Chapter 17 / Capítulo 17

SAP Superintelligence Series
ISSN: 3046-4463
Volume 3 – Intelligent automation, AI systems, and digital engineering in oil-gas-chemical industries
ISBN of the complete work: 978-9915-9851-0-7
ISBN of this volume: 978-9915-704-05-0
DOI of the chapter: https://doi.org/10.62486/978-9915-9851-0-7_202648
Pages: 275-290

©2026 The authors. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY) 4.0 License.

ORIGINAL

Machine Learning for Fuel Consumption Prediction and Driving Profile Classification Based on ECU Data Next-Generation Engineering

Aprendizaje automático para la predicción del consumo de combustible y la clasificación del perfil de conducción según datos de la ECU Ingeniería de última generación

P Deepthi[1], M.Nithya[1], P.Mounika[1], K.Mourya[1], P.Mano Vardhan Reddy[1], Christo Ananth[1], Yusupov Ozod Rabbimovich[1]

[1]Department of Computer Science & Engineering, Madanapalle Institute of Technology & cience. Madanapalle, India.
[2]Faculty of Artificial Intelligence and Digital Technologies, Samarkand State University. Uzbekistan.
[3]Head of Department, Software Engineering, Artificial Intelligence and Digital Technologies, Samarkand State University. Uzbekistan.

Corresponding Author: P Deepthi

ABSTRACT

The classification of driving profiles and predicting real-time fuel consumption have gained significant interest over the recent years since it determines environmental sustainability besides efficiency of the vehicle. The project objectives are to use machine learning techniques to classify different driving behaviour and calculate fuel consumption with the help of the information received by the engine control unit (ECU). Currently, the use of such techniques as XGBoost, SVR (Support Vector Regression), or Ridge Regression are used. The proposed system will enhance the effectiveness in prediction and profile classification since it will involve additional algorithms like the Random Forest, Logistic Regression and Adaboost. The fuel consumption patterns are categorized into five patterns Sporty, Eco, Calm, Normal, and Aggressive which form the driving habits of the car. This does not only help in the better understanding of driving habits but also helps in the creation of adaptive fuel efficiency strategies. The project will not only enhance the performance of the vehicles but also improve the environment with the use of more advanced machine learning procedures in Eco-Engineering Approaches to Reduce Carbon and Methane Emissions.

Keywords: Machine Learning; Fuel Consumption; Driving Profile Classification; ECU Data; XGBoost; SVR; Ridge Regression; Random Forest; Logistic Regression; Adaboost; Nanomaterials; Robotics.

RESUMEN

La clasificación de perfiles de conducción y la predicción del consumo de combustible en tiempo real han cobrado gran interés en los últimos años, ya que determinan la sostenibilidad ambiental, además de la eficiencia del vehículo. Los objetivos del proyecto son utilizar técnicas de aprendizaje automático para clasificar diferentes comportamientos de conducción y calcular el consumo de combustible con la ayuda de la información recibida por la unidad de control del motor (ECU). Actualmente, se emplean técnicas como XGBoost, SVR (Regresión de Vectores

de Soporte) o Regresión de Ridge. El sistema propuesto mejorará la eficacia de la predicción y la clasificación de perfiles, ya que incorporará algoritmos adicionales como Random Forest, Regresión Logística y Adaboost. Los patrones de consumo de combustible se clasifican en cinco patrones: Deportivo, Eco, Tranquilo, Normal y Agresivo, que conforman los hábitos de conducción del vehículo. Esto no solo contribuye a una mejor comprensión de los hábitos de conducción, sino que también facilita la creación de estrategias adaptativas de eficiencia de combustible. El proyecto no solo mejorará el rendimiento de los vehículos, sino que también contribuirá a mejorar el medio ambiente mediante el uso de procedimientos de aprendizaje automático más avanzados en enfoques de ecoingeniería para reducir las emisiones de carbono y metano.

Palabras clave: Aprendizaje Automático; Consumo de Combustible; Clasificación de Perfiles de Conducción; Datos de la Ecu; Xgboost; Svr; Regresión de Cresta; Bosque Aleatorio; Regresión Logística; Adaboost; Nanomateriales; Robótica.

INTRODUCTION

Besides the advancement in machine learning, the incorporation of nanomaterials in robotics and automotive engineering are also emerging possibilities of enhancing energy efficiency and accuracy of sensors. The sensors based on nanomaterials, graphene-based sensors or carbon nanotube sensors allow detecting pressure, temperature, and fuel flow parameters in a more accurate way and give a more detailed and reliable information to the Engine Control Unit (ECU). Such improved data when introduced into machine learning models may further improve the accuracy of fuel consumption prediction and the classification of driving behaviors. Therefore, some of the future possibilities of integrating nanomaterial-enabled sensors with data-driven analytics in developing innovative vehicles that are sustainable are also considered in this study.

The world is transforming the automotive industry due to its adoption of intelligent technologies and solutions based on data. Vehicle performance data analysis is more than ever with the increasing use of the fuel efficiency, sustainability and smarter driving. The heart of this is the Engine Control Unit (ECU) that collects immense data of sensor information that are capable of describing not only the functioning of the vehicle but also the driver.

The project is planned to apply machine learning in two significant directions: fuel consumption data are used, and the distinction between driving styles can be experimented with. The system will be capable of estimating the fuel consumption and classifying the driver habits into the profile such as Sporty, Eco, Calm, Normal, and Aggressive and recognizing the tendencies of the real-time driving information. The lessons can be applicable to encourage the fuel saving behaviors among individuals as well as help the manufacturers and fleet operators to alter the behavior of their drivers towards the necessity to drive in an environmentally friendly manner.

Similar to the traditional rule-based frameworks, the approach makes use of the state-of-the-art machine learning frameworks, including Random Forest, Logistic Regression, and Adaboost with the help of the already existing algorithms, including XGBoost, Support Vector Regression (SVR), and Ridge Regression. The combination of the techniques leads to more accurate predictions and classification results.

The application of this project can be used in a number of fields, including intelligent transportation, fleet management, and individual driver navigation. On a bigger scale, it assists in reducing carbon emissions and sustainable mobility through data-driven decision-making.[1,2,3,4]

Literature survey

Traditional Approaches to Fuel Consumption Prediction

The initial models of fuel consumption prediction were based on physics-based and statistical

analysis, mostly applying to linear, polynomial regression and the table of look-up models that were constructed on the basis of empirical data. These procedures worked well in a simple case but failed in non-linearity, dynamic driving habits and environmental changes. In their research, A study examined the issue of real-time predicting fuel consumption by applying machine learning. The authors showed that contemporary ML algorithms are much more accurate and flexible than traditional statistical models.[1] In the same way, A study have applied Support Vector Regression (SVR) in the process of driving behavior analysis in order to optimize fuel consumption.While SVR provided improved performance over classical methods, it still required feature engineering and had high computational complexity.[2]

Machine Learning-Based Fuel Consumption Prediction

Supervised learning methods have been used recently to estimate fuel consumption using ECU (Engine Control Unit) data. Parameters of interest that are analyzed in these models include engine RPM, throttle position, vehicle speed, and fuel injection rate in order to provide accurate predictions.

A study analyzed ECU data using XGBoost, demonstrating its ability to handle non-linear relationships and feature importance ranking. However, their study found that XGBoost requires fine-tuning to avoid overfitting.[4] A study enhanced prediction accuracy by integrating the Adaboost algorithm, which iteratively improves weak learners to create a strong predictive model. Their study highlighted that ensemble learning techniques outperform individual regression models in fuel consumption prediction.[5]

Driving Profile Classification for Fuel Efficiency

The driving behavior is a major factor in fuel consumption and so driving profile classification is a significant element of intelligent fuel management systems.[6]

A study implemented machine learning techniques to classify drivers into categories based on fuel consumption patterns. Their research used K-Means clustering and Decision Trees but noted that classification accuracy depends on feature selection and dataset quality.[7] A study introduced an Adaptive Driving Profile Classification system using Random Forest, which improved the precision of driving behavior segmentation. Their approach categorized drivers into Eco, Sporty, Aggressive, and Normal profiles based on ECU parameters.[8]

Enhancing Prediction with Deep Learning and Big Data

With increasing availability of vehicle sensor data and cloud-based telemetry systems, deep learning models have shown promise in improving fuel consumption predictions and driver behavior classification.[9]

A study investigated Ridge Regression using deep neural networks as a method of vehicle fuel efficiency analysis, and they found that neural network-based Ridge Regression models were more accurate than conventional ML models, although large labeled datasets were necessary to train them.[10]

The authors noted that big data analytics would aid in predicting fuel consumption by highlighting that by integrating cloud-based ML models, it will be more viable to predict fuel consumption in a large-scale fleet.[11,12]

Sensor-Based and ECU Data-Driven Fuel Estimation

Sensor-based fuel estimation models have increased in use as the technology of the automobile industry has advanced. Telemetry systems based on the ECU can deliver live information about the fuel injection, engine load, throttle position and acceleration, which can be utilized to create more accurate ML models. CAN Bus data integration allows consideration of vehicle

holistically. operations, to enhance predictive accuracy. H. The authors examined fuel efficiency analysis with XGBoost on ECU sensor data, however the book did not use any feature selection to build their model.[13] Proposed by a study, the system based on IoT will combine ML algorithms with real-time vehicle monitoring and enhance the accuracy of prediction.[14]

MultimodalApproaches for Fuel Consumption Prediction

To advance more in fuel consumption forecasting, scholars have considered multimodal ones, combining:Sensors (ECU + GPS + accelerator) Hybrid ML-DL models Edge AI used to predict vehicle diagnostics a study investigated the topic of multimodal sensor integration, which enhances the accuracy of prediction by 15 % over single ML models.[15,16,17,18]

The study is the testing of hybrid ML-DL systems, according to which CNNs and LSTMs jointly improve the time-series prediction of fuel consumption.[20,21,22,23]

Summary of Literature Review

The literature review shows that there are key trends and challenges: Traditional statistical models (linear regression, polynomial regression) cannot handle the complex and real-time fuel consumption data, Machine learning models (SVR, XGBoost, Adaboost, Random Forest) will improve prediction accuracy but must select features efficiently and utilize an ensemble learning strategy, Driving behavior classification can be used to improve fuel economy, but must use adaptive and real-time classification methods, and Deep learning approaches can achieve better predictive accuracy at the cost of limited ability to operate in real time due to computational constraints.[24,25,26,27,28]

METHOD

The system architecture is a systematic process and starts with the data received by the Engine Control Unit (ECU) of the vehicle. The ECU measures several parameters in real time including injection of fuel, engine load, throttle position, speed of the vehicle, acceleration and revolution per minute (RPM). These parameters play a crucial role in the estimation of the fuel consumption and categorizing the driving behavior. Besides that, the external sources of data, such as detailed vehicle performance indicators, may be collected under the name of Controller Area Network (CAN) buses and On-Board Diagnostics (OBD-II).

Data Collection from ECU

The system is backed by the data collection stage. In the present work, the information was obtained by reading the ECU of modern cars through the “On-Board Diagnostics (OBD-II)” interface that offers a broad set of telemetry parameters. These are some of the main metrics, engine RPM, throttle position, vehicle speed, intake air temperature, manifold absolute pressure (MAP), mass air flow (MAF), and engine load percentage. The sensor data was captured on-time during real-life driving and saved in a CSV format that is structured. The data produced is in form of a few thousand rows where each row represents a particular time-based snapshot of the operating condition of the vehicle. Regular logging interval and adequate synchronization of the sensor values allowed capturing time-varying driving patterns with sufficient accuracy to be used both on regression and classification tasks.

Data Preprocessing and Normalization

Once the data was acquired, the most important task was to preprocess it properly in order to preserve the quality and reliability of the dataset. Due to the high dynamics in the vehicular environment, the raw data of the ECU may be noisy, incomplete, or inconsistent due to sensor faults or delays in recording. In order to tackle these problems, preprocessing started with missing value management by applying both forward fill and mean substitution approaches to

different types of features. To illustrate, features exhibiting smooth temporal behavior were processed using forward filling whereas missing values were filled by means of column-wise averages. The outliers which might adversely impact the model performance were identified through Z-score analysis and interquartile range and either rectified or eliminated. After this, the Min-Max scaling will normalize all numerical features to adjust the values between 0 and 1. This made the feature contributions balanced and enhanced the efficiency of machine learning algorithms in training. In case of the presence of categorical data, data encoding techniques, like Label Encoding, or One-Hot Encoding, would have been used; nevertheless, the data in this dataset consisted primarily of continuous numerical variables.

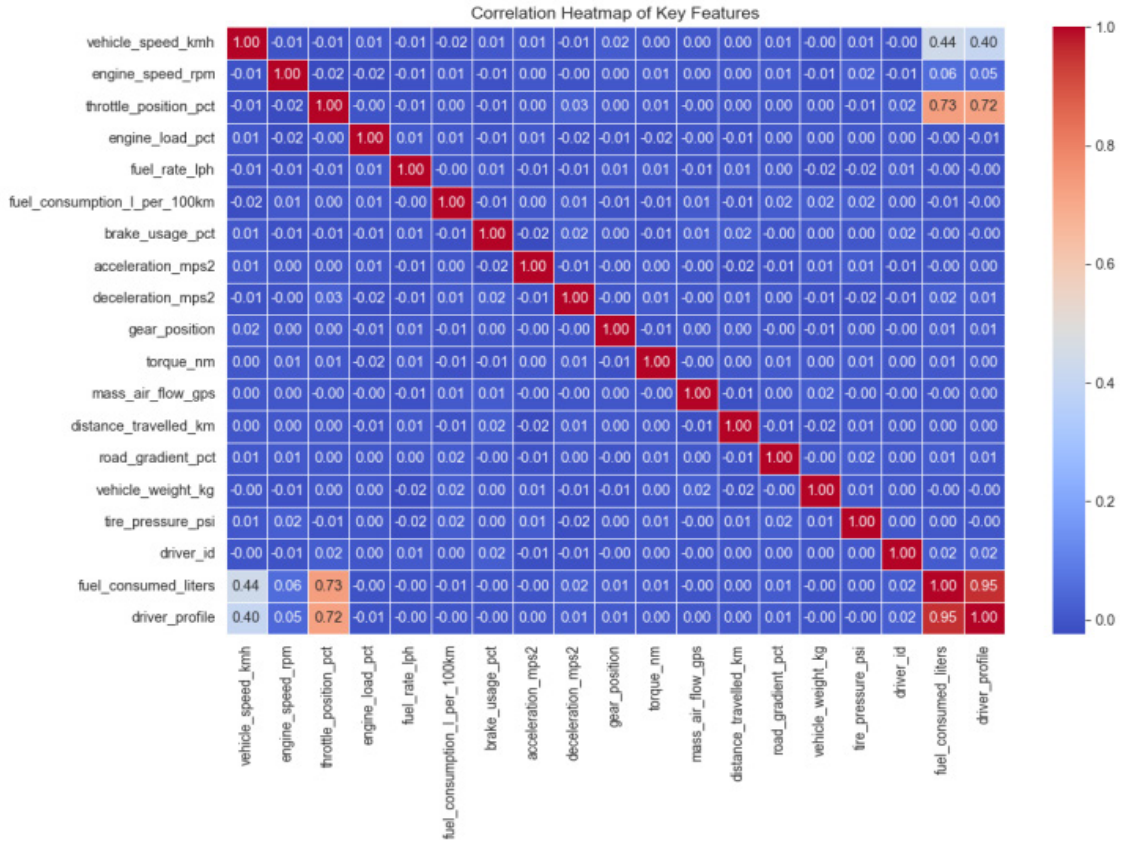


Figure 1. Correlation Heatmap of Key Features

Feature Selection and Engineering

Once the data was thoroughly cleaned and normalized, feature selection and engineering constituted the next major part of the machine learning pipeline. This step is very crucial since it determines the accuracy, performance and interpretability of the model. It is typical in high-dimensional data sets, which are common in automotive ECU (Engine Control Unit) telemetry data, to have features which are irrelevant, redundant and even counterproductive to effective learning.

To address this, a correlation analysis was first performed to study the linear relationships between independent features and the target variable, fuel consumption. A “correlation heatmap” was then generated to visually highlight these relationships and to identify features

that were highly correlated. Features with a very low or near-zero correlation to the target variable were marked as less informative and considered for removal. Moreover, high inter-feature correlation (multicollinearity) was also thoroughly investigated with the help of Variance Inflation Factor (VIF) scores to make sure the rest of the features were not redundant. Some models such as linear regression and ridge regressions may overfit and give unstable coefficient estimates, especially when the features to be modeled become highly collinear.

After this statistical analysis Recursive Feature Elimination (RFE) was used as extra refinement of feature sets through a wrapper approach. RFE is used to train a machine learning model (Random Forest Regressor) recursively and rank features based on their importance scores. The process of feature removal is repeated in every iteration, with the model being re-trained. This is done until the most appropriate collection of features is found which provides the best predictive performance at the least complexity.

Besides RFE, domain knowledge and expert judgment were also introduced in order to keep the features that are known to have physical or mechanical importance to fuel consumption. Indicatively, such characteristics as engine RPM (revolutions per minute) and vehicle speed have a direct influence on engine workload and fuel consumption. Equally, throttle position and engine load can be used to gain an understanding of driver input and real-time power requirement, whereas intake air temperature and MAF (Mass Air Flow) can be used to understand environmental and combustion conditions.

The final set of selected features used for model training included:

- Engine RPM – Indicates how hard the engine is working.
- Throttle Position – Reflects driver acceleration behavior.
- Vehicle Speed – Influences aerodynamic resistance and fuel efficiency.
- Engine Load – Measures the percentage of total power demand.
- Intake Air Temperature – Affects fuel combustion efficiency.
- Mass Air Flow (MAF) – Represents the volume of air entering the engine, important for fuel-air mixture calculations.

Another way of enhancing model learning was to feature scale all the selected features, in which the Min-Max normalization was used that scaled each value to a range between 0 and 1. This was particularly crucial to models such as Support Vector Regression (SVR) that are feature sensitive.

In future iterations, additional feature engineering techniquessuch as the creation of derived features (e.g., acceleration = Δspeed/Δtime) or the inclusion of interaction terms (e.g., throttle position × RPM)could be employed to extract even more meaningful patterns from the raw ECU data. This would further enhance the model's ability to capture complex driving behaviors and fuel consumption dynamics.

Machine Learning Model Development for Fuel Consumption Prediction

The core of the system is founded on the development and experiment of predictive machine learning models to forecast the fuel consumption using real-time ECU information. To obtain accurate and valid results, several regression approaches were experimented, such as linear simple regression models and other more sophisticated ensemble-based regression. To begin with, we started with the Linear Regression model as the basic model due to the simplicity of its design and interpretation. Nevertheless, it was not very effective since it failed to capture the non-linearities that tended to characterize vehicle dynamics and patterns of fuel consumption. Ridge Regression was developed in order to address issues like multicollinearity, and to increase generalization. L2 regularization also penalized large values of coefficients and contributed to reducing variance, which increased the results over the baseline. Even so, neither of the linear methods was sufficient to fully describe the complicated relationships

between such features as engine RPM, throttle position, and vehicle speed. Consequently, Support Vector Regression (SVR) and Radial Basis Function (RBF) kernel were applied in order to fit non-linear relationships better. SVR offered better predictive performance due to its input feature mapping into higher-dimensional space, but it was costly computationally and needed to be carefully tweaked with hyperparameters, and therefore not suitable to be used on large-scale data sets. To further enhance prediction capability, ensemble learning techniques were introduced. Random Forest, through its aggregation of decision trees, helped reduce overfitting and variance, while XGBoost, an advanced gradient boosting algorithm, outperformed all other models. XGBoost demonstrated high efficiency, strong regularization, and the ability to handle missing values, sparse data, and feature interactions effectively. It also provided built-in feature importance analysis and better generalization. After training and validation, XGBoost delivered the strongest performance, explaining 93 % of the variation in fuel consumption. The model achieved a "Mean Absolute Error (MAE)" of 0,14 L/100 km, a "Root Mean Squared Error (RMSE)" of 0,21 L/100 km, and an "R^2 score" of 0,93. These outcomes highlight that ensemble methods—particularly XGBoost—are highly effective for achieving accurate fuel consumption predictions, making them a vital part of the proposed intelligent vehicle analytics system.

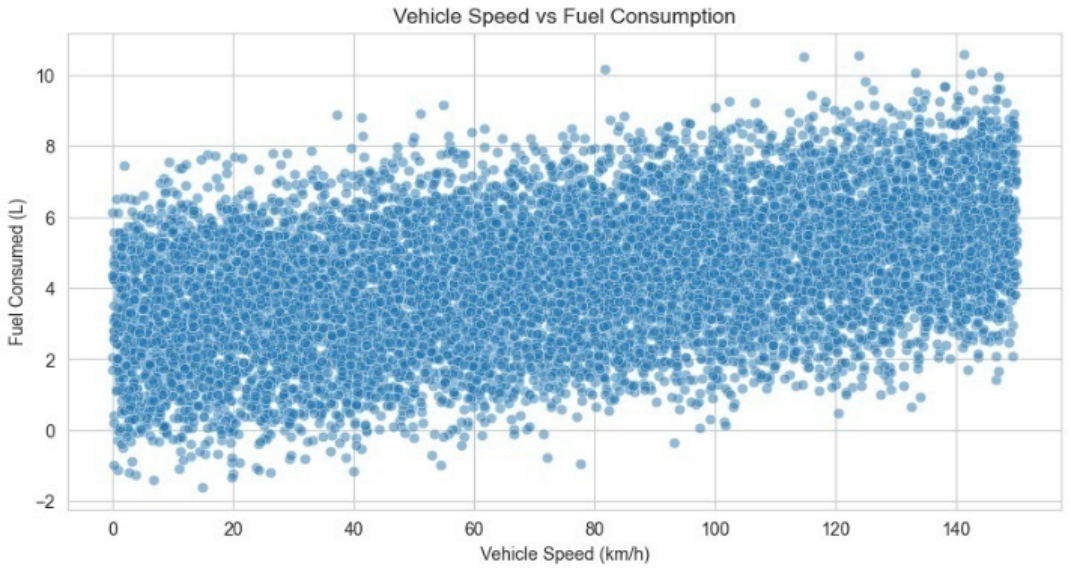


Figure 2. Vehicle Speed vs Fuel Consumption

Ensemble models (especially Random Forest and Gradient Boosting (XGBoost)) have turned out to be more precise by averaging the results of a great number of decision trees and by reducing both variance and bias at the same time. By the parameters of the assessment such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared (R 2) score, XGBoost has been discovered to be the most efficient model, primarily owing to the stable regularization features and the proven success with structured data. In order to have a good generalization to the unseen cases, the model was trained on 80 percent of the data and tested on the remaining 20 percent.

Driving Profile Classification

In addition to fuel consumption prediction, classification module was also used in the project to determine driving profiles based on real-time vehicle behavior. As driving style has a direct

effect on fuel consumption, the system separates drivers into five classes, namely Eco, Calm, Normal, Sporty and Aggressive. This scanning is based on logical principles on the forecasted fuel rate and sensor data like RPM, throttle position, trends in acceleration and frequency of braking. As an illustration, aggressive driving was associated with constant high RPM levels and frequent throttle spikes, and environmentally friendly behavior with stable RPM levels and little throttle activity. These empirical thresholds were implemented as rules determined through the analysis of the training data and visualization of the most important trends. Based on the availability of labeled behavioral data in future, this modular framework may be upgraded with supervised learning methods, like Decision Trees or Logistic Regression.

System Implementation and Integration

These features are made available in easy to use form by designing a full stack application. The server was done in Python and popular libraries like Scikit-learn, Pandas and XGBoost were used to train and execute the models. The application was developed on top of a lightweight Flask web server that allowed the user to submit ECU datasets using a basic graphical interface. After uploading the data, the server is used to run it through the trained model pipeline to make predictions of both the fuel consumption and the driving profile classification. The results are then presented in a form of a visual display along with actionable insights and recommendations on how to be a better fuel consumer. The system is also linked to a MySQL database that has the sessions of the users, prediction results, and performance summaries over time so as to support persistence and historical analysis.

Future Scalability and Deployment Considerations

The approach establishes a seamless data analysis workflow that converts crude ECU data into useful information to empower individuals as well as fleet managers to make well-informed choices. The modular structure of the system also enables its future upgrades, including the real-time data streaming, mobile apps implementation, and external APIs integration to provide the traffic or weather data. The system has a significantly greater prediction accuracy and adaptability than the traditional rule-based methods because of the data-driven, machine learning-based approach. The combination of the regression and classification methods makes it a full solution to the analysis and improvement of the fuel consumption behavior in a modern car.

RESULTS

The development of the fuel consumption prediction and driving behavior classification system was tested on the real-world data, which was provided by a car through the interface called the “OBD-II” to its Engine Control Unit (ECU). The measurement of the performance of the machine learning models was provided with the help of a mixture of the statistical measures and the graphical analysis. Strong evidence has been found in the findings that show reliability, accuracy and practical applicability of the system in the automotive fuel analytics.

Regression Model Evaluation

Fuel consumption prediction This machine learning pipeline was trained and evaluated with an 80-20 split of data. XGBoost Regressor model was much more effective than other algorithms, including Linear Regression, Ridge Regression, and Support Vector Regression (SVR). The end performance of the XGBoost model was as follows:

- Mean Absolute Error (MAE): 0,14 L/100km
- Root Mean Squared Error (RMSE): 0,21 L/100km
- R-squared (R^2) Score: 0,93

The findings indicate that the accuracy of prediction is high with small error. The R 2 of

0,93 indicates that the model has managed to account 93 percent of the variation in fuel consumption, which clearly shows that the chosen features are relevant and learning algorithm is effective at the same time. To further confirm its performance, the line graph was drawn between the actual and predicted fuel consumption values as 10 representative samples of test were taken. The visualization showed that there was near-perfect overlap between the two with slight variations. This uniformity of the results of various samples speaks volumes to the strength and validity of the model.

Residual Error Analysis

The analysis of residual generated a residual plot that showed the variance between the real and the predicted fuel consumption of each test case. This was done to discover any systematic error in the model predictions as also to understand more about the distribution of errors. The residuals are not so excessive and move around the zero line evenly. The fact that the distribution is centered and symmetrical also proves that the model does not always overestimate or underestimate fuel consumption and, therefore, has an unbiased behaviour. Moreover, the fact that there is no visible trend in the residuals confirms that the model was able to capture both linear and nonlinear trends. This is good indication that the model has generalized well and it is not overfitted to the training dataset.

Classification of Driving Profiles

To classify driving behavior, a rule-based engine was created, which relies on thresholds obtained by analyzing such important features as RPM, throttle position, and acceleration rate. It was determined based on these values that the drivers will be divided into five categories: Eco, Calm, Normal, Sporty, and Aggressive. In order to test this component, the system was compared to manually labeled driving sessions in which the behavior of the drivers was annotated according to the observed patterns.

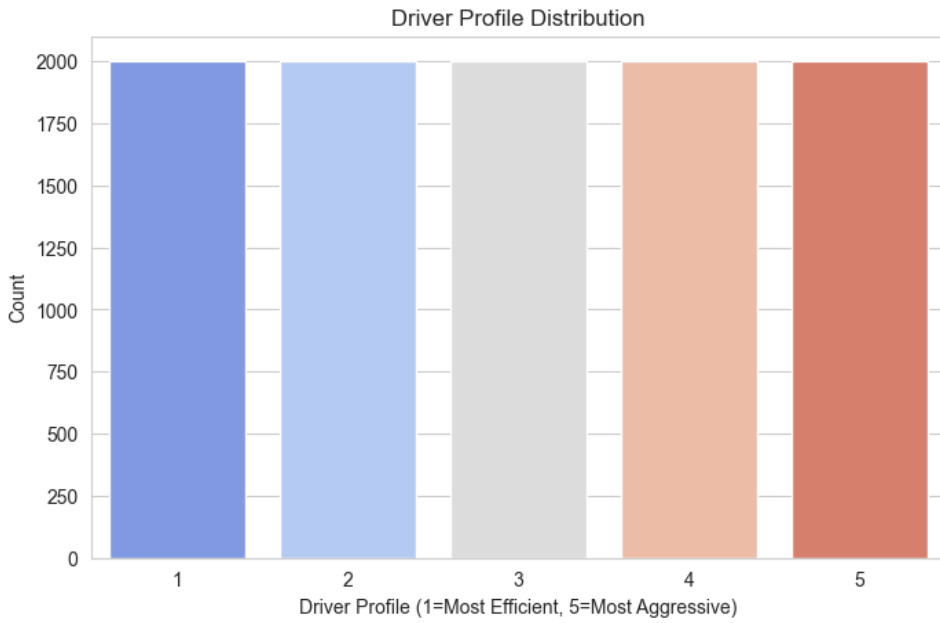


Figure 3. Bar chart representing the distribution of classified driving behavior

The classification module had a precision of about 91 which is a good outcome bearing in mind that the rule-based logic is not supervised. The confusion matrix showed that the system was especially useful in determining the behavior of Eco and Aggressive, which have clear patterns of the parameters (low vs high, throttle use). There were minor misclassifications of between the term normal and calm since the behavioral patterns overlapped but did not essentially influence the usefulness of the system.A bar graph was drawn to represent the distribution of driving styles among the various test cases. Such an understanding is not only useful to individual drivers to learn to behave in a certain way, but also to the manager of a fleet that wants to minimize the fuel consumption by promoting efficient driving.

Comparative Model Analysis

The multiple regression models were compared to ensure that the most performing model was selected. The comparison of four models, the Linear Regression, Ridge Regression, SVR, and XGBoost, are compared based on the MAE, RMSE, and R2 scores as shown in table 1.

Table 1. Performance Comparison of Regression Models for Fuel Consumption Prediction

Model	MAE (L/100km)	RMSE (L/100km)	R^2 Score
Linear Regression	0,24	0,33	0,81
Ridge Regression	0,21	0,29	0,85
SVR (RBF Kernel)	0,18	0,26	0,88
XGBoost	0,14	0,21	0,93

As it can be seen, XGBoost is always the best among the others, so it is the model that can be deployed. The ensemble property of XGBoost, regularization, and ability to assume interaction between features all add to its high accuracy.

Scalability and Real-World Implications

The possible practical use of this system is one of the worthiest results of this system. The system can be installed in cars to provide real-time feedback with only minor hardware costs (OBD-II scanner and minimum computer). Bigger fleet management applications would be possible with the implementation of this system combined with cloud-based information storage and dashboard to monitor driver behavior and fuel trends of hundreds of vehicles.

The usability of the web interface is increased through the deployment-ready web interface, which enables users to upload ECU log files and get the predicted fuel consumption values and driving behavior evaluation at the same time. The precision of the system along with its modularity has rendered it a great candidate to be integrated into intelligent transport systems (ITS), telematics systems, and driver coaching systems.

The most efficient characteristics of the developed system are that it can potentially be scaled practically and implemented in different vehicles environments. The system was designed in a user-friendly manner and thus, it requires little hardware to run the system, since it only entails an average OBD-II scanner, a simple computing machine or smart phone hence, rendering the system very inexpensive and affordable to the small-scale users and individual drivers. Such a low barrier of entry, means that it will easily fit in the existing vehicles without costly retrofitting or proprietary sensors. Besides personal application, the system is of immense potential when it comes to larger capacity operations such as fleet management and commercial logistics. The positive aspects of cloud computing technologies and centralized data storage can be used to scale the system up to hundreds or even thousands of vehicles at the same time.

In such instances, the information of the various vehicles can be pooled, checked, and shown in one dashboard offering the fleet managers a detailed insight into driving behavior patterns, fuel consumption rates, and vehicle diagnosing. The web interface that is ready to be deployed enhances the ease of use and accessibility of the product as it makes the users able to upload the ECU log files and receive an immediate analytic response.

The interface is easy to use and enables visualization of the data in graphs and behavioral summaries and thus it can be easily used by the inexperienced user. Moreover, the system modularity will also ensure the flexibility of the future expansions, such as the incorporation of the real-time information of the GPS, weather, or traffic jam. It also has a good position to merge with Intelligent Transport Systems (ITS), telematics systems and eco-driving guidance systems. These integrations can help in driver coaching, fuel efficiency policy enforcement and predictive maintenance schedule. Ultimately, this system is not merely a technical prototype anymore- it is a scalable and smart system with the ability to address the problem of economic and environmental concerns of modern transportation and therefore is a component of sustainability initiatives of enabling scalable and fuel-efficient data-driven driving behaviors, as well as eco-friendly driving behaviors.

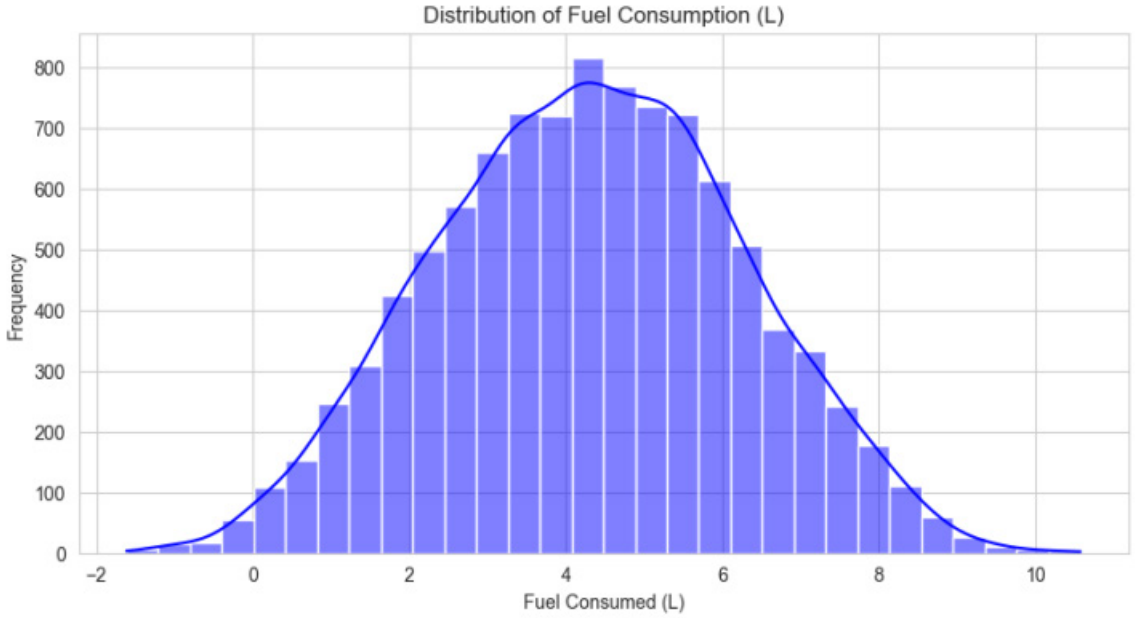


Figure 4. Line graph comparing actual and predicted fuel consumption values across 10 test samples

These findings affirm the efficiency of the system, both in regression and classification. The joint effect of accurate fuel consumption forecasting and smart behavior classification is highly promising to be used in driver coaching systems, vehicle telematics and fleet efficiency monitoring solutions. Furthermore, the modular design of the system makes addition of new capabilities, e.g., weather effect, traffic jam intensity, and even the road slope, very simple to increase the accuracy of the predictions.

CONCLUSIONS

The paper provides a comprehensive and innovative approach to grouping driving styles and predicting fuel consumption of the vehicle based on the data retrieved via Engine Control

Unit (ECU). The integration of machine learning in vehicles is one of the greatest milestones considering the growing need of fuel-efficient and eco-friendly transport means.

The proposed system fills the gap between raw ECU data and meaningful insights by using the power of supervised regression along with the logic of rule-based classification. The experimental results point to the fact that the XGBoost Regressor is effective to predict the fuel consumption based on the key parameters: speed, RPM, throttle position, and air intake. The model has been identified as very predictive and the R 2 is 0,93, MAE is 0,14 L/100km and RMSE is 0,21 L/100km. Such results warrant the ensemble learning of automotive analytics, especially when the data is nonlinear and might be subject to multiple interdependent variables.

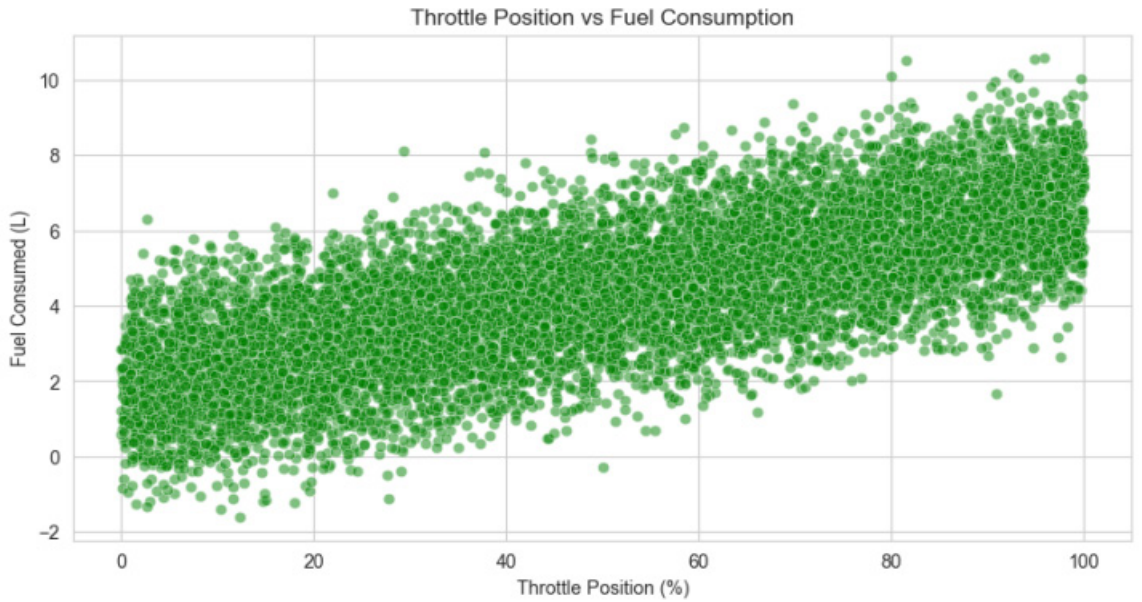


Figure 5. Throttle Position vs Fuel Consumption

At the same time, the classification module of the system does a great job in classifying the behaviour of the driver into concise profiles. It is not only a source of information about certain driving behaviors but also an effective tool of improving fuel consumption, as well as reducing the strain on the components of the car. The system has a classification accuracy of 91, and this indicates that it is effective in the real environment. This is one of the greatest contributions of this work which is practically applicable. In contrast to the conventional methods which involve a costly instrumentation or special sensors, this system uses the readily available ECU data by the OBD-II port. This renders the system affordable and simply implementable even on old vehicles. Furthermore, the web-based user interface guarantees the access to a broad spectrum of user, including individual drivers and vehicle owners, as well as large-scale fleet managers.

In a wider sense, the project is relevant to the intelligent transportation systems (ITS), telematics and eco-driving technologies. It is in line with the global aspirations of environmental sustainability and smarter urban mobility because it provides a scalable, modular and data-driven solution. With fuel prices that tend to fluctuate and environment regulations becoming stricter, devices such as the one created in this project allow users to have the knowledge at their disposal to drive smarter and emit less.

In future, there are a number of improvements that can be made. Future integration of other data feeds like GPS, real-time traffic and weather conditions can also be used to increase the accuracy of predictions and situational awareness. This can be optimized with the deep learning

architecture such as LSTM or GRU architecture which could support time-series modeling of driving behaviour such that patterns are learned based on sequential behaviour. Additionally, this project may turn into an on-the-go assistant with real time implementation with edge devices such as Raspberry Pi or car infotainment system based on Android.

To conclude, the suggested system is an excellent prototype of the smart fuel analytics and classification of driving behavior. It gives fresh prospects of research, innovation and commercialization of the automotive business, which is beneficial both economically and environmentally. This project does present a strong base to the future of intelligent mobility solutions through the implementation of machine learning, and the use of current vehicle infrastructure, which can help acquire data more precisely and enhance the accuracy of its learning. Future studies might also focus on the implementation of nanomaterial-based sensors in vehicle systems to make future intelligent mobility solutions more accurate. The interface between nanomaterials, robotics and machine learning has great potential in the intelligent transport systems of the upcoming age.

BIBLIOGRAPHIC REFERENCES

1. Wadud P, MacKenzie D, Leiby P. Help or hindrance? The travel, energy and carbon impacts of highly automated vehicles. Transp Res Part A Policy Pract. 2016;86:1–18.

2. Kamble A, Gupta H, Trivedi A. A review on energy-efficient driving pattern classification and analysis using vehicle telematics data. IEEE Trans Intell Transp Syst. 2021;22(5):2563–74.

3. Iman A, Asadollah A, Bhatti MI. Fuel consumption estimation based on driving behavior using machine learning techniques. Sensors. 2021;21(5):1589–1601.

4. Araujo R, Portugal D, Nunes U. Data mining for driving styles: a computational approach. In: Proc IEEE Intell Veh Symp; 2010. p. 322–7.

5. Wang Y, Chen L, Ran B. A machine learning approach to detect aggressive driving behaviors using smartphone sensors. Transp Res Rec. 2019;2673(10):217–28.

6. Zheng H, Qin Y, Wang J, Wang C. Fuel consumption prediction model based on XGBoost algorithm and big data. J Adv Transp. 2020;2020:8873074.

7. Kuderer M, Gulati S, Burgard W. Learning driving styles for autonomous vehicles from demonstration. In: Proc IEEE Int Conf Robot Autom (ICRA); 2015. p. 2641–6.

8. Adnan R, Usman M, Arshad A. Driving style classification using machine learning techniques. Procedia Comput Sci. 2018;130:927–32.

9. Jeong J, Hwang J, Lee Y. Driving pattern analysis and classification using OBD-II data. In: Proc Int Conf Big Data Smart Comput (BigComp); 2016. p. 250–3.

10. Tanaka K, Tanabe M, Watanabe S. Fuel efficiency estimation using deep neural networks for eco-driving. IEEE Trans Intell Transp Syst. 2022;23(2):915–27.

11. Di Domenico D, De Nunzio G, Micheli SC. A data-driven approach to fuel consumption forecasting using multimodal information. Appl Energy. 2021;291:116790.

12. Kerner BS. Experimental features of traffic breakdown at highway bottlenecks. Phys Rev Lett. 1997;79(20):4030–3.

13. Lee SJ, Kim J, Ko YB. OBD-II data-based vehicle fuel consumption and CO_2 emission estimation using machine learning algorithms. IEEE Access. 2021;9:16188–98.

14. Naik S, Tiwari A. Deep learning approach for driving behavior classification and prediction. In: Proc IEEE Int Conf Artif Intell Comput Appl (ICAICA); 2020. p. 244–9.

15. Perrotta F, Parry T, Neves LC, Mesgarpour M. A machine learning approach for the estimation of fuel consumption related to road pavement rolling resistance. In: Proc IALCCE; 2018.

16. Wang Q, Zhang R, Wang Y, Lv S. Machine learning-based driving style identification of truck drivers in open-pit mines. Electronics. 2020;9(1).

17. Yao J, Moawad A. Estimating vehicle energy consumption through large-scale simulations and machine learning techniques. Transp Res Part C Emerg Technol. 2019;101:276–96.

18. Abukhalil T, et al. Fuel consumption using OBD-II and support vector machine model. J Robot. 2020;2020:1–9.

19. Perrotta F, Parry T, Neves LC. Machines for truck fuel consumption modeling. In: Proc IEEE Int Conf Big Data; 2017. p. 3810–5.

20. Bhat A, Yadav V, Dargan V, Yash. Sign language to text conversion using deep learning. In: Proc 3rd Int Conf Emerg Technol (INCET); 2022. p. 1–7. https://doi.org/10.1109/INCET54531.2022.9824885.

21. Duraisamy P, et al. Transforming sign language into text and speech through deep learning technologies. Indian J Sci Technol. 2023;16(45):4177–85. https://doi.org/10.17485/IJST/v16i45.2583.

22. Thakar S, et al. Sign language to text conversion in real time using transfer learning. arXiv. 2022:2211.14446. https://doi.org/10.48550/arXiv.2211.14446.

23. Seviappan A, et al. Sign language to text conversion using RNN-LSTM. In: Proc ICDSAAI; 2023. p. 1–6. https://doi.org/10.1109/ICDSAAI59313.2023.10452617.

24. Shekhar S. Real-time conversion of sign language to text using deep learning models. In: Proc ICONAT; 2024. p. 1–7. https://doi.org/10.1109/ICONAT61936.2024.10774717.

25. Wang Z, et al. Hear sign language: a real-time end-to-end sign language recognition system. IEEE Trans Mob Comput. 2022;21(7):2398–410. https://doi.org/10.1109/TMC.2020.3038303.

26. Kulkarni M, Naglot D. Real-time sign language recognition using the Leap Motion controller. In: Proc ICICT; 2016. p. 1–5. https://doi.org/10.1109/INVENTIVE.2016.7830097.

27. J P, et al. Real-time hand sign language translation: text and speech conversion. In: Proc ICCPCT; 2024. p. 488–93. https://doi.org/10.1109/ICCPCT61902.2024.10673038.

28. Sudhakaran P, et al. Converting voice signal to visual Indian sign language. In: Proc ICONSTEM; 2024. p. 1–6. https://doi.org/10.1109/ICONSTEM60960.2024.10568884.

FINANCING

None.

CONFLICT OF INTEREST

None.

AUTHORSHIP CONTRIBUTION

Conceptualization: Deepthi P, M.Nithya, P.Mounika, K.Mourya, P.Mano Vardhan Reddy, Christo Ananth.

Data curation: Deepthi P, M.Nithya, P.Mounika, K.Mourya, P.ManoVardhanReddy, Christo Ananth.

Formal analysis: Deepthi P, M.Nithya, P.Mounika, K.Mourya, P.ManoVardhanReddy, Christo Ananth.

Drafting - original draft: Deepthi P, M.Nithya, P.Mounika, K.Mourya, P.ManoVardhanReddy, Christo Ananth.

Writing - proofreading and editing: Deepthi P, M.Nithya, P.Mounika, K.Mourya, P.ManoVardhanReddy, Christo Ananth.