An IoT-based Data Collection and Analytics Framework using Bluetooth Proximity Beacons

Background :

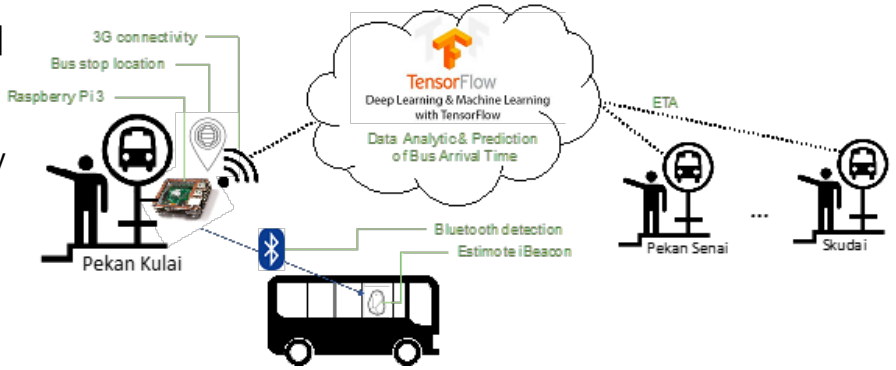
Tracking of public bus location requires a GPS device to be installed, and many bus operators in developing countries do not have such a solution in place to provide an accurate estimated time of arrival (ETA)

This project proposes an innovative IoT solution to track the location of buses to collect transportation data without requiring the deployment of GPS devices. It uses Bluetooth Low Energy (BLE) proximity beacon to track the journey of a bus by deploying an *Estimote* proximity beacon on the bus

Targets:

Bus operators in Johor (PAJ Bus) in Malaysia and School bus operator in Malang Indonesia

Benefit general public who commutes everyday by public transport



Speaker:

Sharul Kamal bin Abdul Rahim
Universiti Teknologi Malaysia (UTM)

An IoT-based Data Collection and Analytics Framework using Bluetooth Proximity Beacons

Project Members :

Members	Affiliation
Sharul Kamal bin Abdul Rahim (Project Leader), Abu Sahmah bin Md Supa'at, Jafri Bin Din, Mohd Adib bin Sarijari, Omar bin Abdul Aziz, Siti Fatimah Bte Ausordin, Muhammad Zairil bin Muhammad Nor	UTM, Malaysia
Sye Loong Keoh, Chee Kiat Seow, Qi Cao	UGS, Singapore
Somnuk Phon-Amnuaisuk, Md Saiful bin Haji Omar, Soon-Jiann Tan, Haji Idham Maswadi bin Haji Mashud	UTB, Brunei

Members	Affiliation
Achmad Basuki, Adhitya Bhawiyuga, Eko Setiawan, Agung Setia Budi	UB, Indonesia
Yung-Wey Chong, Mohd Najwadi Yusoff, Noor Farizah Binti Ibrahim	USM, Malaysia
Kok Chin Khor, Mau Luen Tham	UTAR, Malaysia

Project Duration :

June 2022 – May 2024 (2 years)

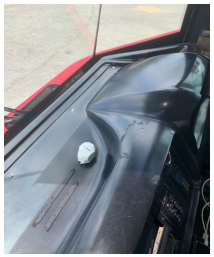
Project Budget:

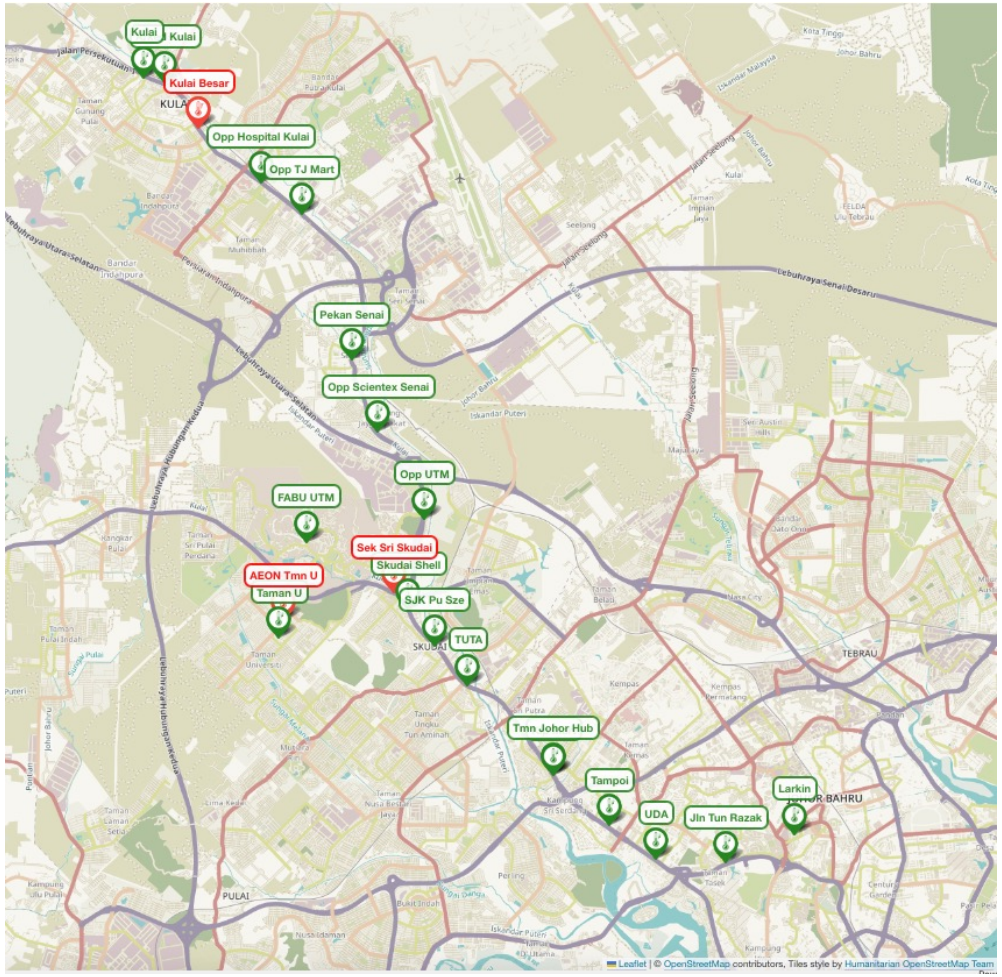
US$80,000

Project Activities: Field Trial & Deployment in Johor

Deployment by UTM and UGS:

- The Raspberry Pi Zero together its battery charging system is deployed on 21 sites spanning across three city councils (Kulai, Iskandar Puteri and Johor Bahru).
- The Raspberry Pi is powered by battery during the day, and the battery charging system taps on the lamp post electricity at night to charge the batteries at night.
- Deployment on two bus services by Johor Public Transport (PAJ) – P411 and P211, with a total of 3 buses.
  - P211 – 26 bus stops
  - P411 – 31 bus stops



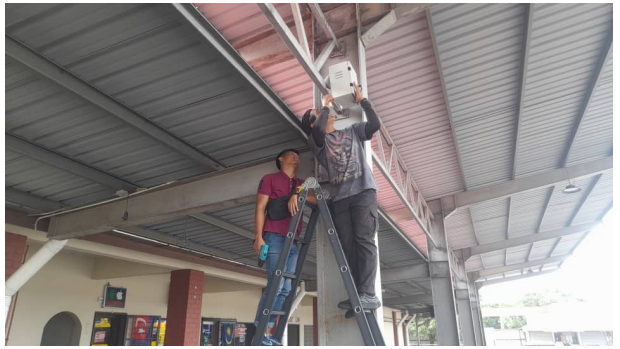


Project Activities: Field Trial & Deployment in Johor

Deployment by UTM and UGS:


Opp UTM Gate

Arked Kulai

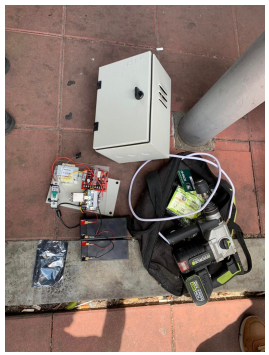
Taman Universiti Terminal

FABU UTM


Skudai Shell Petro Station

SJK Pu Sze





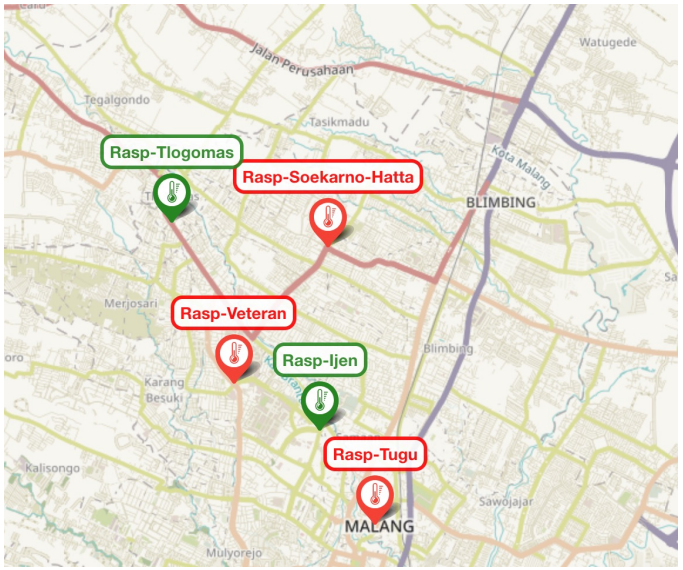
Larkin Sentral Terminal

Battery Charging System

Project Activities: Field Trial & Deployment in Malang

Deployment by UB: Universitas Brawijaya

- Installation of Raspberry Pi 4 on CCTV poles or Road Side Units (RSU), with continuous electricity power supply. A total of 8 RSUs have been installed. A 4G modem is used to provide Internet connectivity to RPI.
- Deployment is used to track School Bus – Pool A in Malang.
- The Raspberry Pi tracking system is now integrated with the Johor deployment server. Location data can now be logged on the same cloud/DB server and the ETA can be calculated based on fixed journey duration.







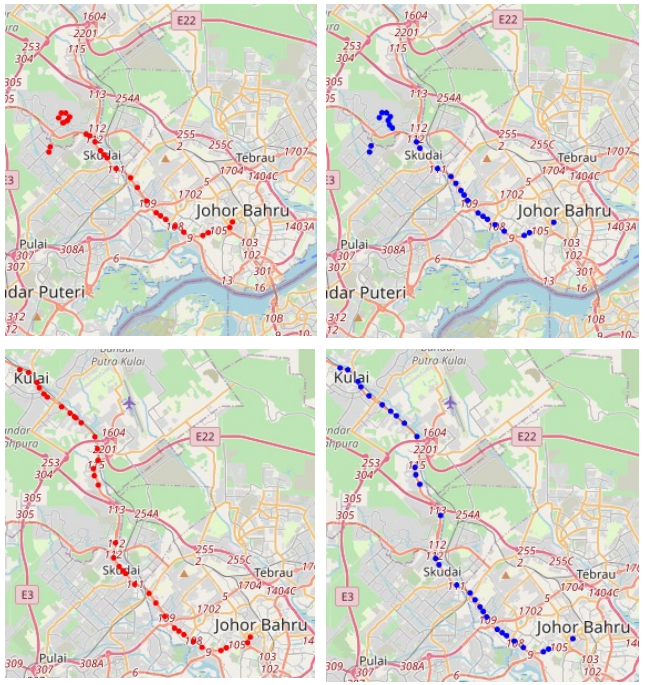
Project Activities: Keynote Speaker

2023 International Conference on Multidisciplinary Applications of Information Technology (ICOMIT, Solo, Central Jawa, Indonesia**,**
17 JULY 2023

Project Activities: Data Analytics and Machine Learning

Data Cleaning, Analytics and Machine Learning: UNIVERSITI TEKNOLOGI BRUNEI | UNIVERSITI SAINS MALAYSIA | UTAR

- Estimate the travel duration between bus stops based on historical GPS dataset of P211 (Taman University <-> Larkin Sentral) and P411 (Kulai <-> Larkin Sentral) using three approaches: statistical central tendencies, nonlinear regression techniques, and Artificial Neural Network(ANN) predictive model.
- Dynamics of traffic scene might not be fully captured, e.g., congestion due to accidents or weather conditions.
- For the non linear regression model, the following 3 different models were used: (i) K-Nearest Neighbour Regressor, (ii) Decision Tree Regressor, (iii) Random Forest Regressor.
- Two previous duration spans were employed to predict the duration to the next bus stop, where the prediction is the output of ANN. The ANN model was trained using Keras API with 2 hidden layers, four input nodes and one output nodes using ReLU activation function, while other nodes used Tanh activation function.
- 15,470 data samples for P 211 and 26,610 data samples for P411, trained using 70:30 split.

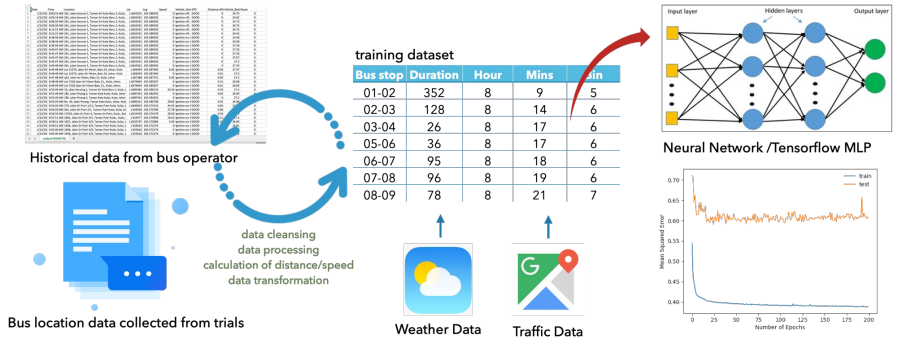


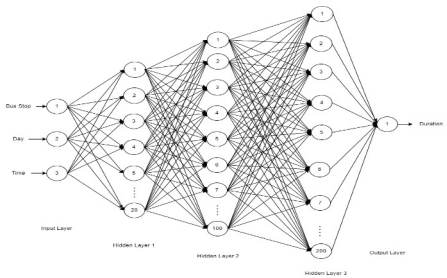
Visualisation of dataset P211 (Top)
And P411 (Bottom)

Project Activities: Data Analytics and Machine Learning

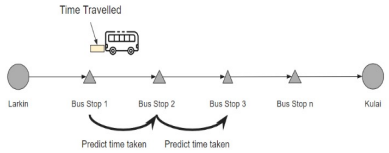
Data Cleaning, Analytics and Machine Learning: University of Glasgow

- **Approach 1:** Based on the GPS datasets of 2019 for (Kulai <-> JB Sentral) provided by S&S. The dataset was cleaned, pre-processed to filter out missing and erroneous data.
- Data entries are mapped to the nearest bus stop to approximate speed and compute journey duration between two bus stops.
- The pre-processed dataset is used to build a base model using MLP and MLP Regressor for predicting journey duration. The MLP uses the rectified linear unit function (ReLU), which is defined as f(x) = max(0, x). The MLP Regressor uses a linear activation function, which is defined as f(x) = x.
- **Approach 2:** Based on the GPS datasets of 2021 and 2022 for (Kulai<-> Larkin Sentral) provided by PAJ. The dataset was cleaned, pre-processed.
- Used Artificial Neural Network (ANN) to train the journey prediction model using (bus stop 1, bus stop 2, time, month , duration), validated using the dataset in 2023.





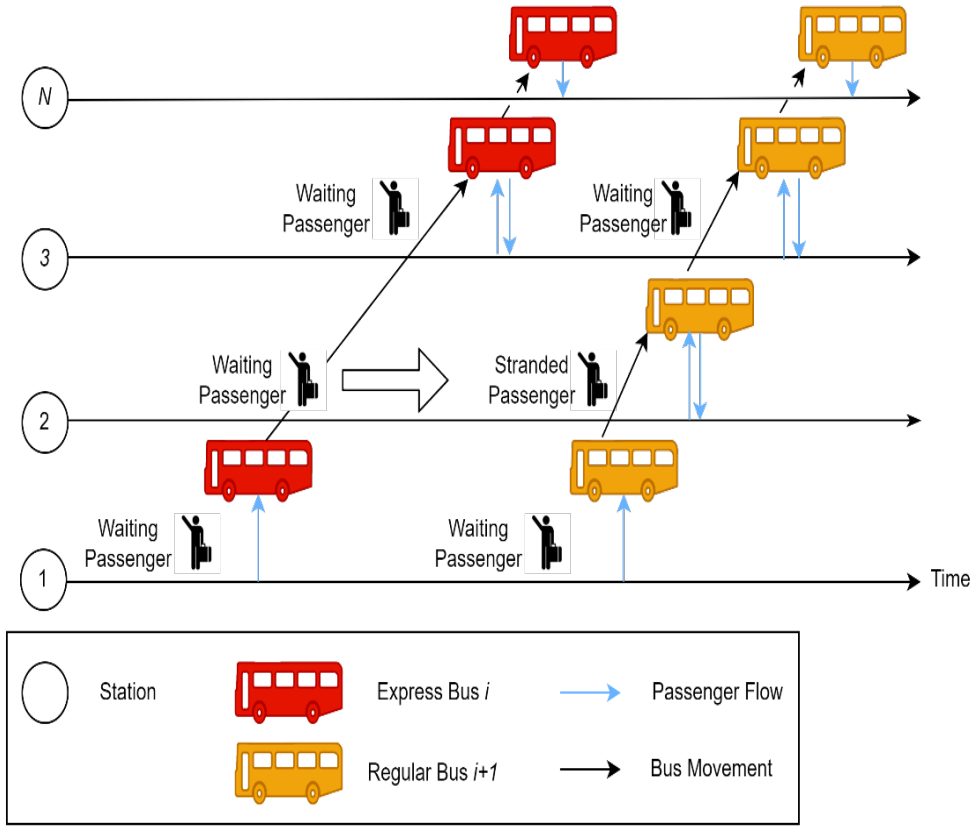
Approach 1: MLP / MLP Regressor



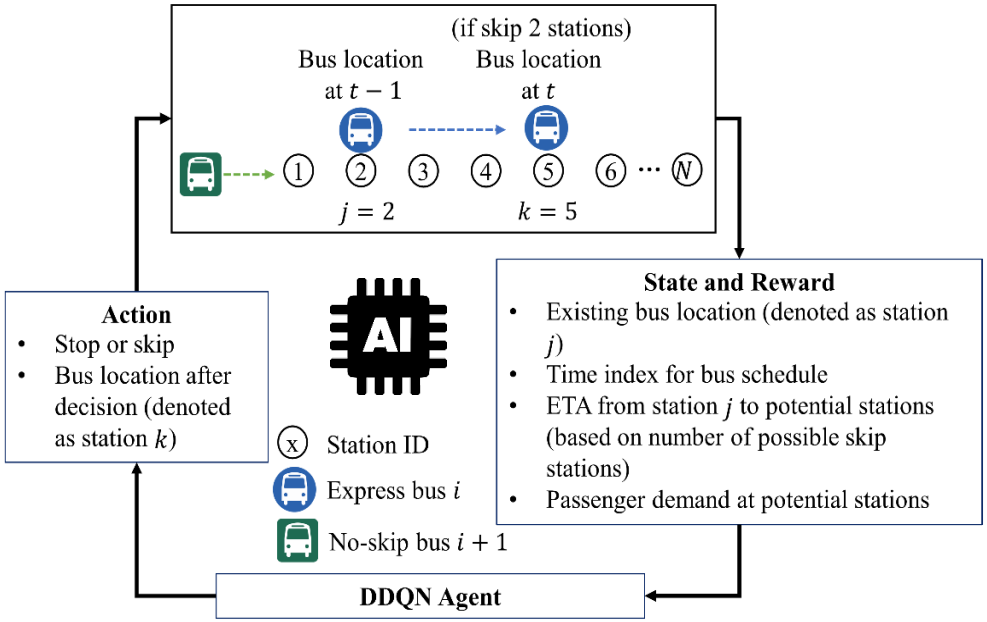
Approach 2: ANN

Project Activities: Bus Route Planning based on Stop-Skipping Strategy

The bus stop-skipping scheme starts with express bus *i* and is followed by a no-skip bus *i+1*.



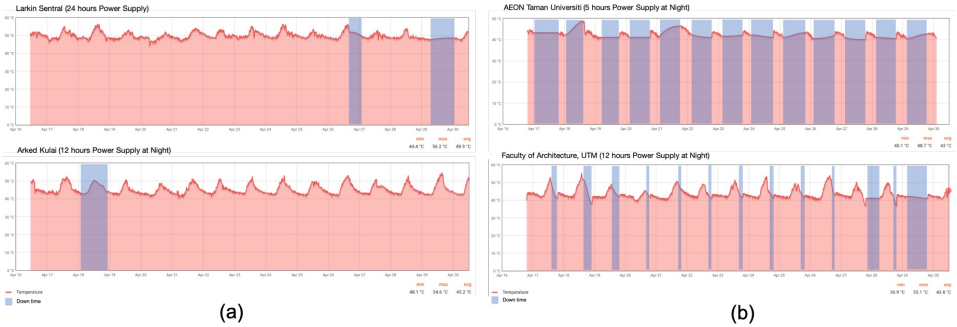
The strategy is decided by Double Deep Q-Network (DDQN).



R&D results: Field Trial & Deployment in Johor

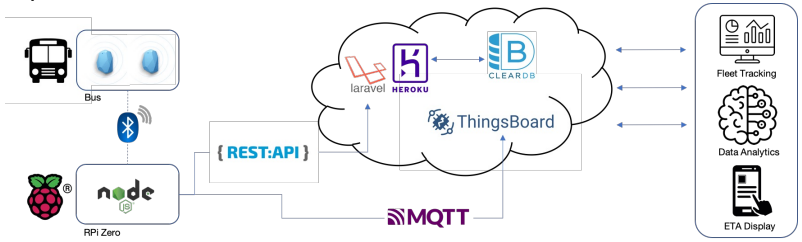
Deployment by UTM and UGS:

- Reliability of Raspberry Pi - mostly affected by the availability of power supply at night. (a) consistent at Larkin/Arked Kulai (b) inconsistent due to only 6 hours of power supply at night at AEON.

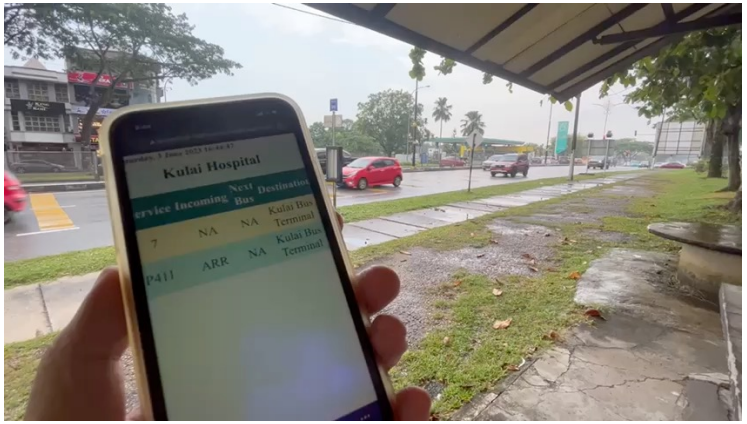


(a) (b)

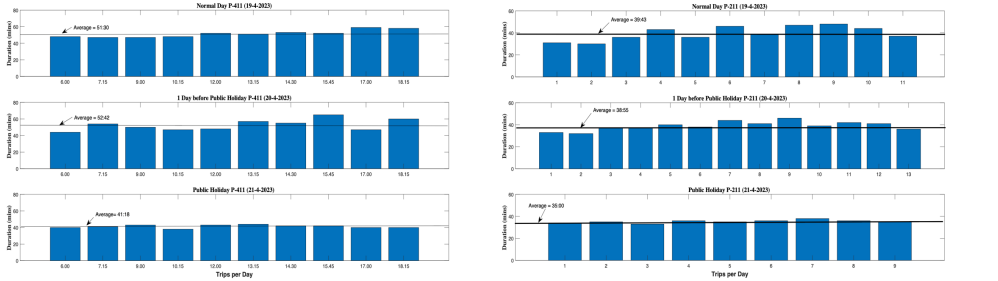
- Implemented software architecture, pending integration with the journey prediction model and ETA display at bus stop.



- Validation of the detection of BLE beacon and the display of ETA on mobile app / web app.



- Traffic pattern and journey duration between normal day and public holiday in Johor.

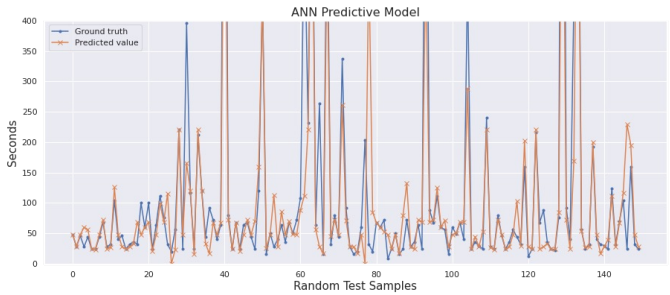


(a) (b)

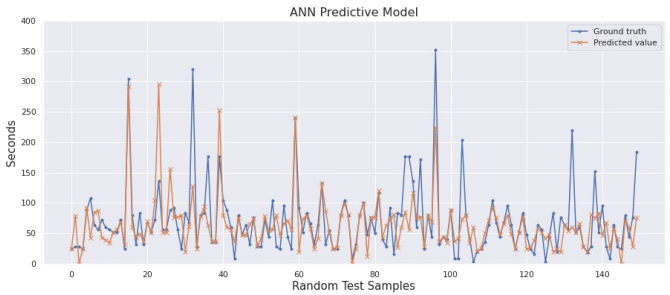
R&D results: Data Analytics and Machine Learning 1

Data Cleaning, Analytics and Machine Learning: UNIVERSITI TEKNOLOGI BRUNEI · UNIVERSITI SAINS MALAYSIA · UTAR

- Statistical central tendency approach – expected duration between two bus stops is based on the global picture from the whole dataset.
- Nonlinear regression approach – relationship between duration between two bus stops, day of the week, time of the day. It provides a natural way to fuse contextual information into the model.
- ANN predictive model – implicitly exploits the dynamics of the traffic environment via previous duration spans observed in the same routes.



P211



P411

TABLE I
SUMMARY OF PERFORMANCE FROM TEST DATASETS

	Regressors			ANN
	Decision Tree	*Random Forest*	*K-Nearest Neighbor*	*Tri-gram Predictor*
RP2 [a]				
MSE (μ)	341.11	340.64	356.67	129.27
MSE (σ)	103.91	104.11	94.42	6.60
R^2 (μ)	0.212	0.215	0.115	0.573
R^2 (σ)	0.081	0.082	0.143	0.033
parameters	DT: no depth limit for tree RF: employ 300 estimators KNN: use 7 neighbors Dataset size: 15,780 samples			[b] 15,470 samples
RP4 [a]				
MSE (μ)	123.77	123.81	129.62	132.64
MSE (σ)	39.31	39.29	37.42	5.10
R^2 (μ)	0.738	0.738	0.714	0.670
R^2 (σ)	0.117	0.117	0.114	0.026
parameters	DT: no depth limit for tree RF: employ 300 estimators KNN: use 7 neighbors Dataset size: 53,044 samples			[b] 46,610 samples

[a] Test results are averaged over 40 runs
[b] Test results averaged over 20 runs, See parameters in Section C. Predictive Models

R&D results: Data Analytics and Machine Learning 2

Data Cleaning, Analytics and Machine Learning: University of Glasgow

- **Approach 1:** MLP Model outperformed MLP Regressor according to MAE, MSE and RMSE.
- Prediction error is lower for MLP. These results show that MLP model is more accurate and reliable for small datasets with limited data availability, which is common in real-world applications.
- MLP model showed bias towards underestimating trip duration, indicating a need for improved pattern recognition
- Outliers observed in plots with journey durations ≥ 4000 seconds, likely due to bus breakdowns or road closures causing delays

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

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

- **Approach 2:** ANN model trained for 20 epochs at a learning rate of 0.00001, achieved an MAE of 0.6% (approximately 18s) and RMSE of 1.4% (approximately 40s).
- Additional features such as travelling time of the preceding journey, day of the week, day of the month and distance of the journey do not seem to improve the prediction accuracy.

Table 1. Validation of Additional Features

	MAE	RMSE
Base Features	0.0067	0.0139
Additional Features		
Travelling times of preceding journey	0.0066	0.0139
The day of week	0.0066	0.0140
The day of month	0.0067	0.0141
Distance of journey	0.0056	0.0132

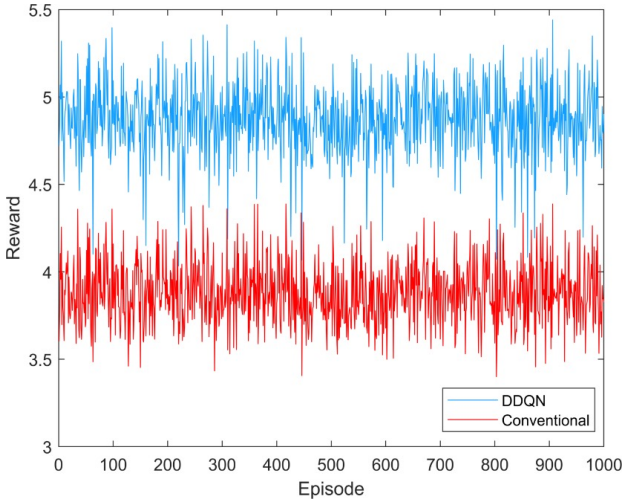
- Though the model was trained using 2021 and 2022 dataset (covid-19), when trained to include the 2023 datasets collected for Jan, Mar, Apr and May, the journey duration prediction remains accurate with low errors.

	MAE	RMSE
Original Model	25.7s	46.2s
Model Trained with 2023 Dataset	24.8s	42s

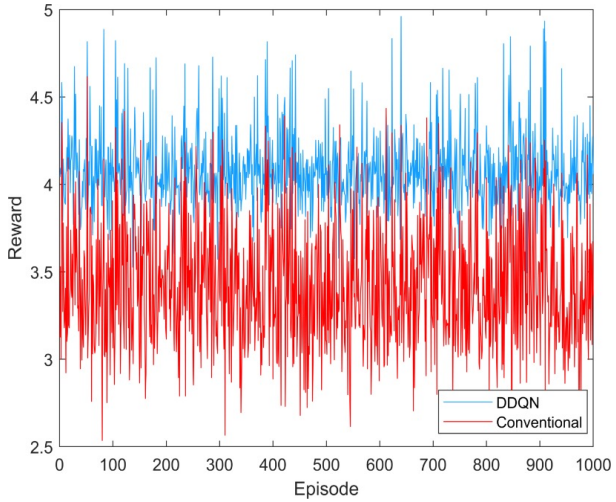
R&D results: (Max. 3 slides)

10 stations are considered.

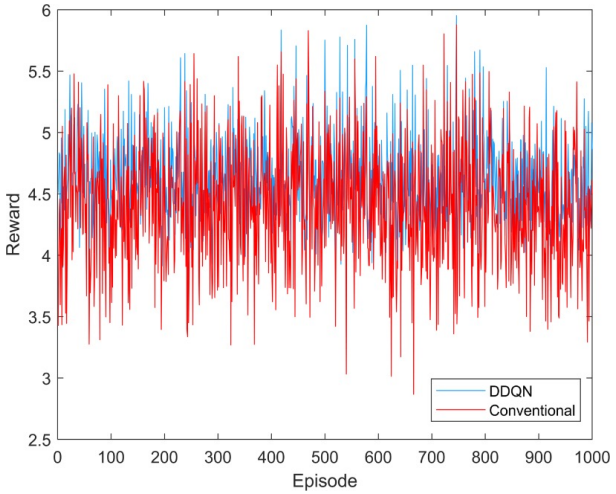
Conventional method adopts two no-skip routes.



Static passenger demand



Dynamic passenger demand varied with time



Random passenger demand

Scientific Contribution:

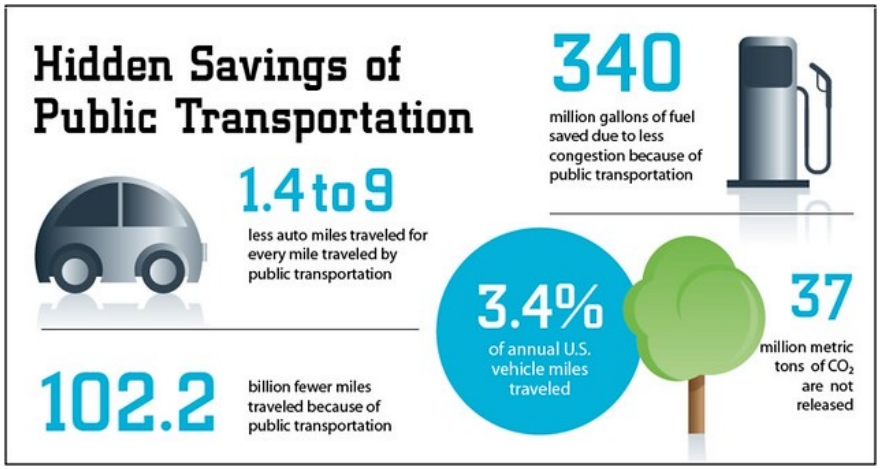
Presentations at International Conferences:

No:	Paper title:	Author names	Affiliation	Conference name:	The date of the conference	The venue of the conference
1	Non-GPS-based ETA Models Constructed from Historical Data and Traffic Contexts	S. Phon-Amnuaisuk[1], S.J Tan[1], K.C Khor[2], M.L. Tham[2], Y.W. Chong[3], S. Omar[1], N.F. Ibrahim[3], M.N, Yusoff[3], I. Mashud[1]	[1.] Universiti Teknologi Brunei [2.] Universiti Tunku Abdul Rahman [3.] Universiti Sains Malaysia	8th International Conference on Business and Industrial Research (ICBIR 2023)	18-19 May, 2023	Bangkok, Thailand
2	Towards Prediction of Bus Arrival Time using Multi-layer Perception (MLP) and MLP Regressor	X. Xu[1], S.L Keoh[1], C.K Seow[1], Q. Cao[1], S.K. Abdul Rahim[2]	[1.] University of Glasgow Singapore [2.] Universiti Teknologi Malaysia	8th International Conference on Business and Industrial Research (ICBIR 2023)	18-19 May, 2023	Bangkok, Thailand
3	Deep Reinforcement Learning Based Bus Stop-Skipping Strategy	M.L. Tham[1], B.S. Tay[1], K.C Khor[1], S. Phon-Amnuaisuk[2]	[1.] Universiti Tunku Abdul Rahman [2.] Universiti Teknologi Brunei	38th International Technical Conference on Circuits, Systems, Computers and Communications	25-28 June, 2023	Jeju Island, South Korea

Societal Impact:

- Aims to improve the use of public transport systems in developing countries through the following:
  - Easy access to the bus departure and arrival data
  - Shorten the passenger waiting time if the ETA is known in advanced
  - Effective monitoring of the bus services from the operator's perspective

- With the increase usage of public transport, it will result in less carbon emission, and less traffic congestion



Source: https://thegreendivas.com/2015/09/28/green-transportation-4-tips-for-navigating-public-transit/

Conclusion:

- Great collaboration among the project members and other project stakeholders:
  - Field trials at Johor and Malang started, and data collection is underway.
  - UTM hosted the visit to Johor deployment site in June 2023, and we are planning to visit Malang deployment in December 2023.
  - Base model for journey prediction established, showing feasibility of using the GPS dataset to predict journey duration between bus stops.

- First year progress:
  - CRDA signed among all parties.
  - Procurement for the first year completed.
  - Bi-monthly virtual meeting to update project progress.

- Some of the Issues faced:
  - Un-realiable power supply and Internet connectivity at both Johor and Malang sites.
  - Datasets collected from Raspberry Pi could be missing due to unavailability of RPI and Internet, hence would impact the data analytics.
  - Buses could be out-of-service due to repair and maintenance, hence no fleet data logged for some times.

Future works:

- Field Trials:
  - Monitor the status of the RPI installation and improve its reliability.
  - Measure the accuracy of the BLE detection.
  - Collect bus fleet data based on the detection of RPI.
  - Develop a passenger information display at the bus stop showing bus arrival time.
  - Further engage bus operators to be part of the ecosystem.

- System Development:
  - Refine the Mobile App and Passenger Information Dashboard.
  - Integrate the ETA on web app with the baseline Prediction Model.

- Data Analytics:
  - Refine the prediction model.
  - Train the model based on the collected dataset from this project.

- Knowledge Sharing:
  - Organize a field trip to Malang in December 2023.
  - Run a workshop on data analytics for knowledge sharing.
  - Publish results in journals and conferences.