2 scores.With the integration of Spark Structured Streaming, the setting up dynamic mode for the whole on-the-move system where data transactions can be treated and reacted to real-time updates would be created [20]. This will allow instantaneous updates for route optimization, battery management, and scheduling for transit operators.

The final visualization and analysis of that data could preferably be through interactive dashboards of Power BI or Tableau in an intuitive interface to further enhance the understanding and decision-making of stakeholders based on these predictions. Thus, it is end-to-end implementation as on-demand and scalable for public transportation management for energy efficiency.

6. Algorithm

1. Linear Regression : Predicting energy consumption (continuous variable) based on input features like speed, weather, passenger load, and route conditions.

$y^=\beta0+\beta1x1+\beta2x2+\cdots+\beta nxn$

Where:

$y^$ = Predicted energy consumption

x_i = Input feature (e.g., average speed, traffic delay)

β_i = Coefficients learned during training

Loss Function (MSE):

2.Random Forest Regression

Ensemble learning for robust prediction of energy under varied conditions (non-linear relations).

Prediction:

$$\hat{y} = \frac{1}{T}\sum_{T=1}^{T} ht(x)$$

Where:

$h_t(x)$ = Prediction from the t-th decision tree

T = Total number of trees

3.Gradient Boosting Machines (e.g., XGBoost)

More accurate modeling of energy consumption with feature interactions.

Objective Function:

$$L = \sum_{i=1}^{n} l(y_i y_i^{(t-1)} + f_t x_i)) + \Omega f_t$$

Where:

l = Loss function (e.g., squared error)

$\Omega(f_t)$ = Regularization term to penalize complexity

4.LSTM (Long Short-Term Memory Networks)

Use Case: Time-series prediction for energy usage patterns considering traffic and temporal patterns.

Core Equations:

Forget gate:

$ft=\sigma (Wf \cdot[ht-1,xt]+bf)$

Input gate:

$it=\sigma(Wi\cdot[ht-1,xt]+bi)$

Cell update:

$Ct=ft *C t-1+it*Ct$

Output gate:

$ot=\sigma(Wo\cdot[ht-1,xt]+bo)$

$ht=ot*\tanh(Ct)$

7. Results

The setting up of the relevant environment through the importation of Python libraries as shown in Fig. 8.1 would be the first step in this process. The whole process is important, in essence, for it guarantees all the dependencies needed to perform data manipulation, machine learning, and visualization. For their part, the users would be shown the Home Screen the main interface from which all functionalities may be accessed.

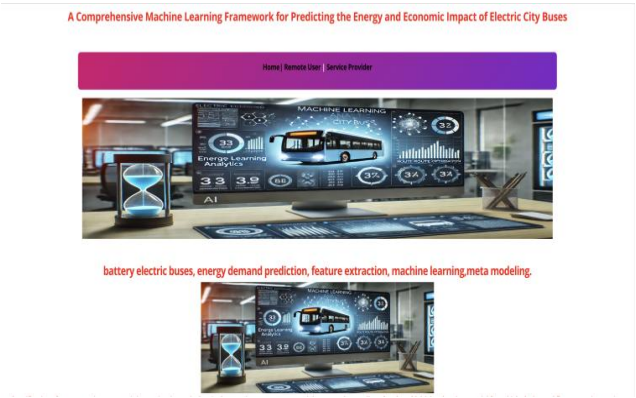

Fig 1: Home Screen

From this screen, the users can log in and view predictions and information relevant to them. This screen is the Remote User Login Page to log in and the Register Details Screen to register. The latter grants them access to the energy economy prediction options, which go along with their order.

Fig 2: Prediction Of Energy Economy Prediction Screen

Once the user is registered and logged into the system, users are allowed access to input real-time data for energy prediction is the Prediction of Energy Economy which shows the results from machine learning models applied on these data. The output presents a generalized prediction of energy consumption with respect to routes, traffic, and bus load.

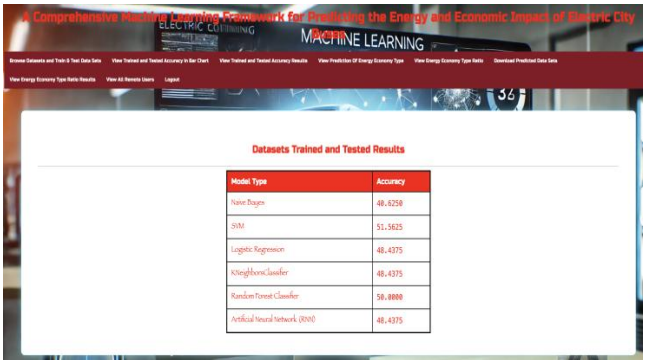

Fig 3: Datasets Trained and Tested Results

For data analysis purposes, shows bar graphical representation and visualization of energy consumption for different bus routes which can be helpful for fleet management. On the other hand, the Datasets Trained and Tested Results indicate how well the trained models are performing and the accuracy of predicting energy usage.

Fig 4: Bar Chart

At last, is a line graph that shows energy consumption trend over a time window, thus facilitating their capability to monitor and predict energy trends while planning for efficient usage of resources in real-time operations. These outputs and visualizations serve as a whistle of better decision-making and honed energy consumption policy formation.

8. Conclusion

In the present world, the combination of machine learning and big data frameworks like Apache Spark has made imbibed changes in the way we analyze energy consumption in electric public transport. The work gives a detailed outline of a scalable and intelligent architecture that is capable of huge amounts of sensor and operation data and able to predict energy economy in electric city buses. Moreover, the system exploits the unique feature of real-time heterogeneous streaming data processing to enable dynamic and data driven decision making that would improve operational efficiency and lifespan of the batteries while lowering total energy costs.

This process is further enhanced by the distributed processing and in memory computations of Apache Spark that give the necessary performance for treatment with endless streams of data and to allow speedy learning of the model to ensure timely and accurate prediction. The framework gives the insight on energy use as influenced by traffic condition, gradient of the route, weather changes and vehicle load by application of machine learning algorithms.

This forecasting ability serves as a basis for route planning, predictive maintenance, and lean and green operations, consistent with the global smart mobility vision. The proposed solution is an environmental win as well as a wallet win for transportation agencies. In conclusion, this research develops a strong data-driven tool for optimizing energy consumption in electric buses that paves the way forward for smarter and greener cities.

9. Future scope

The system proposed is being laid down as a firm groundwork for energy prediction in electric mobility, yet ample grounds lie for its further improvement. One such improvement could be to have real-time data streaming considered, whereby systems like Apache Kafka and Spark Structured Streaming can be employed to allow system dynamics to be in sync with not only traffic situations but weather events and driving behavior alike, thus putting energy prediction on truly reactive modes.

Another interesting option is to explore more deep learning models such as LSTM or reinforcement learning that capture complex temporal patterns and accentuate long-term predictive accuracy. Such models are also capable of tuning prediction results in view of vehicle- or route-specific attributes.

Dashboards formulated in Tableau or Power BI will make richer visual analytics to support fleet-performance monitoring by non-tech users. Cloud-native instances hosted over AWS, Azure, or Google Cloud will allow better scalability and cost optimization once the fleet is on the increase.

Additionally, Natural Language Processing finds application in extracting insights from qualitative text documents, such as maintenance logs, driver feedback, and commuter reviews—thus supplementing quantitative analysis in the predictive system. The endeavor will, therefore, equip the predictive system

to be intelligent, flexible, and relevant towards an intelligent energy use and sustainability of public transport in the future.

References

1. Mustafa, A. H. T., & Badr, S. S. (2021). Big Data Analytics for Competitive Intelligence: A Framework for Real-Time Decision Making. Journal of Cloud Computing and Security.
2. Jaiswal, A., & Soni, H. (2021). Big Data Processing with Apache Spark: Advanced Analytics for Competitive Intelligence. CRC Press.
3. Soni, H., & Gohil, P. (2020). Real-Time Big Data Analytics with Spark Streaming. Springer.
4. Patel, P., & Dave, M. (2020). Real-Time Data Analytics with Apache Spark. Apress.
5. Gajendran, K., & Ramaswamy, P. (2020). Building Real-Time Analytics Solutions with Apache Spark and Kafka. Wiley.
6. Srinivasan, S., & Rajaraman, S. (2020). Data-Driven Approach to Competitive Intelligence: Leveraging Big Data Analytics. Springer.
7. Xu, M., & Li, Z. (2019). Big Data Analytics for Competitive Intelligence. Elsevier.
8. Kim, Y., & Lee, K. (2018). Competitive Intelligence in the Age of Big Data: Leveraging Advanced Analytics to Gain Business Insights. Springer.
9. Chambers, B., & Zaharia, M. (2018). Spark: The Definitive Guide: Big Data Processing Made Simple. O'Reilly Media.
10. Wen, J. R., & White, T. (2018). Apache Spark for Data Science. O'Reilly Media.
11. Manogaran, G., Shanmugam, M., & Babu, S. (2017). Big Data Analytics and Competitive Intelligence: Applications, Opportunities, and Challenges. Springer.
12. Chakrabarti, S. (2017). Apache Spark for Data Engineers. Packt Publishing.
13. Laskowski, J. (2017). Mastering Apache Spark: A Beginner's Guide to Big Data Analytics with Apache Spark. Packt Publishing.
14. Goyal, M. A., & Sharma, J. (2016). Exploring Competitive Intelligence with Big Data Analytics. Springer.
15. Marr, B. (2015). Big Data in Practice. Wiley.
16. Guller, M. (2015). Big Data Analytics with Spark: A Practitioner's Guide to Using Spark for Large-Scale Data Analysis. Packt Publishing.
17. Awad, E. M., & Qamar, U. (2014). Competitive Intelligence: A Framework for Web Mining and Data Mining in Competitive Intelligence. Springer.
18. Zikopoulos, P., & Eaton, C. (2011). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill.
19. Kumar, S., & Mishra, S. (2018). Data Analytics for Competitive Intelligence: An Introduction. Springer.
20. De, A., & Saha, A. (2019). Machine Learning and Data Analytics in Competitive Intelligence. Springer.