Technical Report

The Fifth Research Dive on Transportation

February 2018

Australian Government

PULSE LAB JAKARTA

Summary

With the world's population increasing, building a sustainable transportation system is no longer just a priority for megacities as citizens nowadays are more mobile. Rapid urbanisation has contributed to growth in economic activities within and between cities and with increasing connectivity across developing countries, there's now greater demand for efficient, affordable and convenient transportation services.

In developing a national transportation system though, one of the caveats of rapid urbanisation that governments ought to consider is urban congestion. In Indonesia, for instance, the transportation sector faces many related challenges, which are linked to underutilised public transportation modes and insufficient infrastructure for road transport. Notwithstanding, prospects for an improved, inclusive transportation system that can boost productivity have increased with technological advancements and the proliferation of digital data in recent years.

To enhance researchers' familiarity with some of these emerging datasets, Pulse Lab Jakarta invited 16 participants (comprised of academics and researchers) and three senior lecturers as advisors to partake in its fifth Research Dive for Development focussed on transportation. This research sprint ran from November 19-22, 2017. Participants were divided into four groups and assigned the following tasks: (1) modelling highway traffic based on toll data, (2) understanding irregular traffic patterns from traffic condition reports, (3) estimating travel time from geotagged social media data, and (4) analysing traffic connectivity and public transportation efficiency.

This report outlines the findings from the research conducted and is structured as follows:

1. The first paper details background information on the different datasets that were assigned to each group.
2. The second paper explores the performance of different payment systems used at toll roads. Specifically looking at Jagorawi toll roads, the study found that congestion occured at majority of the toll gates that did not use a non cash payment (GTO) system.
3. The third paper provides time-series and spatial aspects analysis in relation to Jakarta's traffic patterns using Waze incident reports. Various statistical modelling approaches were employed, including ARIMA, S-H-ESD and spatial descriptive model, to understand the traffic patterns based on the type of incident reports.
4. The fourth paper investigates the possibility of using geotagged data from social media data to estimate travel time. The geographic information which consists of time and location was analysed and used to determine origin-destination pairs as well as travel time patterns.
5. The fifth paper examines the performance of public transportation in Jakarta based on availability and accessibility. The results were used to propose potential locations for Transit Oriented Development in Jakarta in order to improve public transportation services.

Pulse Lab Jakarta is grateful for the cooperation of Indonesia Toll Road Authority of the Ministry of Public Works and Public Housing (BPTJ), the Greater Jakarta Transportation Agency (BPTJ), Institut Teknologi Bandung, Institut Teknologi Padang, Institut Teknologi Sepuluh Nopember, Institut Teknologi Sumatera, King Abdulaziz University, Universitas Andalas, Universitas Atma Jaya Yogyakarta, Universitas Gadjah Mada, Universitas Islam Riau, Universitas Sebelas Maret, Universitas Negeri Semarang, Universitas Syiah Kuala, Universitas Sebelas Maret, Universitas Islam Negeri Sultan Syarif Kasim Riau, World Resources Institute and the Ministry of National Development Planning.

Research Dive

Advisors

Prof. Ofyar Z Tamin	Institut Teknologi Sumatera/Institut Teknologi Bandung
Dr. Nanang Prayudyanto	GIZ / Institute for Transportation & Development Policy
Dr. Renni Anggraini	Universitas Syiah Kuala

Researchers

Group 1 - Modelling Toll Traffic Patterns: The Jagorawi Toll Case Study

Yosritzal	Andalas University
Siti Nurlaela	Institut Teknologi Sepuluh Nopember
Muhamad Rizki	Kementrian BAPPENAS
Herika Muhammad Taki	King Abdulaziz University

Group 2 - Understanding Irregular Traffic Patterns from Traffic Condition Reports

Anggun Pratiwi	Institut Teknologi Padang
Ilham Nurdayat	Institut Teknologi Padang
Martina Cecilia Adrian	World Resources Institute Indonesia
Ulfah Aliifah Rahmah	Institut Teknologi Bandung
Zakiya Pramestri	Pulse Lab Jakarta

Group 3 - Estimating Travel Time from Geotagged Social Media Data

Alfa Narendra	Universitas Negeri Semarang
Lydia Novitriana Nur Hidayat	Sebelas Maret University
Okkie Putriani	Universitas Atma Jaya Yogyakarta
Ibnu Fauzi	Universitas Atma Jaya Yogyakarta
Imaduddin Amin	Pulse Lab Jakarta

Group 4 - Analysing Traffic Connectivity and Public Transportation Efficiency in Jakarta

Yori Herwangi	Universitas Gadjah Mada
Cahyono Susetyo	Institut Teknologi Sepuluh Nopember
Tri Sefrus	Universitas Gadjah Mada
Muchammad Zaenal Muttaqin	Islamic University of Riau
Muhammad Subair	Pulse Lab Jakarta

Advisors Notes

Analysing Big Data and Developing Analytical Approaches

It is an honour to be an advisor for PLJ's Research Dive. This event is a good venue for academics and researchers to come together, analyse different datasets and develop analytical approaches. All the participants and advisors had an opportunity to discuss important issues and challenges surrounding transportation in developing countries, including Indonesia. Beyond this one in particular, Research Dive is a really convenient get-together where interested researchers can absorb immense information on the topic in question, as well as explore related issues from the ground up.

Prof. Ofyar Z Tamin

Prof. Ofyar is a senior lecturer on Transportation at the Institut Teknologi Bandung (ITB). He also currently serves as rector at Institut Teknologi Sumatera (ITERA). Prof. Ofyar has published books on transportation and modelling theory, and his work on related topics have also been published in international and national journals. He received his PhD in Transport Planning and Modelling from the University College London. He completed both his Master's and Bachelor's degrees at Institut Teknologi Bandung in 1985 and 1982, respectively.

A Step Towards More Collaborative Transportation Research

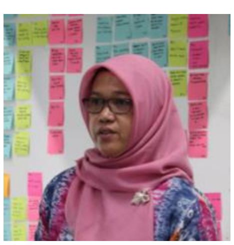
During PLJ's Research Dive, participants from diverse professional backgrounds and with diverse competencies joined forces to accomplish different research tasks that were assigned. The discussions, led by experts that were invited to the event, helped to stimulate the participants' thoughts on the topics and ultimately shaped the research questions. I was impressed with the final presentations as the findings seemed promising, particularly as the Indonesian Government continues to work towards formulating solutions to manage traffic during Idul Fitri and also the everyday traffic based on origin-destination data.

Dr. M. Nanang Prayudyanto

Dr. Nanang currently serves as a senior advisor at Bappenas in the areas of transportation planning, traffic management, public transportation development and sustainable transportation. Previously, he worked for the Coordinating Ministry of Economic Affairs and the Institute for Transportation and Development Policy. He obtained his PhD, Master's and Bachelor's degrees from the Institut Teknologi Bandung in Transportation Engineering

Big Data for Urban Transport Development

When thinking about development, an effective transportation system is one of the things to consider. Transportation infrastructure plays a major role in connecting regions within a country and making human mobility easier. Research Dive is a somewhat new research collaboration format for Indonesian academics due to its hackathon style. However, based on the flow of the event, from the introduction to the datasets to the presentations on findings, I believe many of the participants really enjoyed this format. Using Big Data to understand transportation is a step forward but can indeed be challenging. Still, I believe it will bring many benefits to urban transport planning and development.



Dr. Renni Anggraini, M.Eng

Dr. Renni is a team leader for the Traffic Facility Detail Engineering Design Project in Aceh Province. She is also the coordinator of the Transportation Engineering Management focus for the Master of Civil Engineering Program at Universitas Syiah Kuala, Aceh. Dr. Renni received her PhD in 2009 from the Technische Universiteit Eindhoven, Netherlands, and prior to that completed her Master's at Nagaoka University of Technology, Japan and her Bachelor's at Universitas Syiah Kuala, Indonesia in the field of civil engineering.

Table of Contents

Data Description for Research Dive Transportation

Imaduddin Amin
Pulse Lab Jakarta
Jakarta, Indonesia
imaduddin.amin@un.or.id

Zakiya Pramestri
Pulse Lab Jakarta
Jakarta, Indonesia
zakiya.pramestri@un.or.id

ABSTRACT

Rapid growth in urbanisation over the years has underlined the importance of transportation, particularly as it relates to citizen mobility. Greater Jakarta for example, with its many issues and challenges related to transportation, needs a comprehensive transportation plan and an effective policy framework in order to facilitate mobility among the millions of citizens who move about in the area. To improve the public transportation system, operations management and transport infrastructure, the Government of Indonesia has been developing a master plan with specific targets to achieve a more sustainable transportation system.

Supporting the Indonesian Government's efforts, Pulse Lab Jakarta organised a Research Dive for Development themed on transportation, where 16 researchers were invited to analyse highway transaction data, citizen-generated traffic report data, geotagged social media data, and geospatial data of public transport networks. This paper describes in detail the datasets that were used and helps to contextualise the technical papers that follow. The datasets were provided by Pulse Lab Jakarta with support of Badan Pengelola Transportation Jabodetabek, Jakarta Smart City, and the Institute for Transportation and Development Policy (ITDP) Indonesia.

1 INTRODUCTION

As the world is becoming more urban and more mobile, the transportation needs of citizens have been increasing alongside exponential population growth. In many developing countries, the public transport systems do not adequately meet the demands of citizens, which often lead to an increase in individual vehicle ownership.

Transportation issues are complex and go beyond a one size fits all approach. Hence, a sustainable transportation system that can boost productivity while promoting environmental health and economic well-being is crucial for improving citizen mobility.

Greater Jakarta, also known as Jabodetabek, is one of the most problematic megacities in Asia when it comes to transportation. For example, Jakarta was named the worst city in the world for traffic congestion, according to a new index published by Castrol[1]. With its higher level of population density within a relatively smaller area compared to other cities across the world, Jakarta has a higher level of transportation demands. To help meet these demands, a strategic transportation plan with an effective policy framework are necessary.

Relatedly, the government has come up with several agendas to improve transportation in Greater Jakarta. Badan Pengelola Transportation Jabodetabek (BPTJ, Greater Jakarta Transport Authority) has focused on developing an integrated transport system, which includes integrating intermodal public transportation, transportation networks, transport provider institutions, payment systems and relevant data. BPTJ also outlined a number of targets, such as a 60% modal share of public transportation, a 90-minute maximum travel time for any original-destination pair during peak hours, and a minimum average speed of 30 km per hour during peak hours, among others. In addition, Badan Pengatur Jalan Tol (BPJT, Toll Road Regulatory Agency) has focused on developing an intelligent transport system for highway roads, by implementing toll road electronification, traffic information system, emergency information system, heavy vehicle control programme, asset management system and a control centre room.

One way to support the government's efforts is to contribute an enhanced data-driven approach for developing an intelligent transportation system through the use of new types of data. In this era of digital connectivity, people often leave digital footprints with trends that can be tracked. These data such as public transport transaction data, traffic report data, and social media data can be translated into valuable insights for the government to understand the demand from the ground level. These insights can then be incorporated to better prioritise the implementation of an improved transportation plan.

Pulse Lab Jakarta organised the fifth Research Dive for Development on transportation and hosted 16 academics and researchers from diverse disciplines, including transportation engineers, civil engineers, urban planners, and economists for a research sprint. The objective was to extract insights related to transportation from digital data, which may serve useful for improving the transportation system, transportation networks, and public transportation system throughout Indonesia.

Researchers were given access to toll transaction data, citizen-generated traffic report data, geotagged social media data, and geospatial data of public transportation networks. The participants were divided into four groups and each assigned with a unique task: (a) to model highway traffic based on toll data, (b) to analyse irregular traffic patterns from traffic condition reports, (c) to estimate travel time from geotagged social media data, and (d) to analyse traffic connectivity and public transportation efficiency.

2 DATASETS

In this section, we briefly explain the four types of data that were made available under non-disclosure agreements to the participants for analysis during the Research Dive for Development.

[1] http://time.com/3695068/worst-cities-tra_c-jams/?iid=srlink1

2.1 Toll Transaction Data

In collaboration with Badan Pengatur Jalan Tol (BPJT, the Toll Road Regulatory Agency), we provided highway transaction data for some toll routes, detailing hourly and monthly activities.

The hourly transaction data covers activities from August 7-14 and September 11-24 in 2017 at the entry gates on the Jagorawi route (Jakarta - Bogor - Ciawi). The data is available in CSV format and contains information on: gate name, date, hour, and number of transactions as described in Table 1.

Table 1: Example of Hourly Toll Transaction Data

Column name	Column description	Example of value
gate	gate name (string)	Sentul Selatan
tanggal	date (dd/mm/yyyy)	11/08/2017
jam	time (hh:mm:ss)	01:00:00
total	total transaction (int)	384

The monthly transaction data covers activities from toll routes in Jakarta, as well as some routes in West Java, Central Java, and East Java. The dataset contains information on: route name, year, month, vehicle category, payment type, and number of transactions (from January 2016 to August 2017). Data was also provided on a set of toll routes in Jakarta, which included information on each entry gate for the period June 2016 to July 2017 (example shown in Table 2).

Table 2: Example of Monthly Toll Transaction Data

Column name	Column description	Example of value
rute/ gerbang	route name/ gate name (string)	DALAMKOTA JAKARTA/ KEBONNANAS
tahun	year (yyyy)	2016
bulan	month (mm)	January
golongan	vehicle category (string)	I
jenis	payment type (string)	TUNAI
volume	total transaction (int)	11762918

2.2 Citizen-generated Traffic Report

In partnership with Jakarta Smart City, for the purposes of the Research Dive, we provided aggregated data of citizen-generated traffic reports from Waze, collected from August 2016 to February 2017.

As shown in Table 3, the information contains the date, time (aggregated in 15- minute interval), geolocation (aggregated in 100 meters), report category and number of reports as described in Table 3. The reports are classified under four categories, namely `Jam`, `WeatherHazard`, `Accident`, and `Road closed`, and each category has its own sub-type, for example `JAM_MODERATE_TRAFFIC`, `JAM_HEAVY_TRAFFIC`, and `JAM_STAND_STILL_TRAFFIC`.

Table 3: Example of Citizen-generated Traffic Report Data

Column name	Column description	Example of value
tanggal	date (dd/mm/yyyy)	05/08/2016
jam	time (hh:mm)	13:15
type	report category (string)	`ROAD_CLOSED`
sub-type	sub-type of report (string)	`ROAD_CLOSED_EVENT`
lat	latitude (double)	-6.189
lon	longitude (double)	106.767
total	total reports (int)	987

2.3 Geotagged Social Media Data

We provided the locations of 10.000 twitter users (most active users) and their district level locations based on GPS-stamped tweets that were posted in Greater Jakarta from January 1, 2014 to May 30, 2014. The information contains user identifier, timestamp, and sub district code as described in table 4.

Table 4: Example of Geotagged Social Media Data

Column name	Column description	Example of value
user_id	user id (string)	262450834
timestamp	Datetime (yyyy-mm-dd HH:MM:SS)	2014-01-01 05:00:00
sub_district code	sub-district identification code (int)	3275050
district name	name of district (string)	Bekasi
sub-district name	name of district (string)	Bekasi Selatan

2.4 Public Transportation Network Data

In collaboration with the Institute for Transportation and Development Policy (ITDP) Indonesia, we provided geo-spatial data of (a) Jakarta public transport networks, including public bus routes, (b) KCJ Line, the commuter line network in Greater Jakarta, and (c) TransJakarta route, the Bus Rapid Transit (BRT) system in Jakarta. The data is available in shapefile format. Figure 1 shows an example of the public transportation network map.

3 DATA AND TASK MAPPING

The toll transaction data was assigned to the first group, which used the data to model highway traffic patterns. The second group used traffic condition reports data from Waze to analyse irregular traffic patterns in Jakarta. Using geotagged social media data, the third group estimated travel time between pairs of inferred origin-destination locations. The public transportation network data was shared with the fourth group to investigate traffic connectivity and the efficiency of the existing public transportation system in Greater Jakarta.

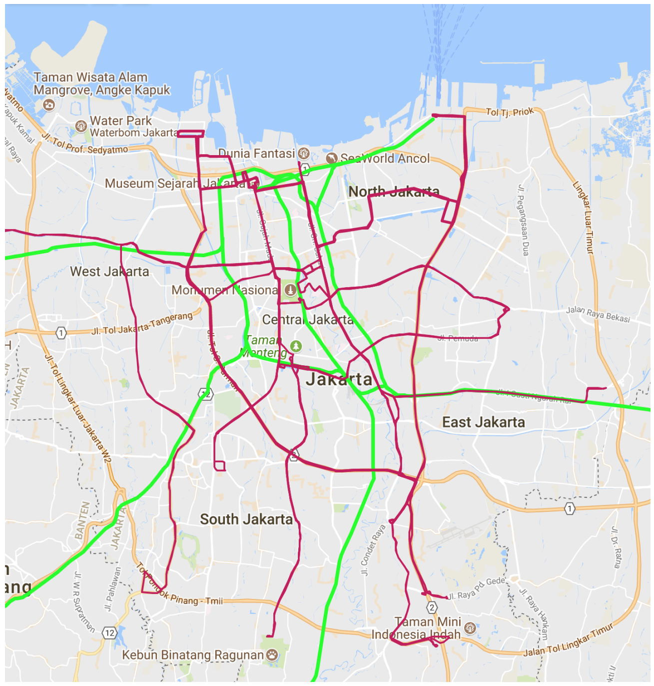


Figure 1: Example of Public Transportation Network (KCJ Line and TransJakarta routes)

Modelling Toll Traffic Patterns: The Jagorawi Toll Case Study

Yosritzal
Andalas University
Padang, Indonesia
yosritzal@eng.unand.ac.id

Siti Nurlaela
Institute Technology of Surabaya
Surabaya, Indonesia
sitnurlael0@gmail.com

Muhammad Rizki
Ministry of National Development Planning
Jakarta, Indonesia
muhammadrizki1404@gmail.com

Herika Muhamad Taki
King Abdulaziz University
Saudi Arabia
htaki0001@stu.kau.edu.sa

ABSTRACT

Toll roads play a significant role in increasing mobility and connectivity. However, the conventional payment system has incurred the problem of long queuing at toll gates, which reduces the overall performance rate. BPJT or Badan Pengatur Jalan Tol in Indonesia has introduced a new non-cash payment (GTO) system and a near-future policy of the multilane free-flow (MLFF) system to reduce the queuing problem associated with payment. Due to different characteristics in toll traffic patterns, there may be differences in the requirement of installment of this new payment system among different toll roads. These requirements include the optimum number of payment gates needed for each toll and the priority of toll locations to be upgraded into GTO and MLFF. To investigate these, we need to understand the impacts of each payment system on the level of queuing, such as the number of cars in queue, the amount of queue time, and the congestion costs due to queuing. The queuing theory was used to compare the performance of different payment systems (conventional, GTO and MLFF) on Jagorawi toll roads as a case study. A model was developed to understand the patterns of queue and the relationship between queuing, the toll 's gate volume, and the number of booth. The study found that congestion occurs at the majority of the toll gates, which is linked to the inefficiency of the booth services. Rather than increasing the number of booth, simulation of queuing models showed that optimisation of toll gate could be achieved with GTO systems. The findings showed that the total number of cars and time in queue at cash payment system is almost 540% higher and annual congestion cost is almost 284% higher than GTO.

KEYWORDS

Queue theory; toll road; traffic pattern; toll performance; payment system

1 INTRODUCTION

Toll roads plays a significant role in increasing mobility and connectivity. Currently, there are 34 toll roads in Indonesia with a total length of 987 km and serving an average of 3.7 million vehicles daily. The National Development Mid-term Plan (2015-2019) has targeted toll road development for 1,807 km in 2019.

While new toll road constructions are progressively developing, current issues are related with the establishment of a new payment system with electronic toll card (ETC), which targeted all toll booths to be cashless by 100% in the end of 2017 and the shift of the payment system into the multilane free flow by 100

The Ministry of Public Works Regulation or Permen PU No. 16/PRT/M/2016 has assigned the minimum service standards on toll accessibility for IndonesiaâĂŹs toll. In relation with the toll payment system, the minimum service standard sets the average transaction speed of maximum 6 second per vehicle on an open transaction, while the close transaction should serves at maximum 5 seconds per vehicle on the entrance and maximum 9 seconds per vehicle on the exit. It is believed that the implementation of ETC and high speed tolling by Jasa Marga and Indonesian Toll Road Authority or BPJT is a direct strategy to achieve these standards. The ETC system is expected to increase the transaction speed by reducing the transaction period, at the service standard of maximum 4 seconds per vehicle on take toll ticket, maximum 5 seconds per vehicle on transaction and no more than 10 vehicles per booth at a normal traffic condition.

The Greater Jakarta tolls currently serve 47% share of toll traffic. Jasa Marga has officially introduced this new non-cash payment system in Jagorawi since 12 September 2017. The implementation of this policy was gradually applied on all Jakarta-Bogor-Ciawi toll gates and was completed on 26 October 2017.

In order to understand the impact of this ETC payment system on toll road performance, the comparison of the new and conventional payment system applied in Jagorawi toll was investigated. This research aims to understand the impact of varied payment systems on toll performance in terms of the queuing level. A model was developed to analyse the relationship between queuing, the tollâĂŹs gate volume, and the number of toll booths. The output of the model will help to identify when and where toll gates are to be upgraded into GTO (the automatic gate) and non-gate or high speed tolling system with MLFF (multilane free flow). The model will also analyse the optimum booth number when the single system of the mix system (hybrid gate) is to be applied.

2 METHODOLOGY

2.1 Research approach

Indonesia's toll road system currently has 261 toll gates that consist of 1,484 toll booths. There are three payment systems, which include the manual gate with cash payment, hybrid gate that combine cash and electronic card, and the automatic gate or GTO that requires tapping electronic card on payment and using e-card plugged in to the OBU.

While the installment of the automatic toll gate (GTO) in the Jabodetabek (the Greater Jakarta) toll roads is currently around 38.5%, the e-payment penetration is still lower at 27.5%. The e-payment issue consists of the adaptability and the acceptability from toll users on e-payment system, which required the gradual implementation and the application of hybrid method on payment and the optimum action plan on full application of ETC and MLFF.

This research investigates the optimum choice of location, timing, and the number of toll booths when the single e-payment or hybrid is to be applied. The approach on this investigation consists of comparing the impact of each and the combination of payment system on the toll performance, especially by considering the minimum service standards. There are three performance indicators to be compared from the application of each payment system, i.e. the number of vehicles in queue, the amount of queue time and the impact queuing has on congestion costs. The input data consisted of one variable: the seven-day toll gate volume patterns at six toll gates in Jabodetabek: Bogor, CIbubur, Taman Mini, Dukuh, Cimanggis Utama and Gunung Putri; from 18 to 25 September 2017 that consisted of hourly gate volume from a total of 69 booths.

A model was developed to investigate the relationship between the queuing level, the minimum service standard on maximum allowable time of transaction period in seconds, and the minimum service standard on maximum allowable number of vehicles in queue. A regression model was developed to predict the optimum number of toll booths as a function of the hourly toll gate volume in each toll booth.

The stages of this research is explained as shown in Fig. 1. Firstly, given the hourly toll gate volume data in the case study, identification of toll traffic patterns based on daily patterns was described by descriptive statistics. Secondly, the queuing model was established based on input data (volume of vehicle per booth per hour or λ), the service rate (an assumption of the transaction time for each payment system, hence refer to the number of vehicles that exit the toll booth or μ), the threshold value when $\lambda = \mu$. The queuing formula will calculate the level of queuing in terms of the number of vehicles in queue per hour. Thirdly, the extension of the model by including other variables, such as the queuing time and the congestion costs of queuing. We assigned some assumptions to produce those variables, such as the queuing capacity in one hour according to the service rate of each payment system and the value of time calculated based on the monthly minimum wage level in Jabodetabek (IDR 3.6 million). Finally, a regression equation was developed to predict the number of required toll booth as a function of the gate volume per hour.

2.2 Queuing theory

The queuing model applied in this research was referred to the formula discussed in Tamin, O.Z (2000) [1]. The basic formula is written as follows:

$$\eta = (\frac{\lambda^2}{\mu * (\mu - \lambda})$$

where is the number of vehicle in queue per hour in each toll booth, $\lambda$is the volume of vehicle per hour entry the toll booth, μ is the capacity of the toll booth to serve vehicle according to the

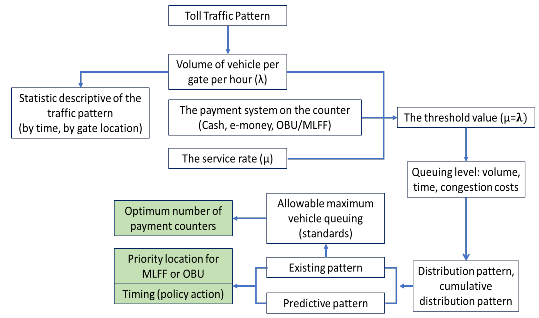


Figure 1: Research Methodology

service rate of each payment system, i.e. 11 seconds per vehicle in the cash system, 8 seconds per vehicle in the GTO system, and 6 seconds per vehicle in the MLFF system.

To simulate the optimum number of payment booth for each toll gate, some constraints was assigned, such as the maximum number of vehicle in queue according to the minimum service standards and the assumption of uniform queuing distribution across all payment booths in one toll gate, hence $\lambda 1 = \lambda 2 = \lambda 3$, where m is the existing number of payment booths.

Figure 2 (a) explained the hypothetical relationship between the number of vehicle in queue and the second component of the denominator in equation 1. The higher positive differences between means the lower number of vehicle in queue. The large negative differences between means the lower space left on toll gate per hour. Figure 2 (b) illustrated the queuing began to accumulate once the threshold value has been achieved, and the remaining vehicle in queue in a given hour t will be accumulated in the hour *t+1*.

The predictive model also developed by a single regression, based on a relationship between the number of vehicle in queue and the volume of toll booth per hour based on data. The predictive model was used to simulate the optimum number of toll booths based on a single or a hybrid payment system in each toll gate and when this particular toll gate should implement the MLFF in the future.

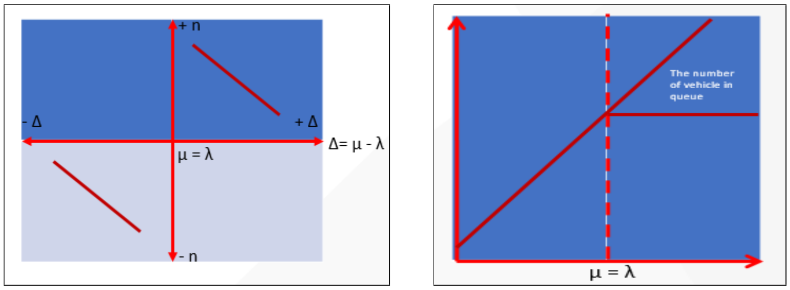
2.3 Study Area

Figure 3 presented the study area, the Jagorawi toll road network that consisted of 14 toll gates overall. The case study specifically was focused on 6 (six) toll gates i.e. Bogor, Cibubur, Taman Mini, Dukuh, Cimanggis Utama and Gunung Putri, accounted for 69 toll booths. Jagorawi toll road network currently covered 59 km length. It served total 550,000 daily traffic on average (2016). The six toll gates being studied have around 220,000 daily traffic (September, 2017). The implementation of cashless payment in Jagorawi was among the first in Indonesia that were started since 18 September 2017, where the first period of cashless payment was gradually implemented during 15 September-28 September 2017.

3 RESULT AND DISCUSSION

3.1 Traffic Pattern at Toll Gate

The pattern of traffic at toll gate Cibubur and Gunung Putri on weekdays from 19 to 25 September 2017 are shown in Fig 4. It can



(a) (b)

Figure 2: Visualisation of the equation 1. (a). Relationship between the number of vehicle in queue, arrival to the queue and the service time. (b). The threshold of queuing time and its impact on the level of queuing

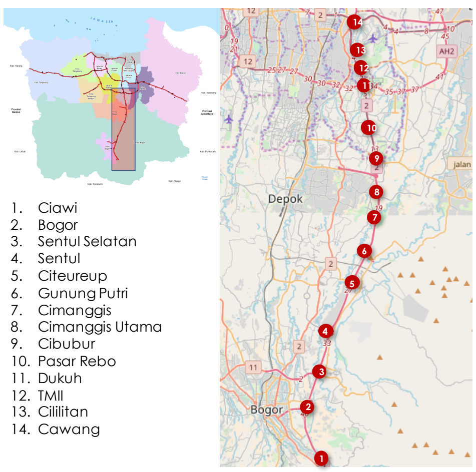


Figure 3: The study area: the Jagorawi toll network

be seen that the pattern for each toll gate is almost similar. Similar findings were found from the other Jagorawi toll gates.

As the pattern for the weekdays is similar, in the next analysis, data on the September 18, 2017 will be use as representation of the weekday data at all toll gates.

Fig. 5 shows the distribution of traffic at each booth at Cibubur Toll Gate on the September 18, 2017 where the traffic was not equally distributed among the booths. One booth was more favorable than the others. This phenomenon may be due to the booth has been upgraded from cash system to automatic payment system. However, no data available to support this assumption.

3.2 Queue Analysis

Using queueing formula in equation (1), the number and the time of vehicle in queue was estimated for each type of payment system i.e. cash, GTO and MLFF at morning and evening peak hour. The estimation was made for six major toll gates such as at Cibubur, Bogor, Gunung Putri, Cimanggis Utama, Dukuh, and Taman Mini. The results were compared to the maximum number of vehicle in queue according to the Minimum Standard of the Service which is 10 vehicles. The results are shown in Fig. 6.

Both graphs are showing similar pattern where cash payment system would not be able to meet the minimum standard of services for the Cibubur, Cimanggis Utama, and Dukuh Toll gates. The number of vehicles queuing even exceeds twice of the maximum number of vehicles allowed in the Minimum Standard of Services. However, at the other toll gates such as Bogor, Gunung Putri, and Taman Mini, the number of vehicle in queue was lower than the maximum allowable.

The implementation of the GTO at the toll gates of Cibubur, Cimanggis Utama and Dukuh was predicted to reduce significantly the number of vehicle in queue. All of the toll gates would meet the requirement mentioned in the standard.

3.3 Congestion Price

Total loss due to the payment system annually was estimated using congestion pricing formula. The result is shown in Fig. 7. The toll gate with the highest total loss was Cibubur, which was estimated to be 84 Billion Rupiahs, followed by Cimanggis Utama with 49 Billion Rupiah and Dukuh with 21 Billion Rupiahs.

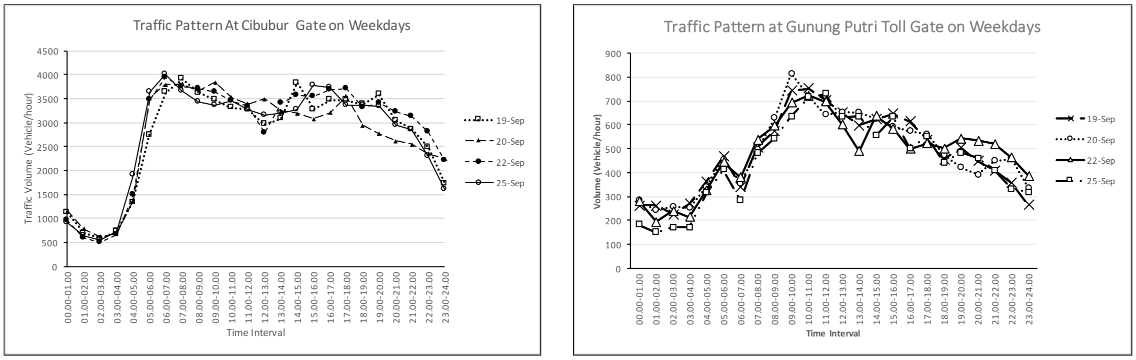
3.4 Modelling and Simulation for Estimation of the Required Number of Booth

In order to estimate the number of required booth, a model was developed. The model was a linier model between number of vehicle and the number of booth required.

Based on the model, the number of booth can be predicted based on the system used and total vehicle enter the gate. For example, when using cash payment system with the number of vehicle enter the gate was 6000, total booth needed would be 20, but if only GTO was used, total number of booth needed would be 15.

The model in Fig. 8 could be used to estimate number of booth required if a combo system will be implemented. If a combination between GTO and cash system will be used, number of counter needed would be depended on the distribution of vehicle entering the payment system. For example, if 2000 from 6000 people will choose GTO, number of required is 5 GTO and around 10.

Based on the model, the booth needed for Cibubur and Cimanggis Utama gates when single system was used for year 2017, 2020, 2025,



(a) (b)

Figure 4: Traffic pattern on Weekdays. (a) at Toll Gate Cibubur and (b) at Toll Gate Gunung Putri

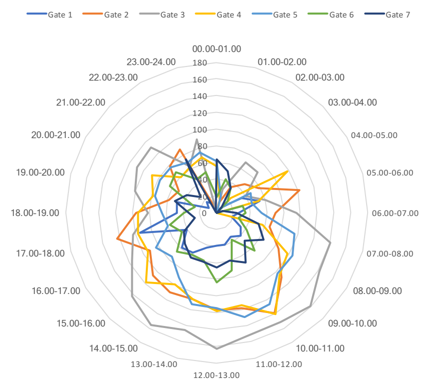
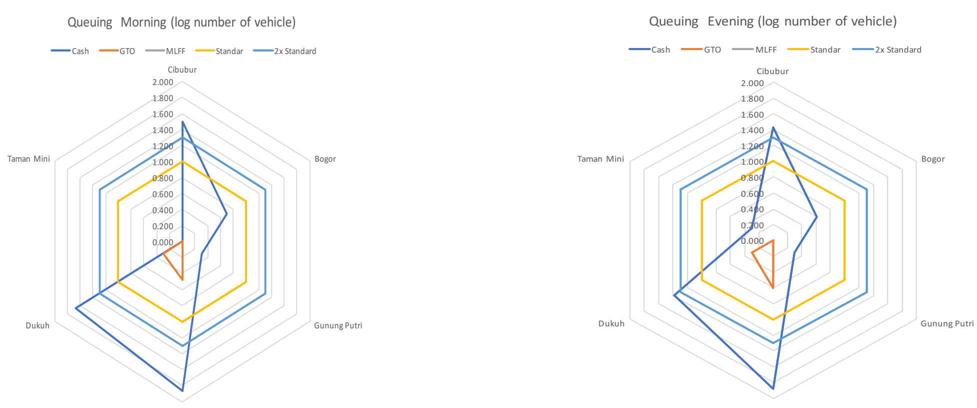


Figure 5: Distribution of traffic at each booth at Cibubur Toll gate



(a) (b)

Figure 6: Traffic pattern on Weekdays. (a) at Toll Gate Cibubur and (b) at Toll Gate Gunung Putri

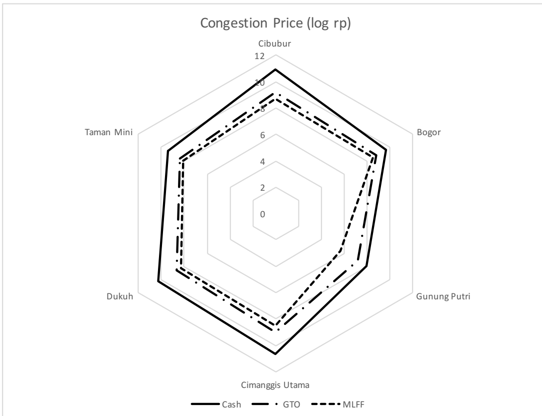


Figure 7: Annual Congestion Price at the Toll Gates

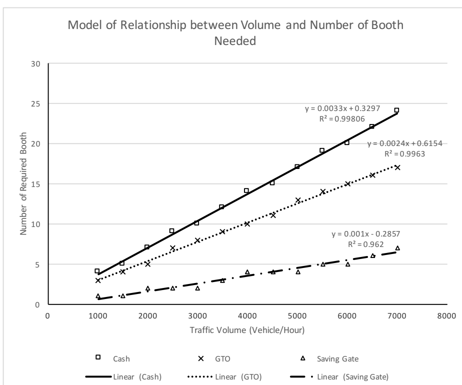


Figure 8: Model of the relationship between Traffic Volume versus number of required booth

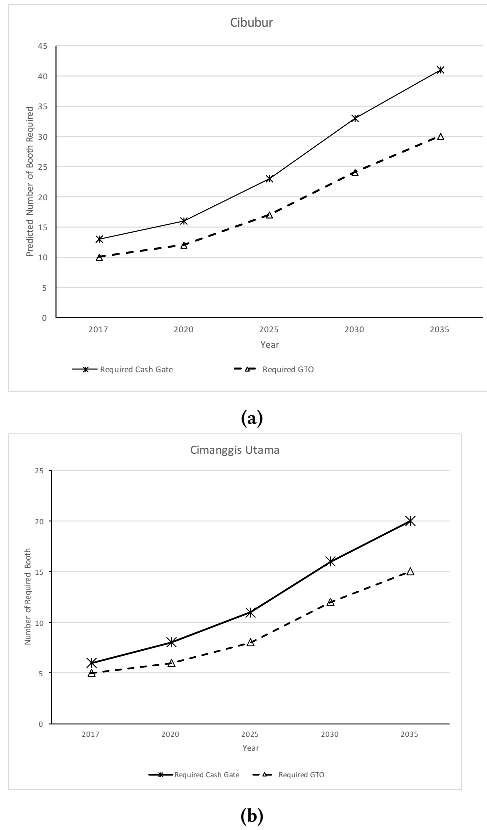


Figure 9: Number of booths needed in the future when single payment system is used. (a) Cibubur, (b) Cimanggis Utama

2030 and 2035 are shown in Fig. 9. The total vehicle enter the gate was assumed to increase by 8% per-year and the number vehicle at each booth was assumed to be uniformly distributed.

The existing booths for the toll gate Cibubur, Bogor, Gunung Putri, Cimanggis Utama, Dukuh and Taman Mini are 17 and 7 respectively. From Fig. 9 we could see that the existing number of booth when using cash payment system would not be enough at Cibubur and Cimanggis Utama in 2020 if the traffic is equally distributed among the counters. When all of the payment system was upgraded to GTO, number of the booths would be enough until 2025.

REFERENCES

[1] O.Z Tamin. 2000. *Perencanaan dan Pemodelan Transportasi, Second Edition.* Institut Teknologi Bandung.

Understanding Irregular Traffic Patterns from Traffic Condition Reports

Ilham F. Nurdayat
Institut Teknologi Bandung
Bandung, Indonesia
ilham.fadhil@sbm-itb.ac.id

Martina Cecilia Adriana
World Resource Institute
Jakarta, Indonesia
martinacecilia91@gmail.com

Ulfah Aliifah Rahmah
Institut Teknologi Bandung
Bandung, Indonesia
ulfahaliifah95@gmail.com

Anggun Pratiwi Jf
Institut Teknologi Padang
Padang, Indonesia
pratiwijf@gmail.com

Zakiya Pramestri
Pulse Lab Jakarta
Jakarta, Indonesia
zakiya.pramestri@un.or.id

ABSTRACT

Traffic pattern has been important subject for transportation study to investigate what and how particular traffic conditions could occurred in an area. Previously, this study limited by lack of data quality that forced researcher to make inference based on limited data. In the advent of technology, high level data quality made available through user-generated content application. In this paper we analyzed time-series and spatial aspect of Jakarta's traffic pattern using Waze incident reports. We employed ARIMA, S-H-ESD, and spatial descriptive analysis to studied traffic pattern categorized in type of incident reports. Our findings showed that, there are possibilities to develop predictive model based on univariate reports pattern, but this model could be developed to incorporated other predictors. Using spatial descriptive analysis, this research conclude that flood event may contribute to regular and irregular traffic condition. However, severity of flood event could not been always reflected from Waze incident reports.

KEYWORDS

traffic pattern, big data analysis, spatial analysis, temporal analysis

1 INTRODUCTION

Jakarta urban traffic has been a prolonged problem. Many stakeholders, urban growth, politics, land use, and travel behavior has made Jakarta urban traffic more difficult to comprehend. In order to understand this phenomenon policy and decision maker need reliable data. Previously, data has been made available through research, survey, and study of researcher and government agencies. Limitation of this approach is, time, manpower, and budget. Only limited observations and studies are generated, as a consequence, decision makers are driven by this limited information. Recent development in hardware and software has generated data in unprecedented volume and detail. This data enabling more microscopic analysis and lower data collection cost [1]. Traffic pattern is microscopic event that required low granularity level to depicts patterns in spatial and time-series.

Road traffic is most observed traffic patterns in transportation study. Road traffic is refers to the interactions between road user and infrastructure caused by movement needs. Road traffic varied by day, yet generate the repeated design that resulting on road traffic pattern such as number of vehicle use, land use, and mode of transport at different times and different locations. By understanding traffic pattern, traffic problems such as congestion can be identified as well as developing the optimal transport network with efficient movement of traffic.

Congestion as part of traffic pattern is the condition where the traffic volume exceed the road capacity causing slower speed than a normal free flow speeds. According to the United States Department of Transport [4], congestion is caused by seven root causes, which often interacting with one another in the road system. The seven primary causes are divided into 3 categorizes i.e.:

Category I: Traffic Influencing Events

(1) Traffic incidents: vehicular crashes, breakdown, and debris in travel lane, both in the shoulder and the road
(2) Work zones: construction activities
(3) Weather: Wet, snowy, or icy roadway surface, and even flood

Category 2: Traffic Demand

(1) Fluctuations in Normal traffic: day to day variability
(2) Special events: special occasion such as Christmas day, Ied Mubarak, and new year

Category 3: Physical Highway features

(1) Traffic control devices: railroad grade crossing and poorly timed signals, etc.
(2) Physical bottlenecks (capacity): number of width lanes and shoulders, merge areas at interchanges, a roadway alignment

In addition to regular traffic pattern, this study also investigate irregular traffic pattern that occurred. Irregular traffic pattern, to our knowledge, still lacks of consensus on definition. Thus, for the purpose of our research, we defines irregular traffic pattern as anomalies from regular pattern. Anomalies could cause by unpredicted or extraordinary event. This research is aimed to explore what caused anomalies, and what pattern could arise from spatial-temporal data.

2 RESEARCH METHODOLOGY

2.1 Data

Dataset is taken from Waze report data from August 2016 to February 2017. The information contains the information of Date, Time (aggregated in 15 minutes interval), Geolocation (aggregated in 100 meters) and report type and number of report as described in Table 1.

Table 1: Data Description

Column Name	Type	Sample
Tanggal	Date (dd/mm/yy)	05/08/2016
Jam	Time (hh:mm)	13:15
Type	String	ROAD_CLOSED
Sub-Type	String	ROAD_CLOSED_EVENT
Lat	Double	-6.189
Lon	Double	106.767
Total	Integer	987

There are four main types of report such as `Jam`, `WeatherHazard`, `Accident`, and `Road closed`. Each type of report will have its own sub-type, for example, `Jam` has sub-types of `JAM_MODERATE_TRAFFIC`, `JAM_HEAVY_TRAFFIC`, and `JAM_STAND_STILL_TRAFFIC`.

2.2 Methodology

Dataset for this analysis featured in 2 characteristics, spatial and temporal. Thus, we separate our methodology to study traffic pattern based on that characteristics. For time-series analysis, we aggregate the location and assumed all Jakarta region have same feature, this data provide us with behavior over time pattern. Seasonal Hybrid ESD (S-H-ESD) and Auto-regressive Integrated Moving Average (ARIMA) were employed to detect anomalies and extract univariate seasonal data to develop prediction model. On the hand, for spatial analysis, we aggregate dataset overtime and generated incident report per location. We use R data analysis software for all of our quantitative analysis in this study [5].

Figure 1: Methodology

2.2.1 Seasonal Hybrid Extreme Studentized Deviate (S-H-ESD). S-H-ESD is a development from ESD that able to detect global and local anomalies. This could be accomplished by conducting time series data analysis by using decomposition and statistical measurement such as median and mean. This method was developed to overcome S-ESD method drawback in large anomalies detection. We employ this method in our analysis assuming traffic pattern have large number of irregularities. This method initially developed for analyzing Twitter anomalies [2].

2.2.2 Descriptive Spatial Pattern. Spatial pattern from Waze incident report gave comprehensive description of what traffic pattern is emerging, and what the antecedents is. We employ this analysis based on anomaly and non-anomaly event to investigate causal and event that may cause the anomaly.

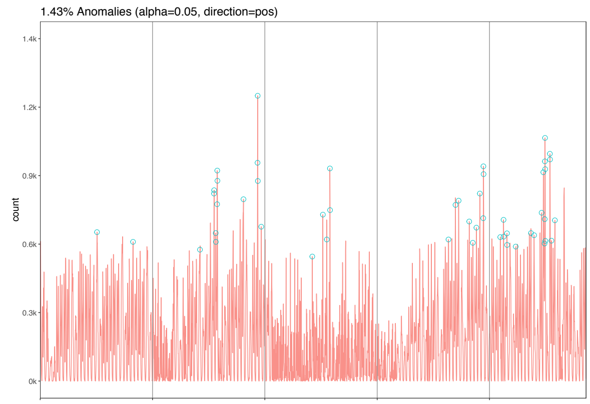


Figure 2: Anomaly Detection

2.2.3 Autoregressive Integrated Moving Average (ARIMA). ARIMA is specified by these three order parameters: (p, d, q). The process of fitting an ARIMA model is sometimes referred to as the Box-Jenkins method. An auto regressive component is referring to the use of past values in the regression equation for the series Y. The auto-regressive parameter p specifies the number of lags used in the model. Mostly ARIMA is used in financial analysis to analyze seasonal data, but recently it find used in other sector such as traffic time-series prediction [6].

3 ANALYSIS

3.1 Seasonal Hybrid Extreme Studentized Deviate (S-H-ESD)

We used JAM_HEAVY and JAM_STAND_STILL in our analysis because this is the two highest incident reported by Waze user. This variables were assumed to represents real condition in Jakarta congestion traffic pattern. Using S-H-ESD algorithm in JAM_HEAVY incident report, we found 1,43% anomalies that equal to 50 data points. Highest anomalies occurred in 15 November 2016, and 11 February 2017, as summarized in table 2 . Cross checking with event that occurred on that particular day, we found that in 15 November 2016 and 11 February 2017 Jakarta experienced flood. On the other hand, flood incident report from Waze were failed to capture this phenomenon. Figure 2 depicts result plot of S-H-ESD analysis.

3.2 Autoregressive Integrated Moving Average (ARIMA)

The analysis aggregated into hourly data for more simple analysis. ARIMA in this research consistent with the previous research that time series data of traffic information formed stationary pattern that enabled univariate short term prediction. ARIMA method is done using R open source data analysis software [3]. Key point of ARIMA is to generate p, d, and q parameter which produced fittest model. This combination should optimized based on several indicator such as Akaike Information Criterion (AIC), and Root Mean Squared Error (RMSE). Based on given data set, we generate optimum ARIMA model in following:

Table 2: Data Anomaly List

timestamp	Report	timestamp	Report
24-09-16 14:00	630	25-01-17 17:00	649
04-10-16 17:00	589	26-01-17 17:00	793
29-10-16 17:00	556	27-01-17 16:00	689
02-11-16 17:00	793	27-01-17 17:00	908
02-11-16 18:00	807	27-01-17 18:00	875
03-11-16 7:00	626	01-02-17 17:00	609
03-11-16 8:00	589	02-02-17 17:00	682
03-11-16 16:00	748	02-02-17 18:00	611
03-11-16 17:00	890	03-02-17 17:00	625
03-11-16 18:00	847	03-02-17 18:00	576
11-11-16 18:00	769	06-02-17 7:00	569
15-11-16 16:00	923	10-02-17 18:00	626
15-11-16 17:00	1206	11-02-17 14:00	617
15-11-16 18:00	846	13-02-17 17:00	712
16-11-16 17:00	653	14-02-17 7:00	883
05-12-16 7:00	527	14-02-17 15:00	582
08-12-16 18:00	704	14-02-17 16:00	685
09-12-16 19:00	599	14-02-17 17:00	929
10-12-16 17:00	899	14-02-17 18:00	1028
10-12-16 18:00	723	14-02-17 19:00	896
17-01-17 17:00	599	14-02-17 20:00	593
19-01-17 18:00	745	16-02-17 7:00	961
20-01-17 17:00	763	16-02-17 8:00	937
23-01-17 17:00	675	16-02-17 17:00	594
24-01-17 17:00	585	17-02-17 18:00	680

Table 3: ARIMA output

	Dependent variable:
ar1	1.725^{***}
	(0.043)
ar2	-0.885^{***}
	(0.054)
ar3	0.068^{***}
	(0.024)
ma1	-1.783^{***}
	(0.039)
ma2	0.787^{***}
	(0.039)
Observations	3,499
Log Likelihood	−21,056.290
σ^2	9,866.803
Akaike Inf. Crit.	42,124.580
Note:	*p<0.1; **p<0.05; ***p<0.01

Results in table 3 generated from `auto.arima()` function on R. This function generated most fittest ARIMA model when value of $p = 3$, $d = 1$, and $q = 2$. Using this ARIMA model, we could predict Jakarta's traffic condition based on JAM_HEAVY_TRAFFIC incident report.

3.3 Spatial Description Analysis

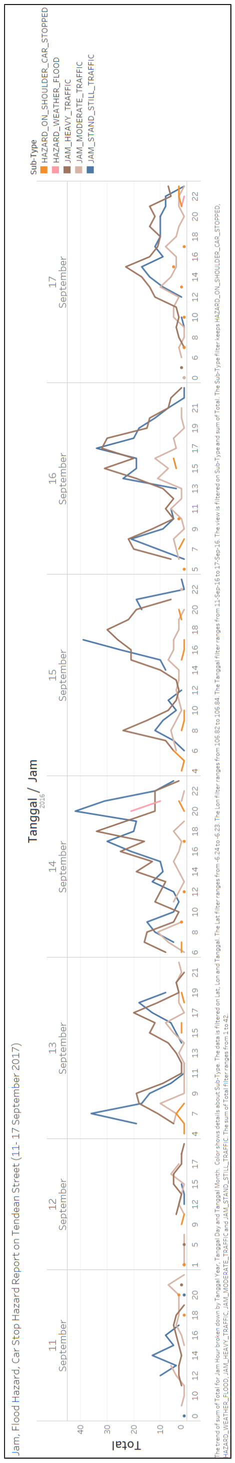
Accident. In spatial description analysis, proxies and assumptions were used to study correlation between accident report pattern to the traffic report. Framework for this analysis depicted in figure 4. Using accident-prone map, we focused on one location with high number of accident in 7 months. Then we count the number of accident incidents report and tabulated it in hourly data. This process indicated Jl. Perintis Kemerdekaan , which have a high number, to be main focus in the analysis.

We observed the peak time of the accident on focused area, then studied 1-week pattern its accident and traffic report. From the analysis we can conclude that, the accident report have spatial correlation to the traffic condition reports. Perintis Kemerdekaan Street has traffic jam regularly at 6 am till 9 am and 3 pm till 7 pm. But if there is an accident in the morning before 10 am, there will be a traffic jam all day up to midnight.

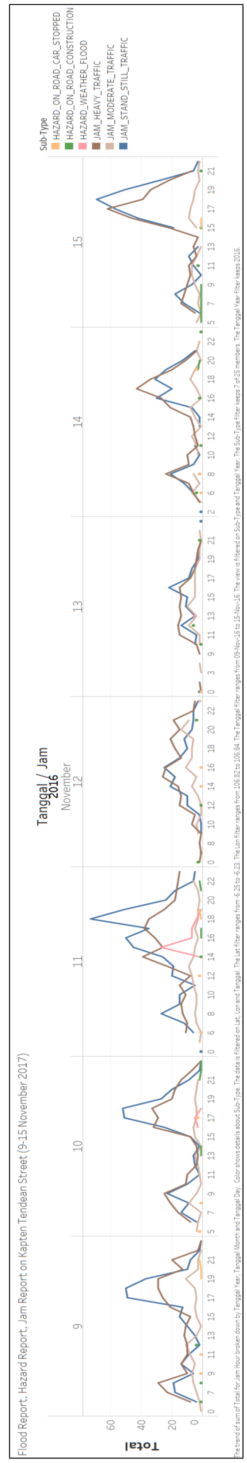
Flood. To understanding the traffic pattern from irregular traffic report, particular example also taken from the flood report. By choosing one case from flood area in specific days, it is expected to observe the correlation between the flood and congestion report. First of all, one area with high frequency and number of report are chosen from Waze data then been validated from the Jakarta flood prone map. Kapten Tendean street are selected due to the high number of people report (80) on the last 7 months. In addition, t-test is also was conducted to validate the area chosen.

After choosing the location, time series analysis is used to identify the exact day and hour when people reporting the flood hazard. It was recorded that in 7 months there are four days where people reporting the flood i.e. 14 September 2016, 24 September 2016, 4 October 2016, and 11 November 2016 with the highest number of report are in 14 September and 11 November 2016.

By knowing the exact date, exact hour of flood in Kapten Tendean street, deeper analysis between flood event and jam report were identified. On 14th September 2016, when the flood occurred at 20:00-21:00, there are significant increase for stand still traffic report. It shows that flood occurrence has impact to the jam report significantly. On the 11 November 2016 (see Figure 3), flood report was recorded on 15:00 until 19:00 followed by the high number of stand still jam report. At 16:00, report decreased then followed by the decline of jam report until 17:00 before it starts to rise again. From 17:00 until 19:00, congestion report increased again and showed the peak at 18.00, which can be identified as the impact of flood occurrence to the peak hour return from the regular traffic pattern. Both days shows the obvious impact of flood as the irregular factor in influencing regular traffic pattern. In addition, some notices to the another report of car stopped on the road occurred during the flood, which indicates that flood also causing the car stopped, which can be worsen the traffic jam. This fact also can proved that congestion sources (flood and car stop) are interconnected in affecting the congestion [4].



(a) 11-17 September 2016



(b) 9-15 November 2016

Figure 3: Volume of Hazard and Jam Report

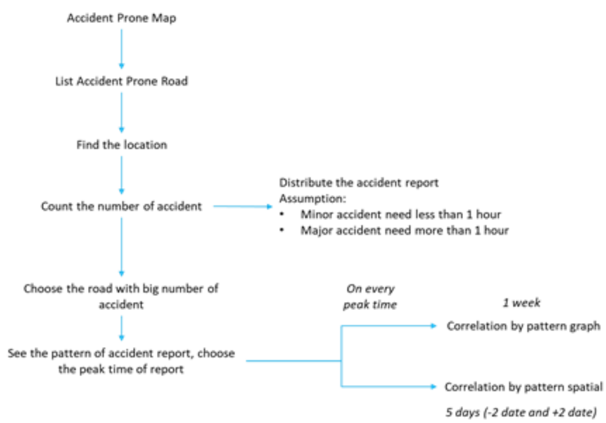


Figure 4: Framework of analysis correlation accident report to traffic report

Kapten Tendean street is a crowded street which prove by the number of people reporting the stand still and heavy traffic. 11 and 17 September 2016 (see Figure 3)are weekend and 12 September 2016 is a national holiday that cause the different pattern from the series. On the regular day, there are an increase for jam report since 13.00 until 20.00. However, on the 14th September the unusual jam report reached its peak at 20:00-21:00, higher than the other days (see Figure 3). On the 9-15 November 2016, the jam report shows the similar pattern except for the 12 and 13 November 2016 (weekend).

4 CONCLUSION

Big data analysis has open up possible attempt to study traffic pattern in unprecedented detail manner. Therefore, researcher and decision maker could gain more comprehension on this subject. In our study, we found that Jakarta traffic pattern can be analyzed in spatial and time-series by using irregular traffic incident report. Time-series model in this research could produce univariate predictive model based on Jakarta traffic condition report.

Although time-series model give insight to Jakarta traffic pattern, ARIMA failed to look into its traffic condition predictor. To understand the predictor, we used anomaly detection and spatial descriptive analysis. Anomaly detection using S-H-ESD able to detected anomaly 1.43% traffic report condition. We cross validate this anomaly with using spatial data and arrived at conclusion that flood may responsible for that condition. Nevertheless,in our dataset, flood condition report do not constantly give valid real world flood condition. In conclusion, despite increasing volume of low granularity data, validation user-generated content report with real world condition is required to ensure fruitful analysis.

REFERENCES

[1] Yuanfang Chen, Mohsen Guizani, Yan Zhang, Lei Wang, Noel Crespi, and Gyu Myoung Lee. 2017. When Traffic Flow Prediction Meets Wireless Big Data Analytics. *arXiv preprint arXiv:1709.08024* (2017).

[2] Jordan Hochenbaum, Owen S Vallis, and Arun Kejariwal. 2017. Automatic Anomaly Detection in the Cloud Via Statistical Learning. *arXiv preprint arXiv:1704.07706* (2017).

[3] Rob J Hyndman and Yeasmin Khandakar. 2008. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software* 26, 3 (2008), 1–22. http://www.jstatsoft.org/article/view/v027i03

[4] Cambridge Systematics et al. 2005. Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation. *Final Report, Texas Transportation Institute. http://ops. fhwa. dot. gov/congestion_report_04/index. htm* (2005).

[5] R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[6] Billy M Williams and Lester A Hoel. 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of Transportation Engineering* 129, 6 (2003), 664–672.

Estimating Travel Time Using Geotagged Social Media Data

Alfa Narendra
Universitas Negeri Semarang
Semarang, Indonesia
alfa.narendra@mail.unnes.ac.id

Lydia N. N. Hidayati
Universitas Sebelas Maret
Solo, Indonesia
lydia.hidayati@staff.uns.as.id

Okkie Putriani
Universitas Atma Jaya Yogyakarta
Yogyakarta, Indonesia
okkieandfriends@gmail.com

Ibnu Fauzi
Universitas Atma Jaya Yogyakarta
Yogyakarta, Indonesia
ibnu.fauzi.civil@gmail.com

Imaduddin Amin
Pulse Lab Jakarta
Jakarta, Indonesia
imaduddin.amin@un.or.id

ABSTRACT

There is possibility of using geolocated social media data for estimating travel time. The main work is developing method for the data mining thus the data is feasible to use. The geographic information which specifies the location and the time were analyzed and resulting travel time of all OD pairs as well as travel time pattern of sub-district interchange trip within the Greater Jakarta area. Data processing were conducted continued by box plot analysis to remove the outliers data before calculating the travel time. Whereas the travel time pattern visualized travel time taken to do sub-district interchange trip with the help of GIS map. The conclusion is that geolocated social media data are practically able to estimate travel time. However, uncertainty derived from the complex process of tweeting should take into consideration.

KEYWORDS

Geolocated Data, Travel Time, Sub-District Interchange Trip

1 INTRODUCTION

Travel time is necessary data in transport analysis. It is significant data for the transport planner in modelling transportation. Travel time also commonly used by the transport authority to determine the performance of transport infrastructure. Whereas for the public or transport user, travel time could help them to arrange their trip efficiently. The conventional method in collecting travel time data is by conducting manual survey. Nowadays, due to mobile technology growing and visible trend in social network application, using geolocated social media data could be a promising alternative in replacing the conventional method. Beforehand Global Positioning System (GPS) is widely used as data source to do transport analysis [1],[5], yet another research [2] stated that there is possibilities in processing and analysing the geolocated social media data to perform spatial analysis, and one of plausible result the usage of geolocated social media data to perform transport analysis can be found in journal Inferring Commuting Statistics in Greater Jakarta from Social Media Locational Information from Mobile Devices [3].

Social media data uploaded by user of social network application, such as Twitter, contain many information. Besides the post or tweet it self, this data may be accompanied by geographic information which specifies the location and the time as the tweet uploaded. By looking into the geolocated social media data of certain user during some period of time, we could get the movement of the user from certain location to another specific location and the time taken to do the trip. The location where user starts the trip will be identified as origin (O) while the later location will be the destination (D). Further, the time taken to do this trip will be calculated as the travel time of a specific OD pair. This method will provide big data of travel time with less time consume and minimum cost. However, there are some disadvantages in performing this method to calculate travel time. The main disadvantage is the lack of accuracy because we could not make sure that the user is tweeting right after arriving at the destination thus the travel time might not be precise.

This paper contributes in giving insight use of geolocated social media data for estimating travel time of all OD pairs within the Greater Jakarta area. Further by looking into the travel time from all origins to a certain destination location, we could get better understanding of trip pattern, particularly the travel time taken, to do sub-district interchange trip within the Greater Jakarta.

2 RESEARCH METHODOLOGY

2.1 Data

Data provided is a set of geolocated data from twitter application in the Greater Jakarta from January 1st, 2014 until May 30th, 2014. The sets of data contain information of users specific location and time as they are tweeting. Covering 1 million unique users with more than 38 million tweets was a big number of data needs to be processed and analyzed for estimating travel time of inter-zone trip within Greater Jakarta. The data processing and analyzing is illustrated in Figure 1.

2.2 Data Processing

Initially, the big data needs to be processed before continuing to the analysis. In the data processing, geolocated tweets which contain specific location of each user is aggregated into a specific zone. The consequences, wherever the tweets is uploaded as long as it is within the specific zone then the data belong to that respective sub-district. In this study, the zones represent 89 sub-districts within the Greater Jakarta area (Jabodetabek âĂŞ Jakarta, Bogor, Depok, Tangerang, and Bekasi). The data is aggregated into sub-district level because sub-district interchange trip give a sufficient detail in describing trip pattern within an area and it also less labor work compared to analyzing it in more detail level.

2.3 Data Analysis

2.3.1 Data Cleansing. Data analysis is performed in several steps in which the first step is data cleansing. Data cleansing aims

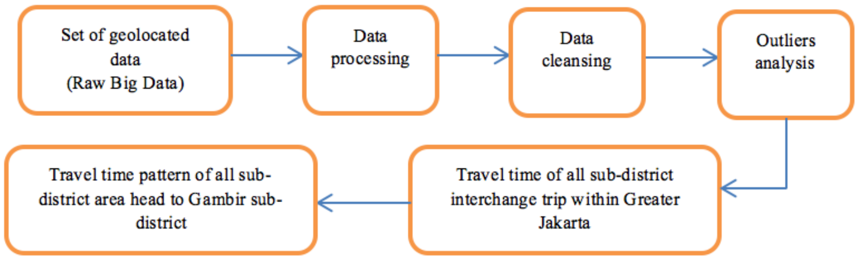
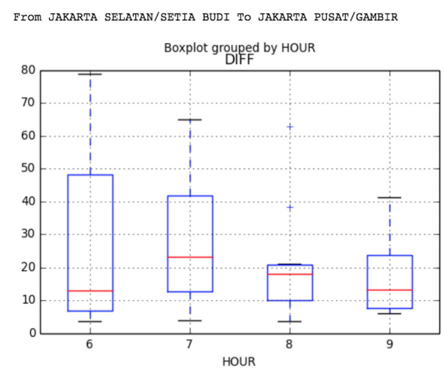


Figure 1: Flow chart of research method

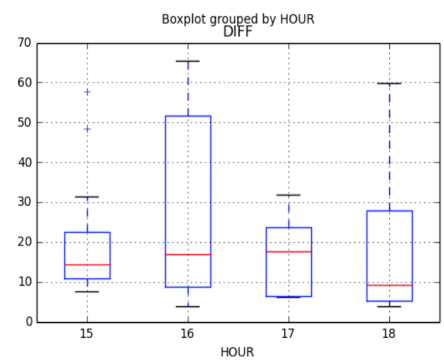
to omit the unused data for further analysis. The are several types of unused data. The first type is intra-zone movement. Intra-zone movement means trip within certain sub-district. This data is omitted because the main interest of this study is to estimate the travel time of inter-zone trip or the trip between zones. The second type of data needs to be deleted is short inter-zoned trip. It is deleted due to the trip distance which less than 1 km, thus its travel time data could not represent the average travel time of sub-district interchange trip. The last type of data cleansing is slow speed user. Slow speed user is user who travels with speed less than pedestrian speed. The threshold of pedestrian speed is referred to Indonesia Highway Capacity Manual (IHCM), 10 kilometers per hour. From the geolocated data and the time stamp, we calculate the speed of the user. If the speed is less than the threshold, then this data will be eliminated. Slow speed user also indicating that the user is not immediately tweeting after arrive in another sub-district location. Therefore, the time that calculated is not the actual travel time since there is additional time that the user has to do something else after arriving at the destination. After cleansing the data, there is only Âś 0,1 percent or around 45,000 tweets left which can be used for estimating travel time of inter-zone trip. Despite its low number compared to the number of data in the beginning, yet this number is still big if it compared to the data which can be obtained with the manual survey. The left data then classified into certain OD pair before carrying out the next stage, outliers analysis.

2.3.2 Outliers Analysis. For certain OD pair, the travel time data then divide into hourly time considering the time when the tweets uploaded. The next step is plotting the data into boxplot to examine the outliers. In this paper, we will only discuss the result of outliers analysis in the morning peak hours between 06.00-09.00 and for afternoon peak hours at 15.00-18.00.

Outliers analysis is performed by using boxplot because it provides a quick and easy visual summary of the dataset [4]. The component of boxplot can be explained as follows; flat red line in the middle of the box is median, the box represents the first quartile on top and the third quartile in the bottom. The outliers data is data which the value is bigger than 1.5 IQR, where IQR is interquartile range. Figure 2 shows the boxplot of hourly travel time data from setia budi sub-district to Gambir sub-district during morning and afternoon peak hours. Based on the median value showed by flat red line, travel time during 06.00-07.00 am has relatively low value. However, by looking into the whisker line, the range of the travel time data during that period of time is large. The larger the range means that there is a great diversity of the data.



(a)



(b)

Figure 2: Travel time of Setiabudi-Gambir OD pair during morning and afternoon peakhours

3 RESULT AND DISCUSSION

3.1 Estimating travel time of Greater Jakarta inter-zone trip

From the Figure 2, the average travel time from setia budi to gambir departed between 06.00 âĂŞ 07.00 am is around 12 minutes and the time increase to 22 minutes if the users travel between 07.00-08.00 am. In general, it can be concluded that the peak hour period

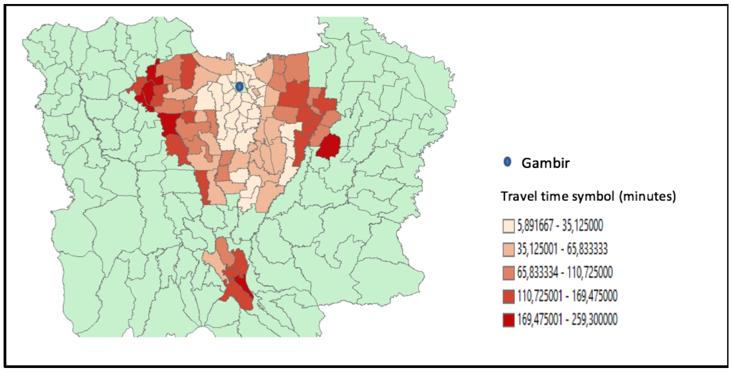


Figure 3: Travel time pattern of Greater Jakarta inter-zone trip to Gambir

happens at 07.00 - 08.00 in the morning. During peak hour period, the traffic volume is maximum, consequently, it resulted in longer travel time. Once the traffic volume is getting down, the travel time will be reduced as well. It is similar to the afternoon peak hour, at 15.00-16.00 the traffic might be in normal condition thus the travel time is quite low. When the time approaching afternoon peak hour at 17.00-18.00, the higher the travel. Then it is getting down again once the peak hour end. After verifying that the travel time is acceptable, then we assume that the average travel time of each period of time is plausible value to represent the travel time from setia budi to gambir of the respective hour.

3.2 Travel time pattern of Greater Jakarta inter-zone trip

The result of the previous stage is the travel time for all inter-zone trip within Greater Jakarta. Further by looking into the travel time from all origins to a certain destination location, we could get better understanding of trip pattern, particularly the travel time taken, to do sub-district interchange trip within the Greater Jakarta. Figure 3 illustrated travel time pattern of other sub-district within Greater Jakarta going to Gambir. Refer to the figure 3, the travel time estimated from geolocated social data seems quite accurate where the further the distance of origin sub-district to gambir the longer the travel time to go to Gambir. In addition, for the Bogor area, we only considering the Bogor city because it has a direct link to Jakarta.

3.3 Discussion

Considering the result of this study, data from the social media platform, Twitter, have high potential to use in analyzing people movement such as inferring OD matrix or estimating travel time due to its abundant data. Travel time estimated from the geolocated data might be beneficial for certain people. For example, transport planner who needs to analysis transport in macro scale or, transport authority who need to make fast decisions. It is because tweeting is tricky process. Poor connections can causes delay thus resulting bias data. Therefore, it is necessary for the user of geolocated tweets to developed more advanced data mining algorithm thus reducing the data biased.

4 CONCLUSION AND RECOMMENDATION

From this study, we can conclude that geolocated social media data are practically able to estimate travel time. However, uncertainty derived from the complex process of tweeting should take into consideration. Future research would be useful in at least three following directions. First, measurement of biased data due to the time delay. Second, improvement of data mining algorithm to increase the accuracy. The last is possibility in using geolocated data for more detail analysis, such as estimating travel time in road link or inferring link flows based on route estimation.

REFERENCES

[1] Antonio Gschwender Marcela Munizaga Mauricio ZÃžÃśiga Cristian E. CortÃĨs, Jaime Gibson. 2011. Commercial bus speed diagnosis based on GPS-monitored data. *Transportation Research Part C: Emerging Technologies* 19, 4 (2011), 695 – 707. https://doi.org/10.1016/j.trc.2010.12.008

[2] Piotr Oleksiak. 2014. Analysing and Processing of Geotagged Social Media. 3, 4 (2014), 250 –260.

[3] Setia Pramana, jonggun lee, Robert Kurniawan, Imaduddin Amin, Ni Luh Putu Satyaning P.P, Yulistina Riyadi, and Jong Gun Lee. 2017. Inferring Commuting Statistics in Greater Jakarta from Social Media Locational Information from Mobile Devices. (04 2017).

[4] Paul Teetor. 2011. *25 Recipes for Getting Started with R.* OâĂŹReilly Media, Sebastopol.

[5] Ruhua Zhang, Tiantian Wang, Hao Wan, Li Lei, and Baiying She. 2011. *Analysis of Jinan BRT Speed Characteristics Based on Vehicle Traveling Data Recorder.* https://doi.org/10.1061/41186(421)97 arXiv:https://ascelibrary.org/doi/pdf/10.1061/41186%28421%2997

Analysing Network Connectivity and Public Transportation Efficiency in Jakarta

Cahyono Susetyo
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
cahyono_s@urplan.its.ac.id

Muchammad Zaenal Muttaqin
Universitas Islam Riau
Pekanbaru, Indonesia
muchzaenalmuttaqin@eng.uir.ac.id

Muhammad Subair
Pulse Lab Jakarta
Jakarta, Indonesia
msubair@gmail.com

Yori Herwangi
Universitas Gadjah Mada
Yogyakarta, Indonesia
yherwangi@ugm.ac.id

Tri Sefrus
Universitas Gadjah Mada
Yogyakarta, Indonesia
sefrus.tri@gmail.com

ABSTRACT

The effectiveness and efficiency of public transport should be a priority for transportation in developing cities. Despite the efforts from the government to improve public transportation, in Jakarta City, Indonesia, there is still an 9.93% increase of private vehicle annually. The highest contributor to the percentage increase is motorcycles with an average annual increase of 10.54%, followed by an increase in the percentage of passenger cars by 8.75%. In contrast, the number of public transportation increased only by 1.74%. This research evaluates the performance of public transportation in Jakarta by their availability and accessibility. The results indicate that the major problems in public transportations in Jakarta City are the coverage area of the service, route connectivity, and access to public facilities. With average walking distance of 300 m, only 7.78% of Jakarta city area is covered by flexible-stop public transportation services. Furthermore, there are some areas that are not passed by public transport routes, which make about 18.5 million people live in blank spot area. Based on the route connectivity aspect, this research proposes potential location for Transit Oriented Developments (TOD) in Jakarta to improve public transportation services.

KEYWORDS

public transport performance, effectivity, efficiency, blank spot

1 INTRODUCTION

As a metropolitan, Jakarta needs good public transportation to support the mobility of its residents. According to studies [4–6], one's ability to travel is related to their opportunities to perform essential activities such as economic, social, cultural, and political activities. Therefore, the availability of good public transport is needed to reduce the use of private vehicles, which have higher negative impacts toward the environment compared to the public transportation. However, the modal share of public transportation in Jakarta has decreased, for example for buses from 38.3 in 2002 to 12.9 in 2010 [2]. Meanwhile, the increase of vehicle ownership from 2011-2015 in Jakarta is 8.75% per year, with the highest contributor is motorcycle (9.14%) followed by passenger cars (8.09%)[1] while public transport fleet (regular bus and transjakarta) only increase by 2.92%[2]

One of the causes of the reduced use of public transport in Jakarta, allegedly is the low performance of current public transport service. Despite the effort of the government to improve public transport service by initiating MRT and new route for TransJakarta BRT, evaluation of the current public transport performance is needed. Thus, the aim of this study is to evaluate the effectivity and efficiency of public transport network in Jakarta. Based on the data availability, the objectives of this paper will be focused on: 1) identifying the coverage area of existing public transport service; and 2) evaluating public transport efficiency based on the potential demand

2 METHODOLOGY AND DATA

2.1 Public Transport Performance Indicator

Public transport performance can be assessed based on its availability, accessibility, affordability, and acceptability. Availability refers to coverage area of route, frequency, operating time and information of available public transport. Accessibility concerns the distance to public transport stops, route connectivity and accessibility from stops to public facilities. Affordability refers to the financial ability of users to use public transport, including tariff and willingness to pay. Lastly, acceptability concerns about the quality of service including safety, comfort, reliability, and regularity.

2.2 Data Availability and Scope of the Study

For this study we use Jakarta public transport network which comprise of TransJakarta bus route and non TransJakarta route in geospatial format. We also use population data, public facilities and TransJakarta station location, land use data, and rail based transport network (MRT, LRT and KRL). Based on the data availability, public transport performance measurement that is conducted in this study is in term of availability (coverage of public transport service), accessibility (route efficiency, distance to public transport stop, access to public facilities), and potential TOD (Transit Oriented Development) area in Jakarta.

[1] https://jakarta.bps.go.id/publication/2017/07/25/2370fe9e77e3c19211832748/statistik-transportasi-dki-jakarta-2015-2016.html

[2] https://jakarta.bps.go.id/statictable/2017/02/23/173/jumlah-kendaraan-umum-lainnya-2005-2015.html

2.3 Public Transport Performance by GIS Software

GIS software can be used to analyze the performance of public transport by combining spatial information with other information to map the various characteristics of the transport system [3]. An example of using this method of analysis to assess the performance of public transport can be found in Currie et al. [1]. In his research Currie measured public transport service and combine it with social and economic data to address the transport "needs-gap" in Melbourne. In this research, two GIS methods were utilized to evaluate the performance of public transportation in Jakarta. The first one is the neighborhood analysis, which generates the coverage area of existing public transportation networks. The second one is the proximity analysis, which identifies overlapping routes of the transportation fleets.

3 RESULT AND DISCSUSSION

3.1 Profile of Public Transport in Jakarta

The development of public transportation in Jakarta has been going on for decades. Several solutions and approaches have been tried, ranging from traditional transportation modes to tram to overcome the transportation problems in Jakarta. Currently, there are evaluations of several types of public transport as well as some options of public transportation alternatives in order to cope with the traffic jam in Jakarta. According their characteristics, the current public transportation in Jakarta can be grouped into two types, namely:

(1) Trans Jakarta Busway began to operate since 2003. It is a mass-transit in which passengers can only get on and off the bus at a designated stop. In this research, this type of public transport is called public transport with fixed stops.
(2) Metromini, angkot, city bus, are types of public transportation in Jakarta which also have designated stops, but are rarely used in practice. Rather, the passenger can get on and off the bus anywhere along the route. Therefore, in this study, this types of public transport are called public transport with flexible stops.

In this study, the entire public transport line of both with fixed stops and flexible stops are evaluated for its performance, especially related to its spatial parameters, such as coverage area and blank spot area.

3.2 Coverage Area and Public Transport Index of Existing Public Transport Service

Coverage area of existing public transport is measured in two ways:

- For public transport with flexible stops (metromini, angkot) by creating a buffer of 300 meter from the street covered by public transport service
- For TransJakarta by creating a buffer of 300 meter, 500 meter, and 1 kilometer from the bus stops.

The area outside the coverage area is identified as blank spot or area without public transport service. For public transportation with flexible stops, the coverage area is about 69%, with the coverage for each of the districts within Jakarta shown in Figure below.

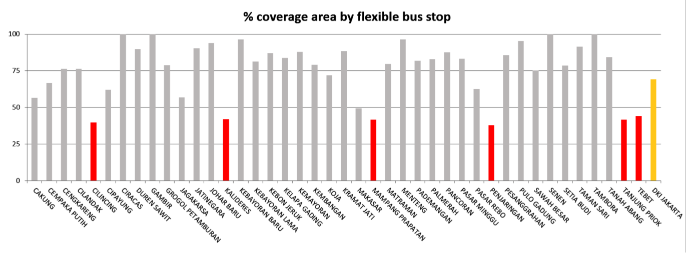


Figure 1: Coverage Area of Public Transportation by Districts

Table 1: Coverage Area of TransJakarta

Assumed walking distance	Coverage area (km^2)	Percentage from total area
300 m	51	7.78
500 m	112	17.08
1000 m	245	37.35

For public transportation service with flexible stops, districts with the lowest public transportation coverage are Cilandak, Johar Baru, Mampang Prapatan, Penjaringan, Tanjung Priok, and Tebet. Based on the coverage area, the size of the districts, and number of routes in each districts, the public transportation index in Jakarta is shown in the figure below.

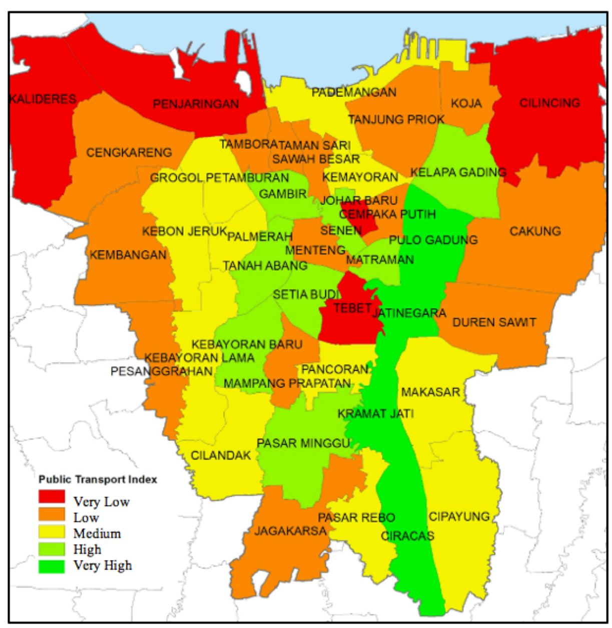


Figure 2: Public Transportation Index in Jakarta

For public transportation service with fixed stops (TransJakarta), the coverage area can be seen in the table below.

3.3 Public Transport Efficiency Based on the Potential Demand and Service Area

In areas that are not covered by public transportation because the walking distance is farther than the accepted standard (300 meters), we evaluate the potential demand on those areas by two indicators. The first one is by population density, and the second one is by population within working age. Based from analysis, there are approximately 3.3 million population live in blank spot area. Figure 3 showed the blank spot area for public transport system in Jakarta related to population density. There are regions that are detected as blank spot area with high density population, that could be considered as potential market for alternative modes of transport, such as paratransit. For instance, Kemayoran and Matraman detected in red colors, which means that region was a high density area in Jakarta, but there are many blank spot located in that region. Public transport system could not reach many areas in these districts because of the narrow path of the residential areas. Therefore, it is assumed that there are a lot of potential demand for on-demand micro-transit modes in this area.

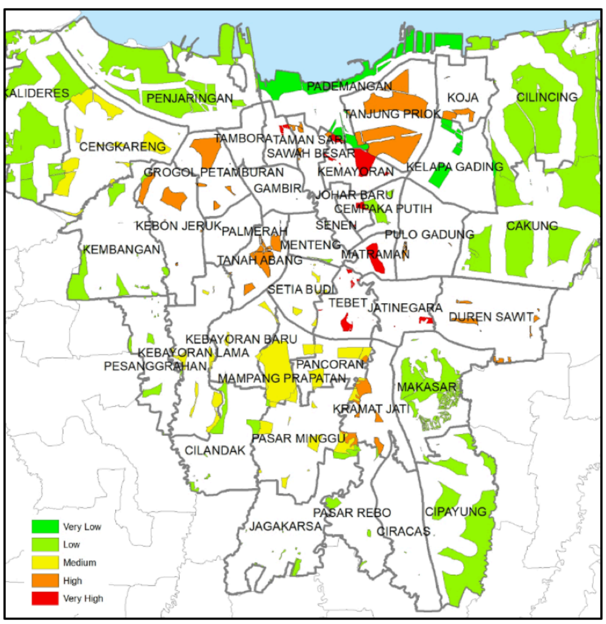


Figure 3: Population Density in Blank Spot Area

Figure 4 showed relationship between blank spot and workforce population. There are more people in their productive age (15 - 64 years old) cannot reached public transport in their areas. However, people in this age group has more commuting activity than other groups. If public transport system could not catch this demand, then other alternatives modes like paratransit (ojek, taxi), or private modes would be a solution to solve their daily transport problem, which is not a desired solution.

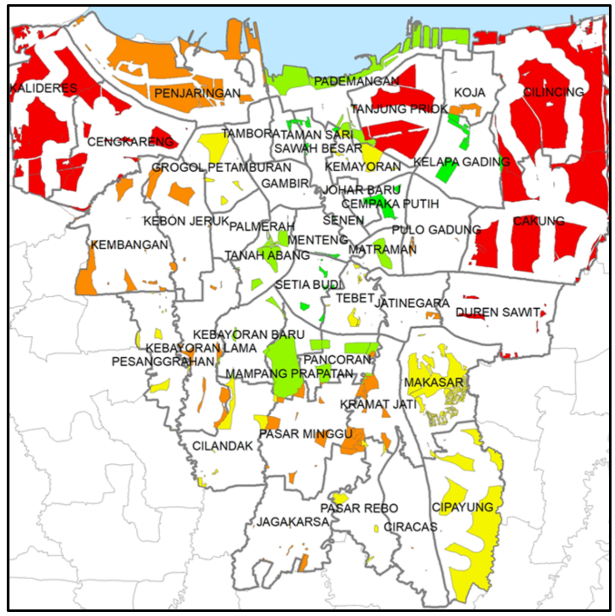


Figure 4: Workforce in Blank Spot Area

3.4 Transit Oriented Development (TOD) Area based on Public Transport Route

The efficiency of public transportation can also be evaluated by the number of overlapping service on the same road segments. This issue is more problematic in Jakarta, because there are different types of vehicle for different route services. Thus, slower vehicles may hinder the faster ones. Overlapping of different services within the same road segment can be seen in the figure below.

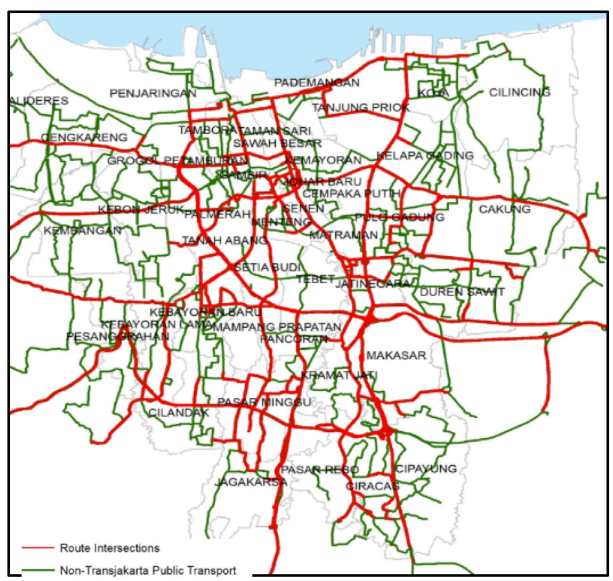


Figure 5: Overlapping Routes

TOD is a type of development where mixed of land uses are built around transportation hub. In this research, potential TOD areas are identified from intersection between TransJakarta routes, MRT planning routes, and also LRT routes. Our proposed locations for TOD can be seen in the figure 6. below. These areas are potential to be developed for new TOD centers based on its connectivity.

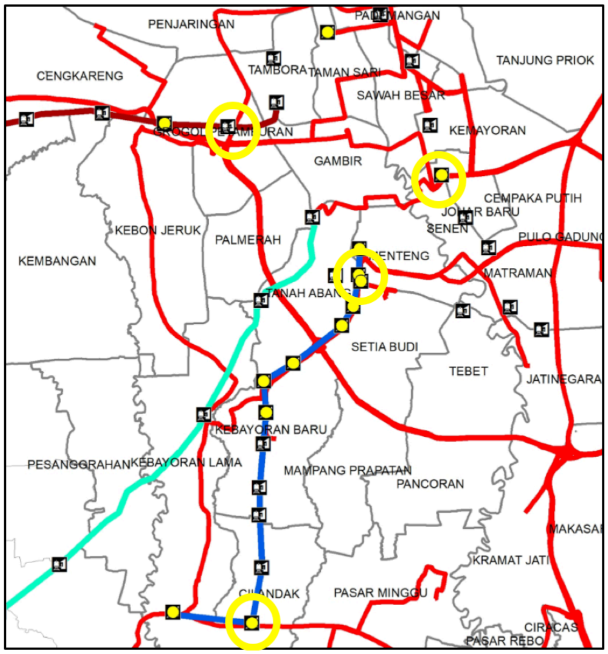


Figure 6: Proposed TOD Locations

Transit-Oriented Development (TOD), is a suitable solution to restrict urban sprawl and stimulate sustainable travel modes for residents around the areas. However, TOD concept is not easy to implement at every location. In high-density city centers a TOD is relatively easy to implement, since density and diversity are already high and most residents have a positive stance toward car alternatives due to self-selection processes [7]. Based from analysis, there are chances to make TOD (Transit Oriented Development) area from the intersection area between Trans Jakarta (current) and MRT station (future) route construction.

- North = Taman Sari District
- Centre = Menteng and Setiabudi District
- West = Grogol Petamburan and Kebon Jeruk District
- East = Kemayoran and Johar Baru District
- South = Cilandak District

The potential TOD areas identified in this study are mostly the centers of community activities in Jakarta, such as Menteng and Setiabudi. So that there are lots of facilities served as the origin and destination of the residents' journey such as housing, social facilities, and commercial center. Therefore, establishing these areas as TOD will encourage people to travel by public transport, especially for a long-distance journey.

4 RESULT

Based on the analysis, it can be concluded:

(1) There are many public transport blank spot areas in Jakarta. Based on the coverage area, TransJakarta has been serving 7.78% of Jakarta area (by using 300 meters as an assumption for walking distance), 17.08% of Jakarta area (assumed walking distance 500 meters), and 37.35% of Jakarta area (assumed walking distance 1,000 meters), while the coverage area for flexible stop transport service is 69% (assumed walking distance 300 meters). So that, there is still ineffectiveness of public transport service in Jakarta.

(2) In term of availability of public transport, there are still many areas with high density population and working age population have not yet covered by existing public transport service.

(3) Approximately 3.3 million people in Jakarta live in blank spot areas.

(4) There are still many important public facilities (in this study we assessed only college and university due to data limitation) have not covered by public transport service.

(5) There are chances to build TOD concept in Jakarta. In the intersection between Trans Jakarta route and future planning route like MRT and LRT, there are regions that are potential to build as TOD area.

5 LIMITATION AND FUTURE RESEARCH

This study has limitations by availability of the data, specifically the lack of service performance data (frequency, load factor, waiting time, in-vehicle time), users'data, land use and location of public facilities data to conduct a comprehensive analysis. There is also an inconsistency of the data from multiple sources. In the future we hope to get these data for further research to see whether the on-demand transit micro services (such as gojek, grab, uber etc.) capture captive or choice users, which groups are using the most, and how much they spend on transit services.

REFERENCES

[1] Graham Currie, Tony Richardson, Paul Smyth, Dianne Vella-Brodrick, Julian Hine, Karen Lucas, Janet Stanley, Jenny Morris, Ray Kinnear, and John Stanley. 2009. Investigating links between transport disadvantage, social exclusion and well-being in MelbourneâĂŤPreliminary results. *Transport Policy* 16, 3 (2009), 97 – 105. https://doi.org/10.1016/j.tranpol.2009.02.002

[2] J.A. Gomez-Ibanez. [n. d.]. Transportation Planning in Rapidly Growing Cities: Lessons from Ho Chi Minh City, Jakarta and Mumbai, Melvin Webber Memorial Lecture. ([n. d.]).

[3] Neil Sipe Jago Dodson and Brendan Gleeson. 2004. *Transport Disadvantage and Social Status: Review of Literature and Methods.* Technical Report. Urban Policy Program, Griffith University. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.531.3606&rep=rep1&type=pdf

[4] Karen Lucas. 2012. Transport and Social Exclusion: Where Are We Now? (2012).

[5] John Preston and Fiona RajÃľ. 2007. Accessibility, mobility and transport-related social exclusion. *Journal of Transport Geography* 15, 3 (2007), 151 – 160. https://doi.org/10.1016/j.jtrangeo.2006.05.002

[6] Sebastian Ureta. 2008. To Move or Not to Move? Social Exclusion, Accessibility and Daily Mobility among the LowâĂŘincome Population in Santiago, Chile. *Mobilities* 3, 2 (2008), 269–289. https://doi.org/10.1080/17450100802095338 arXiv:https://doi.org/10.1080/17450100802095338

[7] Jonas De Vos, Veronique Van Acker, and Frank Witlox. 2014. The influence of attitudes on Transit-Oriented Development: An explorative analysis. *Transport Policy* 35, Supplement C (2014), 326 – 329. https://doi.org/10.1016/j.tranpol.2014.04.004

http://rd.pulselabjakarta.id/

Australian Government

PULSE LAB JAKARTA

Pulse Lab Jakarta is grateful for the generous support from
the Government of Australia