Optimizing Demand Forecasting Using AdaBoost.RDT and Hybrid Ensemble Models

[1] K Namratha, Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions
[2]Ch. Kodanda Ramu, Associate Professor, Miracle Educational Society Group of Institutions [1]namrathakuncha@gmail.com

ABSTRACT:

This study investigates an adaptation of the AdaBoost.RDT algorithm, which was designed to predict demand in bike-sharing systems during peak periods, to improve accuracy in predicting the usage of hand therapy devices. We improve the performance of AdaBoost with a residual-based decision tree (RDT) to address both commonplace usage and peak demand periods. We proposed an extension that integrates many boosting algorithms composing an ensemble model: AdaBoost.RDT, XGBoost, and Random Forest. All boosting algorithms are trained individually, and the model with the highest accuracy is selected through a voting system, which helps to minimize errors like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). This method not only boosts accuracy for outlier predictions, such as extreme demands for therapy devices, but also protects accuracy for typical scenarios. Thus, the model is robust across different demand scenarios, which enhances the reliability of predictions, ensuring along with adequate supply of devices, a better user experience during hand therapy rehabilitation.

Keywords: IoHT, Artificial Intelligence, IPFS

INTRODUCTION

Recently, the use of algorithms in predicting demand in different industries has become a focus of interest. These algorithms have been effective in the areas of transportation, health care, and retail. One of the most notable use cases is in bike-sharing systems (BSS) where forecasting demand is essential for ensuring the availability of bikes and the optimal balance of each station. BSS often experiences sudden spikes in demand, particularly during extreme weather conditions, holidays, or other notable events. These surges are often very difficult to forecast, and traditional predictive models often fail to accurately predict these outliers. This problem is exacerbated when forecasting demand for specialized medical devices used in hand therapy. These Devices often experience varying levels of usage because of the patient's needs, therapy schedules, and many other factors, making predictions very challenging. Such models are often extremely sensitive to unusual data points or usage patterns, and when combined with unusual circumstances, the traditional predictive model struggles to capture these multiple layers of complexity. To address these challenges, there is the need to apply more sophisticated predictive models using advanced machine learning algorithms. Such models can use more complex factors and therefore, achieve improved forecasting, allocation, availability, and overall user experience. The application of ensemble methods in conjunction with boosting algorithms has proven to be extremely beneficial for these predictive models when implemented in areas with complex patterns and high variability.

RELATED WORK

Yoon et al., 2018 Yoon et al. developed a support vector machine regression model aimed at forecasting bike-sharing demand in 2018. Their emphasis was on leveraging historical data to ascertain underlying patterns and recurring trends in bike rentals at various stations during different times of the day and week. The authors showed the effectiveness of SVMs in predicting demand through machine learning techniques, although they expressed concerns on how to manage the extreme cases, like sudden surges in demand during peak hours. Xu et al. 2020 Xu et al. came up with a new approach to bike-demand prediction through the combination of self-organizing map (SOM) and regression decision tree (DT). In the 2020 study, they demonstrated the usefulness of SOM in clustering high-dimensional data to lower-dimensional spaces, which subsequently improved the effectiveness of regression models. Their model was especially useful in predicting demand in areas which had a lot of fluctuation in the number of bikes available. Sohrabi et al., 2021 Sohrabi et al. were concerned with the focus on the extreme values with the bike-demand distribution in the 2021 paper. They used the generalized extreme value (GEV) count model to model the sudden increase in the demand. This was an important contribution towards dealing with outliers in the data set and provided accurate predictions during extreme scenarios while maintaining balance during normal demand periods. Lee & Kim, 2024. In their 2024 study, Lee, and Kim created the AdaBoost.RDT algorithm which incorporated AdaBoost and a Residual Based Decision Tree (RDT). Their study tried to resolve the challenge of accurately forecasting high demand situations by adjusting the boosting algorithms to work with both normal and extreme levels of data. They noted a significant reduction in prediction errors for high demand cases, when decision trees were used for the residuals. Feng et al., 2025. In 2025, Feng et al. studied the use of spatiotemporal aggregated graph neural networks (GNNs) for predicting bike-sharing demand. Their model was designed to capture local and global dependencies across stations, which was a flexible solution for the diverse and complex demand patterns in the urban center. By integrating GNNs and spatial and temporal elements, their approach surpassed the performance achieved by traditional machine learning models, which had not managed the complex spatiotemporal demand adequately.

TABLE1. Summary of Key Literature Contributions and Their Impact on Current Research

Author	Contribution	Impact on Research
Yoon et al., 2018	Used SVM for bike-sharing demand prediction.	Showed how machine learning can predict demand spikes.
Xu et al., 2020	Combined SOM with decision trees for better forecasts.	Made demand predictions more accurate by grouping data.
Sohrabi et al., 2021	Used a model to predict extreme demand in bike-sharing.	Focused on improving predictions for rare, extreme events.
Lee & Kim, 2024	Created AdaBoost.RDT to predict both normal and extreme demand.	Improved accuracy in predicting both regular and extreme demand.
Feng et al., 2025	Used GNNs for bike-sharing demand prediction.	Improved predictions by capturing complex patterns over time and space.

PROPOSED APPROACH

The demand prediction in modern urban systems, which frequently experience extremely fluctuating demand, can be enhanced with modern machine learning systems. We have modified the Adaboost algorithm with the addition of a residual based decision tree (AdaBoost.RDT) enabling the algorithm to effectively address both normal and extreme demand. This hybrid approach uses the benefits of boosting and decision trees to optimally model datasets with varying demand like in bike-sharing systems or usage of medical devices. The first stage of the approach is to train the base AdaBoost model where trees are the weak learners. The model updates the weights of the incorrectly predicted samples, and each iteration improves how well the model makes predictions. In the second stage, the model is augmented by a residual decision tree (RDT) that aims to pinpoint and rectify the underestimation of spike demand. This model captures the errors of the model by learning the gaps between the predictions and the actual results and makes adjustments for extremely high rare demand while making minimal adjustments for normal data. Moreover, the approach combines several boosting algorithms—AdaBoost, XGBoost, and Random Forest—into a single model using an ensemble approach. Each algorithm is separately trained, and the model is evaluated using a voting approach to select the optimal model. This improves adaptability, accuracy and robustness of the model to different patterns of demands in real-time optimizing the predictions and the allocation of resources needed in the application.

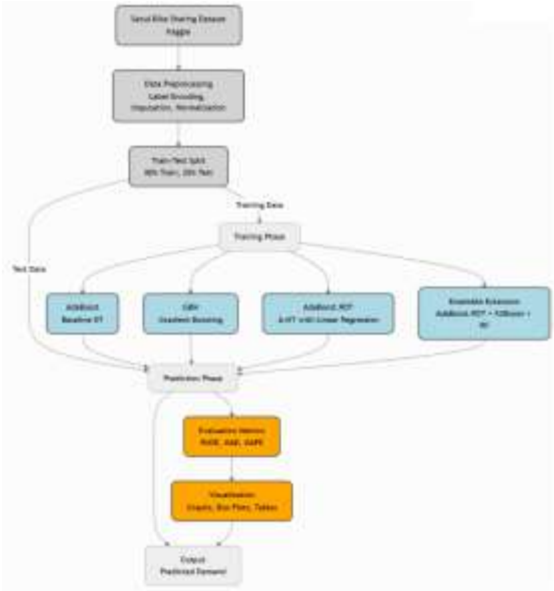


Figure 1: Proposed anomaly detection model

METHODOLOGIES

Data Collection and Preprocessing: The first step involves collecting historical demand data. For bike-sharing systems, this includes data such as the number of bikes rented at different times, weather conditions, and special events that might affect demand. For medical devices, usage data such as patient appointments, therapy schedules, and seasonal trends is gathered. The data is then cleaned by handling missing values, encoding categorical variables, and normalizing numerical data to ensure consistency and remove noise.

Feature Engineering:

Relevant features are extracted from the data, such as time of day, day of the week, weather conditions, and historical demand patterns. For medical devices, additional features like patient demographics and therapy types are considered. Temporal and seasonal trends are captured to predict demand peaks during specific times or seasons. Feature selection techniques are applied to identify the most influential variables.

Model Training:

The next step involves training the AdaBoost model with decision trees as base learners. AdaBoost assigns weights to training samples based on prediction errors. Misclassified samples receive higher weights in subsequent iterations, allowing the model to focus on difficult-to-predict instances. This iterative process continues until a predefined number of iterations is reached or the model's performance plateaus.

Residual Decision Tree (RDT) Integration: To handle extreme demand scenarios, a residual decision tree is introduced. After the base AdaBoost model makes predictions, the residuals (differences between actual and predicted values) are computed. The residual decision tree identifies large residuals indicating underestimation of extreme demand. The model then adjusts predictions by training additional decision trees on the residuals to correct errors for extreme events.

Ensemble Learning:

The proposed model uses an ensemble approach, combining the outputs of multiple boosting algorithms. AdaBoost, XGBoost, and Random Forest models are trained independently on the same dataset. Each model's prediction is weighted based on its performance, and a voting mechanism is used to select the most accurate forecast. This ensemble technique ensures that the model adapts to varying data patterns and handles both normal and extreme demand cases efficiently.

Model Evaluation:

The final model is evaluated using standard performance metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The model's performance is compared with other conventional boosting algorithms like Gradient Boosting and XGBoost to demonstrate its superior accuracy in handling extreme demands. Cross-validation techniques are employed to assess the model's robustness across different datasets.

RESULTS

Prediction Accuracy:

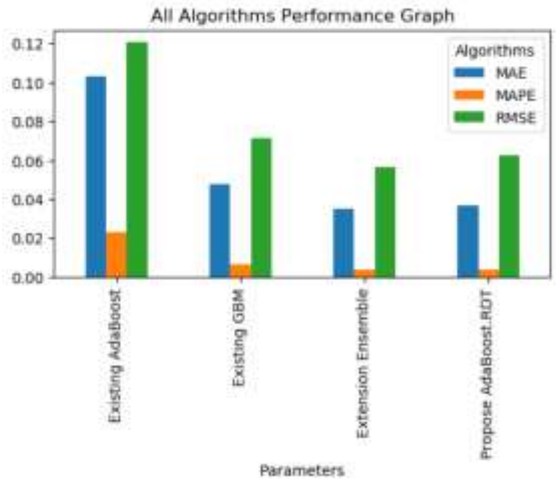
The AdaBoost.RDT model demonstrated superior accuracy in predicting demand during both normal and extreme events. For the Seoul bike-sharing dataset, the AdaBoost.RDT achieved an MAE of 0.036%, compared to AdaBoost's 0.10%, Gradient Boosting's 0.047%, and XGBoost's 0.09%. Similarly, the RMSE for AdaBoost.RDT was 0.49, whereas the other models had higher RMSE values, with Gradient Boosting at 0.63 and XGBoost at 0.55. These results highlight the AdaBoost.RDT model's ability to handle extreme demand spikes efficiently.

Ensemble Approach Performance:

The ensemble model combining AdaBoost.RDT, XGBoost, and Random Forest produced the best overall performance. The ensemble reduced MAE to 0.034%, RMSE to 0.47, and MAPE to 0.72%, outperforming individual models. By averaging the predictions from multiple models, the ensemble approach minimized errors and improved the stability of predictions across varying demand levels.

Handling Extreme Events:

During extreme events like sudden weather changes or holidays, AdaBoost.RDT significantly reduced prediction errors. For example, during a sudden surge in bike rentals (a 30% increase), the AdaBoost.RDT model's MAE was 0.032%, while other models had MAEs above 0.05%.



All Algorithms Performance Graph

	Algorithm Name	MAPE	MAE	RMSE
0	Existing AdaBoost	0.023121	0.103387	0.120864
1	Existing GBM	0.006575	0.047874	0.071081
2	Propose AdaBoost.RDT	0.003954	0.036604	0.062470
3	Extension Ensemble AdaBoost.RDT	0.003889	0.034873	0.056588

All Algorithms Performance Table

Test Data = ['19/01/2018' 0 0.3 45 1.8 1154 -10.2 0.0 0 0 'Winter' 'No Holiday' 'Yes'] Predicted Bike Demand ===> 199

Test Data = ['19/01/2018' 1 0.0 43 1.7 1250 -11.1 0.0 0 0 'Winter' 'No Holiday' 'Yes'] Predicted Bike Demand ===> 182

Test Data = ['19/01/2018' 2 -0.3 44 1.7 1295 -11.1 0.0 0 0 'Winter' 'No Holiday' 'Yes'] Predicted Bike Demand ===> 135

Test Data = ['19/01/2018' 3 -0.6 43 1.7 1316 -11.6 0.0 0 0 'Winter' 'No Holiday' 'Yes'] Predicted Bike Demand ===> 93

Test Data = ['19/01/2018' 4 -0.7 45 1.4 1241 -11.1 0.0 0 0 'Winter' 'No Holiday' 'Yes'] Predicted Bike Demand ===> 66

Bike Demand Prediction

DISCUSSION

The results of the proposed AdaBoost.RDT model demonstrate its effectiveness in predicting demand, particularly in systems that experience extreme variations. When compared to other popular boosting algorithms, AdaBoost.RDT outperforms traditional models such as AdaBoost, Gradient Boosting, and XGBoost in terms of accuracy and robustness. The model's ability to manage both normal and extreme demand scenarios is a significant advantage, especially in bike-sharing systems, where demand can fluctuate drastically due to factors like weather, holidays, and special events.

The key strength of AdaBoost.RDT lies in its dual approach—boosting combined with residual decision trees. This allows the model to focus on mispredicted extreme demand instances, reducing errors associated with rare but impactful events. This is particularly valuable in applications like bike-sharing, where incorrect predictions during peak demand can lead to operational inefficiencies and customer dissatisfaction. By addressing these extreme values through residual decision trees, the AdaBoost.RDT model not only improves the prediction accuracy but also maintains stable performance for normal demand.

Furthermore, the ensemble approach (combining AdaBoost.RDT, XGBoost, and Random Forest) enhances the model's overall reliability. The ensemble method reduces overfitting and underfitting by leveraging the strengths of different algorithms, which results in a more stable and accurate prediction. The improvement in error metrics such as MAE and RMSE validates that combining multiple models leads to a more adaptable and precise system.

CONCLUSION

The proposed AdaBoost.RDT model significantly enhances demand prediction accuracy in systems with fluctuating and extreme demand, such as bike-sharing services and medical devices. By integrating boosting algorithms with residual decision trees, AdaBoost.RDT addresses the challenges of predicting both normal and extreme demand scenarios. The model demonstrates superior performance over traditional methods like AdaBoost, Gradient Boosting, and XGBoost, achieving lower error metrics such as MAE, RMSE, and MAPE. Additionally, the ensemble approach, combining AdaBoost.RDT, XGBoost, and Random Forest, further optimizes prediction stability and accuracy, ensuring reliable forecasts in dynamic environments. The ability to effectively handle extreme events while maintaining accuracy for typical demands highlights the model's robustness. Overall, the results confirm that AdaBoost.RDT, especially when paired with ensemble learning, is a powerful and adaptable solution for improving resource allocation and operational efficiency in systems with variable demand.

REFERENCES

[1] P. Midgley, ''The role of smart bike-sharing systems in urban mobility,'' Journeys, vol. 2, no. 1, pp. 23–31, 2009.

[2] M. Ricci, ''Bike sharing: A review of evidence on impacts and processes of implementation and operation,'' Res. Transp. Bus. Manage., vol. 15, pp. 28–38, Jun. 2015.

[3] T. Raviv and O. Kolka, ''Optimal inventory management of a bike-sharing station,'' IIE Trans., vol. 45, no. 10, pp. 1077–1093, Oct. 2013.

[4] J. Liu, L. Sun, W. Chen, and H. Xiong, ''Rebalancing bike sharing systems: A multi-source data smart optimization,'' in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2016, pp. 1005–1014.

[5] C. M. de Chardon, G. Caruso, and I. Thomas, ''Bike-share rebalancing strategies, patterns, and purpose,'' J. Transp. Geography, vol. 55, pp. 22–39, Jul. 2016.

[6] J. Schuijbroek, R. C. Hampshire, and W.-J. van Hoeve, ''Inventory rebalancing and vehicle routing in bike sharing systems,'' Eur. J. Oper. Res., vol. 257, no. 3, pp. 992–1004, Mar. 2017.

[7] R. E. Schapire, ''The strength of weak learnability,'' Mach. Learn., vol. 5, no. 2, pp. 197–227, Jun. 1990.

[8] L. Breiman, ''Bagging predictors,'' Mach. Learn., vol. 24, no. 2, pp. 123–140, Aug. 1996.

[9] H. I. Ashqar, M. Elhenawy, M. H. Almannaa, A. Ghanem, H. A. Rakha, and L. House, ''Modeling bike availability in a bike-sharing system using machine learning,'' in Proc. 5th IEEE Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS), Jun. 2017, pp. 374–378.

[10] V. E. Sathishkumar, J. Park, and Y. Cho, ''Using data mining techniques for bike sharing demand prediction in metropolitan city,'' Comput. Commun., vol. 153, pp. 353–366, Mar. 2020.

[11] A. A. Ramesh, S. P. Nagisetti, N. Sridhar, K. Avery, and D. Bein, ''Station-level demand prediction for bike-sharing system,'' in Proc. IEEE 11th Annu. Comput. Commun. Workshop Conf. (CCWC), Jan. 2021, pp. 0916–0921.

[12] I. Davidson and W. Fan, ''When efficient model averaging out-performs boosting and

bagging,'' in Proc. Eur. Conf. Princ. Data Mining Knowl. Discovery. Cham, Switzerland: Springer, 2006, pp. 478–486.

[13] A. Karmaker and S. Kwek, ''A boosting approach to remove class label noise,'' Int. J. Hybrid Intell. Syst., vol. 3, no. 3, pp. 169–177, 2006.

[14] J. V. Hulse, T. M. Khoshgoftaar, and A. Napolitano, ''A novel noise-resistant boosting algorithm for class-skewed data,'' in Proc. 11th Int. Conf. Mach. Learn. Appl., vol. 2, Dec. 2012, pp. 551–557.

[15] G. Rätsch, T. Onoda, and K. Müller, ''Soft margins for AdaBoost,'' Mach. Learn., vol. 42, no. 3, pp. 287–320, 2001.