International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(4s) 2026 pp. 334-345
https://doi.org/10.70917/ijcisim-2026-2519
Article

A SHAP-BASED INTERPRETABLE ETA DELAY PREDICTION FRAMEWORK FOR LAST-MILE LOGISTICS

Bora ÖÇAL

1Süleyman Demirel University, School of Civil Aviation, Isparta, Turkey. e-mail: boraocal@sdu.edu.tr, Orcid:0000- 0003-0654-0131

Abstract: The precise estimation of the Estimated Time of Arrival (ETA) is a key aspect of last-mile logistics. This directly impacts customer satisfaction, efficiency, and decision-making in a transport environment. The classical ETA estimation models are distance-based and have limitations in capturing dynamic functional and real-world conditions such as congestion due to traffic, peak timing, multiple delivery issues, and weather. This work aims to devise a machine learning framework with strong predictive performance for the estimation of delivery delay severity (regression model) and the probability of delay occurrence (binary classification model) using a simulated dataset of logistics shipments. Five regression models – Linear Regression, Ridge Regression, LASSO Regression, Decision-boosting Regression, and Decision-forest Regression models – and five different classification models – Logistic Regression Classifiers, Decision-boosting Classifier, Decision-forest Classifier, RFB-Support Vector Machine Classifier, and K-Nearest Neighbour Classifier – are trained and tested on this dataset using multiple figures of merit. The performance on this dataset reveals that linear models are superior classifiers compared to decision-based models in estimating delay duration with a mean absolute error value of 4.8 minutes for all models. The delay probability classifiers perform extremely well on this dataset with Logistic Regression achieving well-balanced characteristics. In addition, SHAP-based Explainable AI Analysis integration is also included for understanding model output at a global and local perspective. Results and findings concerning Explainability Analysis show that traffic intensity, peak conditions, and stops are identified as top variables responsible for ETA Delay.
Keywords: Logistics Delay Prediction, Estimated Time of Arrival (ETA), Machine Learning, Explainable Artificial Intelligence (XAI), SHAP

1. Introduction

However, due to the globalisation of trade, the growth in e-commerce, and the evolving demands in consumer satisfaction, on-time performance in logistics has evolved from being a source of improved competitiveness to a key determinant in sustainable efficiency. The issue of delay in logistics directly Influences the rest of the supply chain in terms of inventory charges, customer dissatisfaction, and ultimately the profitability of businesses. For instance, not only does delay in logistics result in customer dissatisfaction, it also results in an Increase in storage costs, as confirmed in direct field research [1].

The Estimated Time of Arrival (ETA) calculation for delay prediction is one of the most important roles for logistics planning and also for evaluating the performance of freight transport. The traditional calculation of ETA, depending only on the distance between the two points of transfer and the route to be followed, cannot accurately express dynamic conditions such as environmental densities (traffic density, climate), peak hours, as this makes difficult optimizing the estimated time according to actual performance. This is more substantial in intermodal networks and road systems [2].

Recently, artificial intelligence, a subset of machine learning, has been offering promising solutions for ETA prediction and delay forecasting. This is because ML has the ability to process past data, real-time traffic, and environmental data altogether, promising a higher degree of accuracy compared to traditional models. In recent EMS studies in intelligent transport systems, ML-based ETA prediction has been able to not only offer estimates for travel times but build more robust logistic systems by reducing prediction errors. Studies within EMS for intelligent transport systems have indicated that ETA estimates produced by ML-based predictions are reliable and updated regularly [3].

Studies regarding ETA and delay prediction that have appeared in academic literature have gained much interest for diverse logistical scenarios. For instance, the relevance of using ML-based ETA models in intermodal transport networks has been explored, indicating that such methods make it possible for decision-makers to actively inform about the impact of delays [2]. In addition to this, it is also seen that the comprehensive review on prediction of delivery delays in the field of logistics from literature indicates that the methods now concentrate on optimizing performance but also present some deficiencies from the perspectives of practicability and adaptability when applied to actual scenarios. [4].

While delay prediction modelling studies in logistics applications have gained considerable attention in the literature, comprehensive academic studies are still scant, and ones covering both performance enhancement methods and XAI approaches are particularly new. The application of artificial intelligence in logistics operations not only increases prediction accuracy but also enables decision-makers to understand the reasons for delays and make operational improvements. With this feature, XAI goes beyond merely producing 'results' from logistics performance prediction models, making cause-and-effect relationships transparent.

This study aims to compare ML models that predict delay duration and the probability of delay occurrence by examining logistics shipment delay prediction in both regression and classification contexts. It also evaluates the prediction performance and practical adequacy of the developed models using comprehensive metrics and highlights the contributions of the outputs to the decision support process through interpretable artificial intelligence techniques.

The original contributions of this study to the literature and practice are summarised below:

• A scalable machine learning (ML)-based prediction framework has been proposed for the logistics delay prediction problem, combining realistic operational and environmental variables.

• Factors affecting ETA delay have been analysed both globally (feature importance ranking) and locally (example-based explanations) using a SHAP-based explainable artificial intelligence (XAI) approach.

• The effects of operational variables (e.g., number of stops, service type, vehicle type) and environmental variables (e.g., traffic density, precipitation) on delay were examined comparatively, and critical conditions that increase delay risk were reported.

• The proposed models were analysed using different performance metrics and scenario-based evaluations to produce interpretable and actionable outputs for decision-makers in logistics operations.

The second section of the study contains the Related Works section, which examines similar studies with comparable characteristics and objectives that are relevant to the subject and scope of this study. The third section contains the Material and Methods section, which introduces the machine learning methods and data set used in the study. The fourth section of the study contains the numerical findings obtained from the study and the discussion section. The final section, Conclusions, presents the final conclusions of the study and the research objectives for future studies.

2. Related Works

The domain related to Estimated Time of Arrival (ETA) and Delivery Delay has seen a vast shift from classical distance calculation models to more advanced data-driven models using machine learning (ML) and deep learning (DL) models, as it is recognized that correct predictions are crucial for sustainable operational efficiency, maintaining low storage costs, and, more importantly, for customer satisfaction with regard to global trade chains . Although classical models are appreciable due to their low computation cost, their inefficiency lies with their inability to adapt to dynamism, as they require constant navigation factors; for this reason, models with high cardinality variables and non-linear patterns have been emphasized by recent works .

The recent literature focuses on the popularity of ensemble models like XGBoost, LightGBM, and Random Forest because of their effective capability in dealing with tabular data and analyzing complex patterns in transportation data compared to other machine learning models . Research works conducted on urban traffic patterns and autonomous vehicle shuttles and the field of supply chain management have indicated that gradient boosting techniques tend to perform better than conventional statistical regression and even certain types of neural network models . For example, studies conducted on order delivery delay modeling for supply chain disruption management have indicated that Random Forest techniques can yield results with outstanding predictive accuracy, represented by an AUC of 0.98 . Additionally, XGBoost has been effectively employed for vessel ETA prediction in restricted waterways with a mean absolute percentage error of 5% .

Nevertheless, the increased complexity in the longitudinal trajectory has encouraged a growing number of investigations employing deep learning techniques such as LSTM and GRU in particular to address the sequential patterns of travel routes . Specialized models such as 1D-CNNs and WaveNet have also found significant application in utilizing the voyage patterns of ships in bulk ports for multi-step prediction of travel times . Hybrid models are recently on the rise that integrate the feature extraction abilities of CNNs with the sequential processing abilities of LSTMs and attention-based models that are geared towards utilizing noisy sources such as Automatic Identification Systems (AIS). For instance, the Vessel Arrival Time Prediction (VATP) system incorporates a combination of CNNs, LSTMs, and attention-based models in processing AIS simultaneously with maritime weather parameters .

An emerging general consensus among scholars is that it is essential for ETA models to expand their kinematic information by incorporating both space-temporal characteristics, environmental, and operational factors . More advanced models are already incorporating real-time telemetry information with maritime-specific weather variables such as wind speed, wave period, and temperature, as well as ship-specific information such as Gross Tonnage . In terms of air transport, it has been shown by PETA that rotation reactionary delay and differences between reported times and actual times can be forecasted by propagating forecasts through flight rotation cycles . Moreover, for those cases in which access to full trajectory information is restricted, route恢etric approaches are presently under exploration for use as alternative inputs for ML models, albeit with difficulty in capturing real-world traffic information as precisely as their original counterparts . The challenge is compounded by more complex transport networks, which involve both schedule-based and non-schedule-based routes, requiring coordination between these two types of routes .

Despite their accuracy, deep learning and ensemble models are often criticized as "black-box" systems, which hinders their adoption in critical decision-making processes where user trust is paramount. This has fuelled a new research direction focused on Explainable AI (XAI) to enhance transparency. SHAP (SHapley Additive exPlanations) and LIME are the two most prominent methods used to demystify these models by providing both global feature importance rankings and local instance-based explanations . For instance, using SHAP analysis, researchers found that delay algorithms are usually driven by traffic indices, peak hour flags, and shortest distance in urban and last-mile logistics. In addition, ante-hoc approaches such as QLattice, which are inherently explainable using symbolic regression techniques, are also being proposed to provide transparent mathematical expressions in the decision-making process of transport. In this context, the proposed framework in this study progresses from existing works by utilizing a model based on SHAP .

A systematic analysis of the state of current research work shows that although most of these are binary classifiers (delay vs. on-time), there is a dearth of regression-based models that try to forecast the value of the delays themselves. There is also an imminent requirement for integrated DSS that translate the results of prediction into operational recommendations, such as adaptive routing or proactive risk communication . Future efforts are encouraged toward exploring federated learning in ensuring data privacy across collaborative supply chain networks and investigating the generalization of models across different industrial sectors and geographic regions . Moreover, researchers draw significant attention to pathfinding algorithms and reinforcement learning in enhancing robustness within highly congested maritime routes .

Despite the promising performance of deep learning approaches in ETA prediction, this study focuses on interpretable and computationally efficient machine learning models. Given that one of the main objectives is not only to predict delivery delays but also to explain the factors influencing these predictions through explainable artificial intelligence (XAI), models offering greater transparency and interpretability were preferred over more complex deep learning architectures.

3. Material Methods

The creation, use and technical details of the dataset used in this study are explained in a separate section. The machine learning and explainable machine learning algorithms used in the study are introduced in another section.

3.1.*Dataset Description*

For this purpose, a synthetic logistics shipment dataset was generated to analyse estimated time of arrival (ETA) delays using explainable artificial intelligence techniques. The dataset was designed to model the effects of variables such as traffic density, weather conditions, distance, number of stops, and service type on shipment delays, while supporting both predictive modelling and subsequent explainability analysis.

Each record in the created dataset represents one shipment. Records include the time of pickup, planned arrival, and actual arrival; the delay duration has been calculated based on these timestamps. The main target variable in this dataset is the calculated delay duration, defined in minutes and kept in the delay_min field. However, since delay is a critical risk indicator in logistics operations, a binary label on whether the delay duration value surpasses a threshold value has also been generated. In this respect, the is_delayed variable equals 1 for shipments that meet the condition of delay_min ≥ 15 minutes, while 0 otherwise. This way, the dataset is going to be usable for both regression-based delay prediction and classification-based delay risk analysis.

There are both operational and environmental variables in the dataset to incorporate the complex issue of shipment delays. Operational variables comprise shipment distance in kilometers (distance_km), stops made during the delivery of the shipment (stops_count), service type provided to increase the shipment speed (service_type), and type of vehicle used for the delivery purpose (vehicle_type). In addition to these variables, region from which the shipment originated and region to which the shipment is being delivered are also entered as variables in the dataset to incorporate the space-related properties of the logistics system. On the environmental front, traffic index, which is entered as traffic_index, is used to measure traffic density on a scale of 0 to 1, whereas rainfall amount in millimeters (rain_mm) and temperature in degrees Celsius (temp_c) variables measure the environmental impact of delays.

In the data generation phase, the computation of the expected arrival time considers variables such as the distance of the shipment and the type of vehicles, and subsequently, the delay time is generated using variables such as traffic density, amount of rainfall, peak hours, and number of stops. In this regard, traffic density is expected to influence delays more, considering that delays are also dependent on the chances of delays that are expected to be higher during rainfall and peak hours, which might include the morning and afternoon peak hours, respectively. On the other hand, the operation procedures of the express service type are also expected to increase the speed of the journey, hence reducing delays. The dataset used in this study and the definitions of these variables are presented in Table 1.

Variable	**Type**	**Description**
shipment_id	Identity	Shipment unique identifier
pickup_time	Time	Shipment departure time
planned_arrival_time	Time	Scheduled arrival time
actual_arrival_time	Time	Actual arrival time
distance_km	Numeric	Delivery distance (km)
stops_count	Numeric	Number of stops
service_type	Categorical	Standard / Express
vehicle_type	Categorical	Motor / Van / Lorry
region_from	Categorical	Exit area
region_to	Categorical	Destination area
traffic_index	Numeric	Traffic congestion index (0–1)
rain_mm	Numeric	Rainfall amount (mm)

temp_c	Numeric	Temperature (°C)
delay_min	Target (Reg.)	Delay time (min)
is_delayed	Target (Cls.)	Delay label (≥15 min)
Variable	Type	Explanation
shipment_id	Identity	Shipment unique identifier

Table 1. Variables used in the data set and their descriptions

The design of the dataset in this manner supports the objective of developing machine learning models capable of predicting shipment delays with high accuracy and analysing the decision-making mechanism of the developed models using explainable artificial intelligence methods to interpret the factors causing delays. This aims to obtain interpretable and actionable outputs that can offer operational improvement opportunities to decision-makers, rather than an approach focused solely on prediction performance.

3.2.*Machine Learning Algorithms*

This paper tackles two different types of problems: the first one involves modelling the delay in logistics shipments through a regression method to approximate the duration of the delay in minutes, while the second involves classifying whether there is a delay through a classification method. In this context, different machine learning models have been trained on operational and environmental variables derived from the dataset. The model was also compared. In the model construction phase, the categorical variables are transformed to numerical variables through one-hot encoding. The variables involving data on times are used to create new variables such as hour, weekday/weekend, and peak hour. This is in a bid to encompass all factors involved in the delay formation.

The solution to the regression problem was based on the delay time target variable (delay_min), and five different regression algorithms were used: Linear Regression, Ridge Regression, Lasso Regression, Gradient Boosting Regressor, and Random Forest Regressor. Linear Regression was adopted as a baseline model to capture a simple linear relationship. On the other hand, Ridge and Lasso regression methods aimed to enhance generalizability of the model by incorporating regularization methods. Tree-based models, such as Gradient Boosting and Random Forest, were considered as alternative models on the regression side, which allow non-linear pattern learning. For evaluating regression models, MAE, MSE, RMSE, and MAPE metrics have been used. These metrics enable interpreting model performance more comprehensively by analyzing prediction errors in different dimensions.

During the process of solving the classification task, the target variable is the status of the delay (is_delayed), and the objective is to classify the shipments into the delayed or non-delayed categories. On this matter, five different classifiers were utilized: Logistic Regression, Gradient Boosting Classifier, Random Forest Classifier, Support Vector Machine (with RBF kernel), and K-Nearest Neighbours (KNN). Logistic Regression is assessed for its high interpretability as a linear classifier, Gradient Boosting and Random Forest classifiers for their tree-based structure and flexibility for dealing with nonlinear boundaries. Additionally, the model SVM-RBF is preferred for its capability of dealing with nonlinear boundaries. KNN is used for its simplicity, as it is a method for classification that relies on the similarity of samples. The objective of the measures that will be utilized is the accurate classification of shipments—Accuracy, Precision, Recall, and F1-score. These measures provide a well-balanced view concerning the cost of FN and FP errors, especially for logistics.

Apart from model accuracy, an explainable artificial intelligence, also referred to as an XAI, was also employed using the SHAP (SHapley Additive exPlanations) technique to enable interpretation of the prediction tasks by decision-makers. SHAP is an addition to various other model interpretability tools that has been employed to ensure that predictions made by models can be interpreted by allowing features to impact predictions in their own distinct manner, according to values provided by game theory. In this study, SHAP was employed at two levels: at a global level, where feature importance was determined, and at a local level, where specific scenarios, including an individual ship, could be interpreted. Hence, apart from ensuring accuracy within the modeling process, this study employed a technique that offers an addition within decision support mechanisms within logistics operations by ensuring that features that impact delays within an involved connection are quantified.

4. Findings

This section will interpret and discuss the performance metrics obtained from machine learning models. For clarity and a more detailed explanation, the section is divided into three parts. The first part evaluates the performance metrics demonstrated by machine learning models for Predicting Shipment Delay Duration. The second section evaluates the performance metrics demonstrated by machine learning models for Delivery Delay Status Detection. The final section evaluates the performance demonstrated by the SHAP algorithm to clarify the criteria influencing the models' decision-making and ensure transparency.

4.1. Estimated Delivery Delay Period

In this study, five different regression models (Lasso, Linear Regression, Ridge, Gradient Boosting, and Random Forest) were trained to estimate the shipment delay time (delay_min), and the models were compared using the MAE, MSE, RMSE, and MAPE metrics. The numerical results are shown in Table 2. The results obtained show that linear-based models (Lasso, Linear Regression, Ridge) produced lower error values and demonstrated more successful prediction performance in this problem. Specifically, the Lasso model achieved the lowest error level with values of MAE=4.7895, MSE=36.0361, and RMSE=6.0030, and was the model that provided the best performance across all metrics. The Linear Regression and Ridge models also produced similar error levels, and the close performance of these three models suggests that the generated dataset can largely be represented through linear relationships.

When examining the minor differences between linear models, it is observed that Lasso provides only a very limited improvement over Ridge and Linear Regression. This suggests that although Lasso achieves a simpler model structure by suppressing the weights of some features through its L1 regularization effect, a large proportion of the explanatory variables in the dataset make a meaningful contribution to the lag prediction. Similarly, the fact that the performance of the Ridge model is almost identical to that of Linear Regression indicates the presence of a stable pattern based on linear relationships rather than a significant overfitting problem in the dataset.

When examining tree-based methods, it is observed that the Gradient Boosting (MAE=4.8982, RMSE=6.1187) and Random Forest (MAE=5.1387, RMSE=6.4530) models produce higher errors compared to linear models. This result can be explained by the fact that, although tree-based methods generally provide an advantage in non-linear relationships, in the current data set, the delay time production mechanism is largely constructed with linear components (traffic density, rainfall, number of stops, peak hour effect, etc.). Therefore, the use of more complex models in this problem did not provide a significant increase in prediction performance; on the contrary, it led to an increase in error values. The Random Forest model producing the highest error values can be attributed to the fact that the model works on the average of different trees, causing it to estimate delay values in a more 'smoothed' manner in some cases and remain at a disadvantage compared to linear models in capturing linear trends.

When examining the MAPE results, it is observed that very high values are obtained for all models. This situation can be explained by the nature of the MAPE metric becoming unstable due to some samples having very low (even close to zero) delay values in the delay time variable. MAPE can exaggerate the error rate because the denominator shrinks as the actual value approaches zero, and therefore it can be misleading, especially in problems involving values close to zero, such as 'delay time'. In this context, it is considered that scale-sensitive metrics such as MAE and RMSE are more reliable indicators for a healthier performance comparison in the study. However, while reporting MAPE provides benefits in terms of comparability due to its prevalence in literature, the effect of near-zero delay values in the dataset must be considered when interpreting it.

Model	MAE	MSE	RMSE	MAPE(%)
Lasso	4.789482	36.036082	6.003006	4.902482e+06
Linear Regression	4.789843	36.045809	6.003816	4.894213e+06
Ridge	4.791385	36.048703	6.004057	4.937425e+06
GradientBoosting	4.898225	37.439028	6.118744	6.144978e+06
Random Forest	5.138716	41.641024	6.452986	6.949842e+06

Table 2. Prediction performance metrics for shipment delay duration

Overall, the results in Table 2 show that the delay prediction problem in this dataset can be modelled with high accuracy using linear models, achieving an average prediction performance of approximately 4.8 minutes MAE. These error levels indicate an accurate level that is applicable for the early detection of delays in logistics operations and the improvement of operational planning. Furthermore, the simpler and more interpretable structure of linear models will enable the transparent identification of factors causing delays through explainable artificial intelligence (XAI) analyses in the next phase.

4.2. Determination of Shipment Delay Status

In this study, five different classification algorithms (Logistic Regression, Gradient Boosting, Random Forest, SVM-RBF, and KNN) were trained to predict shipment delay status (is_delayed), and the models were compared using the Accuracy, Precision, Recall, and F1-score metrics. The results show that all models performed well and that the delay classification problem contained clearly distinguishable patterns in the dataset. In particular, the fact that Recall values were quite high in all models indicates that delayed shipments (positive class) could be largely captured successfully.

Upon analyzing the outputs, Logistic Regression achieved the best overall balanced performance among all classifiers, with Accuracy=0.9050, Precision=0.9247, Recall=0.9679, and F1-score=0.9458. The strong recall score of the Logistic Regression model is a remarkably advantageous factor in logistics operations as it ensures the model does not misclassify delayed shipments as ‘non-delayed’ in the case of delayed delivery. Also, in terms of interpretability with SHAP values in future analysis, the linear form of the Logistic Regression model would seem to be a good option.

The result of the Gradient Boosting model was very close to that of Logistic Regression, with an Accuracy of 0.8992 and an F1 score of 0.9428. The fact that precision and recall were both relatively balanced implies that this model can handle both false positives and false negatives. Given the ability of Boosting-based algorithms to handle non-linear relationships, it can be inferred that a few nonlinear patterns may be present in this dataset, but their importance is less compared to Logistic Regression.

The results for the Random Forest algorithm showed that it had excellent recall skills with values for Accuracy of 0.8908, Precision of 0.9059, Recall of 0.9737, and F1 of 0.9386. This shows that the algorithm performs well on the delayed shipments feature; nonetheless, due to its lower precision value compared to the Logistic Regression algorithm, there may be a higher possibility of predicting normal shipments as delayed. It could result in unnecessary actions being initiated for logistics planning.

In comparison, the model with the highest delay capture rate was SVM-RBF with Recall = 0.9874, indicating that the model minimized the probability of a missed delayed shipment. However, the low accuracy of Accuracy = 0.8783 and Precision = 0.8841 may be linked to the fact that the SVM model generates a boundary with a higher aggressiveness in terms of estimating the positive classes. Although the model may perform superior in its application in 'Early Warning System' towards delay issues, its effect may be negative in terms of resource utilization.

The KNN model, on the other hand, showed the worst-performing results in comparison to the other models, since it attained values of Accuracy=0.8567, Precision=0.8567, Recall=1.0000, and F1=0.9228. The reduction in precision and accuracy values indicates that the KNN model is likely to predict on-time shipments to be delayed at a very high rate. The fact that the recall measure of KNN is 1.0, in other words, all test delayed shipments are predicted to be in the positive category, shows that the KNN model is very strong in terms of not missing delays. However, on the other hand, it might not be a very good model in terms of having a very high cost of false alarms.

Model	Accuracy	Precision	Recall	F1-Score
LogisticRegression	0.905000	0.924721	0.967899	0.945817
GradientBoosting	0.899167	0.917203	0.969844	0.942790
RandomForest	0.890833	0.905882	0.973735	0.938584
SVM_RBF	0.878333	0.884146	0.987354	0.932904
KNN	0.856667	0.856667	1.000000	0.922801

Table 3. Performance metrics for detecting shipment delay status

Overall, the results in Table 3 show that high recall performance is a common feature across all models in the delay prediction problem and that delayed shipments can be clearly distinguished. In this context, model selection may vary depending on operational usage objectives: if avoiding delays is critical (minimum false negatives), models that produce high recall, such as SVM-RBF or KNN, may be preferred; if the cost of false alarms is high (minimum false positives), Logistic Regression or Gradient Boosting, which offer higher precision values, provide more balanced solutions. Furthermore, Logistic Regression's high performance, coupled with its greater interpretability, will provide an advantage in subsequent stages of the study in terms of understanding the factors causing delays through explainable artificial intelligence analyses.

4.3. Explanatory Nature of Delay Duration and Delay Status Determination

In this paper, SHAP is used as the XAI layer for the model developed to predict ETA delay. The primary motivation for using SHAP is to interpret the contribution of every single feature in the model output, both at global and local levels: global for an overall feature importance across the whole dataset and local for case-based reasoning to explain the results for a specific shipment. By doing so, the proposed framework can go one step beyond black-box predictions and provide actionable insights into the logistics decision-making process.

Firstly, a major ranking analysis based on average absolute SHAP values was performed to identify which variables contributed the most to the model in terms of their global impact. The bar chart in Figure 1 indicates that the traffic_index has the most influence on the model result. The results indicate that traffic_index emerged as the most influential predictor of delay within the generated dataset, suggesting that traffic-related conditions play a major role in the model's delay predictions. The next variables with a high level of influence are identified to be is_peak and stops_count. This suggests that peak-hour deliveries and multi-stop delivery operations are strongly associated with higher predicted delay values within the generated dataset. In addition to this, it is also identified that the variables service_type_express, vehicle_type_truck, rain_mm, vehicle_type_motor, and distance_km also implicate a certain degree of influence on delay computation. This emphasizes that service factors and distance factors also implicate a secondary degree of influence on delay generation.

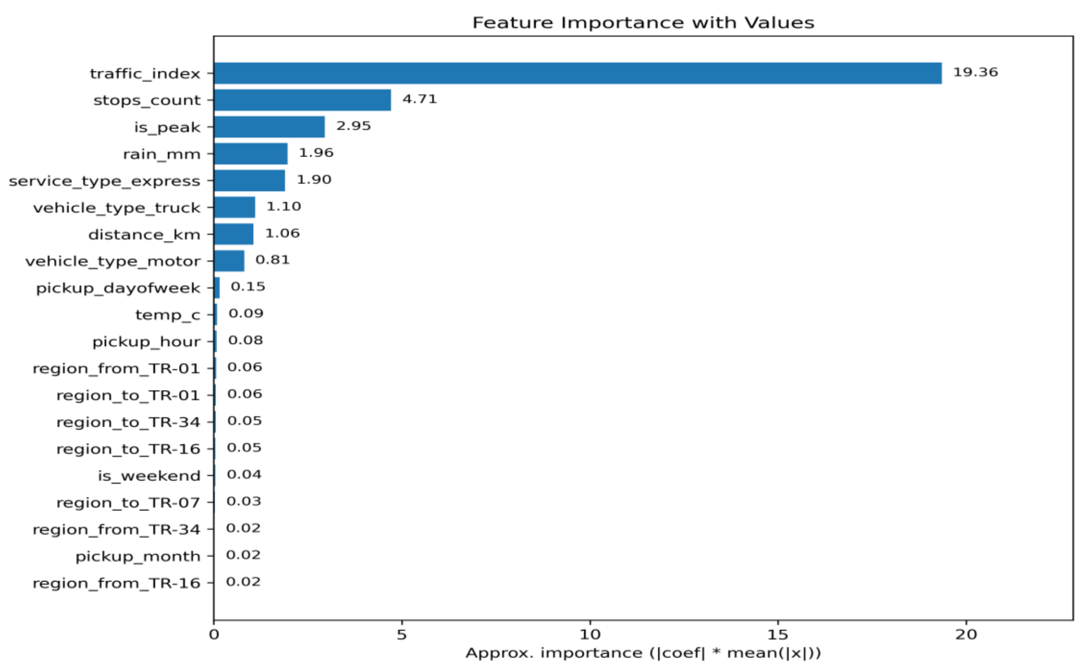

Figure 1. Global feature importance based on mean absolute SHAP values.

After identifying the relevance, based on the ranking of global importance, SHAP summary (beeswarm) plot was used to analyze the way variables influence the result and the extent to which these influences are reflected by ranges of values. As shown by the data represented in Figure 2, it is evident that high values of the traffic_index, represented by red dots, are primarily located in positive SHAP values; thus, it is observed that higher traffic density values are associated with higher predicted delay values. However, for low values of traffic, represented by blue dots, these are primarily positioned in negative SHAP values, thus contributing towards decreased delay times. Further, after analyzing about is_peak variables, it is evident that delay times are increased due to peak hour conditions. For the stops_count variable, it is noticed that when the number of stops rises, the SHAP value rises positively and delays the prediction positively. This interpretation implies that when route planning is carried out in multi-stop routes, the delay effect rises. However, in the case of the service_type_express variable, the overall negative SHAP value indicates that express service tends to reduce the predicted delay duration. For the environmental variables, in the rain_mm case, when the value is high, the positive effect is noticed on delay predictions, meaning that when there is rain, delays can be expected due to the inefficiency in the physical environment that results due to such weather conditions. Figure 2 not only explains the order of priority of the variables but also analyzes the decision-making process of the model to delay or reduce delays in respect to which variables and how.

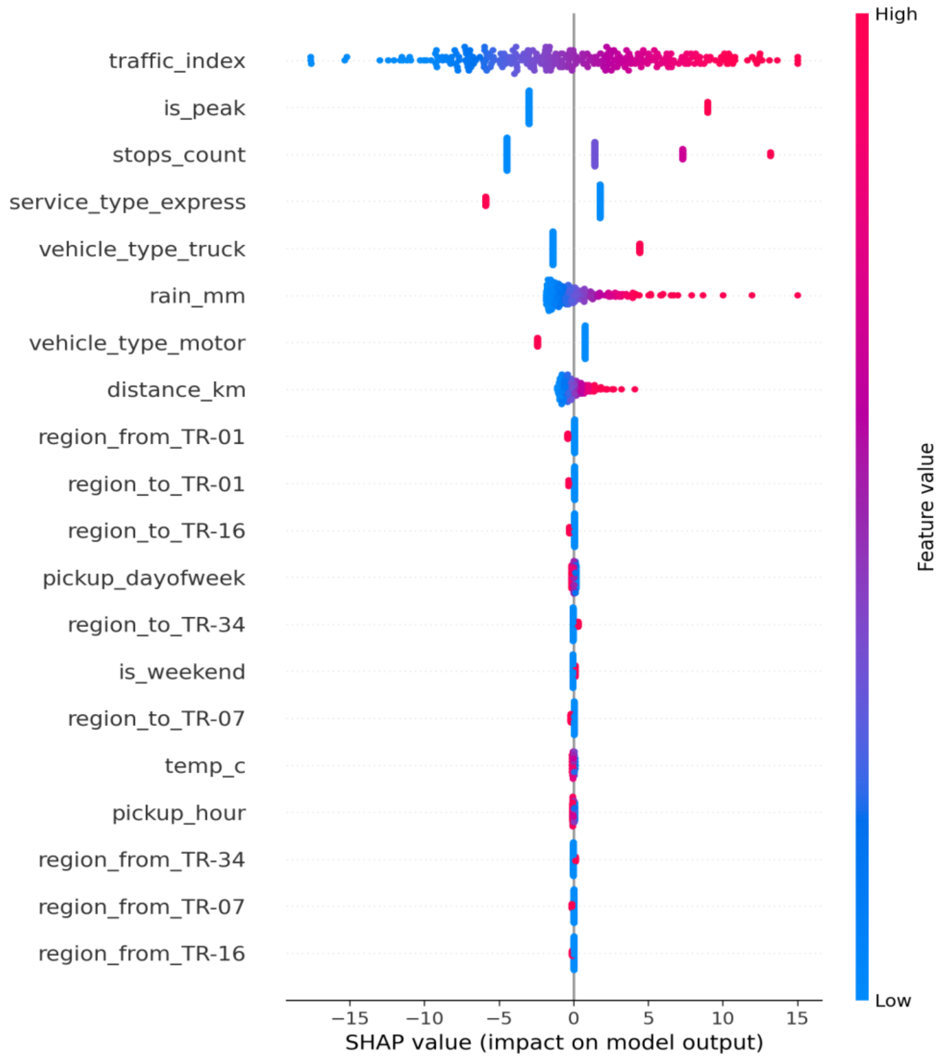

Figure 2. SHAP summary (Beeswarm) plot illustrating feature impact distribution and direction.

In addition to global explanations, local SHAP analysis was performed to demonstrate why the model produced a specific delay estimate for an individual shipment. The waterfall graph presented in Figure 3 shows how the model output for a sample shipment record was updated from the baseline estimate based on feature contributions. In the relevant example, factors such as stops_count=0, traffic_index=0.439, and is_peak=0 are seen to contribute to a reduction in the delay estimate. This indicates that a non-stop delivery scenario, medium/low traffic density, and deliveries made outside peak hours are key factors that reduce delay time. However, in the same example, it is observed that the value of rain_mm=2.74 has an increasing effect on the delay prediction. This local explanation demonstrates that the model produces predictions by consistently combining operational and environmental factors even in a single shipment example and ensures that the model outputs are understood within a ‘cause-and-effect relationship’ framework. Therefore, SHAP-based local explanations enable logistics managers to identify the most influential factors contributing to delay predictions for specific shipments and support preventive actions such as route adjustments, timing improvements, and service type selection.

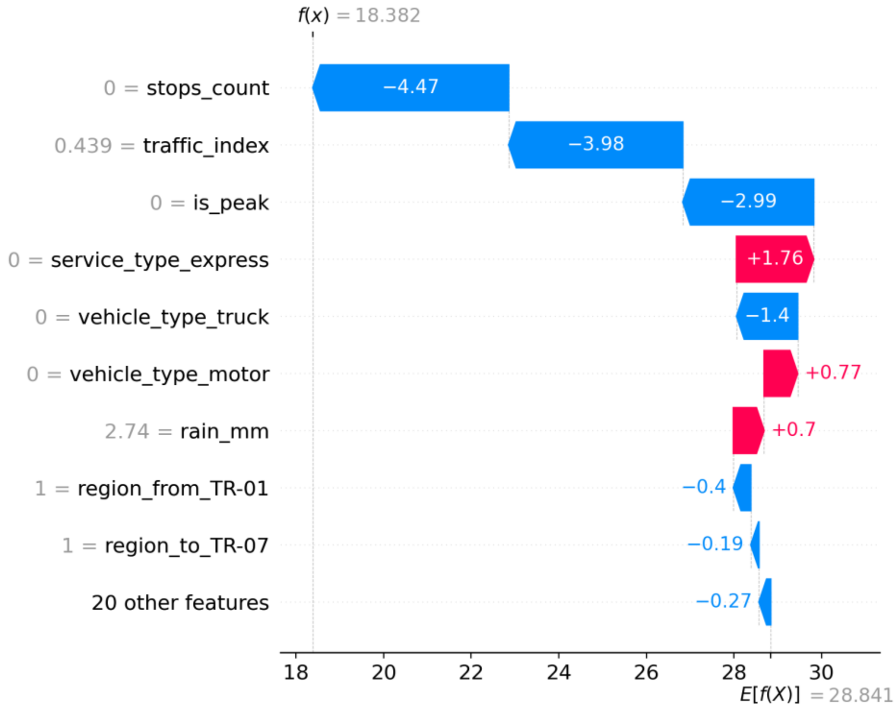

Figure 3. Local SHAP waterfall explanation for a representative shipment.

In conclusion, the SHAP-based explainability analysis revealed that the SHAP analysis identified traffic density, peak-hour conditions, and number of stops as the most influential features contributing to the model predictions. Additionally, it was determined that express service strategies have a delay-reducing effect, while environmental conditions such as rainfall have a delay-increasing effect. These findings demonstrate that the proposed approach not only ensures prediction accuracy but also provides meaningful, interpretable, and actionable insights from a logistics operations perspective. Thus, the study makes a strong contribution in terms of both performance and explainability for integrating delay prediction into decision support processes.

5. Conclusions

For this analysis, an adaptable machine learning framework is proposed to estimate the delay in the ETA for the last-mile logistics environment, covering all aspects related to the estimation of delay duration in minutes (regression) as well as delay risk identification (classification) tasks. For regression tasks, linear models (specifically, the Lasso, Linear Regression, and Ridge algorithm) were observed to have lower error rates than the tree-based algorithms, and the best possible prediction accuracy was reached by them at an MAE value of 4.8 minutes. This analysis showcases the capability to capture the delay dynamics in the dataset through linear components and the strength of highly interpretable techniques as a potent alternative for the problem at hand. For the classification problem, the results were satisfactory for all approaches, and the best-perfaring method was the Logistic Regression algorithm, providing balanced accuracy for all classes.

The value of the research lies in its ability to make decisions in the developed models clear through explainable AI assessment using SHAP value analysis, in addition to its predictive capability. Global and local analysis through SHAP value explainers highlighted the key influential factors that contribute to delay as traffic density, peak hours, and the number of stops; it also explained that decisions related to express delivery services have the capability to generate effects in reducing delay. Thus, the developed mechanism is not only a modelling system that analyzes delay prediction but also a decision support system that provides insights.

Since the dataset was synthetically generated, the SHAP-based feature importance results may partially reflect the assumptions embedded in the data generation process. Therefore, future studies should evaluate the proposed framework using real-world logistics datasets to assess its practical applicability and robustness.

Future studies should validate the findings with real-world data, evaluate the framework across different city and route scenarios, and integrate real-time data streams into the model.

Declarations

Competing Interests: The author declares that they have no competing interests.

Funding: No Funding.

Authors' contributions: Bora ÖÇAL The processes within the scope of the study were carried out by the author.

Disclose editorial: No external editorial assistance was received, and no fee was paid to any third party for the preparation or submission of the article.

References

1. "AI Use Case – Predictive ETA Calculation for Freight." Accessed: Jan. 18, 2026. [Online]. Available: https://www.miloriano.com/ai-use-case-predictive-eta-calculation-for-freight/?utm_source=chatgpt.com
2. A. Balster, O. Hansen, H. Friedrich, and A. Ludwig, "An ETA Prediction Model for Intermodal Transport Networks Based on Machine Learning," Business & Information Systems Engineering 2020 62:5, vol. 62, no. 5, pp. 403–416, May 2020, doi: 10.1007/S12599-020-00653-0.
3. A. Wani, "Ten quick tips for improving estimated time of arrival predictions using machine learning in logistics and transportation systems," PeerJ Comput Sci, vol. 11, p. e3259, Oct. 2025, doi: 10.7717/PEERJ-CS.3259/FIG-4.
4. N. Müller, P. Burggräf, F. Steinberg, C. R. Sauer, and M. Schütz, "An analytical review of predictive methods for delivery delays in supply chains," Supply Chain Analytics, vol. 11, p. 100130, Sep. 2025, doi: 10.1016/J.SCA.2025.100130.
5. A. A. Scheidler et al., "A Review of Vessel Time of Arrival Prediction on Waterway Networks: Current Trends, Open Issues, and Future Directions," Computers 2025, Vol. 14, Page 41, vol. 14, no. 2, p. 41, Jan. 2025, doi: 10.3390/COMPUTERS14020041.
6. Ahmed, I. Kumara, V. Reshadat, A. S. M. Kayes, W. J. van den Heuvel, and D. A. Tamburri, "Travel Time Prediction and Explanation with Spatio-Temporal Features: A Comparative Study," Electronics 2022, Vol. 11, Page 106, vol. 11, no. 1, p. 106, Dec. 2021, doi: 10.3390/ELECTRONICS11010106.
7. E. Antypas, G. Spanos, A. Lalas, K. Votis, and D. Tzovaras, "A time-series approach for estimated time of arrival prediction in autonomous vehicles," Transportation Research Procedia, vol. 78, pp. 166–173, Jan. 2024, doi: 10.1016/J.TRPRO.2024.02.022.
8. A. Thomas and V. V. Panicker, "Application of Machine Learning Algorithms for Order Delivery Delay Prediction in Supply Chain Disruption Management," Lecture Notes in Mechanical Engineering, pp. 491–500, 2023, doi: 10.1007/978-981-99-1665-8_42/FIGURES/6.
9. A. ; ; Sedaghat et al., "Automatic Identification System-Based Prediction of Tanker and Cargo Estimated Time of Arrival in Narrow Waterways," Journal of Marine Science and Engineering 2024, Vol. 12, Page 215, vol. 12, no. 2, p. 215, Jan. 2024, doi: 10.3390/JMSE12020215.
10. S. El Mekkaoui et al., "Deep learning models for vessel's ETA prediction: bulk ports perspective," Flexible Services and Manufacturing Journal 2022 35:1, vol. 35, no. 1, pp. 5–28, Oct. 2022, doi: 10.1007/S10696-022-09471-W.
11. A. Abdi and C. Amrit, "Enhancing vessel arrival time prediction: A fusion-based deep learning approach," Expert Syst Appl, vol. 252, p. 123988, Oct. 2024, doi: 10.1016/J.ESWA.2024.123988.
12. R. Dalmau, A. Trzmiel, and S. Kirby, "Combining Machine Learning Models to Improve Estimated Time of Arrival Predictions," European Journal of Transport and Infrastructure Research, vol. 25, no. 1, pp. 45–66, Jan. 2025, doi: 10.59490/ejtir.2025.25.1.7488.
13. A. Arifi, P. Bouros, and T. Chondrogiannis, "A Study on ETA Prediction using Machine Learning and Recovered Routes," 2024, Accessed: Jan. 18, 2026. [Online]. Available: https://www.openstreetmap.org/
14. Salih et al., "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME," Advanced Intelligent Systems, vol. 7, no. 1, Jun. 2024, doi: 10.1002/aisy.202400304.
15. N. Srivastava, B. N. Gohil, and S. Ray, "Travel time prediction: unveiling the impact of spatio-temporal features on trip duration with eXplainable AI," Transportation 2025, pp. 1–34, Dec. 2025, doi: 10.1007/S11116-025-10701-X.

16. Park, S. Sim, and H. Bae, “Vessel estimated time of arrival prediction system based on a path-finding algorithm,” Maritime Transport Research, vol. 2, p. 100012, Jan. 2021, doi: 10.1016/J.MARTRA.2021.100012.