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

Background :

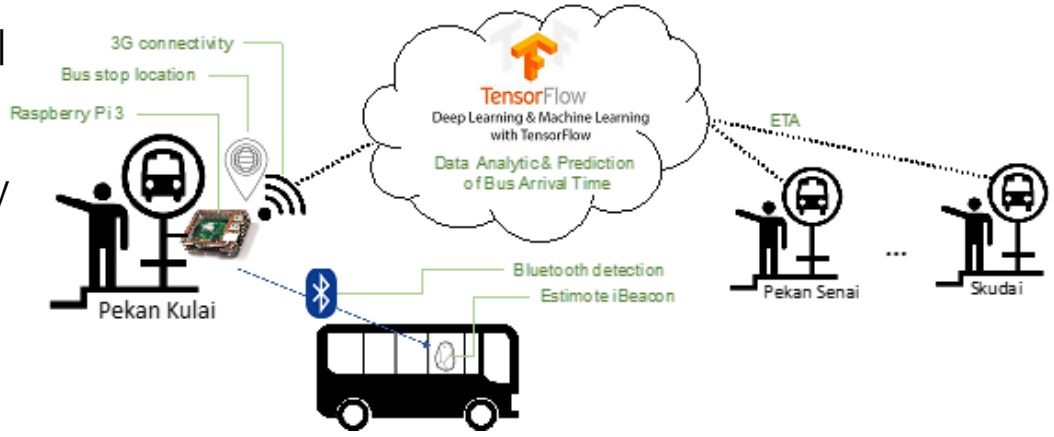
Tracking of public bus location requires a GPS device to be installed, and many bus operators in developing countries do not have such a solution in place to provide an accurate estimated time of arrival (ETA)

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, Olakunle Elijah, 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 2025 (2 years + 1 year Extension)

Project Budget:

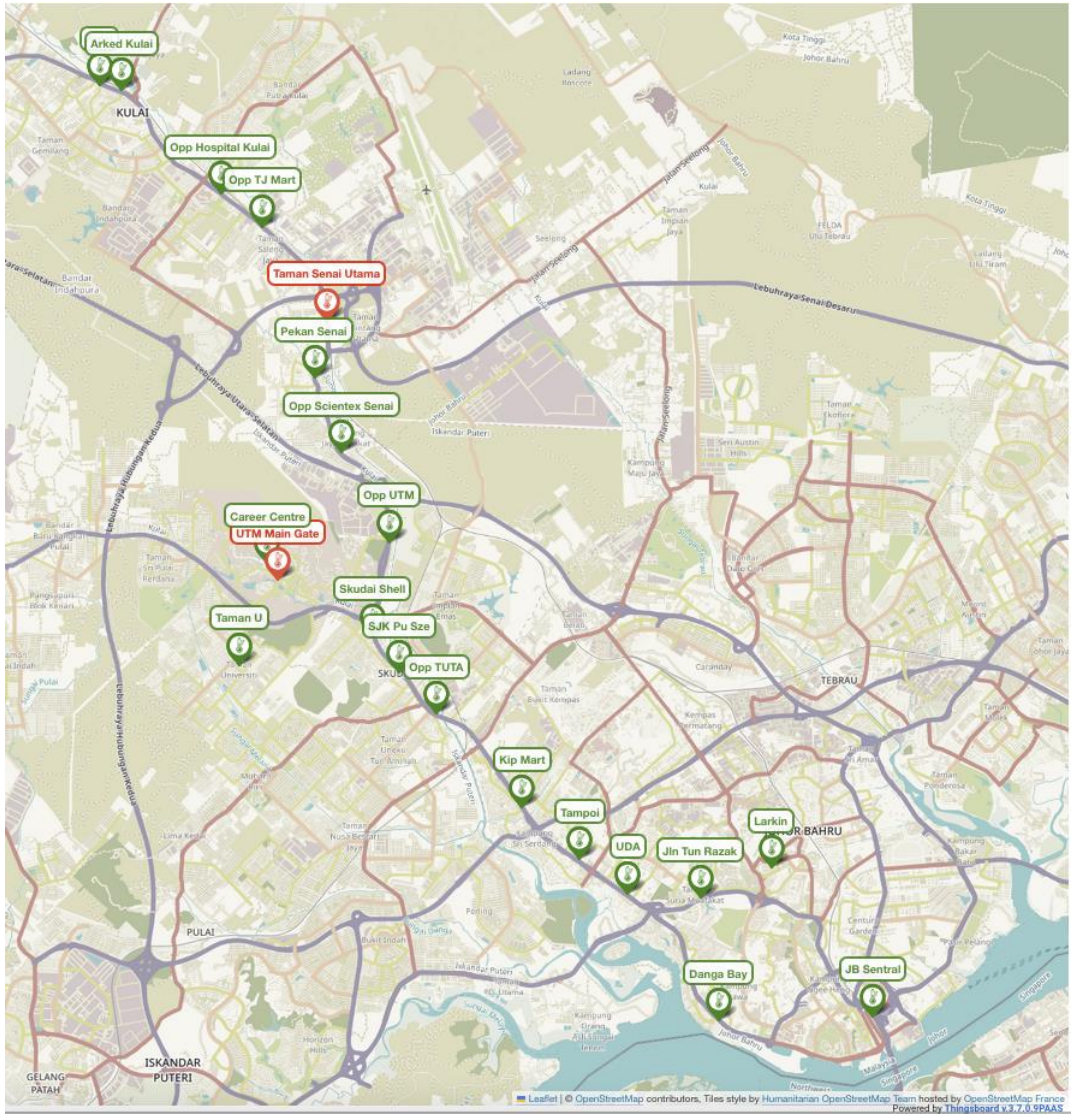
US$80,000

Project Activities: Improving the Reliability of the BLE Sensing System (UGS / UTM / UB)

1. Batteries for the sensing system have been deployed for over 1+ years, and many of them had degraded.

2. Several power outage have caused some batteries to be drained without re-charging. The team had to travel to the sites to replace the batteries.

3. An alternative approach to replace the battery with a 40,000 mAh powerbank is being tested in Johor.

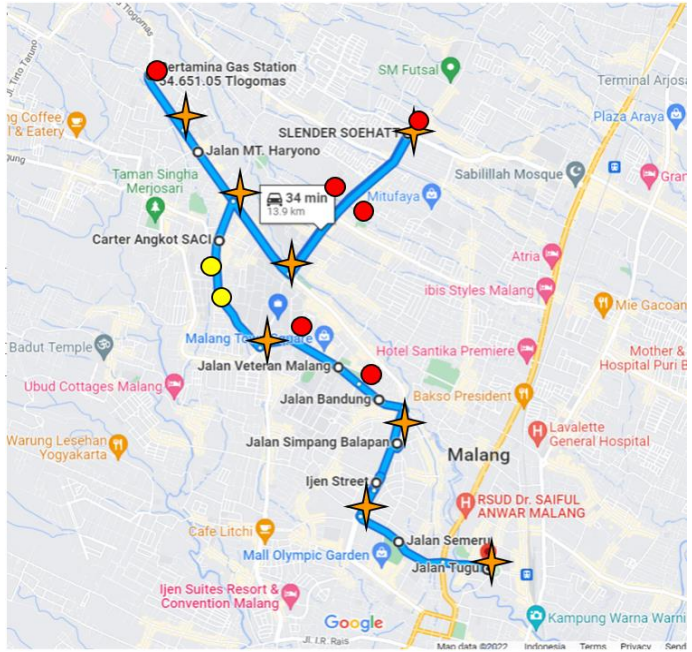


Johor Deployment

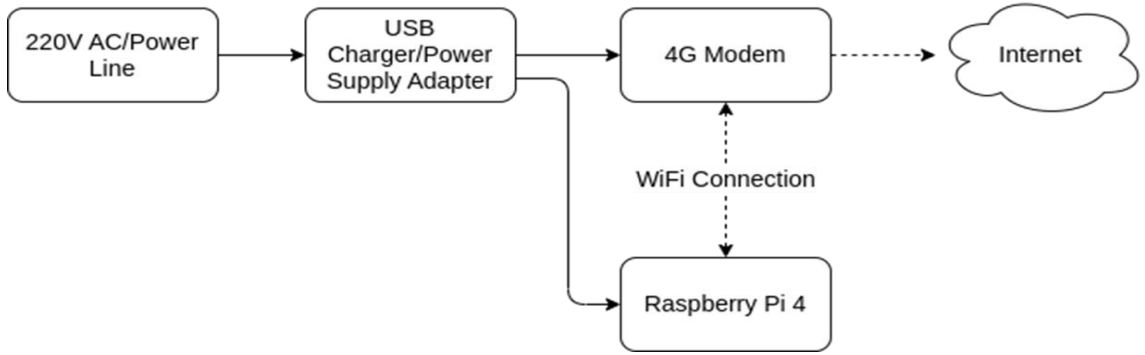
Source: ThingsBoard

Project Activities: Improving the Reliability of the BLE Sensing System (UGS / UTM / UB)

1. Malang's school bus system lacks real-time monitoring.
2. The map shows a typical bus route.
3. Inside metal box, we put Rpi 4 for detection and android phone for internet connection.
4. We use CCTV pole as source of electricity.
5. We put BLE Beacon Inside bus



- Yellow Star = RSU (Installed On CCTV pole
- Red Dot = School Bus Stop & Blue line = Bus Route

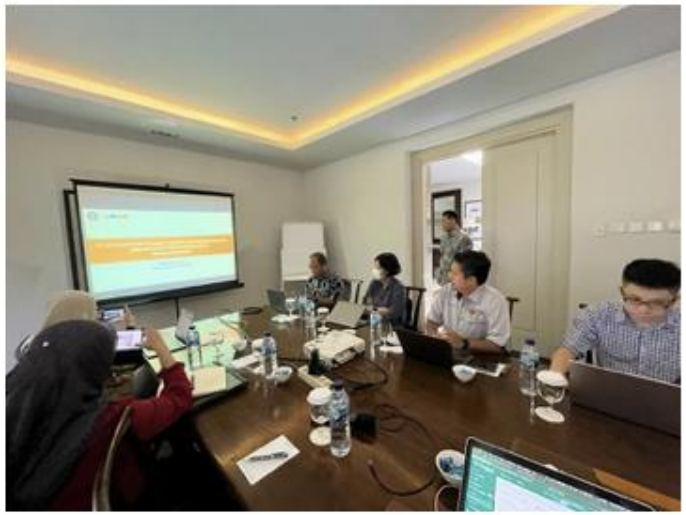


Project Activities: Site Visit to Malang (UB)

1. UB hosted a site-visit meeting on 25-26 Jan 2024 to inspect the installation of RPi for the school bus service location tracking system.

2. We discussed the challenges faced, and the reliability of Internet connectivity.

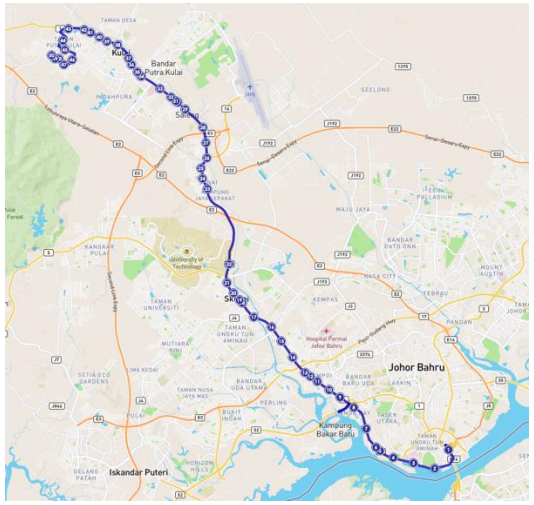
3. The UB team also demonstrated the range/distance of BLE detection on site.



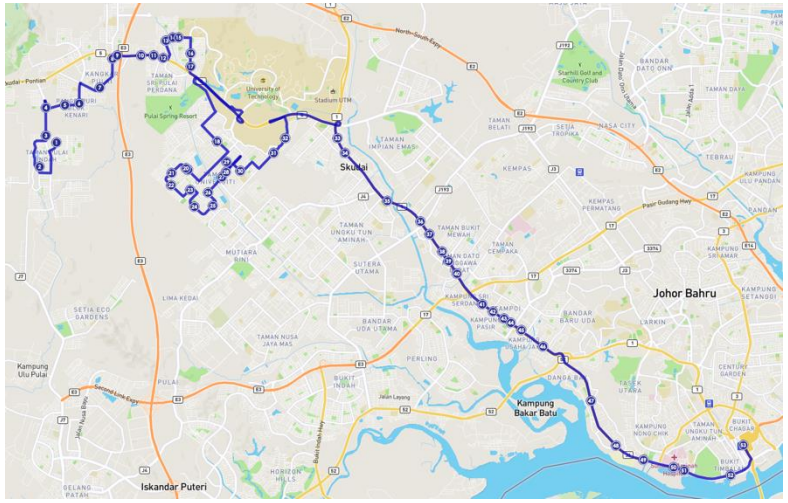
Project Activities: Bus Stop/Station Dashboard Development (UGS / UTM)

1. Developed bus arrival time dashboard's Interface for bus stop and bus terminal.

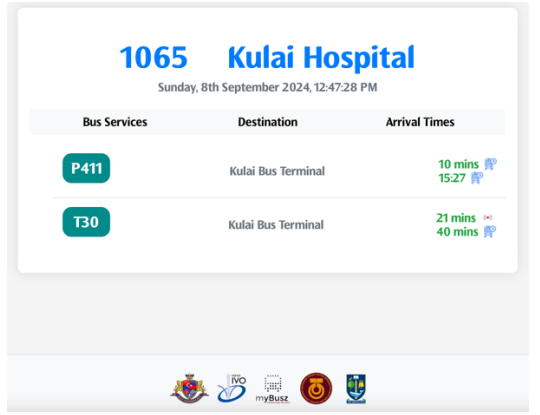
2. Extended collaboration with myBAS Johor to track two additional bus services: T30 and T31.

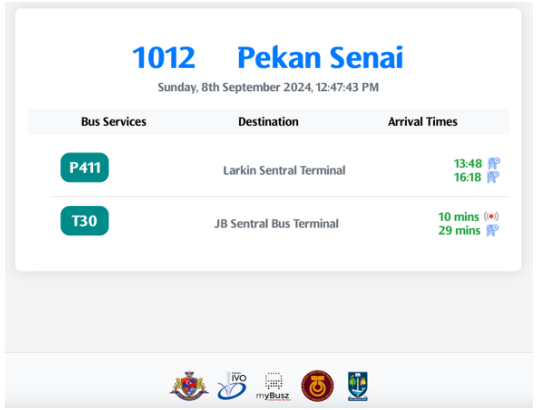


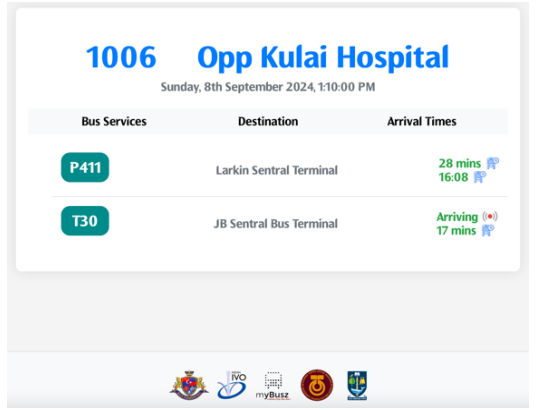
T30



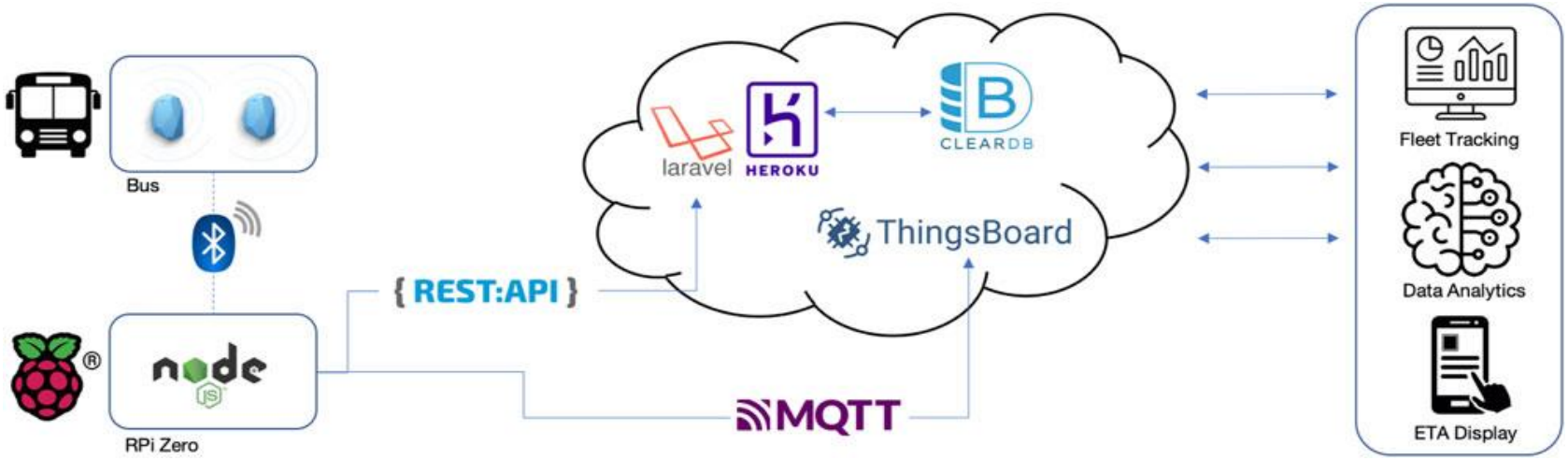
T31







Project Activities: Deployment of Prediction Model in Johor (UGS / UTM)

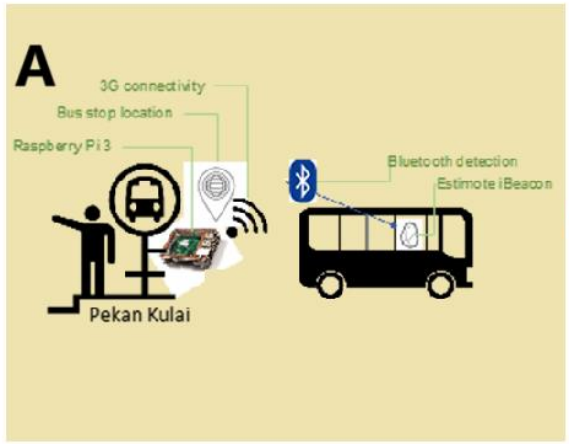


1. Training journey prediction models based on GPS dataset (obtained from PAJ) and BLE dataset (collected from this project) using ANN, LightGBM, XGBoost.

2. Integrate the LightGBM model with the current BLE bus tracking system deployed on Heroku using PHP. The model is deployed using Python and Flask for two bus services P411 and P211.

3. Once the location of the bus is updated, the PHP will query the model to predict the ETA for the remaining bus stops.

Project Activities: ETA Concept Solution Using Data Analytics (UTB/UTAR)

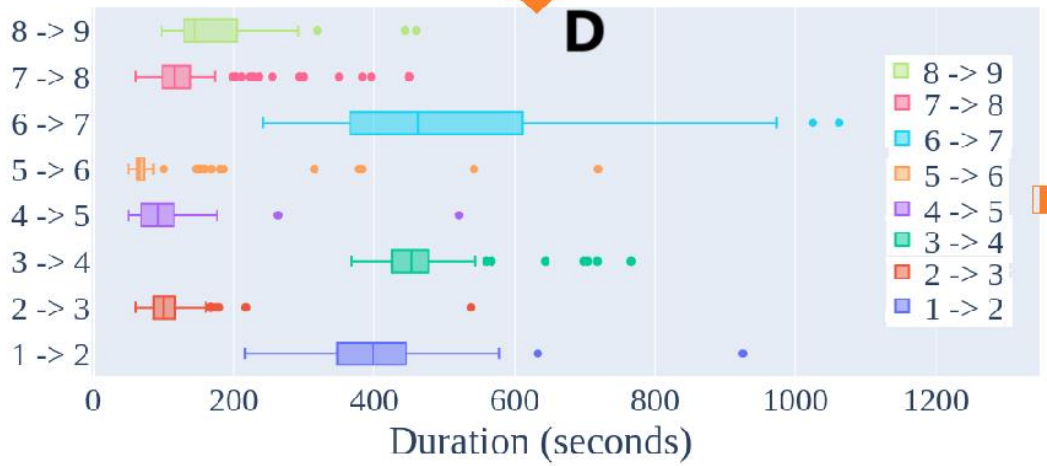


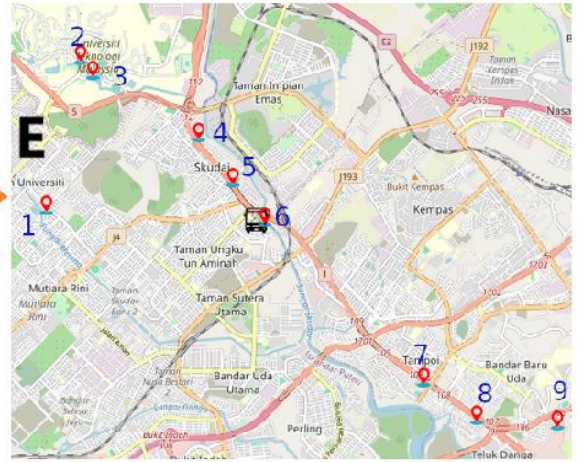
B

bus_id	route_id	place1	timestamp1_in	timestamp1_out	place2	timestamp2_in	timestamp2_out	place3	timestamp3_in	timestamp3_out	place4	timestamp4_in	timestamp4_out	place5	timestamp5_in	timestamp5_out	place6	timestamp6_in	timestamp6_out	place7	timestamp7_in	timestamp7_out	place8	timestamp8_in	timestamp8_out	place9	timestamp9_in	timestamp9_out
																	6	06:02:49	06:02:49									
31	8													5	06:03:40	06:03:40												
31	8										4	06:04:28	06:04:28															
31	8	1	06:10:39	06:10:39	2	06:14:19	06:14:19	3	06:15:38	06:15:40	4	06:22:10	06:22:11				6	06:24:49	06:24:49	7	06:30:41	06:30:41	8	06:32:17	06:32:17	9	06:34:39	06:35:13
31	8	1	07:18:39	08:10:38	2	08:15:37	08:15:40	3	08:17:08	08:17:11	4	08:23:45	08:23:45	5	08:26:02	08:26:02	6	08:27:05	08:27:05	7	08:34:30	08:34:30	8	08:36:22	08:36:40	9	08:38:37	08:39:12
31	8	1	09:41:39	10:11:09	2	10:17:16	10:17:16	3	10:19:09	10:19:11	4	10:26:30	10:27:14	5	10:28:40	10:28:40	6	10:29:47	10:29:47	7	10:38:54	10:39:11	8	10:41:42	10:41:42	9	10:44:26	10:44:26
31	8	1	12:10:40	12:10:40	2	12:17:57	12:17:57	3	12:19:33	12:19:33	4	12:27:06	12:27:06	5	12:29:19	12:29:19	6	12:30:40	12:30:40	7	12:41:36	12:41:40	8	12:43:11	12:43:11	9	12:46:49	12:46:56
31	8	1	14:10:41	14:10:41	2	14:17:17	14:17:17	3	14:19:12	14:19:12	4	14:26:49	14:26:49				6	14:29:00	14:29:00	7	14:41:26	14:41:40	8	14:43:39	14:43:41	9	14:47:34	14:47:34
31	8	1	16:02:11	16:11:41	2	16:17:56	16:17:56	3	16:19:22	16:19:22	4	16:26:59	16:27:36	5	16:29:08	16:29:10	6	16:30:13	16:30:13	7	16:40:22	16:40:40	8	16:42:37	16:42:39	9	16:46:47	16:46:56
31	8	1	18:19:42	18:19:42	2	18:27:06	18:27:06	3	18:28:34	18:28:41	4	18:35:42	18:35:42				6	18:37:53	18:37:53	7	18:48:47	18:48:47	8	18:50:53	18:50:53			
31	8	1	20:14:13	20:14:13				3	20:21:58	20:22:12																		
31	8				2	20:22:45	20:22:46				4	20:29:19	20:29:19	5	20:30:45	20:30:45	6	20:31:53	20:31:53									
31	8				2	20:33:08	20:33:12													7	20:39:27	20:39:27						
31	8	1	21:57:45	22:15:17	2	22:21:30	22:21:30	3	22:22:43	22:22:43	4	22:30:31	22:30:40	5	22:32:07	22:32:07	6	22:33:17	22:33:17	7	22:43:18	22:43:18	8	22:46:49	22:46:49	9	22:50:26	22:50:26
31	8	1	23:39:47	23:42:45							4	23:52:31	23:52:31				6	23:54:32	23:54:32									

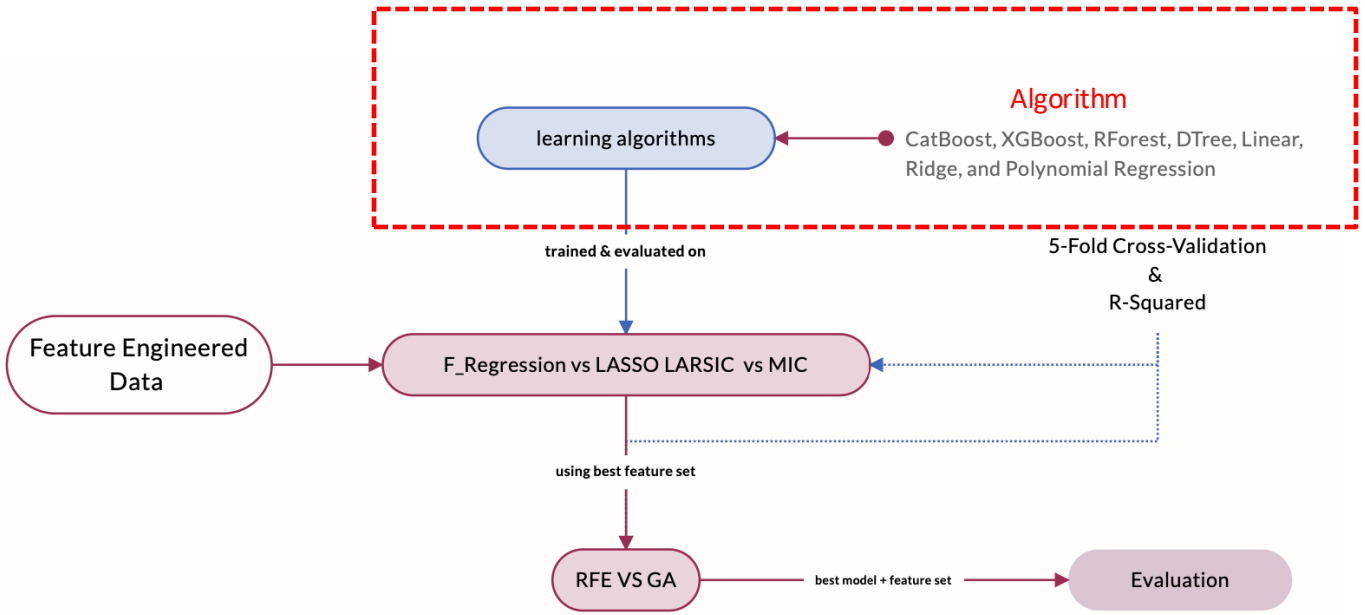
C Apply Non-linear regression models: Decision Tree regressor (DT), Random Forest regressor (RF), and K-Nearest Neighbor regressor (KNN). Apply Artificial Neural Network predictor (ANN).

A. Data collection
B. Data cleaning and conversion to a tabular format dataset
C. Apply regressors (KNN, DT, RF) and an ANN predictor to estimate ETAs
D. Finalize ETA
E. The ETA operation can work as follows: The ETA for all stops is published at the start of the service, and the subsequent ETAs for each stop are adjusted each time the bus arrives at the expected stop.





R&D results: Model Development (Malang Data) – USM Data Analytics Team



Model Design

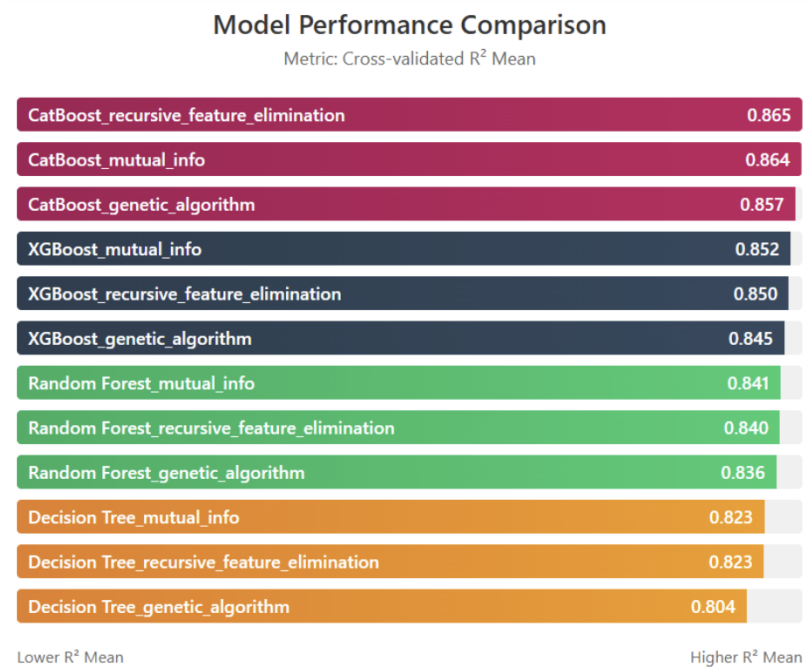
Features	Data Type	
route_id	int64	Sensor Data
distance	float64	
day_of_week	int32	
month	int32	
day_of_week_sin	float64	Cyclical Data
minutes_cos	float64	
temperature	int64	Weather Data
precip_mm	float64	
humidity	int64	
visibility	int64	
pressure	int64	
cloud_cover	int64	
wind_speed	int64	
Speed_Category	int64	Sensor Data
start_stop_encoded	int32	
end_stop_encoded	int32	

1 **Model Design:** 3 layers of feature selection were performed: Correlation-based Filtering (Pearson Correlation), Filter-based and Embedded Selection (F-regression, MIC, Lasso Larsic), and Wrapper-based Optimization (RFE, GA). The selected features were then fed into various machine learning algorithms, including CatBoost, XGBoost, Random Forest, Decision Tree, Ridge Regression, and Polynomial Regression.

2 **Model Features:** Multiple datasets—sensor data, cyclical data, and weather data—were extracted and derived for feature selection in model development

R&D results: Model Development (Malang Data) – USM Data Analytics Team

MODEL AND FS METHOD EVALUATION 2



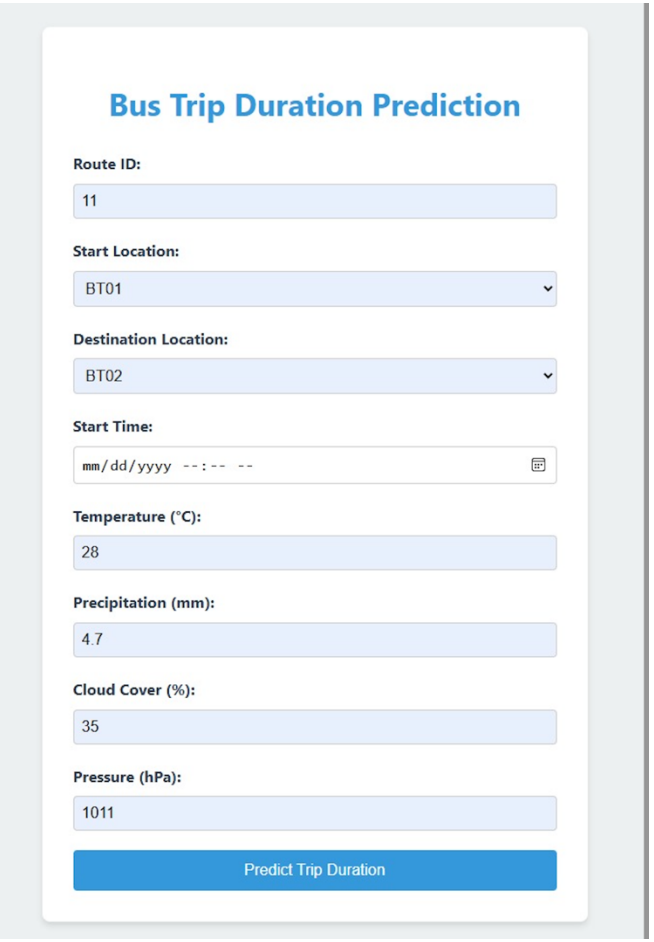
3 **Model/Algorithm Evaluation:** CatBoost with Wrapper Based Optimisation (Recursive Feature Elimination (RFE)) with the R^2 of 0.865 is chosen as the final solution.

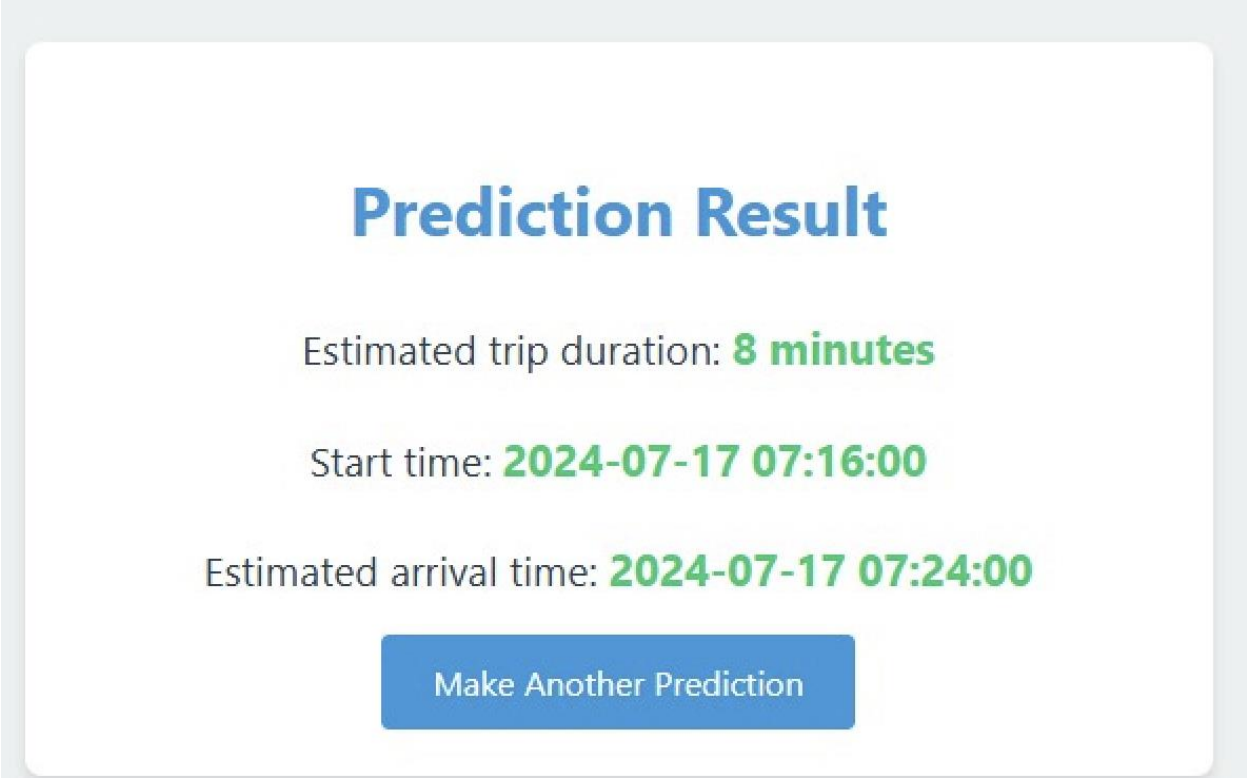
FEATURE IMPORTANCE ANALYSIS

feature	cb_feat_imp	rf_feat_imp
distance	47.56%	71.62%
Speed_Category	11.93%	14.04%
minutes_cos	9.90%	2.78%
start_stop_encoded	9.61%	1.66%
route_id	8.00%	0.81%
precip_mm	3.38%	1.21%
pressure	2.76%	1.57%
end_stop_encoded	2.63%	4.13%
humidity	2.61%	1.37%
day_of_week_sin	1.60%	0.81%

4 **Feature Importance Experiment** to understand the relative impact of different features on model performance: 'distance' is the most important predictor in bus arrival prediction, followed by 'Speed_Category' as the second most significant feature in both models. CatBoost demonstrates a stronger ability to capture temporal patterns, as reflected in the importance of the 'minutes_cos' feature.

R&D results: Prediction Tools (Malang Data) – USM Data Analytics Team





5 **Model deployment: Development** of prediction tool – First page will have list of inputs needed to predict bus arrival times at the bus stops

6 **Output Page:** Prediction results of bus arrival time at the bus stop

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.	An ANN-Based Prediction Model for Public Bus Journey Time	Yong Ting Lim, Sye Loong Keoh, Yung-Wey Chong, Noor Farizah Ibrahim, Sharul Kamal Abdul Rahim	University of Glasgow Singapore, Universiti Sains Malaysia, Universiti Teknology Malaysia	2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)	19-22 Feb. 2024	Osaka, Japan
2.	Predicting Estimated Time of Arrival Using Boosting Models	Say-Hong Kam, Yung-Wey Chong, Noor Farizah Ibrahim, Sye-Loong Keoh, Somnuk Phon-Amnuaisuk, Sharul Kamal Abdul Rahim,	Universiti Sains Malaysia, University of Glasgow Singapore, Universiti Teknology Brunei Universiti Teknology Malaysia	2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)	19-22 February 2024	Osaka, Japan

Scientific Contribution:

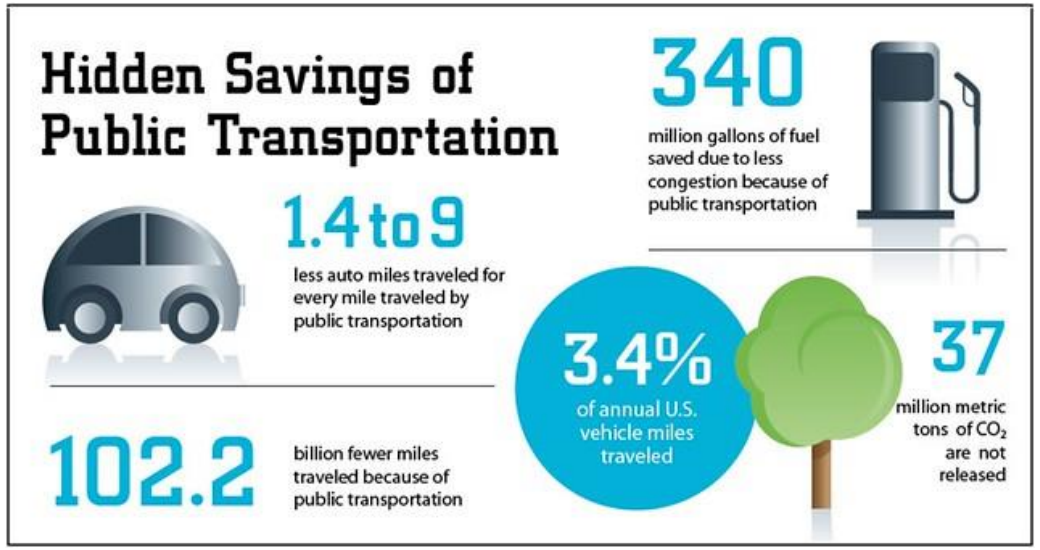
Published Journal Papers:

No:	Paper title:	Author names	Affiliation	Journal name:	The publisher of the Journal	The volume number and Pages
1.	Transforming urban mobility with internet of things: public bus fleet tracking using proximity-based bluetooth beacons	Olakunle Elijah, Sye Loong Keoh, Sharul Kamal Abdul Rahim, Chee Kiat Seow, Qi Cao, Mohammad Adib bin Sarijari, Noor Farizah Ibrahim, Achmad Basuki	Universiti Teknologi Malaysia, University of Glasgow, Universiti Sains Malaysia, Universitas Brawijaya	Frontiers in Internet of Things	Frontiers	Vol 2-2023
2.	Federated Learning for Intelligent Transportation System (ITS): Use Cases, Open Challenges and Opportunities	Yung-Wey Chong, Kok-Lim Alvin Yau, Noor Farizah Ibrahim, Sharul Kamal Bin Abdul Rahim, Sye Loong Keoh, Achmad Basuki	Universiti Sains Malaysia, Universiti Tunku Abdul Rahman, Universiti Teknologi Malaysia, University of Glasgow, Universitas Brawijaya	IEEE Intelligent Transport Systems Magazine	IEEE	Accepted.

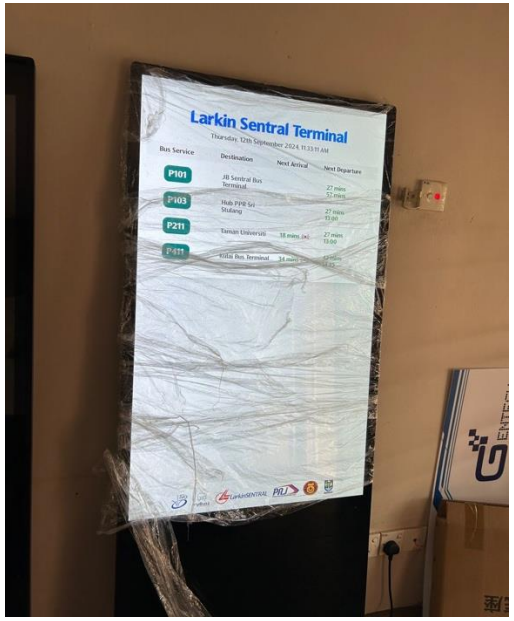
Societal Impact:

- 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 and data analysis.
  - UTM will organize a workshop to engage all the stakeholders, bus operators to present research work in 10-11 December 2025, and we are planning for Indonesia workshop in January 2025.
  - Journey prediction model established, showing feasibility of using the GPS and BLE dataset to predict journey duration between bus stops.

- Second year progress:
  - Procurement for the second year completed.
  - Bi-monthly virtual meeting to update project progress.

- Some of the Issues faced:
  - Un-realiable power supply and Internet connectivity at Johor 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.
  - Display panel purchase phase, working with NICT on equipment procurement

Future works:

- System Development:
  - Expand the system to other routes in Johor and Malang.
  - Expand the system to other cities in Malaysia and Indonesia
  - Expand the system to other ASEAN countries
  - Install the display panels in more bus stop.

- Field Trips:
  - Organize field trips to Johor in 10-112 December 2024.
  - Run a workshop on data analytics for knowledge sharing in Indonesia in 9-10 Jan 2025