International Journal of AI & Data Science Review
Volume1, Issue1, 2025

AI–POWERED FORE CASTING ALGORITHMS TO OPTIMIZE LAST MILE DELIVERY

Sari[1], Tole Sutikno[2]
[1]Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia

Article Info

Received: 28-01-2025 ***Revised:06-02-2025*** ***Accepted:17-02-2025*** ***Published:27-02-2025***

ABSTRACT
Last-mile delivery forecasting is a vital component in preserving an advantage over rivals in the fiercely competitive world of supply chain and e-commerce operations. In order to save operating costs, this study presents a unique machine-learning technique created especially for location-based last-mile delivery forecasting. A variety of machine learning algorithms, such as regression models, decision trees, and neural networks, are used to uncover subtle patterns in the data by using historical demand data, location-specific characteristics, and economic indicators. These models are then compared to traditional forecasting models like SARIMA and ARIMA. The study uses principal component analysis to handle any problems with high-dimensional data. After extensive hyperparameter adjustment, the final model is chosen and assessed using a separate dataset. For easier understanding, a graphic flowchart that summarizes the whole forecasting process is also included. The suggested method demonstrates machine learning's enormous potential for enhancing last-mile delivery predictions, opening the door for lower operating costs and increased overall productivity in supply chain and e-commerce businesses.

INTRODUCTION

Last-mile delivery has become much more important in recent years, particularly in supply chain and e-commerce operations. This is explained by the immediate impact it has on cost optimization, delivery effectiveness, and customer happiness. The term "last-mile delivery" refers to the last stage of the delivery process, in which items are delivered from a transportation hub to their final location. Consequently, accurate demand forecasting for last-mile delivery is essential to guaranteeing smooth operations and helping companies stay competitive in a market that is changing quickly. Accurate demand forecasting has several potential advantages. Businesses may more efficiently organize their workers, save overtime expenses, and avoid employee fatigue by forecasting the amount of packages and

and increases employee satisfaction by enabling firms to proactively modify schedules and hire more staff as needed. With an emphasis on operational cost optimization, this study presents a machine-learning method for location-based last-mile delivery forecasting in light of these advantages. The goal of the study is to determine the best machine learning models for improving the accuracy of last-mile delivery estimates by using historical demand data, location-specific characteristics, and economic indicators. In the end, supply chain and e-commerce businesses may see increased efficiency and decreased operating expenses as a consequence.

1. DATASETANDFEATURES

The study's main dataset is compiled from historical data that shows the demand trends for last-mile delivery in many places. The information, which was collected over a considerable amount of time, provides a thorough examination of the subtleties and dynamics of last-mile delivery operations. The collection contains three different kinds of features: economic indicators, location-specific characteristics, and demand data.

Demand data

Demand data, which includes historical trends of demand for last-mile delivery services in various locations, is a vital component. A strong basis for predicting future demand is provided by this data, which makes it easier to identify trends and patterns across time. In order to take these elements into consideration, the dataset comprises:
1. Data from the previous three years: To record any patterns and changes over time, a significant quantity of historical data is required. 2. Seasonality factors: Variations in demand throughout the year, such as more delivery during holidays or certain shopping occasions, are taken into account. 3. Holidays: The dataset takes into consideration the effect of public holidays and other special events on delivery volume.
4. Day of the week, week of the year, and month: These temporal variables are used to identify any recurring trends in delivery schedules.

Location-specific factors

The research is further deepened by location-specific characteristics. Among the factors taken into account in this category are the geographic dispersion of delivery sites, population density, infrastructure quality, and regional events (such as festivals or local holidays). These elements are important because they often have a big impact on operational effectiveness and delivery expectations. The following location-specific characteristics are included in the dataset: 1. 3-

digit zip codes: 3-digit zip codes reduce the amount of dummy variables and computational complexity when used to identify geographical regions. 2. Delivery hub, fulfillment center, state, and closest city: These characteristics provide light on regional sales trends and the logistical aspects of last-mile delivery. 3. Sales Trends: By identifying regions with greater demand for certain goods or services, regional sales data may be used to improve delivery routes and timetables.

Macro economic factors

Another layer of contextual information that may have an impact on the demand for last-mile delivery services is provided by economic statistics. Indicators that may affect consumer behavior and buying power and, thus, indirectly affect demand patterns include, but are not limited to, GDP, inflation, and employment rates. Together, these three feature categories create a comprehensive and educational dataset. To further enrich the dataset, feature engineering approaches are used, with an emphasis on developing derived variables and interactions that might increase the predictive capacity of the chosen machine learning models. The goal of the study is to provide machine learning algorithms a strong basis on which to learn and predict location-based last mile delivery demand using this meticulously selected and extensive dataset.

2. BENCHMARKMODELS

Benchmark models serve as a vital basis for comparing the efficacy of the machine learning algorithms used in this study. To provide preliminary estimates of last-mile delivery demand, these models—which include a naive approach, seasonal naïve, and a basic moving average model—are used. 1. Naive approach: This is a simple prediction model that makes the assumption that future demand will confirm the most recent finding. This model's main advantage is its simplicity and, under the correct circumstances, its astonishingly accurate projections.
2. Seasonal naïve model: By accounting for any seasonality in the data, this model goes beyond the naive method. It is a useful technique for demand forecasting in situations with recurring seasonal trends because it assumes that future values will follow historical seasonal patterns.
3. The statistical technique known as ARIMA (Autoregressive Integrated Moving Average) is widely used in time series forecasting. This model provides predictions based on previous data by combining autoregression, moving averages, and differencing. 4. easonalAutoregressiveIntegratedMovingAverage, or SARIMA: SARIMA incorporates easonal characteristics into the ARIMA model. The model is better able to capture cyclical oscillations in time series data because to this feature. 5. Weekly Moving Averages of the last n periods: This method provides a smoother depiction of the underlying demand pattern by calculating the average value for the previous n weeks on a rolling basis. This approach is used in the study to illustrate the possible advantages of using cutting-edge methods for location-based last-mile delivery forecasting. These conventional models are essential

for assessing the efficacy of the suggested machine learning technique as they act as the standard. Any gains in predicting accuracy for location-based last mile delivery demand that the machine learning algorithms used in this research make over existing models would be significant.

MACHINELEARNINGALGORITHMS

This research investigates the applicability of several machine learning techniques to improve the prediction of location-based last mile delivery. Every algorithm has distinct advantages for managing various facets of the forecasting assignment.
1. Linear Regression (LR): This approach fits a linear equation to observed data in order to predict the response variable. Estimating the coefficients that minimize the sum of the squared residuals is one of the stages.
DTs divide the data into subsets according to attribute values. They are quite interpretable.
3. Random Forests (RF): RF is an ensemble approach that makes use of several decision trees. It improves accuracy and manages overfitting by producing the mean prediction for regression or the mode of the classes for classification.
Gradient Boosting (GB) is a potent ensemble technique that constructs trees one at a time, with each new tree aiding in the correction of mistakes produced by the one that was previously trained. The model becomes more accurate and lowers the residuals with each additional tree. 5. Support Vector Machines (SVM): SVMs are models for supervised learning that are utilized in regression analysis and classification. Because numerous Kernel functions may be selected for the decision function, they are adaptable and effective in high-dimensional spaces.
6. Artificial Neural Networks (ANN): ANNs are models that estimate functions that rely on a large number of inputs. They are modeled after biological neural networks. Non-linear relationships may be modeled and processed using ANNs.
A thorough hyperparameter tuning and assessment procedure based on mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) is applied to every method. This careful process guarantees that the best algorithm is chosen for location-based last-mile delivery predictions.

3. DATA ENGINEERING,MODEL SELECTION AND EVALUATION

The suggested method includes a number of steps, all of which are essential to creating and choosing the best model for last-mile delivery forecasting.

FeatureEngineering

This step entails getting the dataset ready for analysis. Managing missing data, encoding category characteristics, scaling numerical features, and lowering dimensionality using methods like as Principal Component Analysis are all crucial tasks. These procedures guarantee that the data is appropriate for use in machine learning models and that the patterns present in the data are maintained and emphasized.

Model Selection and Evaluation

In order to guarantee the efficacy of the suggested machine learning technique in location-based last-mile delivery forecasting, the study presents a rigorous procedure. A thorough assessment and selection approach is used in this process to determine which algorithm performs the best. 1. The hyperparameter Tuning: To fine-tune the hyperparameter of the evaluated machine learning algorithms, both train and validation datasets are used in the first step. In addition to improving model performance, this step helps prevent overfitting.

2. Model Selection: Following the tuning procedure, each algorithm's top-performing model is chosen. This choice is based on assessment findings from the validation set and considers variables such as interpretability, accuracy, and computing efficiency. 3. Performance Evaluation on External Dataset: After that, the selected models are tested on an external dataset. This is a crucial stage in assessing the models' generalizability and determining their relevance in the actual world. Recent data often contains important information for time series forecasting. The research uses rolling window predictions for model training in order to take advantage of this. This method is based on an iterative process that uses a shifting window of data to retrain models. This window continuously moves forward in time to include the most current data available, improving the forecasts' precision and applicability. To evaluate the models' performance, a number of important assessment measures are used. These consist of Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). The study guarantees a thorough and impartial evaluation of the models by using these criteria, which take into consideration different facets of prediction accuracy and error magnitude. Results from benchmark models are compared with those from machine learning techniques. This comparison aids in choosing the best algorithm for location-based last-mile delivery predictions, along with a review

of the advantages and disadvantages of each model. In addition to offering high accuracy, the chosen model should be interpretable, computationally efficient, and robust to changes in the underlying data.

4. VISUALWORKFLOWOFPROCESS

To provide a clear and comprehensive understanding of the location-based last-mile delivery forecasting process utilizing machine learning algorithms, a visual flowchart is built.

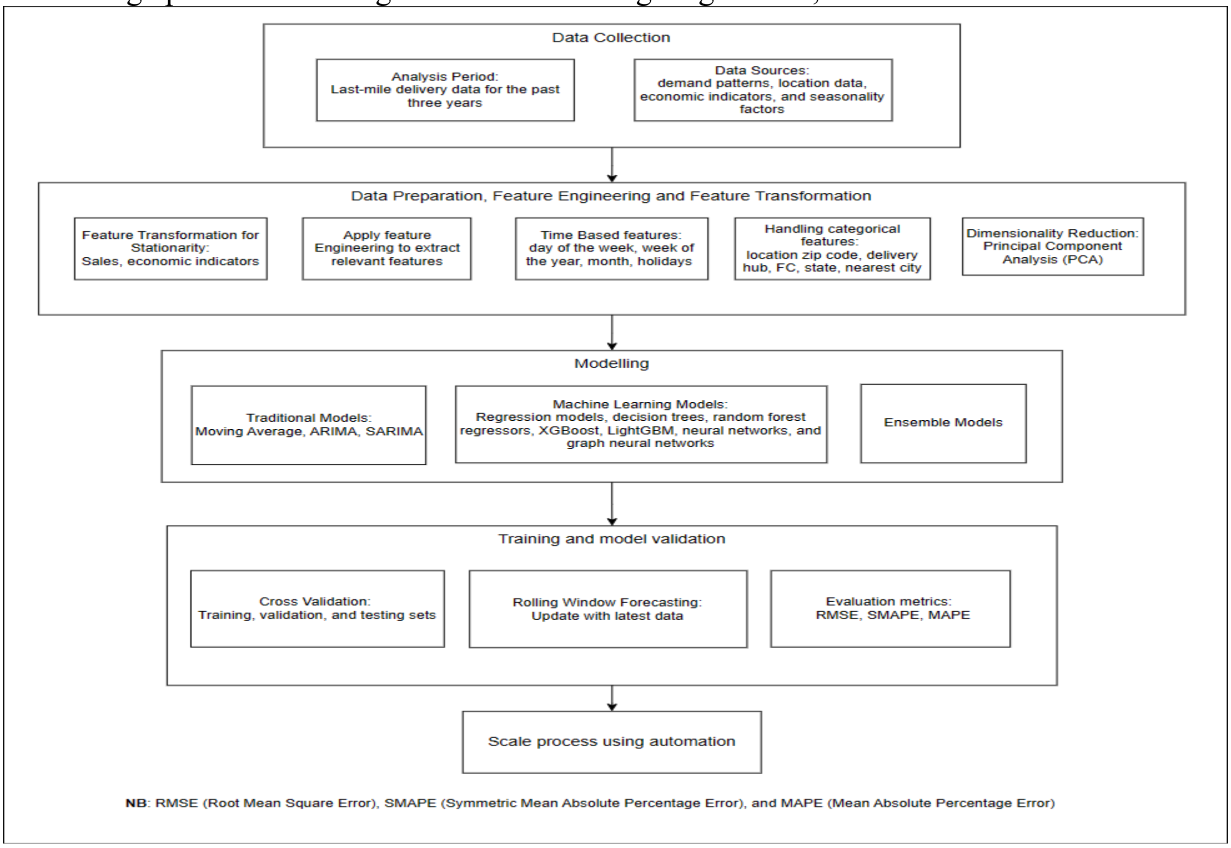

Fig.1: WorkflowforLocation-BasedLast-MileDeliveryForecastingusingMachineLearning Algorithms

A visual flowchart of the suggested machine learning-based method for location-based last-mile delivery forecasting is shown in Figure 1. Data collection, including demand trends, regional data, economic indicators, and seasonality variables during the previous three years, is the first step in the process. After the data is acquired, it moves into a rigorous preparation phase where transformation methods are used to ensure data stationarity and pertinent features are identified via feature engineering. At this level, categorical features like delivery hubs, FC, states, nearby cities, and zip codes of places are also handled. To better capture time-sensitive patterns, the model takes into consideration time-based features such as the day of the week, week of the year, month, and holidays. Dimensionality reduction is carried via using Principal Component Analysis (PCA). This phase resolves any issues brought on by possible multicollinearity and increases computing efficiency. Using cross-validation techniques, the model's training and validation stages focus on separating data into subsets for testing, validation, and training. A variety of measures, including Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE), and Root Mean Square Error (RMSE), are used to assess the effectiveness of the model. Regression models, decision trees, random forest regressors, XGBoost, LightGBM, neural networks, and graph neural networks are among the machine learning models that are evaluated in the process. These are assessed in conjunction with more conventional models like moving averages, SARIMA, and ARGIMA. A rolling window forecasting method is included into the procedure to preserve the models' accuracy with regard to the most recent data. Additionally, the approach supports ensemble models, using the combined capabilities of many models to improve prediction accuracy and resilience. Lastly, automation is used to design the process to run at scale. In real-world applications, this methodical, step-by-step process guarantees effective administration of data and machine learning models at scale.

5. FUTURERESEARCH

This study highlights the enormous possibility for further investigation even though it offers useful insights into the need for last-mile delivery predictions utilizing machine learning methods. More flexible and advanced forecasting models are becoming more and more necessary due to the dynamic nature of e-commerce and the constantly changing behavior of consumers. Future studies might examine the effectiveness of cutting-edge AI technologies like deep learning and reinforcement learning in this context or go further into examining the synergies between various machine learning models. Additionally, real-time forecasting, the incorporation of other data sources like weather patterns or social media trends, and research into the effects of abrupt market disruptions might all be taken into consideration. In the end, our study supports operational efficiency and competitiveness in the fast evolving logistics market by paving the way for more thorough and sophisticated predictive models for last-mile delivery demand forecasting.

6. CONCLUSION

Using a range of machine-learning methods, this study offers a thorough method for forecasting last-mile delivery demand. It has been shown via thorough assessment techniques that these algorithms provide significant gains over conventional models in prediction accuracy, computational efficiency, and resilience. By using rolling window forecasting, the models are continuously updated with new data, improving the accuracy of their predictions. This study demonstrates the potential of sophisticated computational methods in supply chain management and logistics, laying the groundwork for further research in this important area.

REFERENCES

[1] Vakulenko, Y., Shams, P., Hellström, D., & Hjort, K. (2019). Online retail experience and customer satisfaction: the mediating role of last mile delivery. The International Review of Retail, Distribution and Consumer Research, 29(3), 306-320.

[2] Hübner,A.H.,Kuhn,H.,&Wollenburg,J.(2016).Lastmilefulfilmentanddistributioninomni- channel grocery retailing: a strategic planning framework. International Journal of Retail & Distribution Management, 44(3).

[3] Magaña, U., Mansouri, S. A., & Spiegler, V. L. (2017). Improving demand forecasting in the air cargo handling industry: a case study. *International Journal of Logistics Research and Applications*, *20*(4), 359-380.

[4] Dantas, T. M., Oliveira, F. L. C., & Repolho, H. M. V. (2017). Air transportation demand forecast through Bagging Holt Winters methods. *Journal of Air Transport Management*, *59*, 116-123.

[5] Hayel, Y.,Quadri,D., Jiménez, T.,& Brotcorne, L.(2016). Decentralizedoptimization oflast-mile delivery services withnon-cooperative bounded rational customers. Annals of Operations Research, 239, 451-469.

[6] Simoni, M. D., Kutanoglu, E., & Claudel, C. G. (2020). Optimization and analysis of a robot-assistedlastmiledeliverysystem.TransportationResearchPartE:LogisticsandTransportation Review, 142, 102049.

[7] Salama,M.,&Srinivas,S.(2020).Jointoptimizationofcustomerlocationclusteringanddrone-basedroutingforlast-miledeliveries.TransportationResearchPartC:EmergingTechnologies, 114, 620-642.

[8] Bányai, T. (2018). Real-time decision making in first mile and last mile logistics: How smart schedulingaffectsenergyefficiencyofhyperconnectedsupplychainsolutions.Energies,11(7), 1833.