Towards Prediction of Bus Arrival Time using Multi-layer Perceptron (MLP) and MLP Regressor

Xueli Xu*, Sye Loong Keoh*, Chee Kiat Seow*, Qi Cao* and Sharul Kamal bin Abdul Rahim†
*School of Computing Science, University of Glasgow, UK
†Wireless Communication Centre, Universiti Teknologi Malaysia, Malaysia
Emails: 2717922x@student.gla.ac.uk, {SyeLoong.Keoh, CheeKiat.Seow, Qi.Cao}@glasgow.ac.uk, sharulkamal@utm.my

***Abstract*—Intelligent transport systems have been in research and development in recent decades. However, not all countries can afford to deploy such systems for the public usage. Conventional public transport systems such as public buses are still the main mode of public transportation system in many developing countries. Due to the issue of public transportation's inaccurate bus arrival timing, the general public still prefers private transportation. The goal of this study is to investigate the use of machine learning to improve the prediction accuracy of bus arrival timing. Two machine learning models, a multi-layer perceptron (MLP) and a MLP regressor, were compared in terms of their performance on small datasets. The experiment data was collected from Kulai-Johor Bahru Sentral bus route in Malaysia and cleaned to negate errors that influenced the accuracy of the models. The performance of the models were analysed and discussed and we observed that the MLP outperforms the MLP regressor. A limitation of this study is the small dataset that only comprises bus location data collected on a single bus route.**

***Index Terms*—multi-layer perceptron, multi-layer perceptron regressor, intelligent transportation, machine learning models, data analytic**

I. Introduction

In most developing countries without complete intelligent transport systems such as driverless light rail transit, electric and hydrogen buses or autonomous bus rapid transit, people usually rely on public buses to travel from one place to another for work and leisure. However, public buses often suffer from over-crowding and inconsistent travelling frequency. Due to various reasons and constraints, the current transportation infrastructure in most ASEAN countries are still far below the service quality expectations for the overwhelming number of commuters. Additionally, the public bus network in developing countries, especially in the rural areas are mostly unreliable and the service frequency is unpredictable most of the time.

With the advent of Internet of things (IoT) and tracking technologies, such as Global Positioning System (GPS) and LTE/4G/5G cellular based Internet connectivity, bus fleets can now be tracked in real-time [1] or using crowd-sourced applications [2], [3]. This has provided greater certainty to the commuters, allowing them to plan their journey more efficiently and hence, reducing their waiting time. Even so, the estimated time of arrival (ETA) may not accurately reflect the true representation of the actual arrival time, which causes delays in the trips planned by the passenger. Consequently, this causes a decrease in quality of service (QoS) and passenger satisfactions.

The paper presents an approach to exploit the use of data analytic and machine learning techniques to build a journey duration prediction model in order to accurately predict bus arrival time [4]. Coupling with the real-time GPS location tracking, the resulting model can be used to optimise the prediction of the bus arrival timing and hence, improving urban mobility. One specific route in the state of Johor, Malaysia was used as a case study to build the journey prediction model using dataset provided by a local bus operator. We performed data cleaning and data engineering on the dataset in order to prepare an accurate dataset for training. The feasibility of enhancing the accuracy of the journey duration prediction between bus stops was investigated using two machine learning techniques: the multi-layer perceptron (MLP) and the multi-layer perceptron regressor

This paper is organised as follows: Section II presents the background of machine learning and related work on bus arrival time prediction using machine learning techniques. Section III describes the data preparation and machine learning models used for training, while Section IV presents the experiment results and analysis. We conclude the paper with future works in Section V.

II. Related Work

A. Background on Machine Learning and Deep Learning

Machine learning belongs to a branch of Artificial Intelligence (AI), based on the theory that computer systems are able to learn or adapt from data, and make decisions without definite instructions from humans. Machine learning algorithms are classified into three main categories, i.e., supervised, unsupervised and semi-supervised learning.

Supervised Learning is a subcategory of machine learning that requires the training dataset to be labelled, to help the algorithms to derive a function which best represents the relationship between input and output data. This model will adjust its weight accordingly until the model is fitted suitably. Supervised learning is often used in classification and regression problems when historical data is available [5]. As for the subcategory of unsupervised learning, the algorithm is

This work is the output of the ASEAN IVO project, 'An IoT-based Public Transport Data Collection and Analytic Framework using Bluetooth Proximity Beacons', and financially supported by NICT (http://www.nict.go.jp/en/index.html). X. Xu's internship was supported by the University of Glasgow.

given data that is not labelled. It is able to find hidden patterns in the data. Unsupervised learning is frequently used for clustering, anomaly detection and dimensionality reduction [6]. Semi-supervised learning is a combination of supervised and unsupervised learning algorithms. Reinforcement learning is an example of semi-supervised learning as it does not require labelled data to train the model, but it uses an incentive based approach where the expected and negative outcomes are predefined. This approach will maximise the expected outcomes and minimise the negative behaviours [7].

Deep neural network learning is another sub-branch of AI that uses algorithms inspired by the design of the brain neural network, which comprises layers of interconnected nodes. These neural network algorithms use large datasets to learn and make decisions, similar to how a human brain function [8]. The input layer of neural network receives information that will be processed by one or more hidden layers using mathematical computations to make predictions. These predictions are then produced as an output via the output layer. Because of their known flexibility in modeling and generalization potential, neural networks have grown in popularity in a variety of engineering fields, including electrical engineering and mechanical engineering. MLP is a fully connected and feed-forward neural network, which is composed of multiple layers of interconnected nodes. An MLP consists of at least three layers of nodes; input, hidden and output layer. Each node that is not in the input layer, uses a either linear or nonlinear activation function. MLP uses back-propagation, which is a supervised learning method for model training.

B. Related Work

A study on predicting bus arrival time in urban and rural regions of China [9], employed a combination of Support Vector Regression (SVR) and k-nearest neighbours (KNN) in one model to compare against historical data-based prediction methods on a bus route that integrates both urban and rural areas. One of the factors this study accounts for is the stopping time at each stop as the number of passengers alighting and boarding would impact the timing. The results showed a significant difference in relative error between the SVR+KNN model and the historical data-based prediction method, with the average relative error of the SVR+KNN model being 5.74% and the average relative error of the historical data-based being 25.97%. The results showed that the use of machine learning models performed better than that of the predictive analysis.

In another study, the total travel time was predicted instead of the time taken between stops with the use of SVR, which performed better than neural networks for smaller datasets, as it can handle high variances [10]. The authors concluded that SVR is suitable to be used in India as there is high variance in the data due to the traffic conditions and weather conditions in heterogeneous and lane-less traffic.

A study using GPS to collect the bus route data was used to compare the Artificial Neural Network (ANN) and linear regression model. This study only collected six days worth of

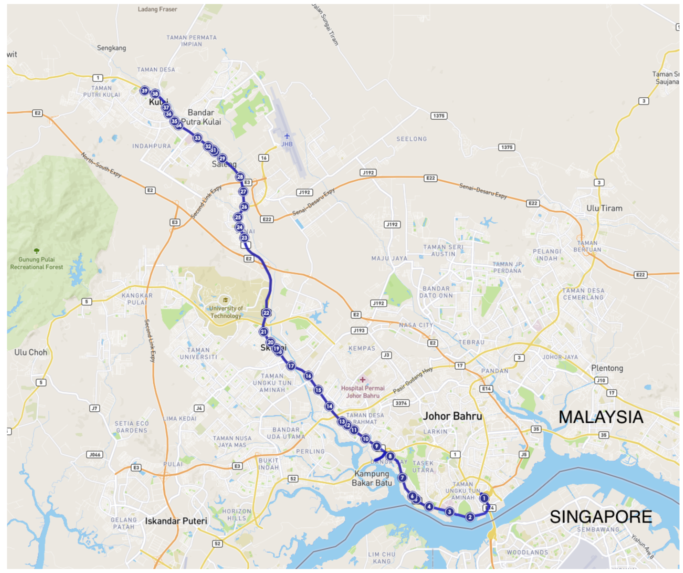

Fig. 1. Route map of Kulai - Johor Bahru Sentral Bus Service

data, and with the small dataset the artificial neural network had a Mean Absolute Percentage Error (MAPE) of 2.09% and the linear regression model had a MAPE of 2.31% [11]. However, the dataset used is rather small and may not be representative.

Another prior work compared five different methods to predict bus arrival timing which includes, Dual-stage Attention-based Recurrent Neural Network (DA-RNN), Long Short-term Memory Recurrent Neural Network (LSTM RNN), MLP, Kalman Filter and SVR [12]. In the study, the best-performing algorithm was the DA-RNN according to the error metrics of Root Square Mean Error (RMSE), Mean Absolute Error (MAE) and MAPE. The prediction error in relation to the increase in the number of stations was tested. The algorithms that involved the recurrent neural network had the lowest MAE and a slower growth rate in MAE.

With a wide variety of studies in the literature and that every city has its own mobility patterns, it is advocated that more research can be conducted to explore the relatively high-performing machine learning models on different cities' dataset, thus validating their performance in different traffic and environmental conditions.

III. Journey Duration Prediction Models

This section describes the methodology for preparing the dataset (i.e., historical GPS data) obtained from a Malaysian bus operator to develop deep learning journey duration prediction models for the route between Kulai and Johor Bahru Sentral in Malaysia. Kulai is situated at the north of Johor Bahru Sentral as shown in the map in Fig. 1. The bus travels via the main trunk road between the two cities with a distance of 34.1 km. There are approximately 40 bus stops along this bus route in each direction, including the terminal stations. Depending on the traffic and weather conditions, the bus

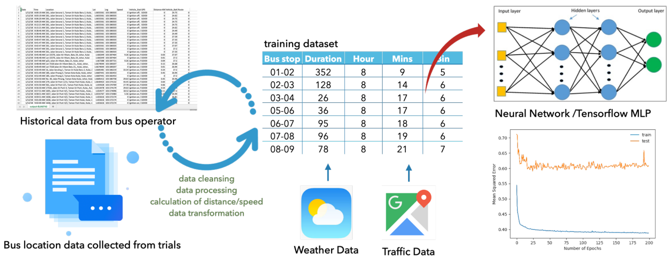

Fig. 2. Methodology for Developing a Journey Duration Prediction Model

journey typically takes between 50 to 70 minutes each way .

Fig. 2 illustrates the proposed approach to build a base model using MLP and MLP Regressor to predict the journey duration for the Kulai-Johor Bahru Sentral route. Currently, the models were built purely based on the historical dataset obtained from the bus operator, with no data on weather and traffic conditions captured by the bus operator. In the future, we plan to collect new set of data based on our IoT-based bus location tracking project [13] together with the real-time weather and traffic data to further improve the prediction models. The prediction models can then be deployed as a web service, fused with real-time bus location data to provide a more accurate prediction of bus arrival time.

A. Historical GPS Dataset

With the portable GPS device being installed on the bus, the Malaysian local bus operator provided four instances of the bus GPS traces for the month of December 2018. The raw dataset contains the following fields: *vehicle number, date, time, location, speed, vehicle status, GPS status* and *distance*. The *location* field was recorded as an address instead of latitude and longitude (lat, lng). Note that, the buses were parked at the bus depots and it had to be driven from the depot to the terminal station in the morning, and at the end of the service, it was driven back to the bus depot. As the bus fleet was rather old and aged, it is observed that some of the location logged by the onboard GPS were not accurate and the address logged could not be determined.

B. Data Engineering

First of all, based on the business knowledge of the stage bus service, the raw dataset were cleaned and pre-processed to filter out missing data. Erroneous data is fixed and transformed into meaningful values.

1) Removal of Incorrect Entries: Firstly, it was noticed that there were generic address such as '*1 Kulai, Johor*', '*1 Senai, Johor*' '*1 Johor Bahru, Johor*' and '*Unnamed Road, Johor*' that needed to be removed as these addresses could not be resolved to a correct (lat, lng) coordinate.

Secondly, based on the route map there were addresses that are too far away from the service route, e.g., '*18, Jalan Dedap 15, Taman Johor Jaya, Johor Bahru*' which is 5.5 km away from the JB Sentral bus terminal and it should not have been recorded for the Kulai-Johor Bahru Sentral route. These addresses needed to be removed. It is likely that these addresses were recorded mainly because the bus could have done a detour due to road closure, or it had to go to the Depot for re-fueling.

Thirdly, while the bus was parked at the bus terminal with ignition on, duplicated entries of the bus location were found particularly at the Kulai bus terminal and Johor Bahru Sentral bus terminal. In this case, only the latest entry of the location data is kept as it was the time at which the bus departed the terminal station.

2) Reverse Geocoding: After the removal of errors, Google Maps API was used for the generation of (lat, lng) coordinates based on the recorded address, and then updated the data records. Addresses that could not be resolved into a valid (lat, lng) coordinate were discarded. Additionally, the direction of travel was encoded, as the dataset contains data for both Kulai-Johor Bahru Sentral direction, and the return trip of Johor Bahru Sentral-Kulai direction.

While validating the dataset, it was also noticed that after the addresses had been geo-encoded successfully, some entries that did not follow the direction of travel correctly. For example, the bus travelled from A $\rightarrow$ B $\rightarrow$ C $\rightarrow$ A' $\rightarrow$ D, in which A' is a location between A and B. Such an erroneous entry could be caused by the unreliable GPS signal due to poor weather conditions. Hence, such entries (A' in this case) were removed from the dataset.

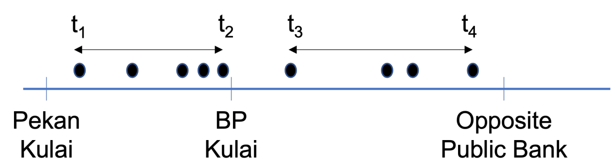

Fig. 3. Approximately the Travel Duration between 2 Bus Stops

3) Approximate the Speed and Duration: Each of the transformed data entry was then mapped to the nearest bus stop along the bus service route based on its direction of travel. This was done in order to approximate the speed of the bus between two consecutive bus stops and subsequently compute the journey duration between them. Fig. 3 shows that there are five GPS data points recorded between the bus stops *Pekan Kulai* and *BP Kulai*, and four data points logged between *BP Kulai* and *Opposite Public Bank* bus stops. The nearest datapoints to the bus stop were selected to derive the speed of the bus between two stops, and subsequently based on the time logged, an approximate speed of travel was derived, i.e., (t_2 - t_1)/$distance_1$ for *Pekan Kulai* to *BP Kulai* bus stop and (t_4 - t_3)/$distance_2$ for bus stops *BP Kulai* and *Opposite Public Bank*. Note that $distance_1$ and $distance_2$ were computed by creating a polyline connecting all the location data recorded between the two stops, and subsequently aggregating the Haversine distance of all the connected points on the polyline.

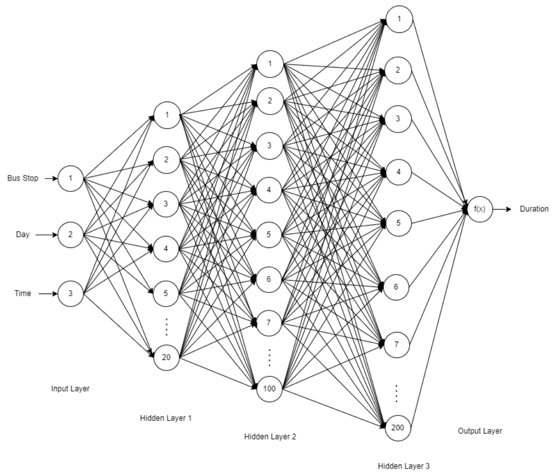

Fig. 4. Architecture of MLP

With the speed of travel derived, the journey duration between the two bus stops can be estimated.

There were also cases where there was no data point recorded between two bus stops, e.g., if there are only two data points in t_1 and t_4, the derived speed between t_1 and t_4 will be used to compute the journey duration for both sets of bus stops instead.

4) Preparing the Dataset for Training: The cleaned dataset was then transformed into a set $S = \{(a_1,b_1,c_1,d_1), (a_2,b_2,c_2,d_2), (a_3,b_3,c_3,d_3) ...\}$, where a_i is the bus stop code of two consecutive bus stops; b_i is the time of the day; c_i is the day of the week; and d_i is the duration of the journey between the two bus stops. As the journey duration can differ greatly within an hour, especially the frequency of the bus service in Malaysia is rather low, i.e., the bus service operates at 15 to 45 mins interval depending on the time of the day. Hence, we defined bins of 30-min timeframe and grouped each data according to the bins.

After data cleaning and engineering, the size of each dataset is shown as follows for the four buses:

- Bus 1 dataset: 8323 rows
- Bus 2 dataset: 3479 rows
- Bus 3 dataset: 7599 rows
- Bus 4 dataset: 7118 rows

Each dataset was split into 80% training and 20% testing for both the MLP and MLP regressor. For a fair benchmarking purpose, both models have three layers of perceptrons, with 20 neurons in the first layer, 100 neurons in the second layer, and 200 neurons in the third layer, as shown in Fig. 4. The main difference between the MLP and MLP regressor is the activation function used in the output layer. The MLP uses the rectified linear unit function (ReLU), which is defined as *f(x) = max(0, x)*; While the MLP regressor uses a linear activation function, which is defined as *f(x) = x*.

IV. Results and Discussion

This section evaluates the results of both machine learning models, i.e., 0LP model and MLP regressor model, on a small dataset with a set of error metrics. To ensure a fair comparison, we chose to measure the R-Squared, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) values of the predictions against the ground truth data for both models. The small datasets enable us to compare the model's performance under limited data availability, which is common in many real-world applications.

A. Model Trained on Multi-layer Perceptron

The results in Table I show the accuracy of the predictions generated by the MLP model for both the training and testing dataset.

TABLE I
Performance of Models Trained using MLP

		R2	MAE	MSE	RMSE
Bus 1	Train	0.18	0.65	0.92	0.96
	Test	-0.05	0.84	1.90	1.38
Bus 2	Train	0.09	1.07	129.71	11.39
	Test	0.24	1.20	30.38	5.51
Bus 3	Train	0.11	0.79	6.50	2.55
	Test	-0.01	0.98	16.82	4.10
Bus 4	Train	0.36	0.70	1.78	1.33
	Test	-0.25	0.73	1.28	1.13

B. Model Trained on MLP Regressor

The results in Table II show the accuracy of the predictions generated by the MLP regressor model for both the training and testing dataset.

TABLE II
Performance of Models Trained using MLP Regressor

		R2	MAE	MSE	RMSE
Bus 1	Train	0.02	57.19	7822.39	88.44
	Test	0.01	55.37	6847.69	82.75
Bus 2	Train	0.56	66.66	36754.47	191.71
	Test	0.01	104.66	2887192.50	1699.17
Bus 3	Train	0.28	63.00	39837.04	199.59
	Test	-0.47	60.25	14607.86	120.86
Bus 4	Train	0.36	50.05	8653.77	93.03
	Test	0.03	51.00	5343.78	73.10

From the results in these two tables, the MAE, MSE and RMSE values show that the MLP model performed significantly better than the MLP regressor model. The MAE, MSE and RMSE values of the MLP model are significantly lower, which means that the prediction error is lower as compared to the MLP regressor model. Under the MLP regressor model, the RMSE value of the test dataset for Bus 2 is 1699.17, which is much higher than its trained dataset of 191.71.

C. Comparison of Predicted vs. Actual Duration

The performance comparisons of the MLP model and MLP regressor model are shown in Fig. 5 and Fig. 6 between the actual and predicted durations for all four buses across both training and testing datasets. The x-axis is the actual duration

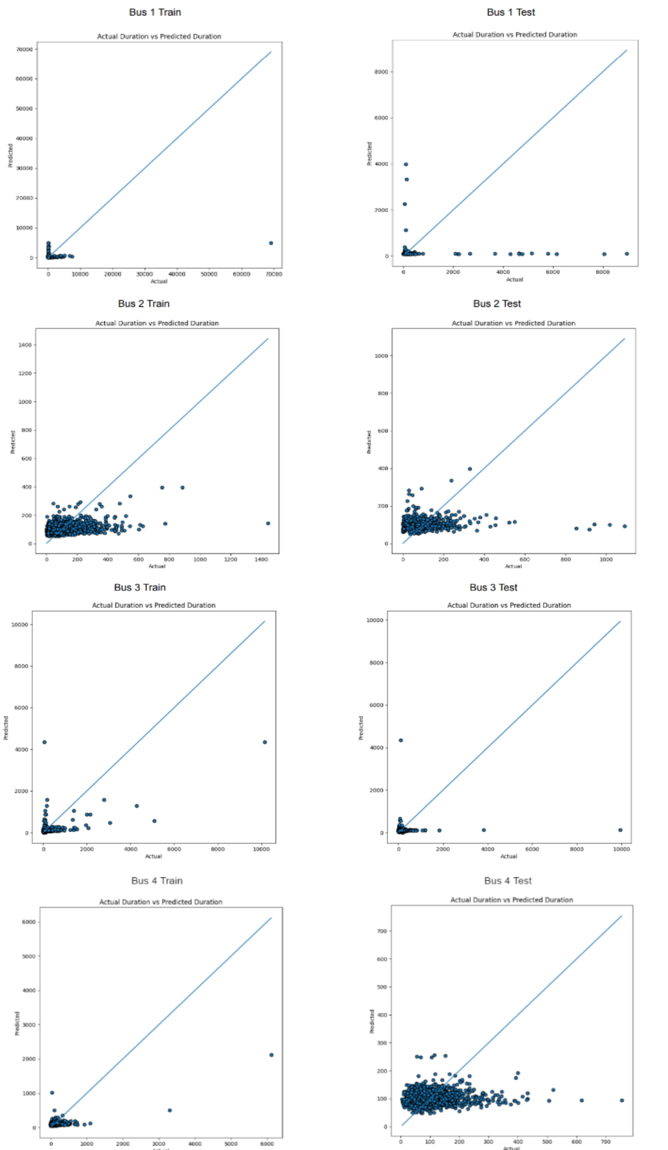

Fig. 5. Performance of MLP Model: Actual v.s. Predicted Duration

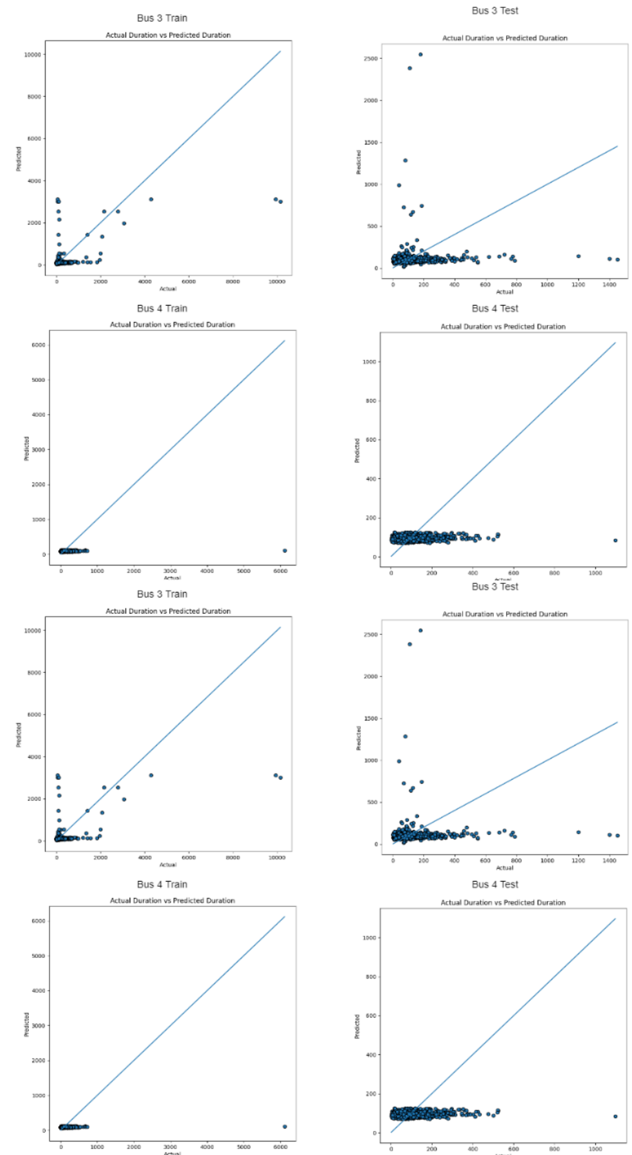

Fig. 6. Performance of MLP Regressor Model: Actual v.s. Predicted Duration

calculated based on the time taken in seconds between bus stops, while the y-axis represents the corresponding predicted journey duration in seconds generated by the machine learning models. It is observed that there are some *outliers* in the plots with a long journey duration, i.e., $\geq$ 4000 seconds (66 minutes). This implies that the journey between two consecutive bus stops was longer than the entire bus service route and this can be classified as an outlier. It could mean that the bus broke down or stalled at the same location for a long time. Alternatively, there could have been a traffic jam or a road closure, causing the time recorded between stops to be much longer.

It is observed in Fig. 5 for all four buses that the MLP model's predicted values are mostly lower than the actual duration, indicating a bias towards underestimating the trip duration. This means that the MLP model is not capturing the patterns in the data accurately and is not performing as well as desired.

As seen in Fig. 6 that for the two larger datasets (bus 1 and 3), the data points are more scattered, indicating high variability and inconsistency in the ability of the MLP regressor to predict the same values. For all four buses, similar to MLP model, the predicted values generated by the MLP regressor are lower than the actual duration.

The MLP model is usually used for classification tasks. It can also be used for regression tasks by configuring the output layer to have a single neuron that predicts continuous output values, such as duration in this work. We used a ReLU activation function to ensure that the predictions are meaningful because the duration cannot be negative. A linear activation will be more appropriate for those use cases with both positive and negative values since ReLU can limit the range of predicted values.

While the MLP regressor model is primarily used for regression tasks, the partial derivative of the loss function concerning the model's parameters is calculated using an iterative approach to update the parameter. As it is intended to build complex nonlinear relationships between inputs and outputs features, it is prone to overfitting, which can explain the significantly higher RMSE value of the testing dataset compared to the training dataset for Bus 2. A possible cause of the overfitting could be the small dataset size, which may have caused the MLP regressor model to fit the noise instead of the actual trend, resulting in inaccurate predictions.

Furthermore, the regressor attempts to map input features to output values using a regression approach. However, the choice of an identity activation function in the output layer, combined with the non-linear ReLU activation function in the hidden layers, forces the output to be a linear function of the first layer. This may not be suitable for modelling complex relationships between input and output features which is non linear in nature.

The current performance of the MLP regressor in this scenario suggests that the relationship between the arrival time and input features may not be linear. Alternatively, the dataset used for training and validation may not be sufficient to capture the underlying relationships between the variables. This highlights the need for more extensive data collection and exploration of alternative modelling techniques.

V. Conclusions

The study on optimising the public transportation system to improve passengers' commuting experience is becoming more important especially with the increased popularity of AI and machine learning approaches. It comes to the question of how AI can play a role in improving the service quality of the public transportation system, and how we can build an accurate prediction model that mimics the behaviour of the public bus services.

In this study, we explore the feasibility of using machine learning to predict the bus arrival time at bus stops and compare two initial models, the MLP model and the MLP regressor model based on datasets collected from the bus route in the state of Johor, Malaysia. The experiment results suggest that the performance of the MLP model is better than the MLP regressor model on the datasets, as observed from the derived error metrics.

This work can be further extended to explore the optimisation of the performance of the MLP model with a larger and broader variety of datasets that take into account varying conditions such as weather and traffic conditions. These factors can significantly impact bus arrival time and should be considered in the predictive model. We are currently working with other bus operators to obtain such datasets, while at the same time, we are deploying our IoT-based fleet tracking system to collect more reliable data. Additionally, experimenting with different activation functions, network architectures, and regularisation techniques could help improve the performance of the MLP regressor model. It may also be beneficial to explore other machine learning models that are more capable of handling non-linear relationships.

In conclusion, this study explored the use of machine learning to estimate and predict bus arrival time, and compared the performance of MLP and MLP regressor models on a limited dataset. Nonetheless, we have illustrated that using a machine learning based approach for bus arrival time prediction has a great potential. It will lead to further research involving varied datasets, different network configurations, regularisation methods, and alternative models to boost prediction accuracy and ultimately enhancing the commuters' travel experiences.

References

[1] J. Biagioni, T. Gerlich, T. Merrifield, and J. Eriksson, "Easytracker: Automatic transit tracking, mapping, and arrival time prediction using smartphones," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 68–81. [Online]. Available: https://doi.org/10.1145/2070942.2070950

[2] S. L. Lau and S. M. Sabri Ismail, "Towards a real-time public transport data framework using crowd-sourced passenger contributed data," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, 2015, pp. 1–6.

[3] W. L. H. Lim, J. T. W. Lum, I. J. W. Yeo, and S. L. Keoh, "A crowd-assisted real-time public transport information service: No more endless wait," in *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, 2016, pp. 1–6.

[4] Y. H. Toh, "Optimisation and prediction of bus arrival timings," Undergraduate Project Thesis, University of Glasgow, March 2017.

[5] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," in *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*. NLD: IOS Press, 2007, p. 3–24.

[6] J. G. Carbonell, R. S. Michalski, and T. M. Mitchell, "1 - an overview of machine learning," in *Machine Learning*, R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, Eds. San Francisco (CA): Morgan Kaufmann, 1983, pp. 3–23. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780080510545500054

[7] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Int. Res.*, vol. 4, no. 1, p. 237–285, may 1996.

[8] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608014002135

[9] Y. Wang, "Intellectualization of the urban and rural bus: The arrival time prediction method," *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 689–697, 2021. [Online]. Available: https://doi.org/10.1515/jisys-2021-0017

[10] K. K. Reddy, B. A. Kumar, and L. Vanajakshi, "Bus travel time prediction under high variability conditions," *Current Science*, vol. 111, no. 4, pp. 700–711, 2016. [Online]. Available: http://www.jstor.org/stable/24908545

[11] J. Amita, S. Jain, and P. Garg, "Prediction of bus travel time using ann: A case study in delhi," *Transportation Research Procedia*, vol. 17, pp. 263–272, 2016, international Conference on Transportation Planning and Implementation Methodologies for Developing Countries (12th TPMDC) Selected Proceedings, IIT Bombay, Mumbai, India, 10-12 December 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352146516307062

[12] X. Zhou, P. Dong, J. Xing, and P. Sun, "Learning dynamic factors to improve the accuracy of bus arrival time prediction via a recurrent neural network," *Future Internet*, vol. 11, no. 12, 2019. [Online]. Available: https://www.mdpi.com/1999-5903/11/12/247

[13] S. Gunady and S. L. Keoh, "A non-gps based location tracking of public buses using bluetooth proximity beacons," in *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, 2019, pp. 606–611.