EXPLAINABLE AI APPROACH FOR IDENTIFYING CRITICAL FACTORS AFFECTING ON-TIME ARRIVAL OF TRUCKS IN LOGISTICS

***Panupong Wanjantuk*[1] *and Ruth Banomyong*[2]**
[1]*Department of Computer Engineering,*
Khon Kaen University, Thailand
[2]*Department of International Business, Logistics and Transport,*
Thammasat Bussiness School, Thammasat University, Thailand

ABSTRACT

Purpose: Effective supply chain management depends on on-time delivery, and knowing what influences on-time arrival can help logistics organizations optimize their processes and improve customer satisfaction. This study explores the key factors that affect the on-time arrival of trucks in logistics operations using an explainable AI technique.
Design/methodology/approach: This study identifies the key factors that have a significant impact on the on-time arrival of trucks using explainable AI techniques and a large dataset made up of historical delivery records, current location, transportation distance, vehicle type, supplier, material shipped, vehicle state, destination state, and other relevant factors.
Findings: The research's conclusions provided clear understandings into the causes of delivery delays by shedding light on the relative significance and interplay of these factors.
Research limitations: The research may face limitations due to the availability and quality of data. Access to comprehensive and up-to-date datasets containing information on various factors that influence on-time arrival of trucks in logistics might be challenging. Insufficient or biased data can affect the accuracy and generalizability of the findings.
Practical implications The logistics sector will be significantly impacted by the results of the study. Logistics organizations may enhance their delivery schedules, manage resources more wisely, and put plans in place to reduce risks by developing a thorough grasp of the essential elements influencing on-time arrival. Cost reductions, increased operational effectiveness, and an overall improvement in logistics performance can result from this.
Originality/value: By particularly applying explainable AI techniques to the logistics context and concentrating on the on-time arrival of trucks, this research makes a contribution to the area. The proposed technique stands out for its transparency and interpretability, guaranteeing that stakeholders can understand the model's decision-making process and develop trust in AI-driven logistics solutions.

Keywords: Explainable AI, On-Time Arrival, Trucks, Logistics, Delivery Delays, Transparent Insights, Operational Efficiency.

1. Introduction

The timely delivery of products and services is fundamental to the operational efficiency and customer satisfaction of supply chain management. The complex network of processes and activities that comprise contemporary logistics operations depends on the capacity to ensure on-time shipment delivery. However, the complex nature of supply chains, which is characterized by diverse variables and dynamic conditions, makes it difficult to consistently meet delivery deadlines.

Understanding the key factors that influence the on-time arrival of a shipment is of crucial importance in this situation. By understanding the complex interaction between factors, logistics organizations can gain valuable insights that facilitate the optimization of operational procedures. Therefore, improved on-time delivery performance not only facilitates the entire supply chain, but also contributes significantly to enhancing customer satisfaction.

This study investigates the influence of these key factors on the on-time arrival of vehicles in logistics operations. To explain the complex relationships at play, the research employs an innovative method based on AI techniques that can be explained. Using the power of explainable AI, the study is to explain the factors that influence on-time arrivals, providing a clear and interpretable view of the decision-making processes.

In the following sections, we conduct a thorough investigation that puts light on the key factors that influence the timely arrival of trucks in the logistics landscape. Through the lens of explainable AI, this investigation aims to provide not only an enhanced understanding of these factors, but also a framework upon which logistics organizations can formulate strategies to optimize their operations and increase customer satisfaction to new heights.

2. Literature review

Predicting transit time and ensuring on-time delivery of shipments are critical challenges in logistics and supply chain management. In recent years, machine learning techniques have gained prominence as powerful tools to address these challenges. This section provides an overview of studies that have employed machine learning for predicting transit time and enhancing on-time delivery of shipments.

(Jonquais and Krempl, 2019) developed a project to determine if Machine Learning and predictive analytics can improve shipment arrival times. A model was developed using Machine Learning computing and historical shipment data, incorporating external factors like holiday seasons and port congestion levels. (Hathikal, Chung and Karczewski, 2020) develops a predictive model for ocean import freight shipment lead time using machine learning methods, considering stakeholder interests. Two terminal criteria are used: empty container return and delivery confirmation at the destination. Real data from an industry partner is used, and multinomial logistic regression is identified as the best classifier. Classifiers like multinomial logistic regression, decision tree, K-nearest neighbors, and support vector machine perform better than Naïve Bayes when categorical variables are binarized or converted into ordinal values. The model offers improved visibility and predictability for shipment lead times, benefiting various parties in the supply chain. (Lin, Chen and Chou, 2023) collected global positioning system travel data from a Taiwanese convenience store chain and proposed machine learning to predict travel times. The model was validated using historical data and a nonlinear regression equation for missing GPS data. The results exceeded 97%, demonstrating the model's potential for logistics fleets to estimate accurate travel times for future delivery tasks and route arrangements. The integrative literature review by (Mugurusi and Oluka, 2021) explores the status of XAI as a solution to AI's black-box problem in Supply Chain Management (SCM). It presents an integrative research typology for XAI in SCM, aiming to align literature conceptions and understand the gap between AI deployment in practice, maturity in SCM, and XAI's extent.

The research on Explainable AI (XAI) in identifying key factors affecting transit time and ensuring on-time delivery of shipments is scarce. Although machine learning techniques have been integrated for transit time prediction and on-time delivery enhancement, the focus on understanding the underlying drivers remains limited. XAI holds immense promise in providing actionable insights for informed decision-making and improved operational efficiency. However, the current research landscape highlights a significant gap in exploring key influencing factors, highlighting untapped potential and the need for further investigation. Researchers and practitioners must recognize this unexplored avenue and work to bridge the gap between XAI and identifying critical factors in transit time prediction and on-time delivery assurance. Addressing this research gap could yield valuable insights for academia and real-world logistics operations, ultimately leading to more reliable and efficient supply chain management.

3. Methodology

Figure 1 shows the whole operational method that this study proposes. At first, the data is pre-processed. The first step in making machine learning models is to divide the data into training, validation, and testing data sets. The training data set is used to build the classification model, the validation data set is used to fine-tune the models' hyperparameters, and the test data set is used to test how well the models work. Once the best model has been found, Explainable AI methods are used to set up additive attributes. These attributes are then used to figure out how important different variables are for injury severity and how different risk factors affect severity mode.

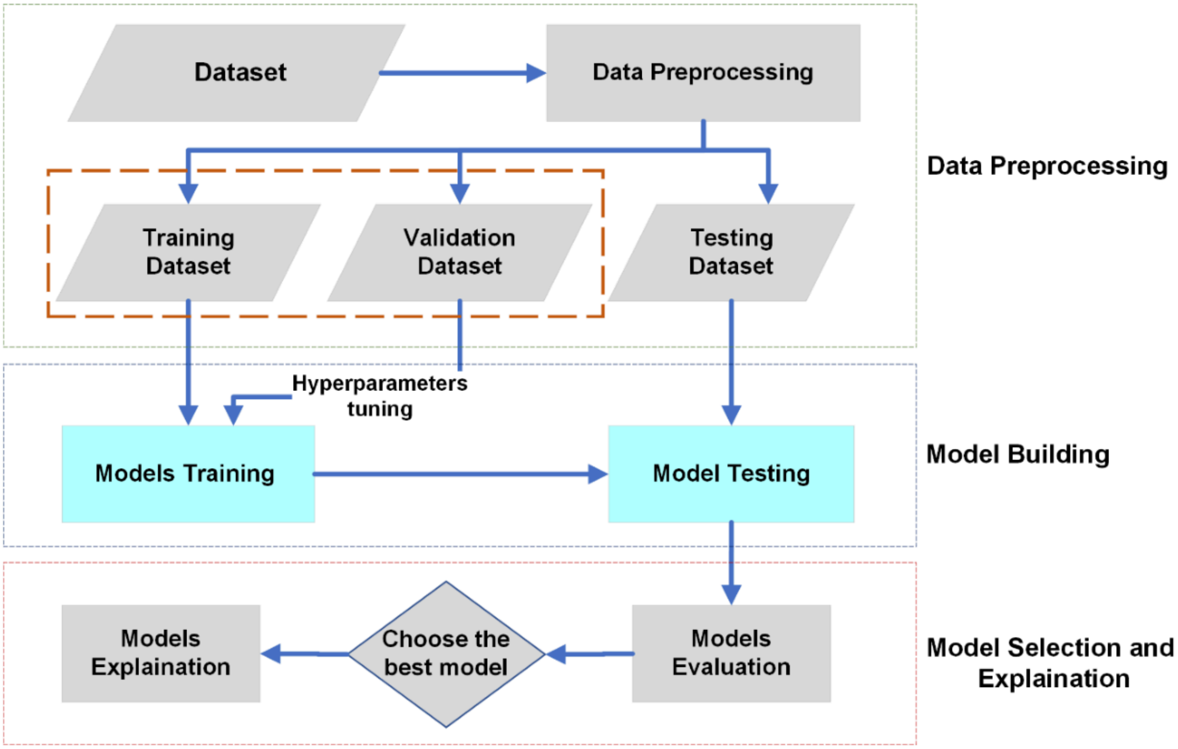


Figure 1: The research framework

3.1. Data Description

In this study, we used a historical delivery records dataset of a logistics company that is publicly available on Kaggle (www.kaggle.com). The dataset contains 6,849 historical delivery records. The description of variables is shown in Table 1.

Variable Name	Description
market_or_regular	Type of trip. Regular - contracted vendors Market - non-contracted vendors
origin_location_code	Origin location code
destination_location_code	Destination location code
transportation_distance_in_km	Transportation distance (in kilometre)
expected_travel_hours	Expected travel time (in hours)

vehicle_type	Vehicle Types
customer_id	Customer ID
supplier_id	Supplier ID
material_shipped	Type of material transported
ontime	On time or Delayed

Table 1: Data Description

3.2. Machine Learning Models

In this study, sixteen machine learning models are used to predicting on-time delivery of a shipment. These models represent a diverse range of machine learning algorithms, each with its own characteristics, strengths, and applications. To train the models, we used the Python packages PyCaret (Ali, 2020). Table 2 shows a short description of the machine learning algorithms used in this study.

No.	Algorithms	References	Short description
1	Logistic Regression	(Cox, 1958)	A classification algorithm used to model the probability of a binary outcome. It estimates the relationship between the independent variables and the probability of a specific outcome.
2	K Neighbors Classifier	(Cover and Hart, 1967)	A non-parametric classification algorithm that assigns a new data point's class based on the majority class among its k nearest neighbors in the feature space.
3	Naive Bayes	(Chan, Golub and LeVeque, 1982)	A probabilistic classification algorithm that applies Bayes' theorem with the "naive" assumption of feature independence to predict the class probabilities.
4	Decision Tree Classifier	(Breiman *et al.*, 1984)	A tree-based classification algorithm that recursively splits the data into subsets based on feature values, creating a decision tree structure to make predictions.
5	SVM - Linear Kernel	(Platt and others, 1999)	Support Vector Machine uses a linear kernel to find a hyperplane that best separates data points of different classes while maximizing the margin between them.
6	Ridge Classifier	(Hastie *et al.*, 2009)	A variant of linear regression that adds L2 regularization (ridge) to the loss function, helping prevent overfitting.
7	Random Forest Classifier	(Breiman, 2001)	An ensemble method that builds multiple decision trees and combines their predictions to improve accuracy and reduce overfitting.
8	Quadratic Discriminant Analysis	(McLachlan, 2005)	A classification technique assuming Gaussian distributions, allowing different covariance matrices for each class, in contrast to Linear Discriminant Analysis.
9	Ada Boost Classifier	(Freund and Schapire, 1997)	An ensemble method that combines multiple weak classifiers to create a strong classifier. It assigns higher weights to misclassified instances.
10	Gradient Boosting Classifier	(Friedman, 2001)	An ensemble technique that sequentially builds decision trees, focusing on the mistakes made by the previous trees to improve overall performance.
11	Linear Discriminant Analysis	(McLachlan, 2005)	A technique that seeks linear combinations of features that best separate classes by maximizing the ratio of between-class variance to within-class variance.
12	Extra Trees Classifier	(Geurts, Ernst and Wehenkel, 2006)	An ensemble method similar to Random Forests, but with additional randomization in selecting splitting points, aiming for improved generalization.
13	Extreme Gradient Boosting	(Chen and Guestrin, 2016)	Extreme Gradient Boosting (Chen and Guestrin, 2016)**: Also known as XGBoost, it's an advanced gradient boosting algorithm designed for improved performance, scalability, and handling missing values.
14	Light Gradient Boosting	(Ke *et al.*, 2017)	A gradient boosting framework that focuses on faster training and high efficiency using histogram-based learning.
15	CatBoost Classifier	(Prokhorenkova *et al.*, 2018)	A gradient boosting algorithm designed to handle categorical features efficiently by encoding them in a way that improves model performance.

No.	Algorithms	References	Short description
16	Dummy Classifier	((Lorena, De Carvalho and Gama, 2008)	A simple baseline classifier that makes predictions based on simple rules like random guessing or class distribution proportions. It's used to benchmark other classifiers.

Table 2: A short description of the machine learning algorithms used in this study.

3.3. Model Evaluation

In this study, the evaluation and selection of the best model are performed using the confusion matrix and its associated metrics, namely accuracy, precision, recall, and F1-score. Additionally, Area Under the ROC Curve (AUC), Cohen's Kappa and Matthews Correlation Coefficient (MCC) are utilized for this purpose. These metrics provide a comprehensive understanding of a classification model's performance, considering different aspects of correct and incorrect predictions.

A confusion matrix (also known as the contingency table) is a tabular representation of the predicted vs. actual class labels for a classification problem. It helps in understanding the performance of a classification model. Figure 2 depicts the confusion matrix, which can be used to compute numerous metrics.

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Figure 2: Confusion matrix

The true positives (TP) and true negatives (TN) have been correctly classified. A false positive (FP) occurs when an outcome is incorrectly classified as yes or positive when it is in fact no or negative. False negative (FN) refers to the incorrect classification of a positive result as negative. The true positive rate (TPR) quantifies the proportion of correctly identified positives, whereas the false positive rate (FPR) quantifies the proportion of incorrectly identified negatives as positives.

Accuracy is the ratio of correctly predicted instances to total instances, evaluating a model's overall performance. A 1 indicates accurate predictions, while a 0 indicates incorrect predictions and incorrect classification. Precision is the proportion of accurately predicted positive instances relative to all correctly predicted positive instances. The higher the value of Precision, the better. The best possible value is 1 (if a model made all correct predictions), and the worst possible value is 0 (if a model made no correct predictions). Recall is the proportion of accurately predicted positive instances in relation to all actual positive instances. It assesses the model's capacity to recognize all positive instances. A high recall score results in fewer false negatives. The best possible value for recall is 1 (model correctly identifies all positive instances and does not produce any false negatives). The worst possible value for recall is 0 (model fails to identify any of the positive instances and produces only false negatives). The F1 score represents the harmonic mean of precision and recall. It maintains a balance between precision and recall. F1 score is useful when classes are unbalanced because it provides false positives and false negatives equal weight. A higher F1 score is better, as it indicates that the model performs well in terms of precision and recall. The F1 score of a perfect classifier is 1. A score of 0 for the F1 measure indicates poor model performance.

Area Under the ROC Curve (AUC) is a binary classification evaluation metric that evaluates a model's ability to distinguish positive and negative instances (Fawcett, 2006). It represents the trade-off between True Positive Rate (Recall) and False Positive Rate (Specificity) for different threshold values. AUC ranges from 0 to 1, with a perfect classifier having a 1 and a random classifier having a 0.5. Higher AUC values indicate better discrimination.

Cohen's Kappa (Fleiss and Cohen, 1973) is a statistical measure that evaluates the level of agreement between observed and expected classifications, considering chance. It is commonly used to evaluate classification models and human raters' agreement in tasks like annotation or labeling. A 1 indicates perfect agreement, a 0 indicates equivalent agreement, and a less than 0 indicates worse agreement than random chance.

The Matthews Correlation Coefficient (MCC) is an evaluation metric that takes into account true positive, true negative, false positive, and false negative predictions to provide a balanced measure of a classification model's performance (Gorodkin, 2004). It ranges between -1 and 1, where 1 indicates perfect predictions, 0 indicates random predictions, and -1 indicates complete disagreement between predicted and actual classifications.

3.4. Model Interpretation

SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017) is a powerful machine learning technique for interpreting complex model predictions. It reveals how individual features contribute to a model's prediction for a specific instance. SHAP values are derived from cooperative game theory concepts (Merrick and Taly, 2019) and seek to allocate each feature's "contribution" to the prediction outcome fairly. The SHAP framework has garnered enormous attention in the field of explainable AI due to its capacity to provide intuitive insights into the decision-making processes of black-box models. It provides a comprehensive and consistent method for understanding the impact of each feature on model predictions, thereby enhancing model transparency and allowing users to utilize machine learning models more effectively.

4. Results

4.1. Performance Assessment

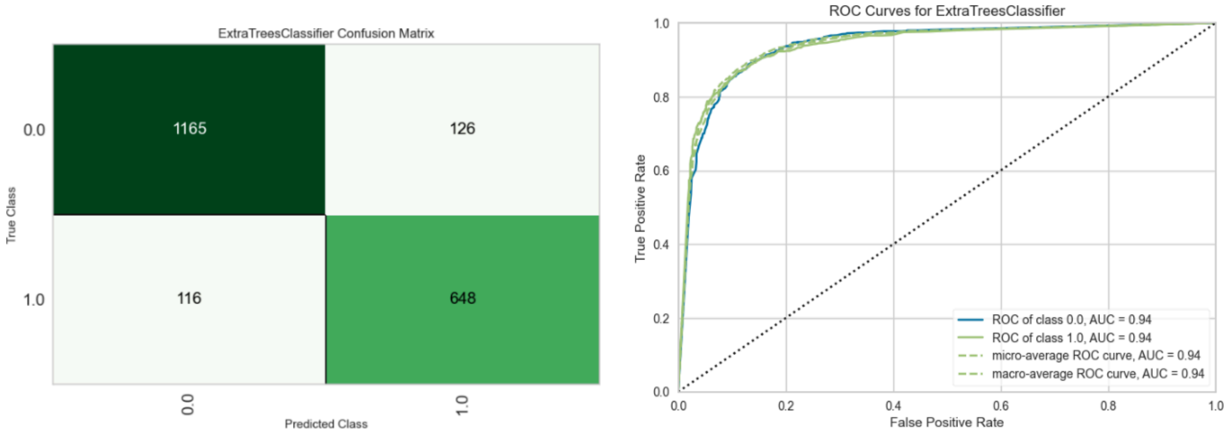
To train the models, we used the Python packages PyCaret (Ali, 2020). Initially, delivery data from the past were collected and pre-processed. For model development, we divided the data into the three subsets listed below: 80% of the data was divided into training and validation subsets, while 20% was used for testing and withheld for model performance evaluation. The training data set was used to develop the models, while the validation data set was used to estimate the performance of the model and tune its hyperparameters. Table 3 shows evaluation metrics for classification model and training time.

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Extra Trees Classifier	0.8828	0.9374	0.8428	0.8434	0.8421	0.7489	0.7501	1.1000
Random Forest Classifier	0.8769	0.9446	0.8479	0.8274	0.8364	0.7378	0.7393	1.1370
Light Gradient Boosting Machine	0.8748	0.9438	0.8372	0.8298	0.8323	0.7326	0.7339	1.0500
CatBoost Classifier	0.8723	0.9464	0.8305	0.8289	0.8283	0.7268	0.7283	2.5110
Extreme Gradient Boosting	0.8694	0.9419	0.8378	0.8182	0.8267	0.7221	0.7236	1.0290
Gradient Boosting Classifier	0.8590	0.9343	0.8137	0.8101	0.8105	0.6983	0.6999	1.0780
K Neighbors Classifier	0.8581	0.9095	0.8187	0.8059	0.8108	0.6975	0.6992	0.9700
Ada Boost Classifier	0.8527	0.9278	0.8204	0.7920	0.8053	0.6870	0.6880	0.9960
Decision Tree Classifier	0.8450	0.8460	0.8058	0.7847	0.7939	0.6698	0.6713	0.9360

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Logistic Regression	0.8440	0.9190	0.7851	0.7938	0.7887	0.6651	0.6659	1.6860
Quadratic Discriminant Analysis	0.8419	0.9084	0.6885	0.8606	0.7638	0.6472	0.6573	0.9340
Ridge Classifier	0.8408	0.0000	0.6925	0.8527	0.7631	0.6454	0.6542	0.9380
Linear Discriminant Analysis	0.8400	0.9138	0.6891	0.8531	0.7613	0.6432	0.6523	0.9920
SVM - Linear Kernel	0.8394	0.0000	0.7862	0.7848	0.7838	0.6562	0.6579	0.9400
Naive Bayes	0.8285	0.8970	0.6347	0.8693	0.7328	0.6112	0.6282	0.9010
Dummy Classifier	0.6283	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9730

Table 3: Comparative evaluation metrics of various machine learning models.

As shown in Table 3, the Extra Trees Classifier model provided the best prediction performance. Accuracy, AUC, recall, precision, F1, Kappa and MCC for the model are 0.8828, 0.9374, 0.8428, 0.8434, 0.8421, 0.7489 and 0.7501 respectively. The model has a high overall accuracy and good precision and recall values. It appears to be performing well in both correctly identifying positive cases and avoiding false positives. Figure 3 shows confusion matrix and ROC curve output of Extra Trees Classifier model.



(a) (b)

Figure 3: Confusion matrix and ROC curve output of Extra Trees Classifier model.

The ROC provide insight into a model's ability to make accurate predictions across different thresholds, with AUC offering a summary measure of performance. From Figure 3 (b), A curve that is close to the top-left corner and has a higher AUC generally indicates better discrimination between classes.

4.2. Model Explanation

The implementation of the SHAP summary evaluation was undertaken in order to facilitate a more detailed analysis of the model. Based on the SHAP summary plot, we have derived a numerical estimation that combines the Shapely values and represents the contributions of variables in the model (refer to Figure 4).

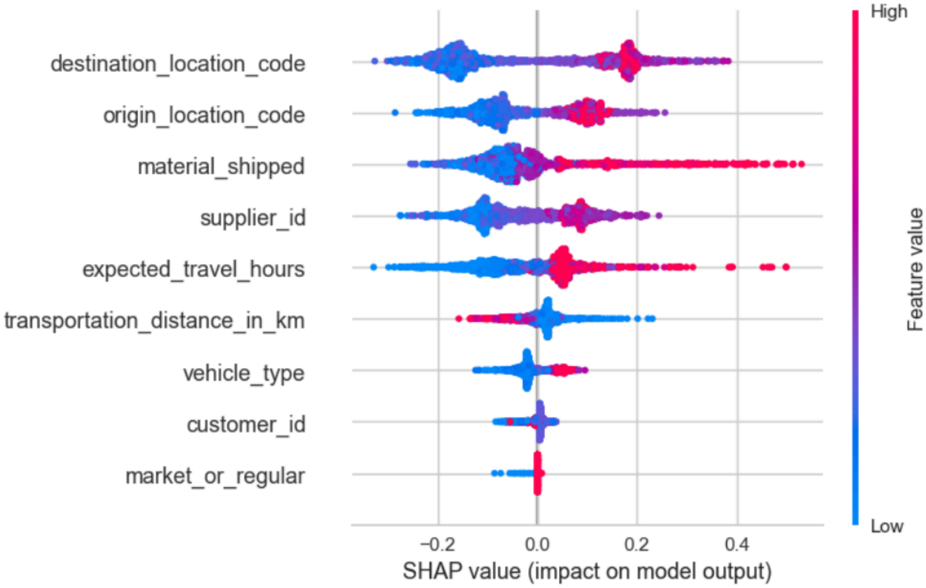


Figure 4: SHAP summary plot for variable importance.

Figure 4 summarizes the Shapely values and gives a graphical representation of the variables' contributions to the model based on the SHAP summary plot. Ascending from most influential to least, the input variables are arranged along the vertical axis. The SHAP value is shown by the horizontal axis, and the significance level of the variable, from low to high, is shown by the color scale. There is a higher degree of correlation between the input variable and the target variable if there are more data points within a particular range of SHAP values. From SHAP summary plot, the most important input variable in determining the on time delivery is destination_location_code, which is ranked first in the summary plots, followed by origin_location_code and material_shipped.

5. Conclusions

In this study, machine learning models and Explainable AI (XAI) were used to identify key factors that have a significant impact on the on-time arrival of vehicles. Typically, accurate models provide a comprehensive depiction of the fundamental relationship between on-time and factors. The Extra Trees Classifier model outperforms other machine learning models in terms of predictive accuracy, AUC, recall, precision, F1-score, Kappa, and MCC in this study. The model provides an additional viable option for modeling on-time vehicle arrival. XAI can be utilized as a tool to find and understand the key factors that affect the outcome of a machine learning model. In this study, SHAP (SHapley Additive exPlanations) was used to interpreting machine learning models and making their predictions more understandable. The analysis revealed that the top five factors that are more likely to affect have a significant impact on the on-time arrival of vehicles are destination location (destination_location_code), origin location (origin_location_code), type of material transported (material_shipped), supplier (supplier_id) and expected travel time (expected_travel_hours).

The destination location and origin location are the key factors that have a significant impact on the on-time arrival of vehicles. The factors concern with traffic congestion, road infrastructure and delivery windows. Urban areas or busy routes can experience heavy traffic congestion, leading to delays. Vehicles originating or destined for such areas might face challenges in maintaining on-time arrivals due to potential traffic jams. The quality of roads and infrastructure on the route can impact travel times. Poor road conditions, construction zones, or detours can lead to unexpected delays. Certain delivery destinations might have specific delivery windows or time restrictions. Adhering to these time constraints is crucial for on-time delivery and customer satisfaction. For the material transported factor, the nature of the material being transported can impact the mode of transportation chosen, handling requirements, and potential delays (e.g., perishable goods might require faster transportation). For the expected travel time factor, as the travel time increases, the amount of uncertainty also increases. Accurate expected travel time estimates allow for more realistic predictions of when a vehicle will reach its destination. If these predictions are precise, it enables better planning and scheduling of vehicle routes.

References

Ali, M. (2020) *PyCaret: An open source, low-code machine learning library in Python*. Available at: https://www.pycaret.org.

Breiman, L. *et al.* (1984) ‘Classification and regression trees (Wadsworth, Belmont, CA)’, *ISBN-13*, pp. 978–0412048418.

Breiman, L. (2001) ‘Random Forests’, *Machine Learning*, 45(1), pp. 5–32. Available at: https://doi.org/10.1023/A:1010933404324.

Chan, T.F., Golub, G.H. and LeVeque, R.J. (1982) ‘Updating formulae and a pairwise algorithm for computing sample variances’, in *COMPSTAT 1982 5th Symposium held at Toulouse 1982: Part I: Proceedings in Computational Statistics*. Springer, pp. 30–41.

Chen, T. and Guestrin, C. (2016) ‘Xgboost: A scalable tree boosting system’, in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Cover, T. and Hart, P. (1967) ‘Nearest neighbor pattern classification’, *IEEE transactions on information theory*, 13(1), pp. 21–27.

Cox, D.R. (1958) ‘The regression analysis of binary sequences’, *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2), pp. 215–232.

Fawcett, T. (2006) ‘An introduction to ROC analysis’, *Pattern recognition letters*, 27(8), pp. 861–874.

Fleiss, J.L. and Cohen, J. (1973) ‘The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability’, *Educational and psychological measurement*, 33(3), pp. 613–619.

Freund, Y. and Schapire, R.E. (1997) ‘A decision-theoretic generalization of on-line learning and an application to boosting’, *Journal of computer and system sciences*, 55(1), pp. 119–139.

Friedman, J.H. (2001) ‘Greedy function approximation: a gradient boosting machine’, *Annals of statistics*, pp. 1189–1232.

Geurts, P., Ernst, D. and Wehenkel, L. (2006) ‘Extremely randomized trees’, *Machine learning*, 63, pp. 3–42.

Gorodkin, J. (2004) ‘Comparing two K-category assignments by a K-category correlation coefficient’, *Computational biology and chemistry*, 28(5–6), pp. 367–374.

Hastie, T. *et al.* (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer.

Hathikal, S., Chung, S.H. and Karczewski, M. (2020) ‘Prediction of ocean import shipment lead time using machine learning methods’, *SN Applied Sciences*, 2(7), p. 1272. Available at: https://doi.org/10.1007/s42452-020-2951-5.

Jonquais, A. and Krempl, F. (2019) ‘Predicting shipping time with machine learning’.

Ke, G. *et al.* (2017) ‘Lightgbm: A highly efficient gradient boosting decision tree’, *Advances in neural information processing systems*, 30.

Lin, Y.-K., Chen, C.-F. and Chou, T.-Y. (2023) ‘Developing Prediction Model of Travel Times of the Logistics Fleets of Large Convenience Store Chains Using Machine Learning’, *Algorithms*, 16(6), p. 286.

Lorena, A.C., De Carvalho, A.C. and Gama, J.M. (2008) ‘A review on the combination of binary classifiers in multiclass problems’, *Artificial Intelligence Review*, 30, pp. 19–37.

Lundberg, S.M. and Lee, S.-I. (2017) ‘A Unified Approach to Interpreting Model Predictions’, in I. Guyon et al. (eds) *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at:https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

McLachlan, G.J. (2005) *Discriminant analysis and statistical pattern recognition*. John Wiley & Sons.

Merrick, L. and Taly, A. (2019) ‘The explanation game: Explaining machine learning models with cooperative game theory’, *arXiv preprint arXiv:1909.08128* [Preprint].

Mugurusi, G. and Oluka, P.N. (2021) ‘Towards Explainable Artificial Intelligence (XAI) in Supply Chain Management: A Typology and Research Agenda’, in A. Dolgui et al. (eds) *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems*. Cham: Springer International Publishing (IFIP Advances in Information and Communication Technology), pp. 32–38. Available at: https://doi.org/10.1007/978-3-030-85910-7_4.

Platt, J. and others (1999) ‘Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods’, *Advances in large margin classifiers*, 10(3), pp. 61–74.

Prokhorenkova, L. *et al.* (2018) ‘CatBoost: unbiased boosting with categorical features’, *Advances in neural information processing systems*, 31.