



# JuTISI

## Jurnal Teknik Informatika dan Sistem Informasi

---

**Perancangan Desain Antarmuka Website Sekolah Menengah Atas Menggunakan Metode Goals Directed Design**  
Aldito Doni Pasha, Ayouvi Poerna Wardhanie, Endra Rahmawati

**Visualisasi Skyline Query untuk Distribusi Tenaga Kesehatan COVID-19**  
Vega Purwayoga, Muhammad Al Husaini, Hen Hen Lukmana

**Pengembangan Aplikasi Maranatha digital signature**  
Bernard Renaldy Suteja, Mewati Ayub, Kafka Febianto Agiharta

**Analisis Faktor Promosi Online Instagram yang Memengaruhi Keputusan Pembelian Kuliner**  
Doro Edi, Yenni Merlin Djajalaksana, Julianti Kasih

**Penerapan Phase E TOGAF 9.1 pada Rencana Migrasi Sistem**  
Diana Trivena Yulianti, Adelia, Tiur Gantini, Gilberth Giandy Taruna Kesumah

**Penerapan Balance Scorecard pada Perguruan Tinggi di Indonesia: Tinjauan Literatur Sistematis**  
Latifha Aini, Salma Riyanti Hanifah, Fitroh

**Pengembangan Prototype Aplikasi Bimbingan Tugas Akhir Mahasiswa Berbasis Web**  
Meliana Christianti Johan, Erico Darmawan Handoyo, Sulaeman Santoso

**Pengembangan Sistem dan Dokumentasi Kelahiran Bayi**  
Daniel Jahja Surjawan, Sulaeman Santoso, Erico Darmawan Handoyo

**Perancangan Enterprise Architecture Menggunakan The Open Group Architecture Framework  
Architecture Development Method**  
Hendri Irawan, Nurul Mutiah, Ferdy Febriyanto

**Analisis Deret Waktu dari Produk yang Terjual Menggunakan Beberapa Teknik Populer**  
Laras Ervintyana Deni Kartika Sari, Andreas Widjaja, Swat Lie Liliawati

**Rancang Bangun Sistem Pembayaran Kursus Mandarin Dengan Metode Waterfall**  
Teddy Marcus Zakaria, Julianti Kasih, Natanael Halim, Michael Sebastian Gunadi

**Perbandingan Akurasi Model Pembelajaran Mesin untuk Prediksi Seleksi Masuk Perguruan Tinggi Negeri**  
Oktavianus Yopi Wardana, Mewati Ayub, Andreas Widjaja

**Ekstraksi Perilaku Komuter Pada Commuter Line Menggunakan Rule-Based Machine Learning**  
Albertus Indarko Wiyogo, Setia Budi, Hapnes Toba

---



9 772443 221007



9 772443 222004

p-ISSN: 2443-2210

e-ISSN: 2443-2229

Halaman 001-166

# Ekstraksi Perilaku Komuter Pada Commuter Line Menggunakan *Rule-Based Machine Learning*

<http://dx.doi.org/10.28932/jutisi.v9i1.6133>

Riwayat Artikel

Received: 27 Januari 2023 | Final Revision: 30 Januari 2023 | Accepted: 06 Februari 2023

Creative Commons License 4.0 (CC BY – NC)



Albertus Indarko Wiyogo <sup>✉</sup>#1, Setia Budi\*<sup>2</sup>, Hapnes Toba\*<sup>3</sup>

<sup>#</sup>Sub Directorate of Information Technology, PT Kereta Api Indonesia (Persero)

Jl. Perintis Kemerdekaan, No 1, Bandung, 40117, Indonesia

<sup>1</sup>indarko@gmail.com

<sup>\*</sup>Magister Ilmu Komputer, Universitas Kristen Maranatha

Jl. Prof. drg. Surya Sumantri, MPH No 65, Bandung, 40164, Indonesia

<sup>2</sup>setia.budi@it.maranatha.edu

<sup>3</sup>hapnes.toba@it.maranatha.edu

<sup>✉</sup>Corresponding author: indarko@gmail.com

**Abstrak** — Penerapan *Automatic Fare Collection* (AFC) pada KA Commuter Line dapat memberikan pengetahuan baru dalam melakukan navigasi antar jalur KA Commuter Line dengan data perjalanan Commuter Line secara real. Sistem AFC memungkinkan manajemen untuk memperoleh data rinci dalam jumlah besar mengenai rute setiap komuter setiap hari. Salah satu tantangan yang dihadapi dalam menggunakan big data di AFC adalah ekstraksi data perilaku penumpang angkutan. Perilaku penumpang Commuter Line merupakan faktor yang sangat penting bagi operator untuk mengambil keputusan yang tepat. Penelitian ini menggunakan metode *association rules* untuk mengekstrak data AFC guna menghasilkan informasi yang baik dan memahami perilaku komuter Jabodetabek. Hasil penelitian menunjukkan bahwa metode *association rules* dapat mengekstraksi data AFC dan menghasilkan aturan asosiasi yang kuat pada perilaku komuter.

**Kata kunci**— Association Rules; AFC; Machine Learning; Perilaku Komuter.

## *Extraction of Commuter Behaviour on Commuter Line Using Rule-Based Machine Learning*

**Abstract** — The application of *Automatic Fare Collection* (AFC) on Commuter Line trains can provide new knowledge in navigating between Commuter Line train lines and real commuter travel data. The AFC system allows management to obtain large amounts of detailed data regarding the routes of each commuter daily. One of the challenges faced in using big data at AFC is the extraction of data on the behaviour of transporting passengers. Commuter Line passenger behaviour is a very important factor for operators to make the right decision. This study uses the *association rules* method to extract AFC data to produce good information and understand Jabodetabek commuter behaviour. The results showed that the *association rules* method could extract AFC data and produce strong *association rules* on commuter behaviour.

**Keywords**— Association Rules; AFC; Commuter Behaviour; Machine Learning.

## I. PENDAHULUAN

Penerapan *Automatic Fare Collection* (AFC) pada kereta Commuter Line dapat memberikan pengetahuan baru dalam mengeksplorasi korelasi antara jalur kereta dan data perjalanan komuter yang real [1]. Sistem AFC memungkinkan management mendapatkan data terperinci dalam jumlah besar mengenai rute setiap komuter setiap hari. Management kereta Commuter Line dapat menganalisis data tersebut untuk mendapatkan informasi tentang perilaku perjalanan dan perpindahan komuter dalam jumlah besar yang memudahkan operator kereta Commuter Line melakukan evaluasi kualitas layanan dan mengoptimalkan strategi operasional kereta commuter [2], [3].

Beberapa metode digunakan untuk mengekstraksi data dari AFC, Jiang et al [4] mengusulkan metode untuk mengekstraksi data *time series* aliran penumpang dari AFC dengan menggunakan algoritma *agglomerate hierarchical clustering*. Chen et al [5] mengembangkan metode untuk mengekstrak informasi *boarding* dan *alighting stop* dengan menggunakan catatan transaksi yang tersedia berupa informasi rute dasar, waktu transaksi pada bus di Guangzhou, Cina. Alsger et al [6] mengusulkan metode untuk menyimpulkan tujuan perjalanan penumpang dari ekstraksi data smart card dengan mengintegrasikan beberapa atribut dari database.

Namun masalah utama yang dihadapi adalah apakah operator dapat memperoleh pemahaman yang lebih baik dari perilaku komuter dalam jaringan layanan kereta Commuter Line dari *data mining* AFC. Beberapa metode *data mining* digunakan dalam mengukur interaksi antar factor [7], [8]. Metode *data mining* yang sering digunakan dalam permasalahan ini antara lain, klasifikasi, klustering, *decision tree* dan *association rules*. Metode *association rules* umum digunakan untuk menganalisis data dengan tujuan mendapatkan pengetahuan baru, dan telah terbukti dapat mencari data dengan pola *if-then* untuk mengidentifikasi asosiasi yang paling penting [1]. Suchaka mengidentifikasi fitur yang menunjukkan kemungkinan besar customer toko buku online melakukan pembelian dengan menerapkan *association rules* [9]. Lu [10] mengusulkan metodologi pendekatan ketahanan jaringan metro shanghai dibawah insiden operasional sehari hari dengan mengintegrasikan karakteristik topologi jaringan dan volume penumpang. Zhao et al [11] mengembangkan metode *association rules* untuk mencocokkan data metro dan sepeda umum dari satu orang yang sama dalam jaringan sepeda-bus publik.

Penelitian ini menggunakan metode *rule-based machine learning* untuk menganalisis perilaku komuter jabodetabek dalam hal karakteristik distribusi *spatio-temporal*. Data AFC yang digunakan adalah data AFC kereta Commuter Line Jabodetabek, yang mana baru pertama kali digunakan. Penelitian ini bertujuan untuk memperoleh informasi mengenai perilaku perjalanan komuter pada Commuter Line secara umum. Hasil penelitian menunjukkan bahwa studi tentang optimalisasi operasional layanan pada operator transportasi publik telah banyak dilakukan, menurut Guo et al [12] jadwal perjalanan pada beberapa jalur yang tersinkronisasi satu sama lain dapat meningkatkan kualitas layanan secara signifikan dengan menjamin adanya perpindahan penumpang yang lancar serta dapat mengurangi waktu tempuh perjalanan. Strogatz [13] menegaskan bahwa dengan pemahaman terhadap struktur dan dinamika dalam sebuah jaringan adalah hal yang sangat kritis dalam menjaga ketangguhan dan kehandalan sebuah sistem. Jelas, sangat penting bagi operator transportasi publik untuk memahami korelasi jaringan di antara jalur yang berbeda di seluruh titik stasiun transit di mana penumpang melakukan perpindahan jalur dalam melakukan perjalanannya. Jika penumpang dengan stasiun asal-tujuan diidentifikasi melakukan sejumlah besar perpindahan jalur (transit), maka operator harus memprioritaskan konektivitas jalur ini untuk memastikan transit yang tersinkronisasi dan terkoordinasi di antara mereka. Sebagian besar penelitian sebelumnya mengidentifikasi jalur yang mempunyai korelasi tinggi hanya berdasarkan prediksi teoritis dan asumsi perilaku. Namun, tidak satu pun dari hasil penelitian tersebut yang divalidasi oleh data nyata [13].

Implementasi *Automatic Fare Collection System* khususnya pada operator transportasi publik saat ini sudah menjadi hal yang umum dan menghasilkan data yang sangat berlimpah setiap harinya. Banyak penelitian sebelumnya berusaha untuk mendapatkan informasi yang berharga berdasarkan data AFC di atas. Ma et al [14] menganalisis pola perjalanan penumpang menggunakan data *smart card*, sedangkan Langlois et al [15] berhasil mengumpulkan 11 cluster penumpang dengan menggunakan 4 bulan data AFC untuk mengidentifikasikan keberagamannya. Beberapa metode *data mining* sudah menjadi bahan penelitian sebelumnya seperti *decision tree*, *clustering*, *classification tree*, *association rules* dan lain-lain. Di antara metode di atas, teknik *association rules* adalah yang biasa digunakan untuk menganalisis data untuk memperoleh pengetahuan baru, yang bermanfaat dalam mencari data dengan pola *if-then* untuk mengidentifikasi asosiasi yang paling penting. Metode ini dapat mengidentifikasikan *interesting rules*, *strong rules*, dan *frequent itemset* yang dapat membantu proses pengambilan keputusan [16]. Lu [10] mengusulkan metodologi yang mendemonstrasikan pendekatan yang sangat baik menggunakan metode ini pada jaringan Shanghai Metro pada saat terjadi insiden operasional harian. Pendekatan asosiasi pada data AFC dapat menggambarkan klasifikasi penumpang komuter yang mendominasi sebagian besar pengguna transportasi publik. Klasifikasi dimaksud dapat membedakan populasi penumpang dengan pola karakteristik *spatio-temporal* tertentu, dimana berdasarkan statistik pada Beijing AFC terdapat rata-rata 65% penumpang yang melakukan transit dalam melakukan perjalanan menuju tempat tujuannya [1], jumlah yang sangat signifikan dan berperan penting dalam menambah kompleksitas perencanaan kapasitas stasiun transit dan penjadwalan armada transportasi publik.

Penelitian ini bertujuan mengaplikasikan metode *rule-based machine learning* untuk memperoleh informasi mengenai perilaku perjalanan komuter pada Commuter Line secara umum. Yang dimaksud dengan perilaku komuter antara lain pola

trip chain, karakteristik distribusi *spatio-temporal*, aturan asosiasi yang menggambarkan perilaku transit antar jalur yang ditempuh oleh komuter dalam melakukan perjalanannya.

## II. METODE PENELITIAN

### A. Association Rules.

Metode *Association Rules* adalah sebuah metode dalam *data mining* yang sangat efektif untuk menemukan hubungan antara variabel dalam semua database yang sangat besar. *Association Rules mining* adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi item. Analisis asosiasi dikenal juga sebagai salah satu teknik *data mining* yang menjadi dasar dari salah satu teknik *data mining* lainnya. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien yaitu analisis pola frekuensi tinggi (*frequent pattern mining*). *Association Rules* dapat didefinisikan sebagai berikut:

$$X \Rightarrow Y, \text{ dimana } X, Y \subseteq I, \text{ dan } X \cap Y = \emptyset \quad (1)$$

Dimana

$I = \{i_1, i_2, i_3, \dots, i_n\}$  adalah *item set*

$X$  = disebut dengan *antecedent* atau LHS (*Left Hand Side*)

$Y$  = disebut dengan *consequent* atau RHS (*Right Hand Side*)

*Association Rules* adalah aturan yang harus memenuhi batas *minimum support* dan *minimum confident*. *Support* dari suatu *Association Rules* adalah presentasi kombinasi *item* tersebut dalam data, dimana jika mempunyai *item X* dan *item Y* maka *support* adalah proporsi dari transaksi dalam data yang mengandung  $X$  dan  $Y$ . Perhitungan nilai *support* dari dua *item* tersebut adalah sebagai berikut:

$$\text{Support}(X, Y) = P(X \cap Y)$$

$$\text{Support}(X, Y) = (\Sigma \text{transaksi mengandung } X \text{ dan } Y) / (\Sigma \text{transaksi}) \quad (2)$$

*Confident* dari *association rules* adalah ukuran ketepatan suatu *rule*, yaitu presentasi transaksi dalam data yang mengandung  $X$  dan mengandung  $Y$ . Dengan adanya *confident* kita dapat mengukur kuatnya hubungan antar *item* dalam *association rules*. Perhitungan nilai *confident* dari dua *item* tersebut adalah sebagai berikut :

$$\text{Confident}(X, Y) = P(X \rightarrow Y)$$

$$\text{Confident}(X, Y) = (\Sigma \text{transaksi mengandung } X \text{ dan } Y) / (\Sigma \text{transaksi mengandung } Y) \quad (3)$$

### B. Data Collection.

Data yang digunakan pada penelitian ini berasal dari data *Automated Fare Collection* PT Kereta Commuter Indonesia atau biasa disebut KRL Commuter Line dalam kurun waktu Januari hingga Maret 2020 yang menggambarkan dengan lengkap pergerakan penumpang area Jakarta Bogor Depok Tangerang Bekasi Banten pada saat pra-pandemi Covid-19. Selama kurun waktu yang dimaksud Commuter Line mengoperasikan 1057 perjalanan rangkaian kereta setiap harinya dan sistem AFC mencatat lebih dari 57 juta penumpang melakukan perjalanan dengan menggunakan *unregistered smart card* sebagai alat pembayaran di gate masuk dan keluar stasiun. AFC akan mencatat data stasiun asal dan stasiun tujuan, *timestamp in* dan *timestamp out*, id card, card type seperti terlihat pada Tabel 1.

TABEL 1  
SAMPEL RAW DATA AFC

	c_station_in	d_gate_in	c_station_out	d_gate_out	i_card_type	c_card	c_status	c_desc
0	UP	03/02/20 00.00	CTA	03/02/20 00.24	10	1003111602179730	S	N
1	DP	03/02/20 00.00	BOO	03/02/20 00.35	10	1006011802018630	S	N
2	BKS	03/02/20 00.00	BKS	03/02/20 00.00	5	8006081801000050	S	NM
3	BOO	03/02/20 00.00	BOO	03/02/20 00.04	5	8006081801000050	S	NM
4	BKS	03/02/20 00.00	BKS	03/02/20 00.00	5	8006081801000050	S	NM
...	...	...	...	...	...	...	...	...
5024111	GDD	07/02/20 23.59	BKS	08/02/20 00.45	10	1003011802179060	S	N

5024112 rows  $\times$  8 columns

### C. Data Preparation.

Langkah ini bertujuan untuk memeriksa dan memverifikasi data AFC sehingga didapatkan data penumpang Commuter Line dengan format yang sudah sesuai dengan standar dan kebutuhan dalam penelitian ini. Proses ini perlu dilakukan karena pada proses *data mining*, semua data yang disimpan dalam AFC akan digunakan dan dalam *raw data* AFC masih ditemukan data dengan format yang salah, *record* kosong dan berbagai kesalahan lainnya.

1) *Memuat raw data dan data trim*: Raw data AFC pada Tabel 1 merupakan data yang cukup besar dengan ukuran file 35GB. Format data yang digunakan pada data AFC adalah format CSV (comma-separated values). Untuk mendapatkan data sesuai kebutuhan diperlukan pemilihan data sesuai dengan field atau kolom yang dibutuhkan. Pada proses memuat data mentah sekaligus dilakukan cleansing atau pembersihan data sesuai tipe datanya. Untuk data dengan tipe datetime dilakukan pembersihan dengan proses trim dan pengecekan format data sesuai dengan format ISO 8601. Jika data tipe datetime tidak sesuai dengan format ISO 8601, maka baris data yang bersesuaian akan dibuang. Untuk kolom dengan tipe INT dilakukan trim dan pengecekan konversi data dari text ke bilangan integer. Jika data tidak dapat dikonversi ke integer maka baris yang bersesuaian akan dibuang.

2) *Pengambilan data yang diperlukan dan proses masking*: Data yang diperlukan dalam penelitian ini adalah data penumpang pada hari kerja dalam satu minggu dan penumpang dari stasiun pemberhentian kereta Commuter Jabodetabek. Pemilihan data dilakukan dengan proses filter dari data mentah sesuai dengan tanggal tersebut pada kolom *d\_gate\_in* dan *d\_gate\_out*. Pada data AFC terdapat data yang bersifat rahasia atau sensitif berkaitan keamanan, sehingga informasinya perlu disembunyikan atau disamarkan. Untuk menjaga kerahasiaan data dilakukan proses masking atau menyembunyikan data penumpang. Data yang disembunyikan adalah data nomor kartu penumpang.

3) *Filter transaksi penumpang*: Data AFC merupakan data transaksi tap gate pada stasiun yang melayani pemberhentian kereta Commuter Line. Untuk itu perlu dilakukan pemeriksaan pada kode stasiun dimana data tersebut merupakan data dari stasiun yang melayani perjalanan kereta Commuter Line. Pemeriksaan dilakukan dengan mencocokkan data kode stasiun dengan data stasiun pemberhentian kereta dari data GTFS KCI. Selain itu, gate tidak hanya digunakan oleh penumpang, tetapi juga digunakan oleh petugas stasiun atau personel lain yang mempunyai kepentingan khusus di stasiun. Untuk itu perlu dipisahkan antara data penumpang dengan data yang bukan penumpang. Pada proses ini dilakukan pengecekan data transaksi penumpang. Data transaksi penumpang harus memenuhi syarat penumpang dan tidak memiliki konflik internal. Filter transaksi dilakukan untuk mengidentifikasi data transaksi yang bukan data penumpang komuter, seperti data yang tidak ada stasiun tujuan, data yang tidak ada nomor kartu, data yang dikategorikan transaksi tidak sukses, termasuk data dari transaksi kartu petugas kereta Commuter Line.

4) *Filter Anomali Penumpang*: Pada tahap ini dilakukan dengan cara mengecek konsistensi data. Pada kereta Commuter Line, diperlukan pemeriksaan data logic namun tidak masuk akal dan data yang *kontradiktif*. Data transaksi penumpang mempunyai kaidah seperti kewajaran waktu penumpang masuk dan keluar stasiun, satu nomor kartu dengan banyak perjalanan serta tap in dan tap out pada stasiun yang sama.

5) *Menambah atribut data baru*: Pada tahap ini akan dilakukan penambahan atribut pada data transaksi AFC yaitu, pertama atribut *in date – out date*, kedua atribut *in hour – out hour*, ketiga atribut *trip chain* dan keempat atribut *travel time*. Pada data transaksi penumpang AFC, data *in* dan *out date* serta data *in hour* dan *out hour* masih disatukan dalam satu kolom. Hal ini nantinya akan mempersulit proses analisis data, sehingga data tersebut perlu dipecah sendiri. Proses pemecahan data tanggal dan waktu transaksi membutuhkan atribut baru untuk menampung data pecahan tersebut.

6) *Deteksi outlier*: Dari proses sebelumnya sudah dapat dilihat adanya anomali dari nilai minimal dan maksimal *travel time* transaksi penumpang. Data outlier seperti waktu perjalanan yang terlalu cepat maupun terlalu lama serta interval yang terlalu pendek atau panjang antara waktu masuk setiap kartu. Tahap ini dilakukan untuk mengenali sebuah perjalanan penumpang yang belum lengkap dan tidak dapat digunakan pada pengolahan data selanjutnya. Pada proses awal deteksi outlier ini menggunakan teori Box Plot, sehingga dibutuhkan data perjalanan terpendek dan terpanjang. Oleh karena itu, data *travel time* transaksi penumpang AFC dianalisis menggunakan distribusi normal untuk melihat data perjalanan terpendek dan terpanjang penumpang komuter.

#### D. Pemilihan Data Komuter dari Data Penumpang.

1) *Fitur komuter*: Pada Commuter Line, penumpang hanya menggunakan kartu saat masuk dan keluar stasiun, sehingga komuter dapat diidentifikasi dengan menggunakan atribut *spatial-temporal* dengan melihat keteraturan komuter dalam menggunakan kartu pada setiap stasiun. Dari sisi atribut *spatial* penumpang yang memiliki 2 atau lebih transaksi masuk dalam kategori komuter. Pada atribut *temporal* masuk dan keluar stasiun penumpang yang memiliki 2 atau lebih trip (perjalanan) per hari diidentifikasi sebagai kategori komuter.

2) *Filter Komuter*: Penumpang yang telah diidentifikasi sebagai komuter pada tahap sebelum ditapis lagi berdasarkan jumlah perjalanan unik yang dilakukan selama periode tertentu. Hanya data komuter yang melakukan perjalanan lebih 4 hari dalam satu minggu yang akan diolah lebih lanjut.

#### E. Pembuatan Trip Chain.

Data AFC digunakan untuk menganalisis *association rules* dan untuk memverifikasi kelayakan metode yang diusulkan. *Association rules* digunakan untuk mengidentifikasi *item trip chain* yang diambil oleh para komuter. Dalam penelitian ini kami mengambil 2 dan 3 *item set* diantara *trip chain* komuter dengan menggunakan algoritma *apriori* dan perspektif *spatial-temporal*.

F. Analisis Data Menggunakan Association Rules.

1) *Membuat Item Set*: Data *trip chain basket* diproses menggunakan algoritma *apriori* untuk mendapatkan *item set*. Dalam penelitian ini nilai minimal *support* adalah 0.000001 karena data *trip chain* yang digunakan dalam penelitian ini cukup besar. Data *item set* kemudian diproses dengan *association rules* untuk mendapatkan *rules* atau aturan.

2) *Analisis 2 Frequent Item Set*: Pada tahap ini akan diambil *item set* yang memiliki frekuensi kemunculan 2 kali per hari. *2 frequent item set* digunakan untuk melihat pola atau hubungan antara item tersebut, serta untuk menghitung probabilitas munculnya item tersebut dalam suatu transaksi atau kejadian yang akan datang. Proses pertama dalam tahap ini adalah menentukan *item set* proses ini telah dilakukan sebelumnya, kemudian dari *item set* yang sudah didapatkan dilakukan perhitungan berapa kali kemunculan masing masing *item set* dalam data set. Kemudian akan di analisis berdasarkan distribusi *spatial-temporal*.

3) *Analisis 3 Frequent Item Set*: Pada tahap ini akan dianalisis *item set* yang mempunyai frekuensi kemunculan 3 kali per hari. Analisis *3 frequent item set* bertujuan untuk melihat pola atau hubungan yang lebih kompleks dari perilaku komuter Jabodetabek dan dianalisis pola atau hubungan yang terjadi antar *item set*.

III. HASIL DAN PEMBAHASAN

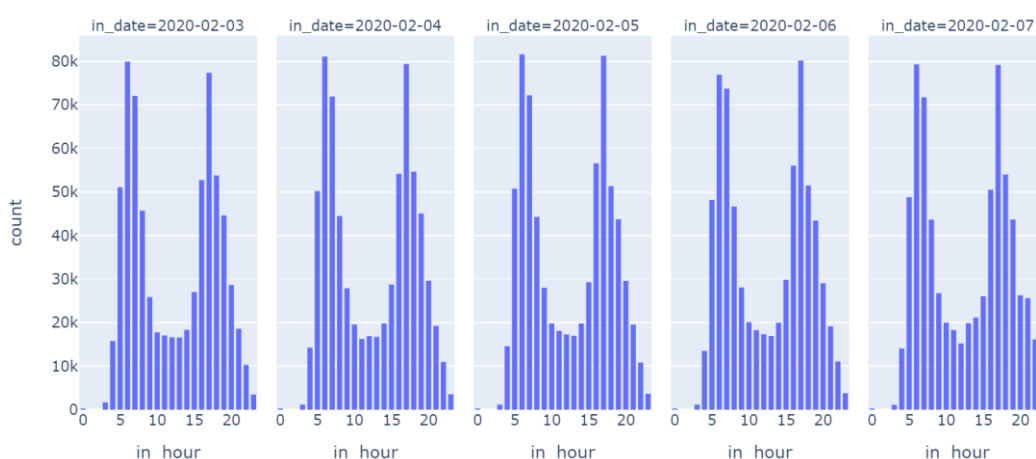
Tahap selanjutnya merupakan tahap *data mining* menggunakan *association rules*. Data AFC dari Commuter Line digunakan untuk menganalisis aturan asosiasi dan digunakan untuk memverifikasi kelayakan metodologi yang diusulkan dalam penelitian ini. Disamping itu, penelitian ini digunakan untuk mengetahui *item set 2* dan *item set 3* yang paling sering diambil oleh penumpang dengan menerapkan teknik *data mining* yang dilihat dari perspektif *spatial-temporal*.

A. Pemilihan Data Komuter.

1) *Fitur Komuter*: Pada Commuter Line, penumpang hanya menggunakan kartu saat masuk dan keluar stasiun, sehingga komuter dapat diidentifikasi dengan menggunakan atribut *spatial-temporal* dengan melihat keteraturan komuter dalam menggunakan kartu pada setiap stasiun. Secara khusus, dari hasil identifikasi penumpang tercatat ada 3334493 perjalanan selama periode 3-7 Februari 2020 dengan 2021289 merupakan transaksi valid penumpang dan 1269204 melakukan dua kali transaksi. Tabel 2 menjabarkan secara rinci transaksi harian selama tanggal 3–7 Februari 2020.

TABEL 2  
JUMLAH TRANSAKSI PENUMPANG

Tanggal transaksi	Jumlah Trip	Jumlah Penumpang unik	Jumlah penumpang unik > 2	Prosentasi penumpang unik > 2 trip
3/2/2020	657673	403591	245943	60.94 %
4/2/2020	667741	400453	258633	64.59%
5/2/2020	672992	403266	260859	64.69%
6/2/2020	666573	400360	257639	64.35%
7/2/2020	669514	413619	246130	59.51%

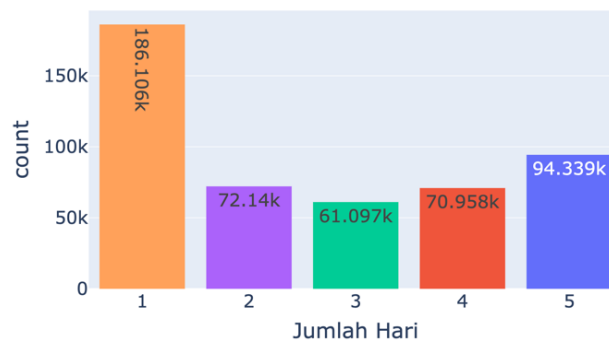


Gambar 1. Distribusi penumpang masuk stasiun

Distribusi waktu masuk stasiun untuk penumpang yang memiliki dua atau lebih perjalanan perhari dapat dilihat pada Gambar 1. Dari data terlihat bahwa sebagian besar penumpang yang bepergian memulai perjalanannya pada pukul 05.00–07.00 pagi dan 16.00–18.00 malam, dan ini mewakili trip chain dari komuter.

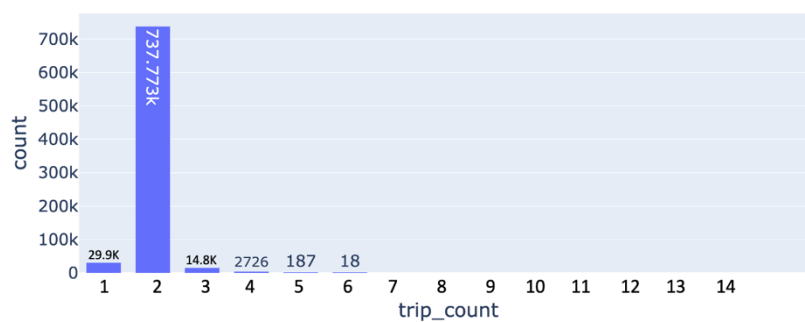
Dalam penelitian ini dipilih 4 atribut sebagai fitur yang melekat pada komuter :

1. Hari dimana perjalanan dilakukan, semakin banyak seseorang melakukan perjalanan harian setiap minggu semakin stabil perilaku perjalanannya.
2. Frekuensi perjalanan per hari menunjukkan perilaku perjalanan harian komuter. Penumpang dikatakan komuter ketika mempunyai minimal 2 kali perjalanan setiap hari.
3. Dimensi waktu masuk dan karakteristiknya menunjukkan kecenderungan keberangkatan perjalanan penumpang. Apabila penumpang memulai di waktu yang sama setiap hari kerja, maka yang terjadi adalah perjalanan perulangan secara berkala setiap hari kerja lebih mungkin terjadi.
4. Jumlah rute yang berurutan dan serupa. Rute yang berurutan dan sama mewakili karakteristik *spatial* dengan dua yaitu: id stasiun asal dan id stasiun tujuan. Sedangkan jumlah urutan rute menunjukkan perjalanan yang berulang pada distribusi *spatial* karakteristik perjalanan komuter ditunjukkan dengan pola loop tertutup dimana stasiun awal merupakan stasiun akhir perjalanan komuter.



Gambar 2. Penggunaan Kartu AFC (2 x Trip) dalam satu minggu

2) *Filter Komuter*: Berdasarkan Gambar 2 dapat diidentifikasi komuter berdasarkan penumpang yang rutin melakukan perjalanan selama 4-5 hari dalam seminggu dengan minimal transaksi AFC 2 kali. Pada proses filter komuter, didapatkan data komuter sebesar 1562011 atau 46.84% dari total perjalanan dalam jangka waktu tersebut.



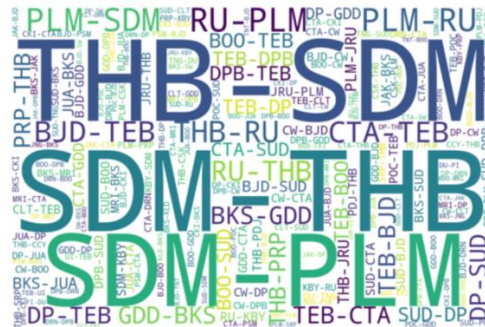
Gambar 3. Jumlah trip per komuter

3) *Membuat Trip Chain*: Dalam penelitian ini diambil 2 dan 3 item set di antara *trip chain* komuter dengan menggunakan algoritma *apriori* dan perspektif *spatial-temporal*. Data transaksi komuter kemudian dibuat data *trip chain set*, untuk mengetahui jumlah *trip* per komuter per hari. Hasil dari *trip chain* kemudian ditampilkan dalam bentuk *histogram* pada Gambar 3. Dapat dianalisis dari *histogram* bahwa komuter paling banyak melakukan *trip* 2 kali sehari dengan jumlah total 737773 dan *trip* dengan jumlah 3 kali sehari adalah 14813. Penumpang komuter dengan jumlah perjalanan 2 kali atau lebih dalam sehari dikategorikan komuter.

### B. Association Rules.

Association rules merupakan metode yang efektif untuk melihat hubungan antar variabel pada data dalam jumlah yang besar [17]. Data *trip chain* yang diperoleh dari proses sebelumnya kemudian di *encoding* untuk mendapatkan *trip chain basket* yang berisi kumpulan atau himpunan data *trip chain* komuter. Data *trip chain basket* ini digunakan untuk menentukan data *item set* menggunakan algoritma *apriori*.

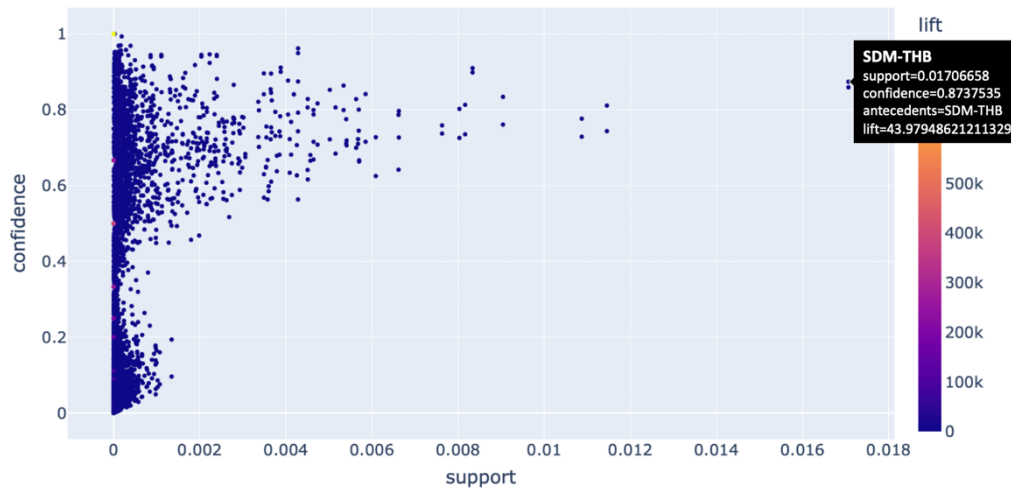
1) *Membuat Item Set*: Data *trip chain basket* diproses menggunakan algoritma *apriori* untuk mendapatkan *item set*. Dalam penelitian ini nilai minimal *support* adalah 0.000001 karena data *trip chain* yang digunakan dalam penelitian ini cukup besar sehingga kombinasi *item set* yang dihasilkan sebanyak 65381.



Gambar 4. Word cloud trip

Gambar 4 menunjukkan bahwa *item set* THB-SDM, SDM-THB, dan SDM-PLM merupakan *item set* dengan *support* yang tinggi daripada *item set* yang lainnya. Hal ini menunjukkan bahwa ketiga *item set* tersebut merupakan *item set* yang paling sering muncul dalam *data set* diikuti dengan *item set* PLM-SDM, RU-PLM, PLM-RU, BJD-TEB, THB-RU, dan CTA-TEB.

Proses selanjutnya adalah membuat aturan untuk setiap *item set*. Pada proses ini kami menentukan batasan minimum untuk *support* sebesar 0.000001, sehingga *item set* yang memiliki nilai *support* kurang dari *threshold* otomatis terhapus. Dari proses *association rules* didapatkan sebanyak 242702 *rules*.



Gambar 5. Hasil output association rules

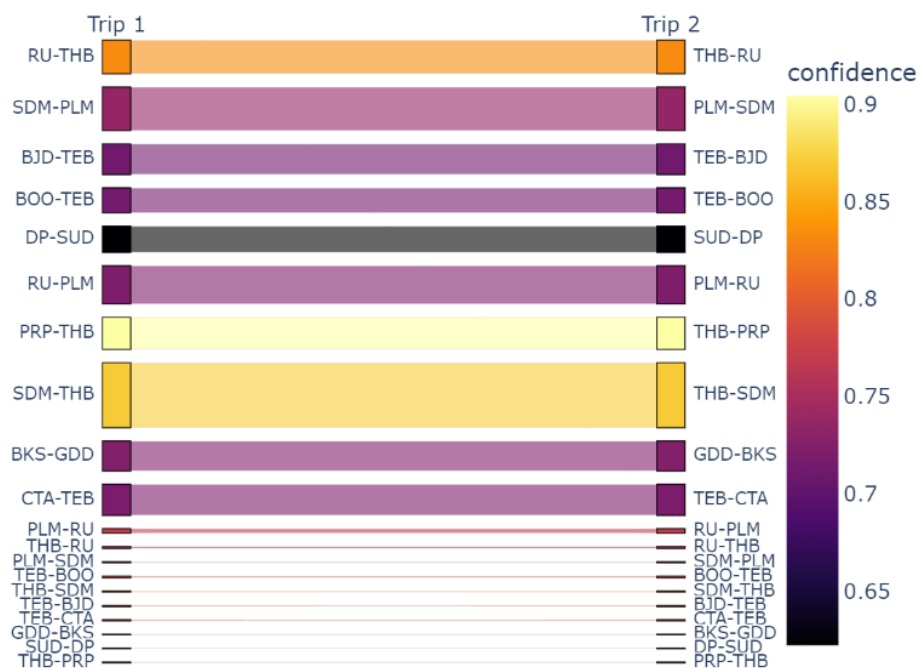
SDM-THB merupakan *item set* dengan *support*, *confident* dan *lift* tertinggi daripada *item set* yang lainnya seperti ditunjukkan pada Gambar 5. Tabel 3 berikut menunjukkan *rules* yang dihasilkan dalam proses ini. Total *rules* yang dihasilkan adalah sebanyak 269748 *rules*, dengan nilai *support* dan *confident* tertinggi yang didapatkan adalah *item set* SDM-THB. Sedangkan *support* dan *confident* terendah adalah *item set* TEB-BKS



TABEL 3  
HASIL RULES, SUPPORT, CONFIDENT DAN LIFT

	antecedents	consequents	support	confidence	lift
0	(SDM-THB)	(THB-SDM)	0.015837	0.869562	46.803858
1	(THB-SDM)	(SDM-THB)	0.015837	0.852427	46.803858
2	(PLM-SDM)	(SDM-PLB)	0.010663	0.801877	55.365540
3	(SDM-PLB)	(PLM-SDM)	0.010663	0.736243	55.365540
4	(RU-PLM)	(PLM-RU)	0.010143	0.720642	54.720377
...	...	...	...	...	...
269747	(TEB-BKS)	(PSM-TEB, JUA-JAK, JAK-CW, BKS-JAK, JAK-GDD, GD...	0.000001	0.000624	529.918902

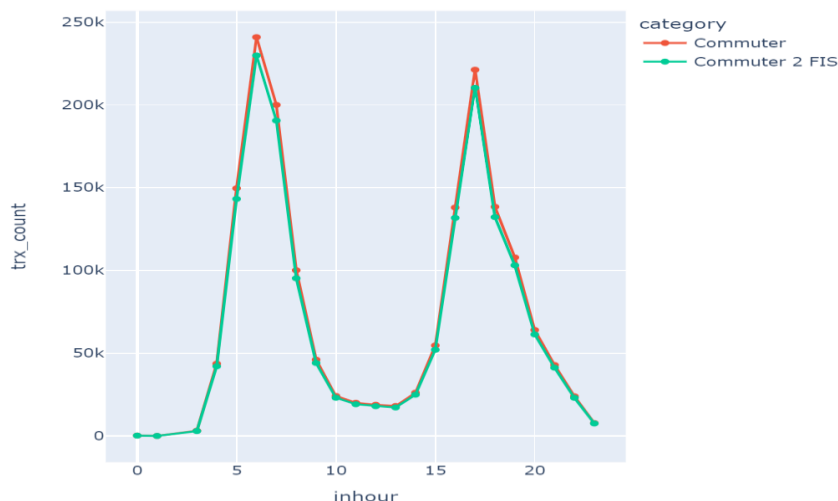
2) Analisis 2 Frequent Item Set: Sebanyak 85526 rules untuk 2 frequent item set dihasilkan dan didapatkan antecedent (SDM-THB) dan consequent (THB-SDM) merupakan item set dengan support, confident, dan lift terbesar. Sedangkan antecedent (KMO-TB) dan consequent (KRI-JNG) merupakan item set dengan support dan confident terendah. Hal ini menunjukkan perilaku Commuter Line, yaitu ketika komuter membeli item set SDM-THB besar kemungkinannya untuk komuter juga membeli item set THB-SDM.



Gambar 6. Distribusi Temporal 2 Frequent Item Set

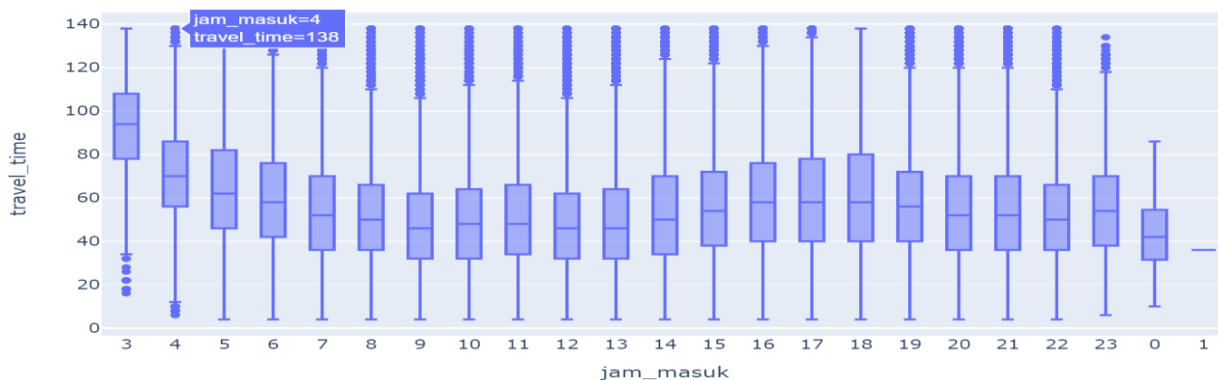
Pada distribusi spatial 2 frequent item set didapatkan bahwa jalur SDM-THB dan THB-SDM mempunyai nilai support paling tinggi dan confident tertinggi jatuh pada jalur PRP-THB dan THB-PRP seperti terlihat pada Gambar 6. Hal ini dapat menunjukkan tingkat kepadatan komuter pada jalur SDM-THB melebihi jalur yang lain. Sedangkan jalur PRP-THB dan sebaliknya memiliki tingkat kepastian tertinggi. Sehingga peningkatan kapasitas transportasi pada kedua jalur tersebut adalah hal yang harus dilakukan agar mendapatkan penjadwalan yang optimal.

Dalam distribusi temporal pada 2 frequent item set, didapatkan semua perjalanan komuter dari stasiun asal ke stasiun tujuan (AT) yang sering terjadi dan paling stabil. Dalam penelitian ini, sebanyak 64% dari komuter pada 2 frequent item set merupakan komuter 05.00-07.00 pagi dan 16.00-18.00 sore.



Gambar 7. Distribusi Temporal 2 Frequent Item Set

Gambar 7 adalah distribusi *temporal 2 frequent item set*, terlihat bahwa jumlah komuter dan komuter dengan 2 *frequent item set* memiliki perbedaan yang tidak terlalu signifikan. Lebih lanjut dapat di analisis bahwa lebih banyak komuter yang melakukan perjalanan di waktu pagi hari. Perilaku komuter ini berkontribusi pada jam sibuk pagi 05.00–07.00.



Gambar 8. Distribusi travel time 2 frequent item set

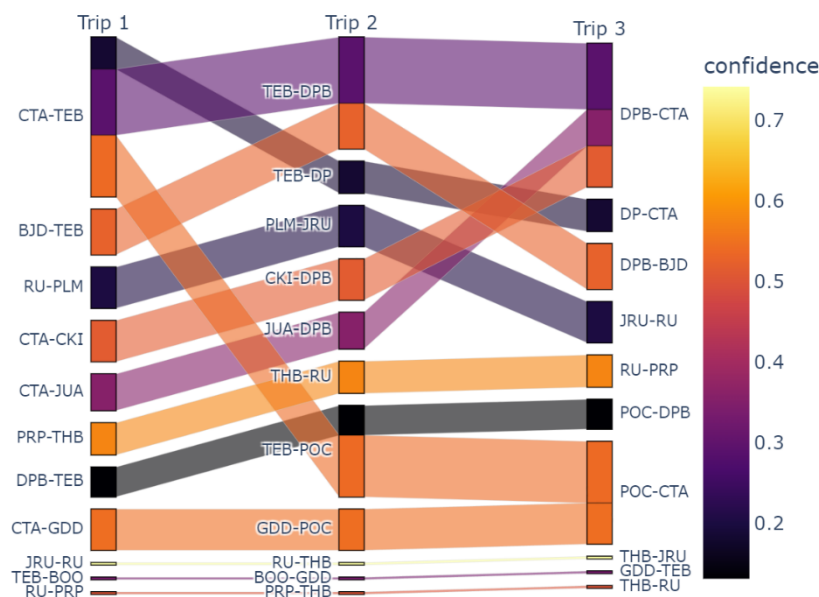
Gambar 8 menunjukkan distribusi *travel time* per jam. Median pada jam sibuk pagi hari 05.00–07.00 mengalami kecenderungan penurunan *travel time*. Data grafik nilai median dari jam 05.00–09.00 berturut turut adalah 63, 58, 52, 50, dan 46 menit. Dengan kata lain, rata-rata *travel time* pada jam sibuk pagi hari mempunyai kecenderungan penurunan. Diasumsikan bahwa komuter pada 2 *frequent item set* sudah memperkirakan waktu tempuh dari stasiun asal ke stasiun tujuan. Sedangkan pada jam sibuk sore hari 15.00–18.00 mempunyai nilai median berturut turut adalah 54, 58, 58, dan 58 menit dan nilai *upper quartile* berturut turut adalah 72, 76, 78 dan 80 menit. Dari grafik menunjukkan bahwa *travel time* mengalami kecenderungan naik. Diperkirakan terjadi penumpukan penumpang sehingga komuter menunggu kereta berikutnya yang lebih longgar karena pada jam tersebut merupakan jam selesai kerja wilayah Jabodetabek.

3) *Analisis 3 Frequent Item Set* : Dengan metode *mining* yang sama dalam penelitian ini juga digunakan untuk menganalisis 3 *frequent item set* yang sering diambil oleh komuter dalam perjalanannya. Sebanyak 52302 *rules* dihasilkan pada 3 *frequent item set*.



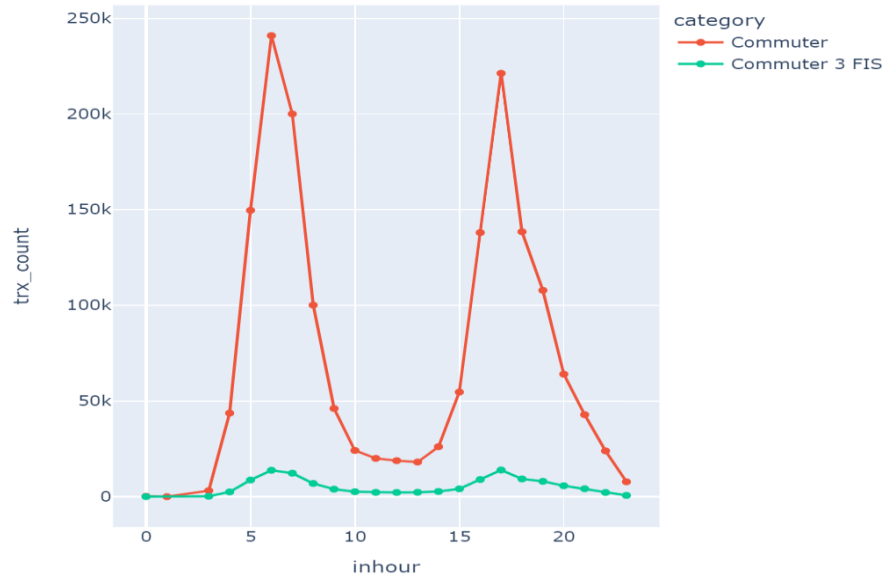
Gambar 9. Word cloud 3 frequent item set

Sebagian besar perjalanan dengan 3 frequent item set terbentuk dalam pola loop tertutup, pola tersebut dapat dilihat contoh pada Gambar 9 dimana komuter berangkat dari SDM ke THB, kemudian dari THB ke JRU, dan terakhir JRU ke SDM. Selain itu terdapat pula pola loop terbuka namun frekuensinya tidak sebanyak pola loop tertutup.



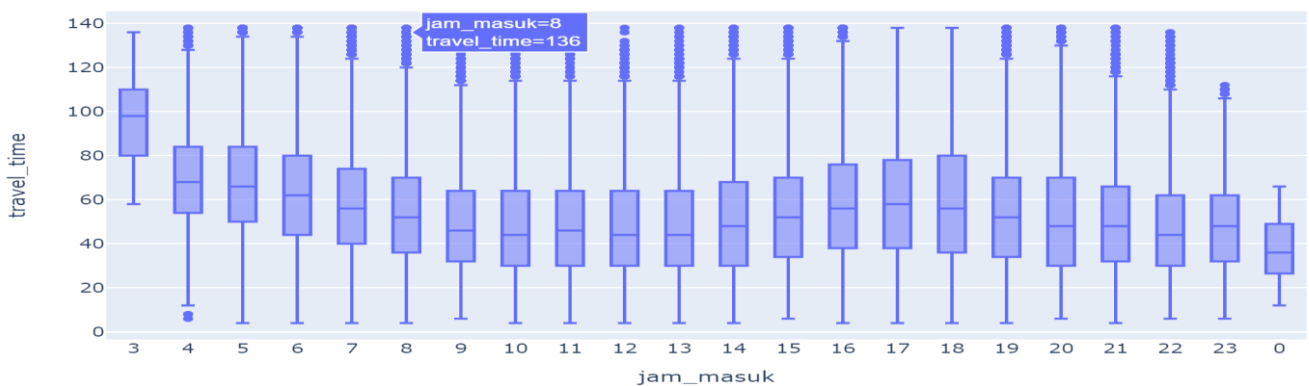
Gambar 10. Distribusi Spatial 3 frequent item set

Gambar 10 menunjukkan distribusi *spatial* pada 3 frequent item set. Lebar pita menunjukkan nilai *support* dan warna menunjukkan nilai *confident*. Pada distribusi ini beberapa *item set* yang memiliki *support* yang tinggi namun nilai *confident* rendah begitu juga sebaliknya. Hal tersebut terlihat pada *item set* CTA-TEB, TEB-DPB, DPB-CTA memiliki lebar pita yang besar yang berarti nilai *support* yang besar namun memiliki warna agak gelap yang berarti nilai *confident* rendah. Hal ini menunjukkan bahwa *item set* CTA-TEB, TEB-DPB, DPB-CTA merupakan *item set* yang banyak diambil oleh komuter namun memiliki tingkat kepastian yang rendah.



Gambar 11. Distribusi Temporal 3 frequent item set

Hasil penelitian pada distribusi temporal 3 frequent item set ini tidak sensitif terhadap waktu masuk, seperti yang terjadi pada distribusi temporal 2 frequent item set. Gambar 11 menunjukkan bahwa terdapat perbedaan yang cukup signifikan antara komuter biasa dengan komuter 3 frequent item set. Jam padat pada komuter 3 frequent item set berkisar antara jam 05.00–07.00 pagi hari dan 16.00–18.00 pada sore hari, namun dengan jumlah transaksi lebih sedikit.



Gambar 12. Distribusi travel time 3 frequent item set

Gambar 12 memperlihatkan distribusi travel time per jam pada 3 frequent item set. Pola yang terlihat hampir sama dengan pola pada 2 frequent item set. Pada jam sibuk pagi hari 05.00–07.00 terjadi penurunan travel time. Rata-rata komuter akan menyelesaikan perjalanan pada jam sibuk pagi adalah 66, 62, dan 58, 52 menit. Sedangkan pada jam sibuk sore hari 16.00–18.00 terjadi peningkatan travel time. Pada jam sibuk sore rata-rata penumpang menyelesaikan perjalanan sekitar 52, 56, 58, dan 56 menit dan upper quartile pada 70, 76, 78, dan 80 menit. Kecenderungan ini hampir sama dengan 2 frequent item set.

#### IV. PEMBAHASAN

Dari hasil analisis menunjukkan bahwa item set SDM–THB merupakan item set yang paling sering diambil oleh komuter. Dari analisis juga menunjukkan bahwa perilaku perjalanan komuter Jabodetabek memiliki jumlah yang besar dan sensitif

terhadap waktu. Hampir semua jenis perilaku perjalanan komuter memiliki karakteristik distribusi waktu yang khas yaitu jam puncak pagi hari 05.00–07.00 dan jam puncak sore hari 16.00–18.00.

Analisis 2 *frequent item set* menunjukkan bahwa *antecedent* SDM-THB dan *consequent* THB-SDM merupakan *item set* dengan nilai *support*, *confident* dan *lift* terbesar. *Item set* ini paling berpotensi terjadi penumpukan komuter. Analisis *spatial* menunjukkan bahwa stasiun SDM dan THB adalah stasiun dengan tingkat kepadatan yang tinggi. Sedangkan hasil analisis *temporal* menunjukkan bahwa sebanyak 64% dari komuter pada 2 *frequent item set* adalah komuter pada pagi dan sore hari, yang menyebabkan jam sibuk pagi dan sore. Analisis *travel time* komuter menunjukkan rata-rata *travel time* pada jam sibuk pagi mengalami penurunan, sedangkan pada jam sibuk sore mengalami kenaikan.

Pada analisis 3 *frequent item set* menunjukkan bahwa sebagian besar perjalanan terbentuk dalam pola loop tertutup, dimana sebagai contoh komuter berangkat dari SDM-THB, kemudian THB-JRU, dan terakhir JRU-SDM. Distribusi *rules* pada 3 *frequent item set* tidak berkontribusi terhadap kepadatan penumpang KRL. Analisis *spatial* menunjukkan beberapa *item set* memiliki *support* yang tinggi namun nilai *confident* rendah begitu juga sebaliknya. Analisis *temporal* menunjukkan tidak ada sensitivitas terhadap waktu masuk, dan jumlah transaksi pada jam sibuk pagi lebih sedikit. Pada jam sibuk pagi terjadi penurunan *travel time*, sementara pada jam sibuk sore terjadi peningkatan *travel time* yang hampir sama dengan 2 *frequent item set*.

Hasil dari analisis *association rules* dapat digunakan untuk meningkatkan kualitas layanan dengan melihat parameter *travel time* komuter. *Travel time* yang disebabkan oleh waktu tunggu komuter di stasiun dapat dikurangi dengan peningkatan kapasitas penumpang kereta Commuter Line. Pemilihan rute dan waktu untuk penambahan kapasitas penumpang dapat mengacu pada nilai *support* dan *confident* dari *association rules* yang tinggi.

## V. SIMPULAN

Dari hasil modeling dan analisis dapat disimpulkan bahwa metode *a priori* dan *association rules* yang digunakan dapat menghasilkan informasi dan pemahaman yang baik tentang perilaku perjalanan komuter Jabodetabek di mana memiliki jumlah yang besar dan karakteristik sensitif terhadap waktu. Hampir semua jenis perilaku perjalanan komuter memiliki karakteristik distribusi waktu yang khas yaitu jam puncak pagi hari dan jam puncak sore hari.

Dari analisis *spatial* pada komuter dengan 2 *frequent item set* didapatkan bahwa stasiun SDM dan THB merupakan stasiun dengan tingkat kepadatan tertinggi, dan berdasarkan hasil analisis *temporal* menunjukkan bahwa sebanyak 64% komuter melakukan perjalanan pada jam sibuk pagi dan sore. Hasil analisis *travel time* komuter menunjukkan rata-rata *travel time* pada jam sibuk pagi hari mengalami kecenderungan turun, diperkirakan karena komuter sudah memperkirakan jarak tempuh dari stasiun asal ke stasiun tujuan. Sedangkan pada jam sibuk sore hari menunjukkan kecenderungan kenaikan karena diperkirakan terjadi penumpukan penumpang sehingga komuter bersedia menunggu kereta berikutnya..

Pada analisis 3 *frequent item set* didapatkan sebagian besar perjalanan terbentuk dalam pola loop tertutup dan tidak berkontribusi terhadap kepadatan penumpang Commuter Line, serta tidak sensitif terhadap waktu masuk. Jumlah komuter dengan 3 *frequent item set* memiliki perbedaan yang sangat signifikan dibandingkan komuter biasa. Sedangkan *travel time* jam sibuk pagi dan sore memiliki kecenderungan yang hampir sama dengan komuter dengan 2 *frequent item set*. Pola *trip chain* dengan 3 *item set* dan pola loop tertutup seperti ini sebetulnya dapat dilakukan juga oleh sebagian besar komuter sesuai dengan asal dan tujuan masing-masing untuk menghindari terjadinya kepadatan penumpang jalur-jalur tertentu.

## UCAPAN TERIMA KASIH

Terimakasih saya sampaikan kepada seluruh *civitas academica* Universitas Kristen Maranatha Bandung, khususnya para dosen pembimbing yang telah mencurahkan ilmu dan pengalamannya hingga Jurnal ini dapat diselesaikan dengan baik. Tidak lupa untuk istri dan anak-anak tercinta yang telah memberikan dukungan penuh hingga akhir, dan terutama kepada Bapa di Surga dengan anugrahNya yang melimpah.

## DAFTAR PUSTAKA

- [1] X. Guo, D. Z. Wang, J. Wu, H. Sun and L. Zhou, "Mining commuting behavior of urban rail transit network by using association rules,," *Physica A: Statistical Mechanics and its Applications*, vol. 559, 2020.
- [2] Y. Wang, L. Zhu, Q. Lin and L. Zhang, "Leveraging Big Data Analytics for Train Schedule Optimization in Urban Rail Transit Systems,," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2018.
- [3] J. Maktoubian, "Analyzing Large-Scale Smart Card Data to Investigate Public Transport Travel Behaviour Using Big Data Analytics,," *Journal of Information Technology & Software Engineering*, vol. 07, no. 04, 2017.
- [4] Z. Jiang and S. Liao, "A method for extracting passenger flow time series feature of urban rail transit,," in *ICTE 2019 - Proceedings of the 6th International Conference on Transportation Engineering*, 2019.

- [5] Z. Chen and W. Fan, "Extracting bus transit boarding and alighting information using smart card transaction data," *Journal of Public Transportation*, vol. 22, no. 1, 2020.
- [6] A. Alsger, A. Tavassoli, M. Mesbah, L. Ferreira and M. Hickman, "Public transport trip purpose inference using smart card fare data," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 123-137, 2018.
- [7] E. Castillo, Z. Grande and A. Calviño, "Bayesian Networks-Based Probabilistic Safety Analysis for Railway Lines," *Computer-Aided Civil and Infrastructure Engineering*, vol. 31, no. 9, pp. 681-700, 2016.
- [8] J. Zhou, W. Xu, X. Guo and J. Ding, "A method for modeling and analysis of directed weighted accident causation network (DWACN)," *Physica A: Statistical Mechanics and its Applications*, vol. 437, pp. 263-277, 2015.
- [9] G. Suchacka and G. Chodak, "Using association rules to assess purchase probability in online stores," *Information Systems and e-Business Management*, vol. 15, no. 3, pp. 751-780, 2017.
- [10] Q. C. Lu, "Modeling network resilience of rail transit under operational incidents," *Transportation Research Part A: Policy and Practice*, vol. 117, pp. 227-237, 2018.
- [11] D. Zhao, W. Wang, G. P. Ong and Y. Ji, "An Association Rule Based Method to Integrate Metro-Public Bicycle Smart Card Data for Trip Chain Analysis," *Journal of Advanced Transportation*, vol. 2018, 2018.
- [12] X. Guo, H. Sun, J. Wu, J. Jin, J. Zhou and Z. Gao, "Multiperiod-based timetable optimization for metro transit networks," *Transportation Research Part B: Methodological*, vol. 96, pp. 46-67, 2017.
- [13] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268-276, 2001.
- [14] X. Ma, Y. J. Wu, Y. Wang, F. Chen and J. Liu, "Mining smart card data for transit riders' travel patterns," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 1-12, 2013.
- [15] G. Goulet Langlois, H. N. Koutsopoulos and J. Zhao, "Inferring patterns in the multi-week activity sequences of public transport users," *Transportation Research Part C: Emerging Technologies*, vol. 64, pp. 1-16, 2016.
- [16] W. X. Xu, L. Xu, X. M. Liu and J. D. Jones, "A new approach to decision-making with key constraint and its application in enterprise information systems," *Enterprise Information Systems*, vol. 2, no. 3, pp. 287-308, 2008.
- [17] E. Varol Altay and B. Alatas, "Intelligent optimization algorithms for the problem of mining numerical association rules," *Physica A: Statistical Mechanics and its Applications*, vol. 540, pp. 123-142, 2020.