Identifying Express Stops for Manila LRT-Line 2: A Comparative and Clustering-Based Approach

Harvy Angelo D. Tan
Department of Computer Engineering
Technological Institute of the Philippines - Manila
City of Manila, Philippines
harvyangelo.tan@gmail.com

Khayzel Anne A. Garcia
Department of Computer Engineering
Technological Institute of the Philippines - Manila
City of Manila, Philippines
khyzlgrc@gmail.com

Mark James P. Abot
Department of Computer Engineering
Technological Institute of the Philippines - Manila
City of Manila, Philippines
markjamesabot@gmail.com

***Abstract*—With rising demand for efficient transit in Metro Manila, this study proposes a data-driven approach to designing an Express Line for the LRT-2 system. A Random Forest model, trained on Kintetsu transit data from Japan, was used to analyze features linked to express stop selection. Using DBSCAN, high-traffic stations were identified based on entry and exit data. Both models identified the same key stations, despite the different methods. The supervised Random Forest model showed viability for using existing mixed local and express lines as basis for proposing express stops to a local only transit line, while the unsupervised DBSCAN algorithm proved suitable for independent line analysis. Although limited by dataset size and granularity, the results show how machine learning can inform better planning decisions. The study offers a framework for future enhancements that prioritize efficiency, accessibility, and rider satisfaction based on data-driven proposals.**

***Index Terms*—transit, express line, data science, random forest, DBSCAN**

I. Introduction

A. Background

Urban mobility plays a critical role in shaping the daily lives of citizens, particularly in densely populated and rapidly urbanizing regions like Metro Manila. As the metropolitan area continues to grapple with chronic transportation inefficiencies, outdated infrastructure, and increasing commuter demand, the urgency for sustainable, data-informed interventions has never been greater. Despite recent infrastructure initiatives, operational delays and long gestation periods mean that millions of commuters still face long travel times, unreliable service, and suboptimal transit planning [1].

The Philippine capital has repeatedly ranked among the lowest globally in terms of urban mobility readiness. Reference [2] had Manila placed 65th out of 70 major global cities. Such a ranking reflects not just physical infrastructure gaps, but also inefficiencies in policy execution, system integration, and real-time response to commuter needs. Local perceptions echo these concerns, with documented widespread dissatisfaction with wait times, vehicle crowding, and general system unreliability [3].

Among the rail networks operating in Metro Manila, the Light Rail Transit Line 2 (LRT-2) stands out as a critical artery connecting eastern and central districts of the region. With over 49 million passengers served in 2023 alone [4], LRT-2 presents a strategic opportunity for improving system efficiency without major infrastructure overhauls. However, like most of the region's networks, LRT-2 adheres to a purely local-stop configuration, forcing all trains to halt at every station irrespective of varying passenger demand.

This study proposes an innovative solution centered on the integration of an express line within the existing LRT-2 framework. While express-stop systems are common in mature rail networks globally, they remain largely absent in the Philippines. In countries like Japan, express operations have proven effective in balancing commuter demand with travel time savings. For example, the Kansai lines of Kintetsu Railway in Japan, which are comparable in size and complexity of service to the entire commuter rail network in Manila, demonstrate how local and express services can be harmoniously deployed within the same corridor [5].

The novelty of this paper lies in its hybrid data science analytics approach to enhancing urban rail systems. It combines Random Forest classification to identify express stop features from the Kintetsu network with DBSCAN clustering to pinpoint high-demand LRT-2 stations. This dual-method framework enables data-driven decisions without major infrastructure overhauls. While LRT-2 serves as the case study, the methodology offers a scalable model for future transit reforms aligned with sustainable mobility goals [6] [7].

B. Objectives

- To benchmark features of transit networks with both Local and Express Lines using Random Forest classification on Kintetsu system data.
- To identify high-demand LRT-2 stations using the DBSCAN clustering algorithm for the formulation of an Express Line.
- To propose data-driven enhancements to LRT-2 operations that reduce travel time and improve service accessibility through express stop planning.

C. Statement of the Problem

Citizens of Metro Manila and neighboring regions experience a transportation crisis on a daily basis. Commuters regularly cite long wait times, outdated systems, and general inefficiencies in public transportation as major daily burdens [3]. These shortcomings directly affect the quality of life for millions of people who rely on public transit in the capital.

Data reinforces the scale of this issue. Reference [8] reports that approximately 60 percent of respondents in the Philippines identify trains and buses as their primary mode of transportation, affirming the importance of efficient mass transit systems. LRT-2 operates exclusively as a local-stop system, with every train halting at each station regardless of fluctuating demand, inherently limiting operational efficiency.

LRT-2 is particularly well-suited as the benchmark rail network for this study. As a medium-capacity, east-west line connecting the heart of Metro Manila to emerging urban centers in Rizal Province, it serves a diverse commuter base and represents the structural characteristics of many other urban rail lines in the country. Moreover, it remains manageable in scope for pilot-level implementations. Its recent expansion to include the Marikina and Antipolo stations in July 2021 also marks a new phase in its operational footprint. However, since May 2022, the Department of Transportation (DOTr) has ceased publishing updated ridership data, limiting transparency and restricting timely access to performance metrics.

To benchmark and analyze the implementation of express-stop service lines—an approach that could reduce passenger travel times and improve operational efficiency—this study draws a comparison to Japan's Kintetsu Railway system, specifically the Kintetsu Kansai Lines. Kintetsu provides a valuable reference model due to its similar network scale and urban-suburban coverage, and successfully operates both Local and Express services within the same network, offering practical, real-world insights into how a mid-sized transit system can balance accessibility with speed. The presence of express operations in Kintetsu allows for the generation of supervised learning models that can later be adapted to other rail networks.

The study will also explore the use of unsupervised learning to independently analyze the available domestic data for comparison and validation of results in both models used.

II. Methodology

A. Conceptual Framework

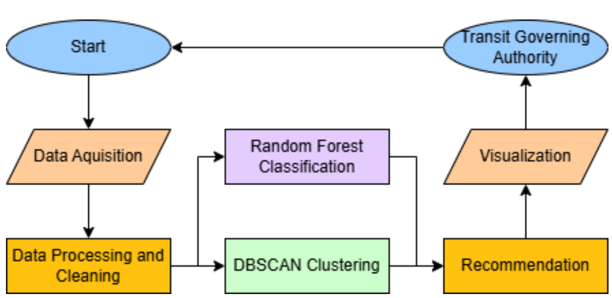

Fig. 1. Conceptual Framework of the Express Line Recommendation System

Fig. 1 shows the conceptual framework designed to analyze transit data and recommend possible express line stops for the LRT-2 system in Metro Manila. This approach makes use of both supervised and unsupervised machine learning models to deliver data-driven insights.

It starts with Data Acquisition, where transit and station data such as passenger volume, station locations, and train metrics, are collected from public sources or government institutions.

In the Data Processing and Analysis stage, the data is cleaned and organized for analysis. This step includes two main processes:

1) Random Forest Classifier: For external validation, a supervised learning model was trained using a benchmark dataset from Japan's Kintetsu Train System, which has established Express and Local line configurations. The model learns patterns from labeled data to classify whether a station should be tagged as an express stop.

2) DBSCAN (Density-Based Spatial Clustering of Applications with Noise): Used for unsupervised learning, DBSCAN identifies clusters of high-activity stations using their `activity_scores`, which are based on the number of passengers entering and exiting. The goal is to internally recognize stations that naturally group together and might qualify for express service.

Both outputs are sent to a Recommendation Module, which merges the insights from both learning models. DBSCAN identifies internal patterns in the LRT-2 dataset, while Random Forest validates station roles using features learned from Kintetsu. The proposed Express Line is developed using both perspectives.

Finally, the system offers a Visualization Layer, displaying the recommendations, clustering outcomes, and model evaluation graphs to help stakeholders better understand the results.

After applying the recommendations to the Transit Governing Authority, the system goes on a loop where it now collects data from the transit lines, given that it already has the applied recommendations. This feedback loop will ensure that the project will recommend the most optimal suggestions based on the data it will be given.

B. Data Gathering

Data collection was done using publicly available sources online, though each dataset had its own set of difficulties. For the Kintetsu-Nara Line, data availability was limited due to corporate restrictions. Luckily, snapshot datasets including average station-level ridership were available. Details like station locations and train specs were publicly accessible.

For LRT-2, the goal was to focus on July 2021 data, around the time the Marikina and Antipolo extensions opened. The proponents obtained ridership data via the Freedom of Information (FOI) online portal. This data covered the period until May 2022. Unlike the Kintetsu data which reported average entry/exit counts, LRT-2 provided total daily figures for 11 months. To align the two datasets, the daily average was computed and rounded for consistency.

The Kintetsu dataset did not specify the number of train sets used on each line. To fill this gap, the proponents estimated 16 train sets based on timetable data. All data was cleaned and formatted to serve the modeling objectives of the project. The following features were selected as necessary to properly assess the potential for an Express Line:

- Station name
- Distance from neighboring stations (in meters)
- Terminal status (whether it is a terminal or not)
- Express stop status (for training data)
- Average daily ridership
- Hours the station is operational
- Train-related features:
- Car capacity
- Number of train sets
- Operating speed
- Maximum speed

C. Model Training

1) Supervised Model: Random Forest Classifier: The proponents used Python libraries like Pandas and Scikit-learn to implement a Random Forest Classifier trained on the Kintetsu dataset. This dataset was chosen due to its mature network that separates Express and Local services. The model used the following features:

- `entry_exit`: Average number of passengers using the station.
- `distance_from_last`: Distance between stations.
- `terminal`: Whether the station is a terminal.
- `express`: Label to indicate if it is an express stop.
- `hours_active`: Number of hours the station operates.
- `train_sets`: Estimated number of trains.
- `operating_speed` and `max_speed`: Train movement speeds.

After training, the model was used to predict possible express stops on the LRT-2 line. The performance of the model was evaluated through: Confusion Matrix with Heatmap: To show how well the model predicted express and local stations.

- Feature Importance Chart: To reveal the most influential data points.
- Correlation Heatmap: To identify overlapping or related features.
- Learning Curve: To detect overfitting or underfitting.
- ROC Curve: To see how well the model distinguishes between classes.
- Precision-Recall Curve: Useful for scenarios with class imbalance.

2) Unsupervised Model: DBSCAN Clustering: The proponents also ran the DBSCAN algorithm on the LRT-2 dataset to detect natural groupings of high-traffic stations. The clustering was based solely on `activity_scores`, which were derived from the average entry/exit data. To visualize these findings:

- A Pie Chart showed the number of stations per cluster.
- A Heatmap displayed the average activity score per cluster.
- A Scatter Plot mapped the station clusters based on activity levels.

III. Results

Having trained the supervised model on the Kintetsu rail network in the Kansai region, the study now proceeded to testing the model on the LRT-2 network based on the analyzed features from the random forest algorithm to identify possible express stops. The DBSCAN algorithm was also used to identify clusters where express stop stations could be selected based solely on the domestic data.

A. Random Forest Test Results

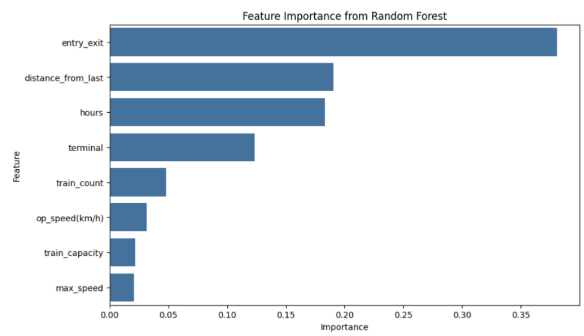

Fig. 2. Feature Importance for Random Forest Classifier

Because only a few features dominate the chart with much higher importance scores, Fig. 2 suggests the dataset's predictive power is concentrated in a few variables, which may indicate skewed feature distributions or that certain features are strongly correlated with the target class. This is largely true for transit networks, as local stops vastly outnumber express stops, making the top features stand out more.

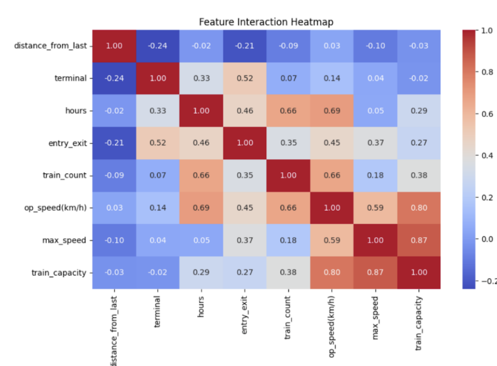

Fig. 3. Correlation Heatmap for Random Forest Classifier

Given the typical imbalance and skew in station data, the learning curve in Fig. 4 indicates that the initial high training accuracy overfits to the majority class consisting largely of local stops. However, the validation score improves as more data is added. This indicates that more diverse data helps the model generalize. But from 50% of the training set size, the validation curve plateaus. The Model has reached its capacity, which means further improvements may require addressing class imbalance or adding new features.

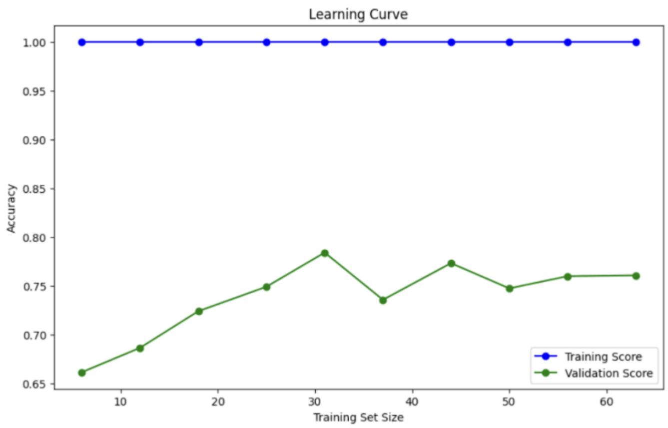

Fig. 4. Learning Curve of Random Forest Classifier

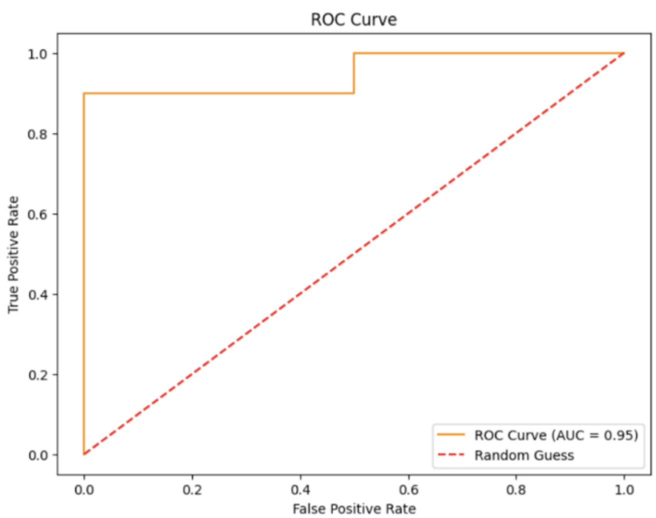

Fig. 5. ROC Curve of Random Forest Classifier

In the ROC Curve show in Fig. 5, the model shows strong discriminators in the features of the model. The model is therefore robust even if the dataset is somewhat imbalanced.

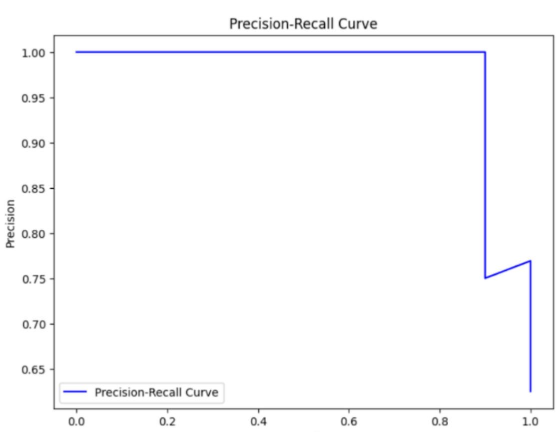

Fig. 6. Precision-Recall Curve of Random Forest Classifier

The Precision-Recall Curve shown in Fig. 6 is is especially informative for imbalanced datasets such as the one used in this study, where local stations far outnumber express stations. Unlike the ROC curve, which can present an overly optimistic view with imbalanced data, the precision-recall curve focuses on the performance for the minority class (express stations). As the graph shows, the model is very good at identifying express stations with few false positives and few false negatives.

```
LRT-2 Stations and Predicted Express Stops:
          stn_name  express_predicted
0            Recto                  1
1          Legarda                  0
2           Pureza                  0
3            VMapa                  0
4            JRuiz                  0
5          Gilmore                  0
6   BettyGo-Belmonte                0
7     Araneta-Cubao                 1
8           Anonas                  0
9        Katipunan                  0
10        Santolan                  0
11        Marikina                  0
12        Antipolo                  1
```

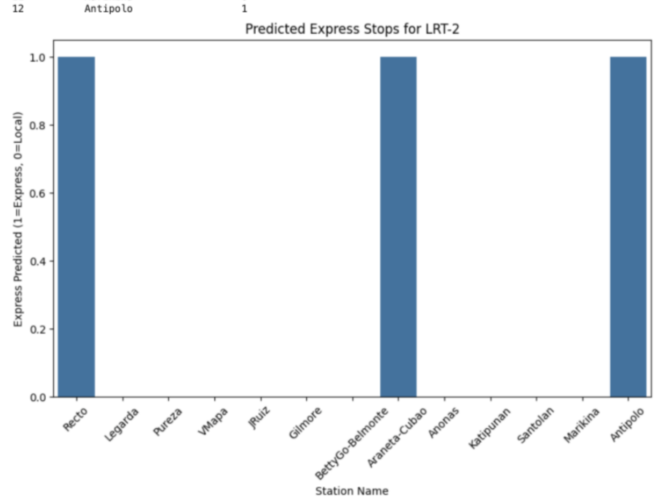

Fig. 7. Predicted express stops for LRT-2 using Random Forest

The Random Forest classifier, trained on the Kintetsu network dataset, achieved an overall accuracy of 93.75%. This indicates strong predictive performance in identifying express stops within a transit network. Feature importance analysis revealed that the most influential predictors were the distance from the previous station, passenger entry and exit data, and whether a station served as a terminal. These variables contributed significantly to the model's classification outcomes.

The model predicted Recto, Araneta-Cubao, and Antipolo as express stops within the LRT-2 system as shown in Fig. 7. These stations are terminals and are characterized by high levels of commuter activity. The identification of these stations aligns with their operational and strategic importance in the network. The relatively high feature importance assigned to entry and exit counts and terminal status suggests that stations with high ridership volumes and terminal functions are key determinants in the classification of express stops.

The application of the trained model to the LRT-2 system demonstrated effective knowledge transfer from the Kintetsu dataset. The high predictive accuracy obtained in this cross-system application supports the robustness of the selected features for similar transit environments. Nonetheless, the model's performance is influenced by the scope and quality of the Kintetsu dataset. Local factors unique to Metro Manila's transit context, which were not captured in the dataset, may limit the generalizability of the results.

B. DBSCAN Results

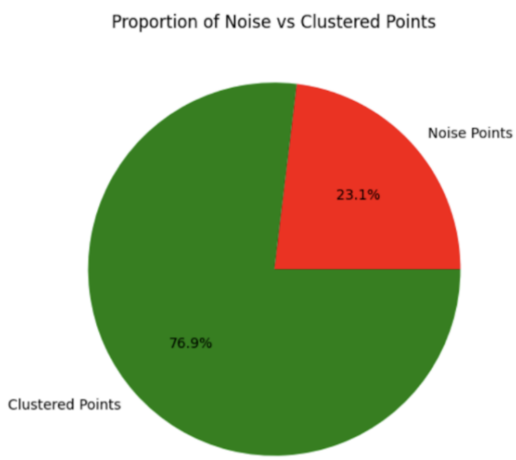

Fig. 8. Noise-to-Cluster Ratio

Based on Fig. 8, nearly a quarter of stations in Manila's LRT-Line 2 do not belong to any significant cluster of activity. These stations are likely outliers—either low-ridership, poorly connected, or spatially isolated. These stations are therefore not strong candidates for express stops. They may be better served by local trains only.

On the other hand, the majority of stations share enough similarity in terms of activity, ridership, or spatial proximity to form clusters. These are the core stations where demand is concentrated. These clustered stations are prime candidates for discriminating express stops.

The noise proportion of 20% to 30% is typical, especially if the system includes suburban or peripheral stations with low demand. It might indicate your DBSCAN parameters are too strict, or your features are not well-scaled. If the proportion of noise were much lower, it could mean the parameters are too loose, and clusters are not meaningful.

The DBSCAN method identified two clusters based on `activity_score`, as illustrated in Fig. 9. Cluster -1 has an average activity score of 1.6, shown in red, while Cluster 0 had a score of -0.48, shown in blue.

Cluster -1 groups stations with significantly above-average activity scores. These are likely the major hubs or high-demand stations in Manila's LRT-Line 2 and are therefore strong candidates for express stops. They represent the core demand centers where express trains would provide the most benefit.

Conversely, Cluster 0 groups stations with below-average activity scores. These are likely regular or low-demand stations. These stations are better suited for local service.

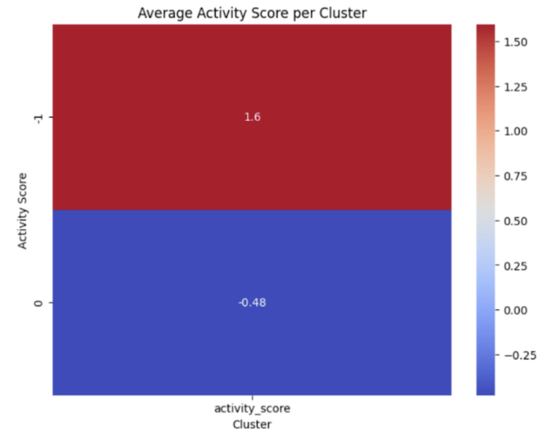

Fig. 9. Average Activity Score per DBSCAN Cluster

DBSCAN has successfully separated stations into two meaningful groups based on activity score. The clear difference in average scores (1.6 vs. -0.48) suggests that your features (activity scores) are effective at distinguishing high-demand from low-demand stations.

The clustering is robust and aligns well with operational goals: prioritize high-activity stations for express service. The model can be used to objectively justify which stations should be included in express operations. In this case, the data has enough variation in activity scores to allow clear clustering. There is a meaningful distinction in station usage patterns, supporting targeted service planning.

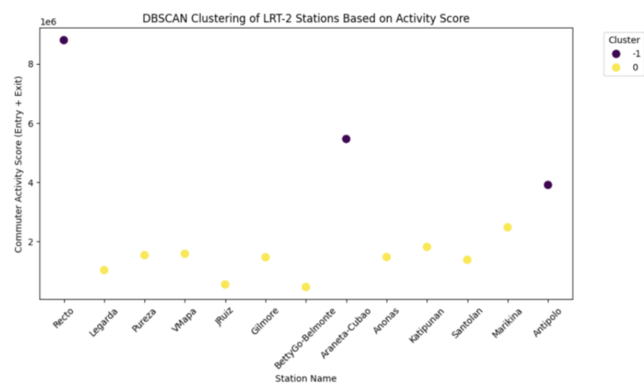

Fig. 10. Clustered express stops for LRT-2 using DBSCAN

The DBSCAN clustering algorithm was employed to identify potential express stops based on patterns of commuter activity. The algorithm grouped stations such as Recto, Araneta-Cubao, and Antipolo into dense clusters, as illustrated in Figure 10. These results corroborated the express stop classifications made by the Random Forest model. In contrast, stations with low levels of activity, such as J. Ruiz and Pureza, were categorized as noise, suggesting their limited suitability for express service designation.

The DBSCAN method operates independently of labeled training data and relies solely on inherent data structure. Its identification of similar express stops as those determined by the Random Forest model lends additional support to the findings. However, the clustering outcomes are sensitive to parameter selection, specifically the values assigned to `epsilon` and `min_samples`. As an unsupervised learning technique, DBSCAN does not produce an accuracy score; thus, interpretation of its results requires contextual analysis of cluster compositions and boundaries.

C. Comparison of Approaches

TABLE I
RESULT OF MODELS USED FOR IDENTIFYING MANILA LRT-LINE2 EXPRESS STOPS

	Random Forest (Supervised)	***DBSCAN (Unsupervised)***
Accuracy	93.75%	No direct metric, but clusters align with expected express stops.
Express Stops	Predicted Recto, Araneta-Cubao, Antipolo	Clustered Recto, Araneta-Cubao, Antipolo
Strengths	High accuracy; transferable insights from Kintetsu data	Data-driven; no training required
Limitations	Reliant on training data quality	Sensitive to parameter tuning; requires interpretation

Table I summarizes the key differences in performance and characteristics between two approaches—Random Forest (Supervised Learning) and DBSCAN (Unsupervised Learning)—as applied in the study.

Based on these results, we can observe that the results of either models are not at all different. This might be the case only particularly for LRT Line 2 of Manila, but used in other networks may yield varying results. Nevertheless, the use of both models strengthens the validity of both of their suggestions for express stops, both on the basis of comparison from existing mixed lines and solely with the data available.

The study primarily employed both approaches to see if selecting express stations could be identified through features or could a model for clustering discern stations apart enough to distinctly identify them without the influence of unreliable external data. With the similar results for both models, this indicates that the strengths of each approach can complement one another; Random Forest provides a grounded, feature-driven prediction based on labeled data, while DBSCAN offers a data-centric validation that reveals natural groupings without prior assumptions. Together, they create a more robust framework for identifying express stops, where supervised insights are reinforced by unsupervised patterns emerging directly from the data.

IV. CONCLUSION AND RECOMMENDATIONS

The analysis and evaluation of potential express stops in the LRT-2 system using supervised and unsupervised machine learning approaches demonstrate the viability of data-driven methodologies in optimizing transit networks.

A. Random Forest Method

The Random Forest Classifier trained on data from Japan's Kintetsu rail system achieved 93.75% accuracy. It successfully identified key predictors of express stop potential: entry and exit data, distance between stations, and terminal status. The model was used on Manila's LRT-2 to reveal potential express stops in **Recto**, **Araneta-Cubao**, and **Antipolo** stations.

However, the model's reliance on a relatively small and domain-specific training set highlights key limitations. While the Kintetsu dataset offers a useful comparison, it may not fully encapsulate the unique operational and socioeconomic dynamics of the LRT-2 system and the other transport options connected tightly or loosely with the rail line. Expanding the dataset with more detailed features would improve generalizability and performance. These features could include temporal ridership trends or inter-modal links.

Additionally, insights from [9] underscore the value of incorporating land use characteristics as predictive features. In their study, metro ridership was significantly influenced by land use densities, particularly government, educational, and mixed-use zones with the same Random Forest approach. Such parameters, if integrated into the current model, could enhance predictive performance by capturing demand-side factors rooted in urban form and functionality. This multi-dimensional approach may provide a more context-sensitive framework for identifying express stops, especially in densely layered urban environments like Metro Manila while retaining the same supervised model.

B. DBSCAN Method

The DBSCAN unsupervised algorithm, applied to station activity data without requiring prior labels, successfully identified **Recto**, **Araneta-Cubao**, and **Antipolo** as high-demand stations, supporting their consideration as express stops. This unsupervised approach proved effective in detecting natural ridership clusters, validating its utility in urban transit planning.

Despite its strengths, DBSCAN's sensitivity to parameters such as `epsilon` and `min_samples` presents a limitation. Variability in these settings can affect clustering outcomes, necessitating domain expertise for parameter tuning and interpretation. Moreover, the absence of a standardized evaluation metric in unsupervised learning highlights the need for qualitative validation alongside statistical results.

The use of DBSCAN in the context of transport stop location optimization, as demonstrated in [10], further supports the method's applicability in transit planning. Their approach, which combines DBSCAN with distance constraints and leverages K-means for parameter selection, illustrates how domain-specific considerations—such as pedestrian accessibility—can enhance clustering relevance. This suggests that integrating spatial constraints or travel behavior characteristics into the DBSCAN process could further refine the identification of suitable express stops, especially in systems where accessibility and connectivity play a pivotal role.

C. Express Stops

Both Random Forest and DBSCAN converged in recommending the same three stations—**Recto**, **Araneta-Cubao**, and **Antipolo**—as ideal candidates for express stops. This alignment between supervised and unsupervised methods strengthens the validity of the findings and illustrates the robustness of machine learning in public transport planning.

Nonetheless, current limitations, such as the exclusion of station-level socioeconomic data and limited temporal granularity, suggest that future iterations should integrate broader datasets for more comprehensive optimization.

D. Recommendations

To enhance the transit systems in the country and establish a data-driven and sustainable transportation model, future projects should consider the following recommendations based on the findings of this study:

1) Expand Data Collection and Accessibility: Future projects should prioritize broader and more accessible data to strengthen the quality of data-driven transit solutions. This includes expanding feature sets to cover temporal ridership trends (such as hourly, daily, and seasonal patterns), socioeconomic factors like population density and income levels near stations, and connectivity metrics that reflect integration with other transport modes such as buses and jeepneys.

2) Broader and Deeper Data: Collaborate with government agencies and corporations to access valuable datasets, allowing more research and innovation in the transportation sector using advanced data science to inform decision-making.

3) Adopt More Advanced Analytical Tools: More researchers should utilize machine learning models and advanced simulations for future transit planning. Analyzing real-time data to dynamically adjust train schedules and express stop selection can be a viable option for a comprehensive upgrade of the existing rail lines. Simulations of various configurations can also inform operations on how to optimize network performance under different conditions. More sophisticated analysis in terms of data used, feature selection, and other supplementary data could help develop better models and more precise predictions.

ACKNOWLEDGMENT

The authors extend their sincere gratitude to their instructor, Engr. Mon Arjay Malbog, whose invaluable guidance greatly contributed to this study. Furthermore, the advice and motivation from Dr. Jennalyn N. Mindoro was instrumental in the submission of this paper. Appreciation is also given to the Technological Institute of the Philippines - Manila for providing the environment and opportunities that enabled the authors to learn, explore, and grow.

The publicly accessible ridership and transit network data on the Manila LRT-Line 2 from the Department of Transportation (DOTr) is valued as the primary source of the data in this study, alongside the Kintetsu Railway data used as the basis for the Random Forest Classifier.

REFERENCES

[1] "Asian Development Outlook December 2023," Dec. 2023. doi: 10.22617/fls230592-3.

[2] G. Thibault, A. Nienhaus, A. Bayen, M. De Clercq, and L. Cartigny, "Urban Mobility Readiness Index," University of California, Berkeley, 2024. Accessed: Mar. 31, 2025. [Online]. Available: https://www.oliverwymanforum.com/content/dam/oliver-wyman/ow-forum/template-scripts/urban-mobility-index-2024/PDF/Mobility-Index-Report.pdf

[3] A. Mijares, M. Suzuki, and T. Yai, "Passenger Satisfaction and Mental Adaptation under Adverse Conditions: Case Study in Manila," Journal of Public Transportation, vol. 19, no. 4, pp. 144–160, Nov. 2016, doi: 10.5038/2375-0901.19.4.9.

[4] "LRT-2 Sets New Ridership Record with Over 49 Million Passengers in 2023 — Light Rail Transit Authority," Jan. 03, 2024. https://www.lrta.gov.ph/lrt-2-sets-new-ridership-record-with-over-49-million-passengers-in-2023/

[5] Kintetsu Railway Co., Ltd., "駅別乗降人員奈良線橿原線天理線— 近畿日本鉄道," 近畿日本鉄道. https://www.kintetsu.co.jp/tetsudo/c.html

[6] Department of Transportation et al., "Decarbonising transport in India and the region," 2024. [Online]. Available: https://www.itf-oecd.org/sites/default/files/repositories/walleastein_sigui_1.pdf

[7] P. Grant, R. Maroso, and S. Popuri, "ASEAN Sustainable Urbanisation Report 2022: Sustainable Cities Towards 2025 and Beyond.," Association of Southeast Asian Nations, 2022. Accessed: Mar. 31, 2025. [Online]. Available: https://unhabitat.org/sites/default/files/2022/12/asean_sustainable_urbanisation_report_final_dec_2022.pdf

[8] Statista, "Popular modes of transportation Philippines 2023," Statista, Feb. 29, 2024. https://www.statista.com/statistics/1338717/philippines-most-used-modes-of-transportation/

[9] A. H. AlKhereibi, T. G. Wakjira, M. Kucukvar, and N. C. Onat, "Predictive machine learning algorithms for metro ridership based on urban land use policies in support of Transit-Oriented Development," *Sustainability*, vol. 15, no. 2, p. 1718, Jan. 2023, doi: 10.3390/su15021718.

[10] Z. Wei, L. Shao, and W. Zhong, "A study on the location of public transport for general studies stops based on DBSCAN algorithm," in Advances in economics, business and management research *Advances in Economics, Business and Management Research*, 2024, pp. 108–113. doi: 10.2991/978-94-6463-570-6$_1$2.