Non-GPS-based ETA Models Constructed from Historical GPS Data and Traffic Contexts

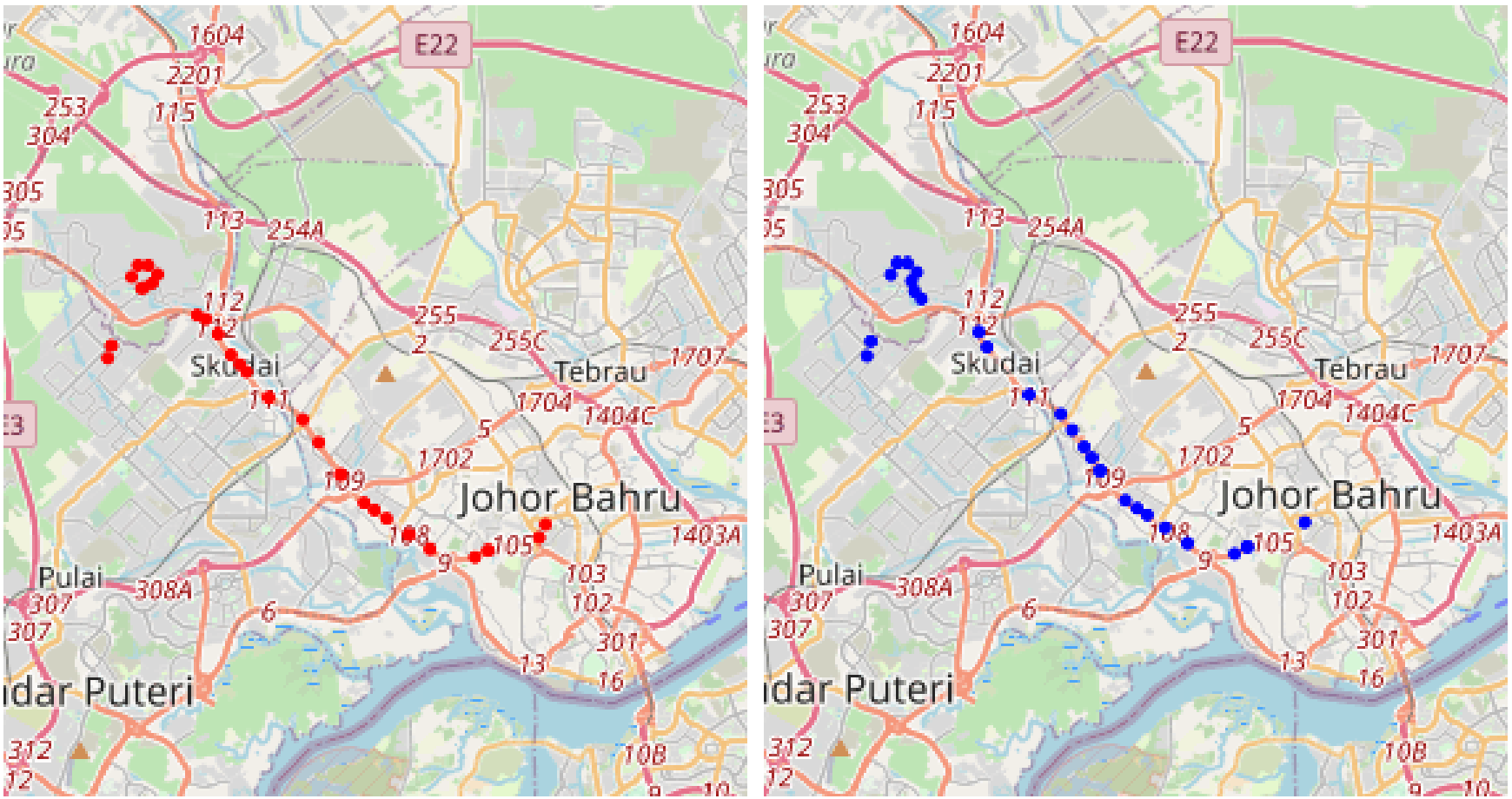
Somnuk Phon-Amnuaisuk

Universiti Teknologi Brunei
ICBIR 2023, 18-19 May 2023
Project:An IoT-based Public Transport Data Collection and Analytics Framework using Bluetooth Proximity Beacons
Members: UTM(MYS), UGS(SNG), UTB(BRN), UB(IDN), USM(MYS), UTAR(MYS)

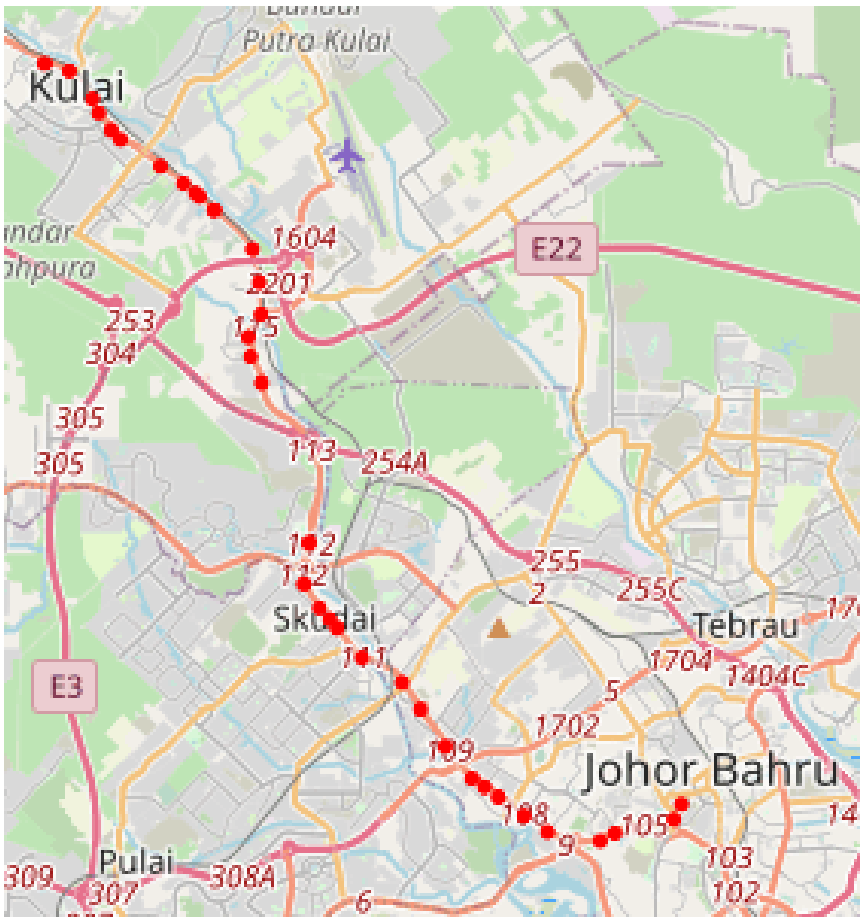
May 24, 2023

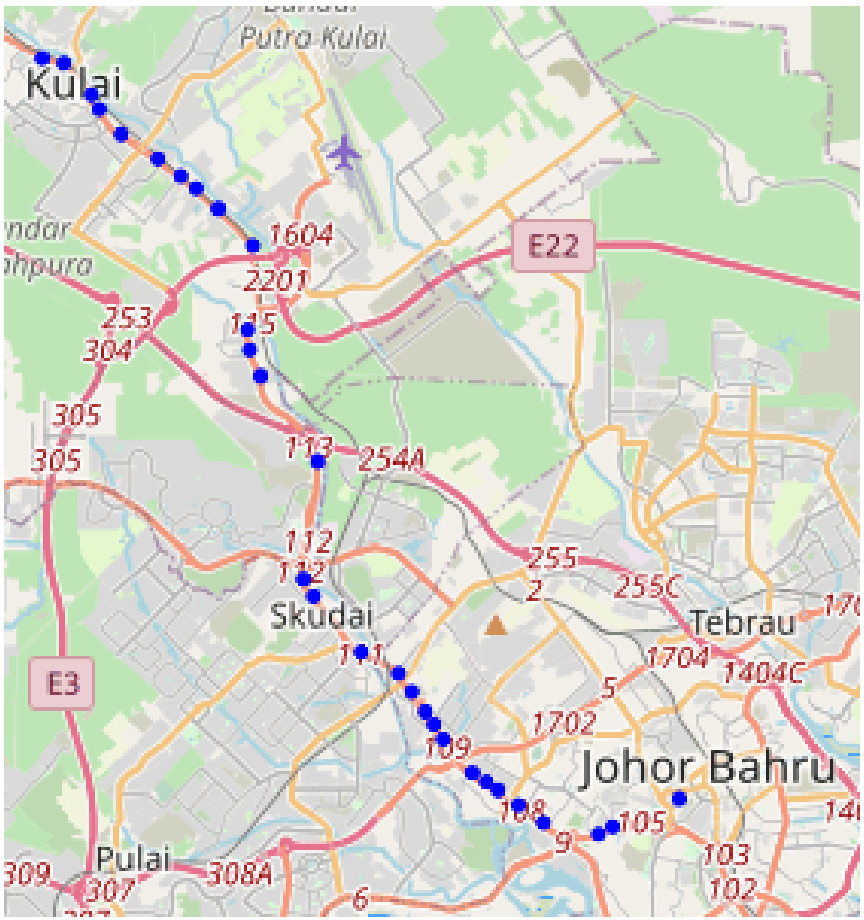
Outline

Public Bus Service: P211



Public Bus Service: P411





Public Bus Service

- P211: Taman University terminal and Larkin terminal
- P411: Kulai terminal and Larkin terminal
- The raw dataset made available to this project comprises bus information (e.g., vehicle ID, license plate), GPS locations of the bus recorded every four seconds and time stamps.
- Based on historical GPS data, the traveled time between any two bus stops for a given day of a week and time of day can be calculated.
- The knowledge distilled then serves as a representative traveled duration between bus stops. The look-up table models are constructed using traditional descriptive statistics and machine learning techniques.

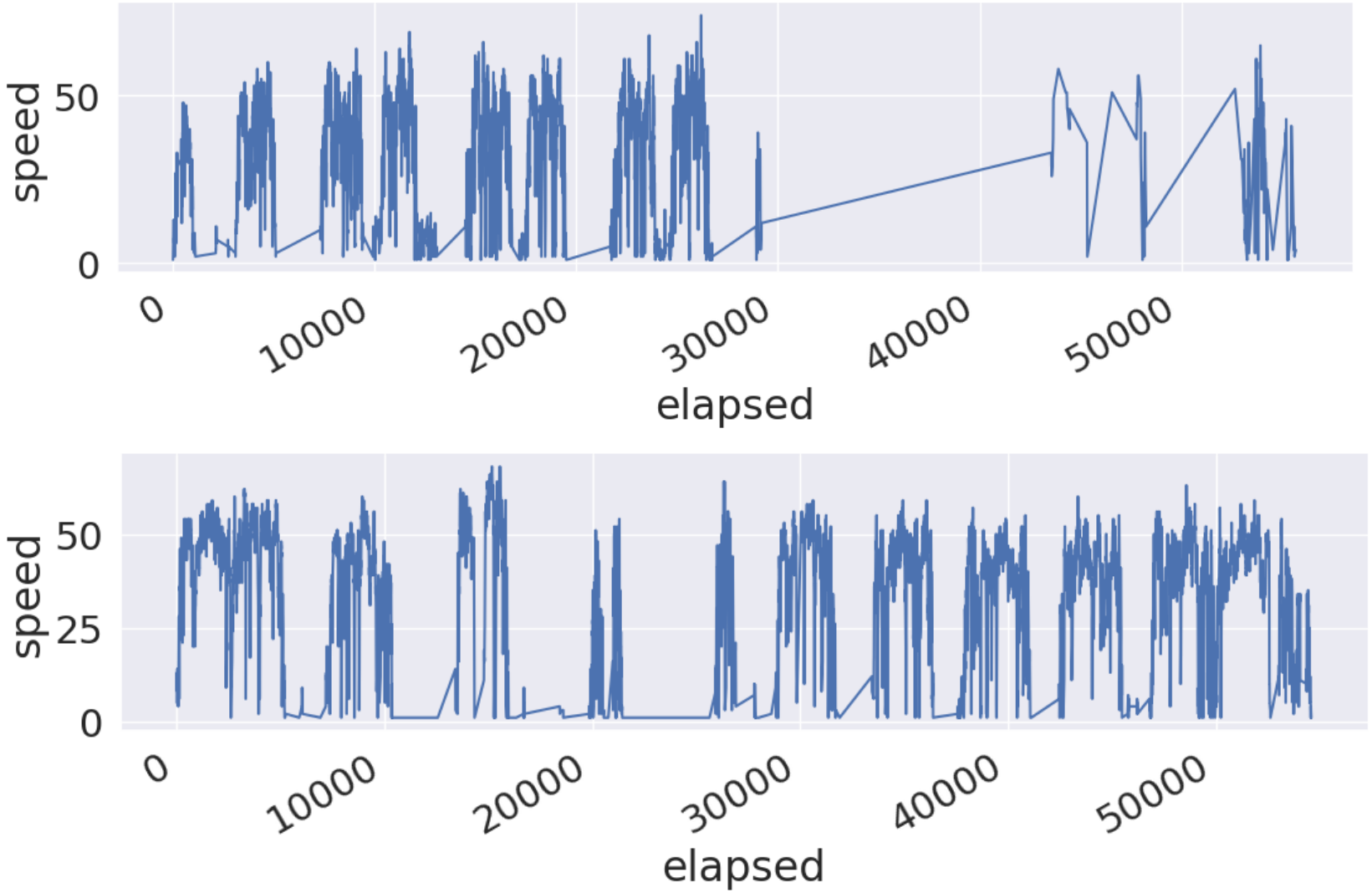
ETA Models using Historical Data

- Historical data for predicting travel times between bus stops:
1. H. Yu, R. Xiao, Y. Du and Z. He, "A Bus-Arrival Time Prediction Model Based on Historical Traffic Patterns," 2013 International Conference on Computer Sciences and Applications, Wuhan, China, 2013, pp. 345–349.
2. X. Zhang and Z. Liu, "Prediction of Bus Arrival Time based on GPS Data: Taking No. 6 Bus in Huangdao District of Qingdao City as an Example," 2019 Chinese Control Conference (CCC), Guangzhou, China, 2019, pp. 8789-8794.
3. L. Ye, P. Thiengburanathum and P. Thiengburanathum, "A Real-Time Bus Arrival Time Prediction System Based on Spark Framework and Machine Learning Approaches: a case study in Chiang Mai," 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering, Cha-am, Thailand, 2021, pp. 243-248.

Preparing Historical GPS Data for ETA Estimation

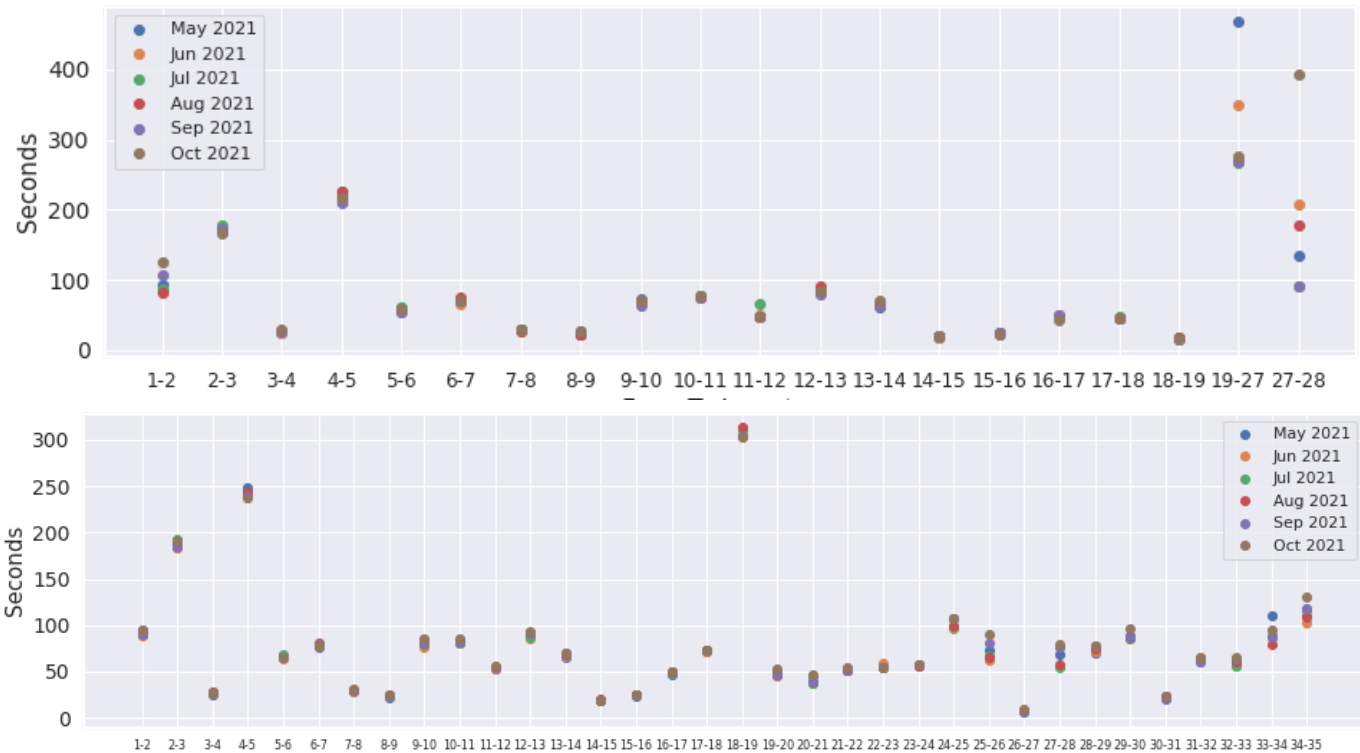
	id	trip_id	bus_line_id	socket_date	socket_datetime	gps_location	bus_vehicle_id	bus_vehicle_plate	bus_station_id	bus_station_code	distance	speed	bearing	status	created_at	updated_at	
1	id	trip_id	bus_line_id	socket_date	socket_datetime	gps_location	bus_vehicle_id	bus_vehicle_plate	bus_station_id	bus_station_code	distance	speed	bearing	status	created_at	updated_at	
2	'159985163	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:0	'1.537463	103.752174'	'59'	'JSX5418'	'334'	'P211-09'	'0.00'	'1.00'	'0.00'	'1'	'2021-05-02 00:17:5	2021-05-02
3	'159985163	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:1	'1.53745	103.752174'	'59'	'JSX5418'	'334'	'P211-09'	'0.00'	'3.00'	'161.00'	'1'	'2021-05-02 00:17:5	2021-05-02
4	'159985163	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:1	'1.537427	103.752174'	'59'	'JSX5418'	'334'	'P211-09'	'0.00'	'2.00'	'161.00'	'1'	'2021-05-02 00:17:5	2021-05-02
5	'159985164	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:1	'1.537405	103.752182'	'59'	'JSX5418'	'334'	'P211-09'	'0.00'	'2.00'	'161.00'	'1'	'2021-05-02 00:17:5	2021-05-02
6	'159985165	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:2	'1.537379	103.75219'	'59'	'JSX5418'	'334'	'P211-09'	'0.00'	'2.00'	'161.00'	'1'	'2021-05-02 00:17:5	2021-05-02
7	'159985166	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:2	'1.537363	103.752205'	'59'	'JSX5418'	'334'	'P211-09'	'0.00'	'2.00'	'161.00'	'1'	'2021-05-02 00:17:5	2021-05-02
8	'159985167	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:3	'1.53735	103.75222'	'59'	'JSX5418'	'334'	'P211-09'	'0.00'	'2.00'	'161.00'	'1'	'2021-05-02 00:17:5	2021-05-02
9	'159985168	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:3	'1.537342	103.752243'	'59'	'JSX5418'	'334'	'P211-09'	'0.00'	'2.00'	'161.00'	'1'	'2021-05-02 00:17:5	2021-05-02
10	'159985169	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:3	'1.537351	103.752258'	'59'	'JSX5418'	'334'	'P211-09'	'0.00'	'2.00'	'161.00'	'1'	'2021-05-02 00:17:5	2021-05-02
11	'159985170	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:4	'1.537394	103.752327'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'8.00'	'56.00'	'1'	'2021-05-02 00:17:5	2021-05-02
12	'159985171	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:4	'1.537447	103.752403'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'9.00'	'56.00'	'1'	'2021-05-02 00:17:5	2021-05-02
13	'159985172	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:5	'1.537544	103.752487'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'12.00'	'28.00'	'1'	'2021-05-02 00:17:5	2021-05-02
14	'159985173	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:5	'1.537642	103.752487'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'9.00'	'352.00'	'1'	'2021-05-02 00:17:5	2021-05-02
15	'159985174	NULL	'48'	'2021-05-01'	'2021-05-01 05:22:5	'1.537723	103.752426'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'10.00'	'320.00'	'1'	'2021-05-02 00:17:5	2021-05-02
16	'159985175	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:0	'1.537772	103.752327'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'10.00'	'284.00'	'1'	'2021-05-02 00:17:5	2021-05-02
17	'159985176	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:0	'1.537732	103.752197'	'59'	'JSX5418'	'334'	'P211-09'	'0.02'	'12.00'	'245.00'	'1'	'2021-05-02 00:17:5	2021-05-02
18	'159985177	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:1	'1.537642	103.752098'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'12.00'	'228.00'	'1'	'2021-05-02 00:17:5	2021-05-02
19	'159985178	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:1	'1.537546	103.751999'	'59'	'JSX5418'	'334'	'P211-09'	'0.02'	'15.00'	'230.00'	'1'	'2021-05-02 00:17:5	2021-05-02
20	'159985179	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:1	'1.537442	103.751884'	'59'	'JSX5418'	'334'	'P211-09'	'0.02'	'14.00'	'232.00'	'1'	'2021-05-02 00:17:5	2021-05-02
21	'159985180	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:2	'1.537344	103.751823'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'11.00'	'199.00'	'1'	'2021-05-02 00:17:5	2021-05-02
22	'159985181	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:2	'1.537224	103.751869'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'13.00'	'151.00'	'1'	'2021-05-02 00:17:5	2021-05-02
23	'159985182	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:3	'1.537106	103.751961'	'59'	'JSX5418'	'334'	'P211-09'	'0.02'	'14.00'	'141.00'	'1'	'2021-05-02 00:17:5	2021-05-02
24	'159985183	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:3	'1.537003	103.752037'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'11.00'	'145.00'	'1'	'2021-05-02 00:17:5	2021-05-02
25	'159985184	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:3	'1.536921	103.752075'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'8.00'	'150.00'	'1'	'2021-05-02 00:17:5	2021-05-02
26	'159985185	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:5	'1.536908	103.75209'	'59'	'JSX5418'	'334'	'P211-09'	'0.00'	'5.00'	'147.00'	'1'	'2021-05-02 00:17:5	2021-05-02
27	'159985186	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:5	'1.536855	103.752129'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'6.00'	'143.00'	'1'	'2021-05-02 00:17:5	2021-05-02
28	'159985187	NULL	'48'	'2021-05-01'	'2021-05-01 05:23:5	'1.536779	103.75219'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'11.00'	'142.00'	'1'	'2021-05-02 00:17:5	2021-05-02
29	'159985188	NULL	'48'	'2021-05-01'	'2021-05-01 05:24:0	'1.536651	103.752258'	'59'	'JSX5418'	'334'	'P211-09'	'0.02'	'14.00'	'155.00'	'1'	'2021-05-02 00:17:5	2021-05-02
30	'159985189	NULL	'48'	'2021-05-01'	'2021-05-01 05:24:0	'1.536508	103.752274'	'59'	'JSX5418'	'334'	'P211-09'	'0.02'	'13.00'	'186.00'	'1'	'2021-05-02 00:17:5	2021-05-02
31	'159985190	NULL	'48'	'2021-05-01'	'2021-05-01 05:24:1	'1.536403	103.752144'	'59'	'JSX5418'	'334'	'P211-09'	'0.02'	'19.00'	'239.00'	'1'	'2021-05-02 00:17:5	2021-05-02
32	'159985191	NULL	'48'	'2021-05-01'	'2021-05-01 05:24:1	'1.536302	103.751839'	'59'	'JSX5418'	'334'	'P211-09'	'0.04'	'33.00'	'248.00'	'1'	'2021-05-02 00:17:5	2021-05-02
33	'159985192	NULL	'48'	'2021-05-01'	'2021-05-01 05:24:1	'1.536179	103.751541'	'59'	'JSX5418'	'334'	'P211-09'	'0.04'	'31.00'	'247.00'	'1'	'2021-05-02 00:17:5	2021-05-02
34	'159985193	NULL	'48'	'2021-05-01'	'2021-05-01 05:24:2	'1.536081	103.751266'	'59'	'JSX5418'	'334'	'P211-09'	'0.03'	'27.00'	'247.00'	'1'	'2021-05-02 00:17:5	2021-05-02
35	'159985194	NULL	'48'	'2021-05-01'	'2021-05-01 05:24:2	'1.536024	103.751152'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'8.00'	'248.00'	'1'	'2021-05-02 00:17:5	2021-05-02
36	'159985195	NULL	'48'	'2021-05-01'	'2021-05-01 05:24:3	'1.535988	103.751053'	'59'	'JSX5418'	'334'	'P211-09'	'0.01'	'13.00'	'247.00'	'1'	'2021-05-02 00:17:5	2021-05-02
37	'159985196	NULL	'48'	'2021-05-01'	'2021-05-01 05:24:3	'1.535861	103.750877'	'59'	'JSX5418'	'334'	'P211-09'	'0.02'	'25.00'	'234.00'	'1'	'2021-05-02 00:17:5	2021-05-02
38	'159985197	NULL	'48'	'2021-05-01'	'2021-05-01 05:24:3	'1.535677	103.750633'	'59'	'JSX5418'	'334'	'P211-09'	'0.03'	'33.00'	'231.00'	'1'	'2021-05-02 00:17:5	2021-05-02
39	'159985198	NULL	'48'	'2021-05-01'	'2021-05-01 05:24:4	'1.535447	103.750343'	'59'	'JSX5418'	'334'	'P211-09'	'0.04'	'38.00'	'231.00'	'1'	'2021-05-02 00:17:5	2021-05-02

Preparing Historical GPS Data for ETA Estimation



Descriptive Statistic Models

- Let $S = [s_1, ..., s_n]$ be a sequence of n bus stops.
- Expected ETA can be computed based on historical data.
- $\mathbb{E}[d(s_i, s_{i+1})] = \frac{1}{N} \sum_{j=1}^{N} (t_{s_i+1} - t_{s_i})_j$



Non Linear Regression Models

- Regression: $d(s_i, s_{i+1}) = f(X, \beta) + \epsilon$, where
- x $= (busstop(i, j), day, time)$ where $busstop(i, j) \in \mathbb{Z}^+$, $day \in \{0, ..., 6\}$, $time \in \{0, .., 4\}$ and d is the duration between bus stops i and j, i.e., $d(s_i, s_j) \in \mathbb{R}$.
- Decision tree regressor
- Random forest regressor, and
- KNN regressor


Decision Tree Regressor
Seconds
0
100
200
300
400
500
0
1000
2000
3000
4000
Sorted Random From-To bus stops Test Samples


Random Forest Regressor
Seconds
0
100
200
300
400
500
0
1000
2000
3000
4000
Sorted From-To Bus Stops Test Samples

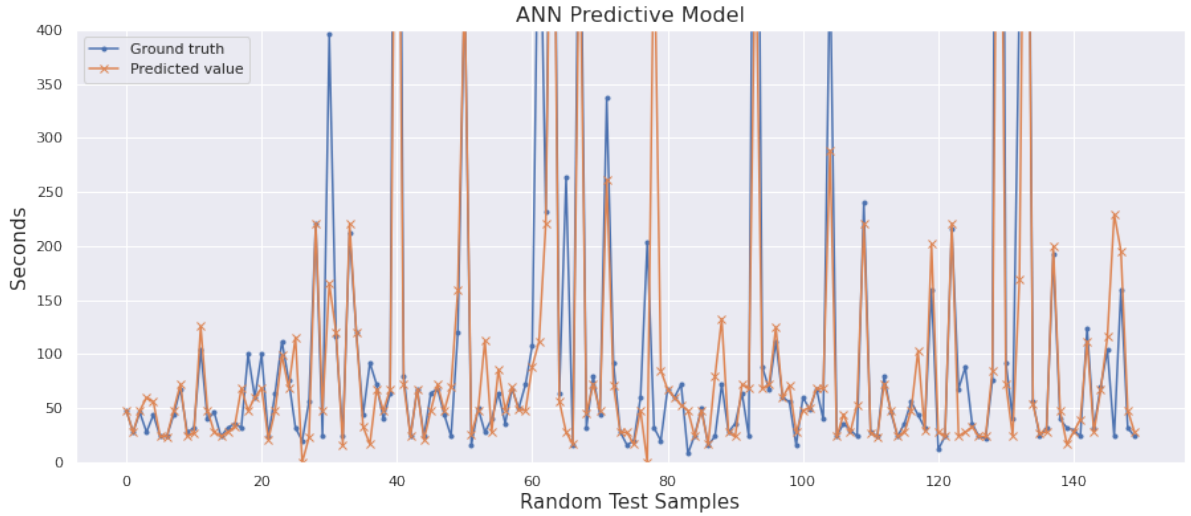
Regression Model Evaluation

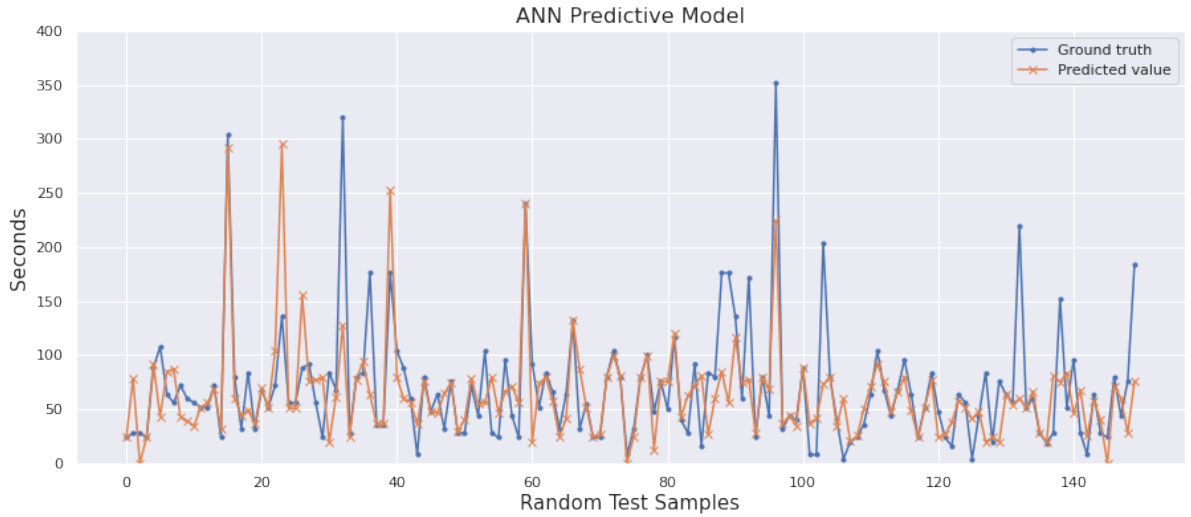
- *MSE* represents the squared distance between actual and predicted values.
  $MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - y_i')^2$
- R^2 score is a metric that tells the performance of the model, not the loss in an absolute sense. $R^2 = 1$ is the best case.
  $R^2 = 1 - \frac{\sum_i (y_i - y_i')^2}{\sum_i (y_i - \mathbb{E}[y])^2}$

Predictive Models

- Since ETA is dependent on the dynamics of the traffic environment, previous duration spans observed from prior bus stops of the same trip can reveal the dynamics of the environment.
- In this paper, two previous duration spans were employed to predict the duration to the next bus stop. This can be expressed as:
  $d(s_i, s_{i+1}) \leftarrow \phi([d(s_{i-1}, s_i), d(s_{i-2}, s_{i-1})], C)$
  where the predicted $d(s_i, s_{i+1})$ is computed from previous duration spans and a given context C i.e., day of week and time of day.

Predictive Models





	Regressors			ANN
	Decision Tree	**Random Forest**	**K-Nearest Neighbor**	**Tri-gram Predictor**
RP2 (P211) [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 (P411) [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

Conclusion 1

- Without GPS data, ETA can be estimated with good precision from historical GPS data with time stamps.
- Three approaches were employed to compute the duration between bus stops:

  (i) statistical central tendencies,
  (ii) nonlinear regression techniques i.e., decision tree regressor, random forest regressor and k-nearest neighbors regressor, and
  (iii) an ANN predictive model.

Conclusion 2

- It is possible to estimate ETA using historical GPS data. The accuracy of the estimated ETA depends on the quality of the collected data, including whether the data has captured relevant traffic contexts.
- The dynamics of the traffic scene might not be fully captured in the current setup (e.g., congestion due to accidents and weather conditions).
- For this setup, it may be more effective to place Bluetooth Low Energy (BLE) sensors at a predetermined distance before each bus stop s_i, rather than installing them directly at the bus stops.

Q & A

This publication is the output of the ASEAN IVO (https://www.nict.go.jp/en/asean_ivo/index.html), project titled: *An IoT-based Data Collection and Analytics Framework using Bluetooth Proximity Beacons* and financially supported by NICT (https://www.nict.go.jp/en/index.html). The authors would like to express their gratitude to ASEAN-IVO and Thai-Nichi Institute of Technology for their assistance and support, as well as to Universiti Teknologi Malaysia for providing the raw traffic data. We would also like to thank the anonymous reviewers for their useful comments, which helped improve the quality of this paper.