

Exploratory Data Analysis terhadap Kepadatan Penumpang Kereta Rel Listrik

<http://dx.doi.org/10.28932/jutisi.v7i2.3700>

Riwayat Artikel

Received: 18 Juni 2021 | Final Revision: 18 Juli 2021 | Accepted: 22 Juli 2021

Feliks Victor Parningotan Samosir^{#1}, Loudry Palmarums Mustamu^{#2}, Erik Dwi Anggara^{#3},

Albertus Indarko Wiyogo^{#4}, Andreas Widjaja^{#5}

[#]Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Kristen Maranatha
Jl. Prof. drg. Surya Sumantri No. 65 Bandung, Indonesia

¹2079002@maranatha.ac.id

²2079003@maranatha.ac.id

³2079004@maranatha.ac.id

⁴2079005@maranatha.ac.id

⁵andreas.widjaja@maranatha.ac.id

Abstract — The existence of the Kereta Rel Listrik Commuter Line as the backbone of transportation in the Jakarta - Bogor - Depok - Tangerang - Bekasi - Banten area has a very important role for commuter mobility around Daerah Khusus Ibukota Jakarta. With an average number of 1.1 million passengers per day, Kereta Rel Listrik is one of the factors supporting Indonesia's economy and growth in various sectors. On the other hand, the Covid-19 pandemic that hit the world caused restrictions on human movement which resulted in a decline in all economic sectors. The purpose of this research is to obtain optimal train schedule recommendations for the operation of the Kereta Rel Listrik Commuter Line in the Rangkasbitung - Tanah Abang service to carry passengers optimally while adhering to the physical distancing protocol set by the Minister of Transportation to prevent the wider spread of Covid-19. With such a large amount of data that must be processed, Exploratory Data Analysis is one of the choices we take to process the above data to get satisfactory results.

Keywords— agile development; exploratory data analysis; jupyterlab; korelasi; Kereta Rel Listrik Commuter Line.

I. PENDAHULUAN

Kereta Komuter *mass rapid transit* merupakan layanan transportasi massal dengan jangkauan lokal yang tersedia bagi siapapun yang membayar biaya yang telah ditentukan dan dirancang untuk memindahkan sejumlah besar penumpang dalam waktu bersamaan. Salah satu bentuk dari *mass rapid transit* adalah kereta komuter, yang membawa penumpang di dalam wilayah perkotaan atau dari kota ke daerah pinggiran. Biasanya kereta ini melayani pada saat jam puncak (*peak hour*).

Ada beberapa keunggulan dan kelemahan dari Kereta Komuter [1]:

Keunggulan kereta komuter:

1. memiliki kapasitas angkut yang lebih besar dibandingkan dengan angkutan umum lainnya, misalnya bus, sehingga dapat memindahkan penumpang dalam jumlah besar dari suatu tempat ke tempat lain,
2. memiliki jalur khusus, sehingga tidak mengganggu pengguna jalan lain,
3. waktu tempuh relatif lebih cepat dibandingkan dengan angkutan lain untuk tujuan yang sama.

Kelemahan kereta komuter:

1. daerah jangkauannya kurang luas, tidak dapat menjangkau daerah-daerah pelosok karena kereta ini hanya diperuntukkan untuk menjangkau daerah- daerah tertentu saja,
2. jadwal kereta, penumpang harus mau menyesuaikan diri dengan jadwal yang ada dan harus menunggu dengan sabar jika kereta tersebut mengalami keterlambatan

Untuk wilayah Indonesia kita dapat menemukan kereta komuter mass rapid transit di KAI Commuter Jabodetabek (atau disebut juga KRL Commuter Line, dulu dikenal sebagai KRL Jabodetabek) adalah jalur kereta rel listrik yang dioperasikan

oleh PT KAI Commuter Jabodetabek, anak perusahaan dari PT Kereta Api Indonesia (PT KAI). KRL telah beroperasi di wilayah Jakarta sejak tahun 1976, hingga kini melayani rute komuter di wilayah DKI Jakarta, Kota Depok, Kota Bogor, Kabupaten Bogor, Kota Bekasi, Kabupaten Lebak, Kota Tangerang, dan Kota Tangerang Selatan.

Dari data yang kami peroleh KRL memiliki 80 stasiun di Jabodetabek dan Banten, memiliki panjang rel 418 km, memiliki 993 perjalanan tiap harinya, 337 juta penumpang pada tahun 2019. Rute KRL yang akan digunakan dalam penelitian kali ini adalah Rangkasbitung – Tanah Abang [2]. Rute KRL tersebut seperti ditunjukkan pada Gambar 1 yaitu mulai dari Rangkasbitung, Citeras, Maja, Cikuya, Tigaraksa, Tenjo, Daru, Cilejit, Parungpanjang, Cicayur, Cisauk, Serpong, Rawa Buntu, Sudimara, Jurang Mangu, Pondok Ranji, Kebayoran, Palmerah, Tanah Abang [2]. Rute KRL tersebut seperti ditunjukkan pada Gambar 1 yaitu mulai dari Rangkasbitung, Citeras, Maja, Cikuya, Tigaraksa, Tenjo, Daru, Cilejit, Parungpanjang, Cicayur, Cisauk, Serpong, Rawa Buntu, Sudimara, Jurang Mangu, Pondok Ranji, Kebayoran, Palmerah, Tanah Abang.



Gambar 1. Rute KRL Rute Rangkasbitung - Tanah Abang [2]

Terlepas dari kelebihan dan kelemahan dari kereta komuter atau KRL, moda transportasi ini masih merupakan salah satu moda transportasi yang sangat diminati masyarakat. Banyak sekali alasan dari masyarakat, terutama kalangan menengah ke

bawah, untuk menggunakannya. Beberapa diantaranya adalah harganya yang murah dan jadwal yang tepat waktu karena tidak terkendala dengan kemacetan mengingat KRL mempunyai jalannya sendiri. Proses bisnis KRL dapat dilihat di Lampiran 1.

Dalam kondisi normal, kondisi di dalam kereta yang dapat dikatakan berdesakkan satu sama lain bukanlah menjadi masalah. Hal ini berbeda sekali dengan kondisi pandemi Covid-19 yang mengharuskan penerapan *physical distancing*, sehingga jarak antara satu penumpang dengan penumpang yang lain diatur minimal 1,5 meter. Dengan penerapan aturan ini, tentu saja kondisi berdesakkan dalam kereta harus diubah. Dalam penelitian kali ini akan dilihat bagaimana cara untuk mengatasi masalah tersebut. Dengan melihat hubungan antara waktu interval keberangkatan antar kereta dengan jumlah total penumpang yang diangkut dalam satu kereta pada kondisi sebelum pandemi, diharapkan dapat menghasilkan sebuah rekomendasi sederhana untuk perubahan jadwal untuk menyebar penumpang.

Dengan tingkat kepadatan penumpang yang sangat tinggi di jam-jam tertentu dan hari-hari tertentu membuat pihak KRL kesulitan dalam mengangkut penumpang. *Physical distancing* dalam masa pandemi Covid-19 saat ini juga menjadi masalah pihak KRL untuk mengangkut setiap penumpang dengan baik. Karena masalah tersebut pihak KRL perlu menyelesaikan masalah tersebut. Salah satu cara yang paling baik yaitu melakukan pengoptimalan jadwal. Untuk mengoptimalkan jadwal yang sudah ada, sebelumnya harus dilakukan analisis variabel apa saja memiliki korelasi paling tinggi terkait kepadatan penumpang. Ketika variabel tersebut sudah ditentukan maka pengoptimalan jadwal dapat dilakukan. Pengoptimalan jadwal dapat membuat penumpang lebih nyaman dalam menggunakan layanan KRL, sehingga membantu pihak KRL untuk dapat mengangkut setiap penumpang.

Variabel yang kita analisis berasal dari dataset perjalanan penumpang Tanah Abang-Rangkasbitung. Akan dilihat korelasi tertinggi dari variabel tersebut untuk memberikan saran rekomendasi penjadwalan KRL yang memperhatikan peraturan pemerintah *physical distancing*. Dari dataset jika telah ditemukan variabel dengan korelasi tertinggi terkait kepadatan penumpang maka akan dilakukan analisis dengan pendekatan *exploratory data* yang akan menghasilkan saran rekomendasi penjadwalan KRL.

Berdasarkan penjelasan tersebut maka dapat dibuat rumusan masalah sebagai berikut :

1. Apakah waktu interval keberangkatan antar rangkaian kereta api mempengaruhi jumlah total penumpang pada satu rangkaian kereta?
2. Rekomendasi seperti apa yang dapat disarankan untuk menerapkan ketentuan *physical distancing* penumpang dalam KRL?

Dalam tahapan perumusan hipotesis ini dimulai dengan pengumpulan fakta lapangan. Berikut beberapa fakta yang berhasil dikumpulkan :

1. Populasi data berasal data tap in penumpang yang diambil dari bulan Januari sampai dengan bulan Februari 2020 sebelum pandemi Covid-19. Dari data tersebut memperlihatkan dalam kondisi asli di lapangan menunjukkan kepadatan penumpang yang cukup memprihatinkan.
2. Jumlah penumpang yang diwajibkan selama masa pandemi dengan menerapkan protokol *Physical Distancing* adalah max. 74 penumpang dalam satu gerbong kereta api [3].
3. Jumlah gerbong yang tersedia dalam satu rangkaian kereta api rata-rata 10 gerbong.

Berdasarkan fakta tersebut disusunlah hipotesis “Waktu interval keberangkatan antar rangkaian kereta api mempengaruhi jumlah total penumpang pada satu rangkaian kereta api.”

Hipotesis inilah yang dapat digunakan sebagai rekomendasi perubahan jadwal keberangkatan kereta api untuk menyesuaikan jumlah penumpang maksimal dalam mempersiapkan kondisi new normal sesuai ketentuan *physical distancing*.

II. METODE PENELITIAN

A. *Exploratory Data Analysis*

Exploratory Data Analysis (EDA) [4] merupakan sebuah tahapan dalam *data science* untuk mengenal dan memahami dataset. EDA juga merupakan teknik pencarian heuristik untuk menemukan relasi signifikan antara variabel pada dataset yang besar [5].

Secara definitif, *Exploratory Data Analysis* mengacu pada proses kritis dalam melakukan investigasi awal pada data untuk menemukan pola, untuk menemukan anomali, untuk menguji hipotesis dan untuk memeriksa asumsi dengan bantuan statistik ringkasan dan representasi grafis. Dengan melakukan EDA, kita dapat lebih memahami kondisi dataset yang kita miliki. Sehingga, kita dapat memulai pembentukan model *Machine Learning* dengan lebih baik kedepannya. 'Memahami kondisi dataset' dapat merujuk pada sejumlah hal termasuk namun tidak terbatas pada poin-poin berikut :

1. mengekstrak variabel penting dan meninggalkan variabel yang tidak berguna,
2. mengidentifikasi pencilan (outliers), nilai yang hilang (missing values), atau kesalahan manusia (human error),

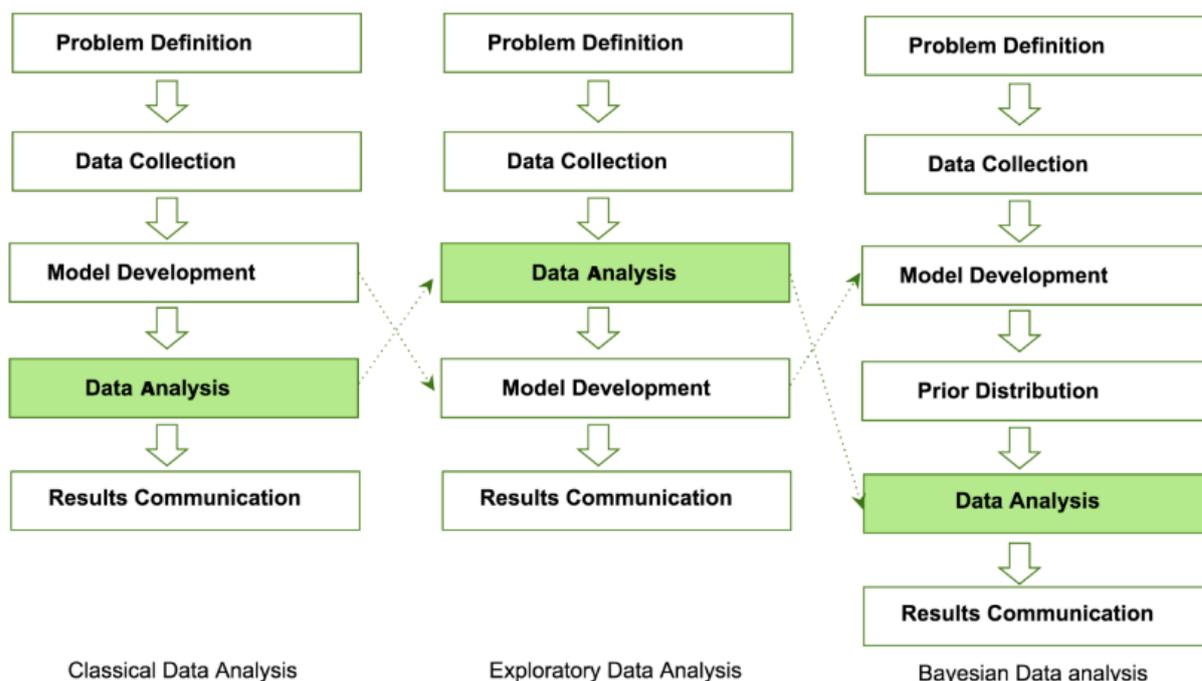
3. memahami hubungan, atau kekurangan, antar variabel,
4. memaksimalkan wawasan yang kita miliki atas kondisi data dan meminimalkan potensi kesalahan di kemudian hari.

Analisis data eksploratif (*Exploratory Data Analysis – EDA*) merupakan metode eksplorasi data dengan menggunakan teknik aritmatika sederhana dan teknik grafis dalam meringkas data pengamatan [4]. EDA banyak digunakan dalam berbagai hal seperti:

1. memaksimalkan telaahan data,
2. mencari struktur data yang tersembunyi (mengungkap misteri yang tersembunyi),
3. mengeluarkan variabel penting,
4. mendeteksi kelainan dan anomali,
5. melakukan test suatu asumsi,
6. membangun model,
7. melakukan optimasi.

Kontribusi utama dari pendekatan analisis data secara eksploratif terletak pada penyajian seluruh ringkasan statistik secara visual. Ringkasan statistik secara numerik saja, dapat mengaburkan, menyembunyikan, atau bahkan salah dalam merepresentasikan struktur data. Apabila ringkasan numerik digunakan secara terpisah dan langsung diterima tanpa pemeriksaan data terlebih dahulu secara visual mungkin akan mengakibatkan pemilihan model tidak tepat. Pemilihan model yang langsung dilakukan secara tergesa-gesa dan mungkin berdasarkan pada asumsi yang salah akan mengakibatkan penarikan kesimpulan yang salah. Untuk alasan tersebut, analisis pendahuluan harus dimulai dengan pemeriksaan secara visual, bukan ringkasan data secara numerik.

Terdapat tiga pendekatan analisis data seperti yang ditunjukkan oleh Gambar 2: Klasik (Classical), Eksploratif (Exploratory (EDA)) dan Bayesian. Dengan demikian, EDA merupakan salah satu dari tiga teknik pendekatan analisis data yang ada. Ketiga pendekatan tersebut mempunyai kemiripan, semuanya dimulai dari teori atau permasalahan umum dan diakhiri dengan suatu kesimpulan. Perbedaannya terletak pada urutan dan fokus dari langkah-langkah perantara.



Gambar 2. Perbandingan Pendekatan Analisis Data : Klasikal, Eksploratif, dan Bayesian.

Jadi, pada analisis klasik, pengumpulan data diikuti dengan penerapan model (normalitas, linieritas, dll) dan selanjutnya adalah analisis, estimasi, dan pengujian yang difokuskan pada model parameter tersebut. Pada EDA, pengumpulan data tidak diikuti dengan penerapan model, melainkan segera diikuti dengan analisis dengan tujuan untuk menentukan model apa yang akan sesuai. Akhirnya, metode pendugaan Bayesian merupakan proses pendugaan dengan mempertimbangkan dua hal, yaitu data yang sekarang kita miliki dan informasi awal mengenai kasus yang sedang kita teliti. Keduanya digunakan

secara bersama-sama untuk membuat suatu kesimpulan atau uji asumsi mengenai model parameter. Pada kenyataannya, analisis data merupakan kombinasi dari ketiga pendekatan di atas (serta pendekatan-pendekatan lainnya). Perbedaan tersebut diatas hanya diuraikan untuk menekankan perbedaan utama dari ketiga pendekatan tersebut. EDA bukan seperangkat teknik. EDA merupakan suatu pendekatan, pola/sikap/filosofi tentang bagaimana kita dalam menganalisis sekumpulan data.

B. Analisis Korelasi

Analisis korelasi [6] [7] adalah metode statistika yang digunakan untuk menentukan kuatnya atau derajat hubungan linier antara dua variabel atau lebih. Semakin nyata hubungan linier (garis lurus), maka semakin kuat atau tinggi derajat hubungan linier antara kedua variabel. Ukuran untuk derajat hubungan garis lurus ini dinamakan koefisien korelasi.

Korelasi Pearson r bernilai $-1 \leq r \leq 1$, dimana apabila nilai $r = -1$ artinya korelasi negatif sempurna; $r = 0$ artinya tidak ada korelasi; dan $r = 1$ artinya positif sempurna atau korelasinya sangat kuat [8].

Korelasi linier Pearson didefinisikan sebagai berikut:

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \quad (1)$$

untuk pasangan data (x_i, y_i) , $i = 1, 2, \dots, n$.

Berikut ini, Tabel 1, adalah pedoman untuk memberikan interpretasi koefisien korelasi adalah sebagai berikut:

TABEL 1
TINGKAT KEERATAN KOEFISIEN KORELASI [8].

Interval Koefisien Korelasi	Tingkat Hubungan
0,00 - 0,199	Sangat rendah
0,20 - 0,399	Rendah
0,40 - 0,599	Sedang
0,60 - 0,799	Kuat
0,80 - 1,000	Sangat kuat

A. Persyaratan Sistem

Analisis dataset dilakukan menggunakan bahasa pemrograman berlisensi open source, yaitu Python. Pengolahan dataset dengan Python akan menggunakan *tools* Jupyter Lab yang ada pada Anaconda. Untuk melakukan EDA pada penelitian ini, Python akan menggunakan modul dari Pandas. Pandas merupakan pustaka Python yang menyediakan *tools* untuk keperluan analisis, pembersihan, dan manipulasi data [9].

Visualisasi data akan dilakukan dengan mengimpor modul dari Matplotlib dan Seaborn. Visualisasi data diistilahkan dengan plot (penggambaran). Pandas memiliki beberapa plot dasar (pandas: Powerful Python data analysis toolkit) di antaranya *bar*, *histogram*, *box*, *density*, *area*, *scatter*, *hexagonal bin*, dan *pie*. Setiap plot memiliki *data requirements* yang berbeda dan peruntukan yang berbeda. Perlu memperhatikan fungsi dan kegunaan dari setiap plot untuk bisa menghasilkan visualisasi data yang jelas dan baik

B. Tahapan Penelitian

Penelitian ini akan melakukan analisis menggunakan *Exploratory Data Analysis (EDA)*. Metode kerja yang digunakan dalam melakukan analisis adalah Scrum [10]. Scrum merupakan salah satu jenis metode pengembangan perangkat lunak *Agile Development* [11].

Agile memiliki prinsip atau nilai utama yang tercantum dalam *Agile Manifesto*:

1. Individu dan interaksinya jauh lebih penting daripada proses dan *tools* yang digunakan
2. Perangkat lunak yang benar-benar bekerja dibandingkan dokumentasi yang lengkap
3. Kolaborasi dengan pelanggan atau pengguna jauh lebih penting daripada negosiasi kontrak
4. Responsif terhadap perubahan daripada mengikuti dengan kaku rencana awal

Metode ini dianggap sangat tepat digunakan untuk penelitian ini karena penelitian ini dikerjakan secara kolaborasi dengan dataset yang masih perlu ditelusuri lebih jauh. Penelusuran ini dinilai akan mempengaruhi rencana pengerjaan yang dirancang dalam penelitian.

Lebih lanjut penerapan Scrum dilakukan dengan menentukan *product backlog* apa saja yang akan dihasilkan. *Product backlog* ini kemudian diturunkan menjadi *Sprint Backlog*. Setiap *Sprint Backlog* akan diberikan *Sprint Points* sebagai estimasi bobot pengerjaan dari setiap *Sprint Backlog*. Mulai dari *Sprint Backlog* yang sudah diberi *Sprint Points* inilah kemudian dirancang tahapan-tahapan pengerjaannya [12]. Penerapan ini dijabarkan secara detail pada Tabel 2.

TABEL 2
PRODUCT BACKLOG, USER STORIES, SPRINT BACKLOG DAN SPRINT POINTS

No	Product Backlog	Sprint Backlog	Sprint Point
1	<i>Hypothetical</i> <i>User Story :</i> Sebagai seorang Peneliti, saya ingin membuat suatu pernyataan atau jawaban sementara dari masalah yang dihadapi yang akan dibuktikan dalam penelitian ini.	Membuat bisnis model	3
		Menentukan Variabel	5
2	<i>Collecting Data</i> <i>User Story :</i> Sebagai seorang Peneliti, saya ingin mendapatkan data perjalanan penumpang selama 2 bulan sebelum masa pandemi (Januari - Februari 2020) untuk mendapatkan gambaran kondisi nyata kepadatan penumpang pada masa normal	<i>Data Collection</i>	7
		<i>Penumpang Tap-In/Out</i> <i>Data Collection & Processing Jadwal Perjalanan KRL</i>	3
3	<i>Preprocessing Data</i> <i>User Story :</i> Sebagai seorang Peneliti, saya ingin mendapatkan gambaran jumlah penumpang keluar dan masuk stasiun pada rute TanahAbang - Rangkasbitung per menit dan pemetaan terhadap kapasitas KRL yang berjalan saat itu	<i>Data processing</i> perjalanan penumpang KRL	8
4	<i>Analysis</i> <i>User Story :</i> Sebagai seorang Peneliti, saya ingin melakukan analisis korelasi untuk menemukan variabel dengan korelasi tertinggi dari setiap variabel yang digunakan serta visualisasinya.	Membuat Variabel Turunan	8
		Mencari korelasi	10
		Visualisasi Data	3
5	<i>Conclusion</i> <i>User Story :</i> Sebagai seorang Peneliti. Saya ingin menemukan kesimpulan akhir dari variabel yang memiliki korelasi tertinggi, serta menyimpulkan validasi kebijakan KRL apa sudah cocok atau belum.	Menentukan Variabel dengan korelasi tertinggi	2
		Validasi kebijakan KRL yang sudah berjalan	2

Ada sepuluh tahapan yang akan dilakukan dalam menganalisis data. Tahapan tersebut adalah sebagai berikut:

- 1) Memuat dataset pertama
- 2) Mengidentifikasi dimensi dan tipe data
- 3) Melakukan pembersihan data
- 4) Mencari jumlah penumpang dari gabungan data tap in dengan jadwal keberangkatan kereta
- 5) Memuat dataset kedua
- 6) Mengelompokkan data
- 7) Membuat visualisasi kelompok data pertama dan kedua
- 8) Memuat dataset ketiga
- 9) Membuat visualisasi
- 10) Menemukan korelasi dan membuat visualisasinya

III. HASIL DAN PEMBAHASAN

A. Pengumpulan, Pembersihan dan Persiapan Dataset

Untuk mendapatkan gambaran keseluruhan dari perjalanan penumpang KRL selama masa sebelum pandemi Covid-19, harus dilakukan pengambilan data perjalanan dari basis data utama KRL yang saat ini tersimpan di Data Center PT Kereta Commuter Indonesia dalam bentuk basis data PostgreSQL dan selanjutnya untuk memudahkan proses pengolahan data lanjutan, maka format data akan diubah menjadi format *Comma Separated Value (csv)*. Proses ekstraksi dan *filtering* data

dimaksud menggunakan query SQL agar lebih mudah dan dapat memaksimalkan fungsi-fungsi store procedure yang merupakan fitur standar dari PostgreSQL. *Entity relationship diagram* (ERD) dari KRL dapat dilihat di Lampiran 2.

1) Pengumpulan Data

Query ke database utama untuk memfilter data antara Januari-Februari 2020, disimpan dalam bentuk csv. Gambar 3 menunjukkan perolehan 53.913.454 row data yang mencerminkan jumlah penumpang KRL dalam 2 bulan di atas.

```
-- RAW DATA GATE JAN-FEB 2020

SELECT * FROM ctm_test.t_d_gate_out A
WHERE
  A.d_gate_out BETWEEN '2020-01-01 03:00:00' :: TIMESTAMP WITHOUT TIME ZONE
  AND '2020-02-29 23:59:59' :: TIMESTAMP WITHOUT TIME ZONE
  AND A.c_status = 'S'
  -- AND A.c_desc IN ( '', 'AC', 'N' )
  -- AND A.i_card_type IN ( 1, 10 )
  -- AND A.i_issuer IN ( 1, 9 )
  -- AND A.c_station_in IN ( SELECT c_station FROM opr.t_m_station WHERE b_active = TRUE )
ORDER BY d_gate_out;
```

Gambar 3. Query filter data perjalanan Januari-Februari 2019

Sample raw data sebagai berikut:

```
"80495398","BJD ","GATE 04 A ","2303 ","1/1/2020 02:57:27","BJD ","GATE 05 B ","2354 ","1/1/2020
03:00:00","31FFD505304636345356","0","0","1","4","8001081801007189","31/1/2020","0","0","0","1/1/2013 00:00:00","
","3","S","NP","2119473206","2001051301010066","0","1/1/2020 03:13:41.842464"
"17106132","KLDB ","GATE 03 A ","2902 ","1/1/2020 02:56:53","KLDB ","GATE 09 B ","2958 ","1/1/2020
03:00:01","31FFD005344D34391971","0","0","1","5","8006081801000058","31/1/2020","0","0","0","1/1/2013 00:00:00","KLDB ","
","3","S","NM","2119474769","2001021301010158","0","1/1/2020 03:18:06.439779"
"80495399","BJD ","GATE 04 A ","2303 ","1/1/2020 02:57:27","BJD ","GATE 05 B ","2354 ","1/1/2020
03:00:04","31FFD505304636345356","0","0","1","4","8001081801007189","31/1/2020","0","0","0","1/1/2013 00:00:00","
","3","S","NP","2119473207","2001051301010066","0","1/1/2020 03:13:41.842464"
```

2) Pembersihan Data

Selanjutnya untuk mendapatkan data penumpang yang melakukan perjalanan spesifik pada lintas pelayanan Rangkasbitung - Tanah Abang maka terhadap data sebelumnya dilakukan query filtering lebih lanjut seperti pada Gambar 4.

```
-- FILTER LINTAS KULON

SELECT * FROM ctm_test.t_d_gate_out A
WHERE
  A.d_gate_out BETWEEN '2020-01-01 03:00:00' :: TIMESTAMP WITHOUT TIME ZONE
  AND '2020-02-29 23:59:59' :: TIMESTAMP WITHOUT TIME ZONE
  AND A.c_status = 'S'
  AND A.c_station_in IN ( 'THB', 'PLM', 'KBY', 'PDJ', 'JRU', 'SDM', 'RU', 'SRP', 'CSK', 'CCY', 'PRP', 'CJT', 'DAR', 'TEJ', 'TIG', 'CKY', 'MJ', 'CTR', 'RK' )
ORDER BY d_gate_out;
```

Gambar 4. Query filter data perjalanan lintas Rangkasbitung-Tanahabang

Berdasarkan filter di atas diperoleh 11.988.352 row data yang mencerminkan jumlah penumpang dalam 2 bulan di atas.

3) Persiapan Data

Terhadap seluruh data penumpang yang melakukan perjalanan di atas dilakukan transformasi menjadi “Jumlah penumpang masuk dan keluar per menit untuk setiap stasiun.” Dengan struktur data sebagai berikut :

- i_gate_permin* : index number
- c_station* : nama stasiun
- d_date* : tanggal
- t_minute* : menit
- vol_in* : jumlah penumpang TapIn
- vol_out* : jumlah penumpang TapOut

Diperoleh 1.641.602 row data yang mencerminkan perjalanan penumpang pada 19 stasiun setiap menit nya selama 2 bulan.

Sampling data hasil pengolahan sebagai berikut :

```
"1554781","THB","29/2/2020","17:00","117","10"  
"1554782","THB","29/2/2020","17:01","100","7"  
"1554783","THB","29/2/2020","17:02","115","19"  
"1554784","THB","29/2/2020","17:03","100","26"  
"1554785","THB","29/2/2020","17:04","112","18"  
"1554786","THB","29/2/2020","17:05","126","15"  
"1554787","THB","29/2/2020","17:06","111","29"  
"1554788","THB","29/2/2020","17:07","112","24"  
"1554789","THB","29/2/2020","17:08","85","6"  
"1554790","THB","29/2/2020","17:09","120","5"  
"1554791","THB","29/2/2020","17:10","113","10"  
"1554792","THB","29/2/2020","17:11","120","8"  
"1554793","THB","29/2/2020","17:12","136","5"  
"1554794","THB","29/2/2020","17:13","126","16"  
"1554795","THB","29/2/2020","17:14","116","16"
```

B. Implementasi dan Eksperimen

Penelitian ini akan menggunakan tiga dataset. Dataset Pertama adalah hasil dari pengumpulan, pembersihan dan persiapan dataset. Dataset pertama ini akan diolah pada Jupyter Lab dan hasilnya akan diolah lebih lanjut pada Microsoft Excel. Hasil pengolahan pada Microsoft Excel akan menghasilkan dua dataset yang kemudian akan disebut Dataset Kedua dan Dataset Ketiga. Dataset Kedua diolah kembali pada Jupyter Lab. Hasil dari pengolahan Dataset Kedua kemudian digabungkan dengan Dataset Ketiga. Hasil penggabungan ini akan diolah dan dianalisis untuk menghasilkan visualisasi dari korelasi variabel. Tahapan pengolahan dan analisis akan dijelaskan lebih rinci dibawah ini.

1) Memuat dataset pertama

Dataset pertama terdiri dari tiga bagian file berformat CSV. Ketiga data ini dimuat pada Jupyter Lab dan digabungkan menjadi satu kesatuan file berformat DataFrame. Gambar 5 menampilkan dataset pertama yang berhasil dimuat.

```
# gabungkan semua csv files yang terpisah ke dalam satu DataFrame dengan mengguna glob, dan pd.concat()  
krl = pd.concat((pd.read_csv(file)  
                for file in stock_files), ignore_index = True)  
krl
```

	t_d_gatein_gateout_arah_perminutes_result	c_station	d_date	t_minute	vol_in	vol_out	selisih	arah
0	1	CCY	1/1/2020	00:00	0	0	0	WRK
1	2	CCY	1/1/2020	00:00	1	0	1	THB
2	3	CCY	1/1/2020	00:01	0	0	0	WRK
3	4	CCY	1/1/2020	00:01	0	0	0	THB
4	5	CCY	1/1/2020	00:02	0	0	0	WRK
...

Gambar 5. Dataset Pertama yang baru saja dimuat pada Jupyter Lab

2) Mengidentifikasi dimensi dan tipe data

Dataset yang sudah dimuat selanjutnya diidentifikasi untuk lebih memahami dimensi dan bentuk dataset. Dataset terdiri dari delapan kolom dengan tipe data masing-masing kolom seperti ditampilkan pada Gambar 6.

```
[ ] kr1.dtypes  
  
t_d_gatein_gateout_arah_perminutes_result    int64  
c_station                                     object  
d_date                                         object  
t_minute                                       object  
vol_in                                         int64  
vol_out                                        int64  
selisih                                       int64  
arah                                           object  
dtype: object
```

Gambar 6. Informasi tipe data dari Dataset Pertama

3) Melakukan pembersihan data

Terdapat beberapa kolom yang perlu dibuang dan bisa gabungkan. Misalnya seperti table *t-d-gatein_getout_arah_perminutes_result* yang berisi hal yang sama seperti *index*. Kolom *selisih* juga dirasa kurang perlu. Kolom *d_date* dan *t_minutes* dapat digabungkan dan tipe datanya diubah menjadi *datetime*. Perlu juga menyeleksi baris yang hanya menampilkan arah == RK.

Hal ini dilakukan untuk membuat dataset lebih ringkas dan hanya menampilkan data yang benar-benar digunakan saja. Hasil dari pembersihan dapat dilihat pada Gambar 7.

	c_station	vol_in	vol_out	arah
timestamps				
2020-01-01 00:00:00	CCY	1	0	THB
2020-01-01 00:01:00	CCY	0	0	THB
2020-01-01 00:02:00	CCY	0	0	THB
2020-01-01 00:03:00	CCY	0	0	THB
2020-01-01 00:04:00	CCY	0	0	THB
...
2020-02-29 23:55:00	TIG	0	0	THB
2020-02-29 23:56:00	TIG	0	0	THB
2020-02-29 23:57:00	TIG	0	0	THB
2020-02-29 23:58:00	TIG	0	0	THB
2020-02-29 23:59:00	TIG	0	0	THB

1654412 rows x 4 columns

Gambar 7. Hasil Pembersihan Data Dataset Pertama

4) Mencari jumlah penumpang dari gabungan data tap in dengan jadwal keberangkatan kereta

Dataset pada Tahapan Ketiga selanjutnya diolah dan dianalisis menggunakan Microsoft Excel. Pada Microsoft Excel akan dilakukan beberapa hal:

- mengekstrak data berdasarkan waktu sample (13 - 19 Januari 2020).
- mencari rata-rata penumpang di setiap menit pada *range* waktu *sample*
- menggabungkan jadwal keberangkatan kereta di setiap stasiun dengan waktu tap in penumpang yang akan menjadi calon penumpang setiap kereta (kode kereta - waktu *tap in*)
- menjumlah penumpang pada setiap waktu tap in yang memiliki kode kereta yang sama dari setiap stasiun.

Gambar 8 menunjukkan potongan hasil dari penggabungan jadwal keberangkatan kereta setiap stasiun dengan waktu *tap-in* penumpang. Hasil ini akan dijadikan Dataset Kedua.

```
[ ] dataset2_df.head(10)
```

	c_station	t_minute	code_train	2020-01-13 00:00:00	2020-01-14 00:00:00	2020-01-15 00:00:00	2020-01-16 00:00:00	2020-01-17 00:00:00	2020-01-18 00:00:00	2020-01-19 00:00:00	Rata- rata
0	RK	00:00:00	1923.0	0	0	0	0	0	0	0	0.0
1	RK	00:01:00	1923.0	0	0	0	0	0	0	0	0.0
2	RK	00:02:00	1923.0	0	0	0	0	0	0	0	0.0
3	RK	00:03:00	1923.0	0	0	0	0	0	0	0	0.0
4	RK	00:04:00	1923.0	0	0	0	0	0	0	0	0.0
5	RK	00:05:00	1923.0	0	0	0	0	0	0	0	0.0
6	RK	00:06:00	1923.0	0	0	0	0	0	0	0	0.0
7	RK	00:07:00	1923.0	0	0	0	0	0	0	0	0.0
8	RK	00:08:00	1923.0	0	0	0	0	0	0	0	0.0
9	RK	00:09:00	1923.0	0	0	0	0	0	0	0	0.0

Gambar 8. Dataset Kedua

Sementara Gambar 9 menunjukkan potongan hasil penjumlahan penumpang pada setiap waktu *tap-in* pada kereta yang sama dari setiap stasiun. Hasil ini akan dijadikan Dataset Ketiga.

	Kode Kereta	Jam Keberangkatan	Jumlah Penumpang
0	1923	04:00:00	1391
1	1931	04:30:00	1139
2	1937	05:00:00	2048
3	1943	05:30:00	2243
4	1949	06:00:00	2987
5	1955	06:30:00	2352
6	1961	07:00:00	1778
7	1967	07:30:00	1530
8	1973	08:00:00	1219
9	1979	08:30:00	873
10	1985	09:00:00	1045
11	1991	09:30:00	676
12	1997	10:00:00	953
13	2003	10:30:00	638
14	2011	11:30:00	635
15	2017	12:00:00	674
16	2023	12:30:00	502
17	2029	13:00:00	553
18	2035	13:30:00	463
19	2039	14:00:00	560
20	2045	14:30:00	627
21	2051	15:00:00	529
22	2057	15:30:00	415
23	2063	16:00:00	627
24	2069	16:30:00	521
25	2075	17:00:00	609
26	2081	17:30:00	361
27	2087	18:00:00	517
28	2093	18:30:00	349
29	2101	19:30:00	238
30	2107	20:00:00	301

Gambar 9. Dataset Ketiga

5) Memuat dataset kedua

Dataset Kedua akan dimuat kembali dalam Jupyter Lab. Selanjutnya dilakukan pembersihan data dari *missing values* dan mengubah tipe data tanggal dari *object* ke *numeric*. Hasil pembersihan tersebut dapat dilihat pada Gambar 10.

```
covid[covid.isna()]
covid.dropna(inplace=True)

cols = ['1/13/2020', '1/14/2020', '1/15/2020', '1/16/2020', '1/17/2020', '1/18/2020', '1/19/2020',
'Rata-rata']

covid[cols]=
covid[cols].apply(pd.to_numeric,errors='coerce', axis=1)
covid[cols] = covid[cols].astype(np.int64)
```

```
# menampilkan isi seluruh dataset
krl
```

	c_station	t_minute	train	1/13/2020	1/14/2020	1/15/2020	1/16/2020	1/17/2020	1/18/2020	1/19/2020	Rata-rata
0	RK	12:00:00 AM	1923	0	0	0	0	0	0	0	0
1	RK	12:01:00 AM	1923	0	0	0	0	0	0	0	0
2	RK	12:02:00 AM	1923	0	0	0	0	0	0	0	0
3	RK	12:03:00 AM	1923	0	0	0	0	0	0	0	0
4	RK	12:04:00 AM	1923	0	0	0	0	0	0	0	0
...
25800	PLM	9:43:00 PM	2107	0	0	0	0	0	0	0	0
25801	PLM	9:44:00 PM	2107	0	0	0	0	0	0	0	0
25802	PLM	9:45:00 PM	2107	0	0	0	0	0	0	0	0
25803	PLM	9:46:00 PM	2107	0	0	0	0	0	0	0	0
25804	PLM	9:47:00 PM	2107	0	0	0	0	0	0	0	0

10396 rows x 11 columns

Gambar 10. Hasil Pembersihan Dataset Kedua

6) Mengelompokkan data

Dataset yang sudah bersih tersebut selanjutnya diolah lagi dengan cara mengelompokkan kereta dan stasiun berdasarkan total jumlah penumpang. Pengelompokan data ini seperti yang ditunjukkan pada Gambar 11, dilakukan untuk melihat kondisi kepadatan jumlah penumpang pada setiap kereta di stasiun-stasiun yang dilewati kereta tersebut.

```
Penumpang = krl.groupby(['train', 'c_station']).agg({'Rata-rata': np.sum})
penumpang = penumpang.reset_index()
print(penumpang)

penumpang.sort_values(by='Rata-rata', ascending=False)
```

	train	c_station	Rata-rata
50	1937	SDM	496
83	1949	PRP	480
86	1949	SDM	385
104	1955	SDM	375
68	1943	SDM	358
...

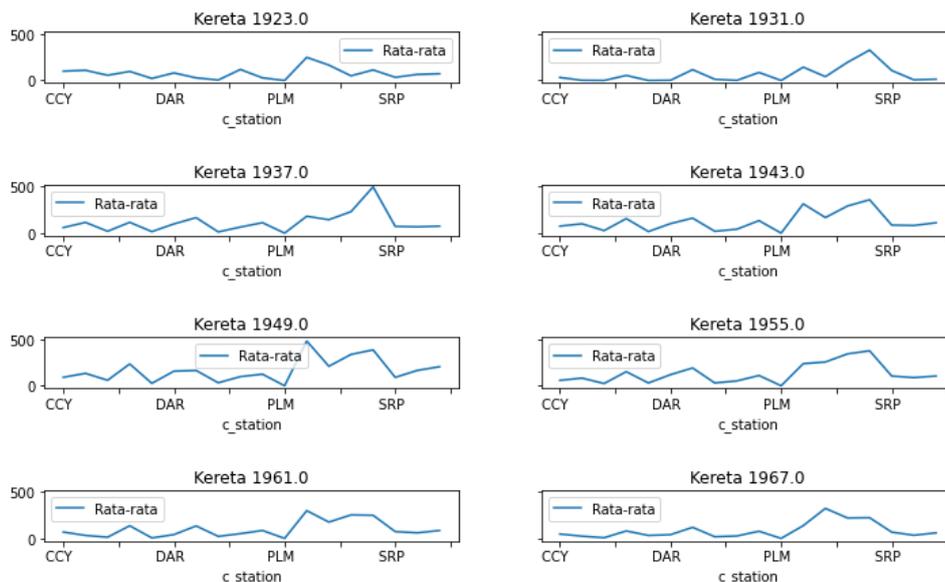
Gambar 11. Hasil Pengelompokan Dataset Kedua

7) Membuat visualisasi kelompok Dataset Pertama dan Kedua

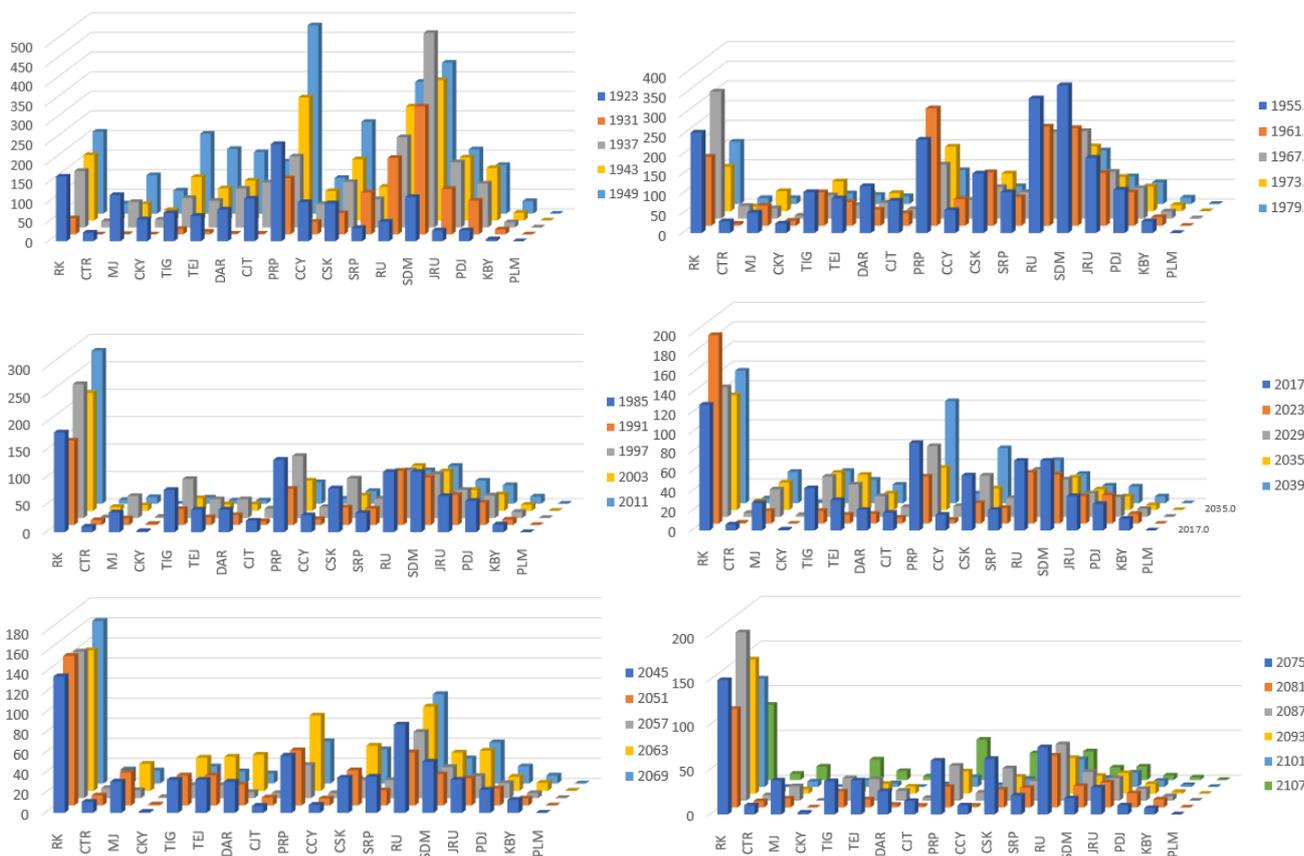
Hasil pengelompokan ini kemudian dibuat visualisasinya untuk memberikan gambaran lebih jelas. Namun visualisasi ini memiliki tidak menampilkan urutan stasiun yang tepat. Ini menjadi salah satu kelemahan dari pengelompokan data yang dilakukan.

```
grouped = penumpang.groupby('train')
ncols=2
nrows = int(np.ceil(grouped.ngroups/ncols))
fig, axes = plt.subplots(nrows=nrows, ncols=ncols, figsize=(12,32), sharey=True)
fig.subplots_adjust(hspace=2)

for (key, ax) in zip(grouped.groups.keys(), axes.flatten()):
    grouped.get_group(key).plot(x='c_station', y='Rata-rata', ax=ax)
    ax.set_title("Kereta {}".format(key))
ax.legend(loc = 'best')
plt.savefig('./dataset/baru/test.png')
plt.show()
```



Gambar 12. Visualisasi terpisah Jumlah Penumpang setiap Kereta pada stasiun-stasiun yang dilewati



Gambar 13. Visualisasi gabungan Jumlah Penumpang setiap Kereta pada stasiun-stasiun yang dilewati. Sumbu vertikal adalah rata-rata kepadatan penumpang, dan sumbu horizontal adalah kode-kode stasiun. Setiap warna mewakili kode kereta.

Gambar 12 dan Gambar 13 menunjukkan visualisasi Jumlah Penumpang setiap Kereta pada stasiun-stasiun yang dilewati. Gambar 12 menunjukkan cuplikan visualisasi kondisi dari setiap kereta sementara Gambar 13 menampilkan visualisasi keseluruhan. Selain stasiun RK (Rangkasbitung) yang merupakan stasiun awal keberangkatan setiap kereta, ada stasiun

yang hampir selalu mengalami peningkatan jumlah penumpang yang tinggi, yaitu SDM (Sudimara). Rata-rata per hari jumlah penumpang yang diangkut dari stasiun SDM adalah:

- Kereta 1937 : 496 penumpang (Pukul 06:25)
- Kereta 1949 : 385 penumpang (Pukul 07:25)
- Kereta 1955 : 375 penumpang (Pukul 07:55)
- Kereta 1943 : 358 penumpang (Pukul 06:55)

8) *Memuat Dataset Ketiga*

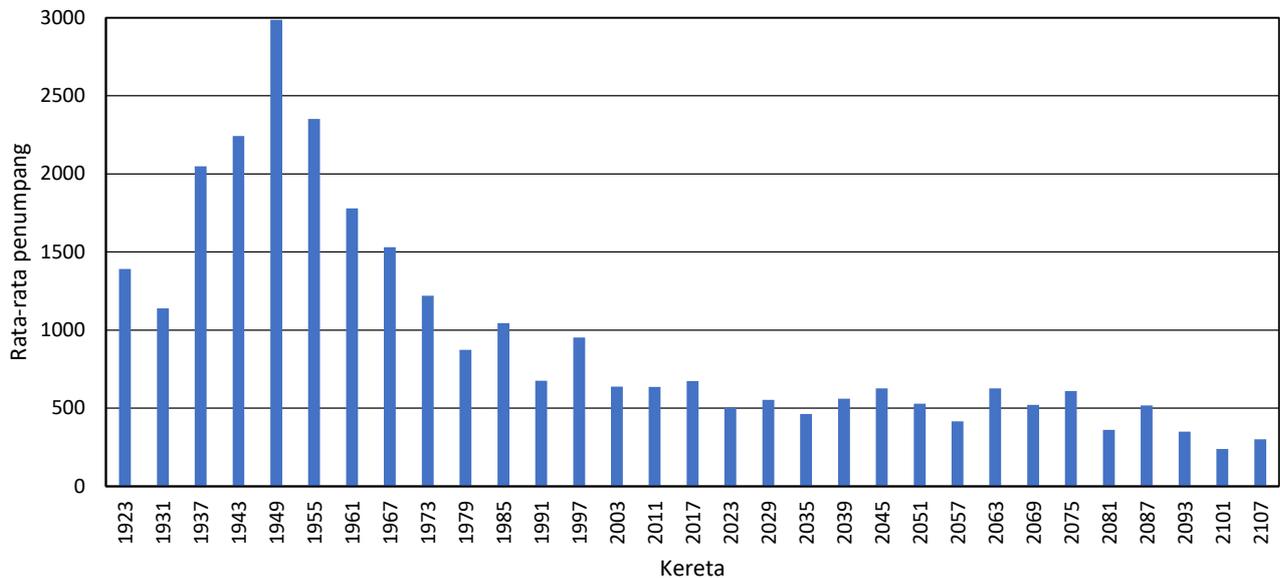
Jika pada Tahapan 6 pengelompokan dilakukan untuk kereta dan stasiun berdasarkan jumlah penumpang, maka pada Tahapan 8 ini pengelompokan yang dilakukan adalah mengelompokkan kereta berdasarkan total jumlah penumpang saja. Hasil pengelompokan ini disimpan dalam dataset bernama *kereta*. Selanjutnya memuat Dataset Ketiga yaitu Jadwal Kereta. Dataset *kereta* digabungkan dengan Dataset Ketiga. Hasilnya ditunjukkan pada Gambar 14.

	train	Rata-rata	departure	interval
0	1923	1391	19	0:00:00
1	1931	1139	20	0:30:00
2	1937	2048	21	0:30:00
3	1943	2243	22	0:30:00
4	1949	2987	23	0:30:00
5	1955	2352	24	0:30:00

Gambar 14. Dataset Ketiga yang dimuat ke Jupyter Lab

9) *Membuat visualisasi*

Tahapan ini akan membuat visualisasi jenis kereta dengan jumlah rata-rata penumpang setiap hari yang diangkut seperti yang ditampilkan pada Gambar 15.



Gambar 15. Visualisasi Rata-Rata Kepadatan Kereta

Pada visualisasi dapat dilihat bahwa kepadatan penumpang melewati angka 2000 dimulai dari kereta dengan Kode 1937. Mencapai puncak tertinggi pada kereta dengan Kode 1949. Kode ini juga menggambarkan waktu kepadatan tersebut terjadi. Kereta dengan Kode 1937 mulai berangkat pada pukul 05:00 WIB sementara Kode 1949 berangkat pada pukul 06:00. Sampai kereta dengan Kode 2003 yang berangkat mulai pukul 10:30, kepadatan penumpang masih diatas 500. Kondisi ini perlu menjadi perhatian khusus jika ingin dibandingkan dengan anjuran pemerintah yang hanya mengijinkan 720 penumpang dalam satu kereta.

10) Menemukan korelasi dan membuat visualisasinya

Tahapan terakhir yang dilakukan pada penelitian ini adalah membuktikan hipotesis dengan mencari nilai korelasi antara jadwal keberangkatan kereta dengan kepadatan atau jumlah penumpang pada setiap kereta. Hal ini dilakukan dengan terlebih dahulu mengubah jadwal kereta ke dalam format kategori yang *numeric*. Setelah itu dilakukan regresi linier [13] dan eksponensial [14] [15] yang hasilnya masing-masing adalah $r = -0.7915$ (linier) dan $r = 0.8975$ (eksponensial) atau $r^2 = 0.6265$ (linier) dan $r^2 = 0.6696$ (eksponensial). Persamaan regresi dengan *hyperparameter*-nya yang didapat adalah:

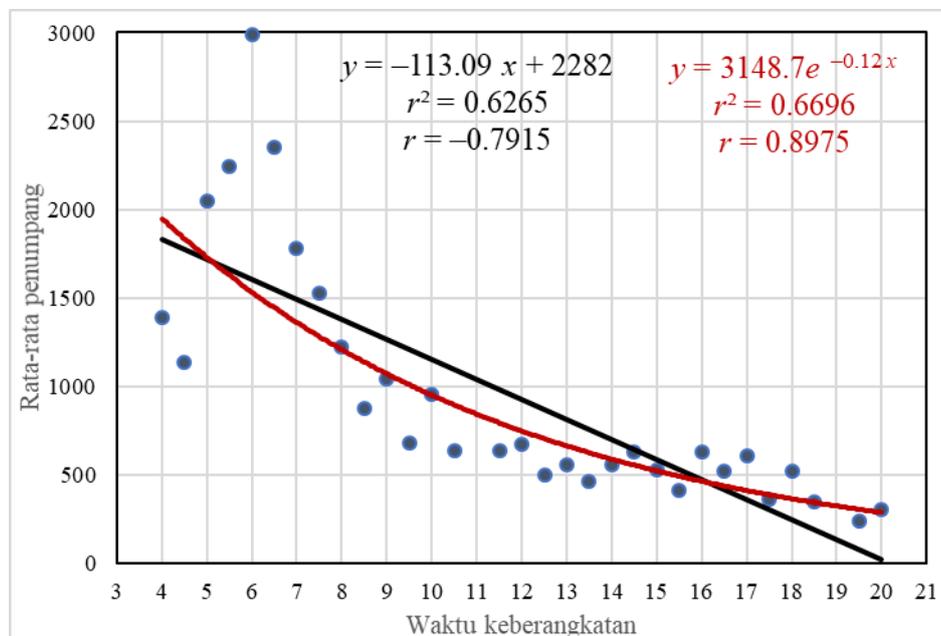
$$y = -113.09 x + 2282 \quad (2)$$

untuk regresi linier, dan

$$y = 3148.7 e^{-0.12 x} \quad (3)$$

untuk regresi eksponensial, dengan x adalah waktu keberangkatan dan y adalah banyaknya penumpang, yang ditunjukkan pada Gambar 16.

Menurut Korelasi Pearson, hasil ini menunjukkan adanya korelasi yang kuat antara jadwal keberangkatan kereta dengan jumlah penumpangnya, dimana tanda negatif pada korelasi linier menunjukkan bahwa semakin sore jumlah penumpang semakin menurun, dan korelasi yang kuat untuk regresi eksponensial.



Gambar 16. Regresi linier (hitam) dan regresi eksponensial (merah) untuk rata-rata penumpang tiap kereta terhadap waktu keberangkatan.

C. Jawaban Terhadap Rumusan Masalah dan Hipotesis

Penelitian ini memberikan gambaran bahwa jumlah total penumpang yang melebihi protokol kesehatan mulai terjadi dari pukul 04:00-10:00. Kepadatan penumpang tertinggi terjadi pada rangkaian Kereta 1949 yang berangkat pada pukul 06:00.

Hasil korelasi antara Jadwal Kereta dengan Jumlah Penumpang sebesar $r = -0.7915$ (linier) dan $r = 0.8975$ (eksponensial). Merujuk Korelasi Pearson [8], hal ini menunjukkan adanya korelasi kuat (linier) dan sangat kuat (eksponensial).

Berdasarkan Permenhub Nomor 18/2020 tentang pencegahan penyebaran Covid-19, maka jumlah penumpang per gerbong dibatasi maksimal 74 orang, sehingga untuk satu rangkaian kereta maksimal 740 orang (10 gerbong). Perlunya penambahan jadwal keberangkatan dari rentang 04:00-10:00 untuk memenuhi persyaratan tersebut. Maka rekomendasi penambahan jadwal yang dapat diberikan adalah seperti yang ditampilkan pada Gambar 17.

04:00	04:30	1 Kereta
04:30	05:00	1 Kereta
05:00	05:30	2 Kereta
05:30	06:00	2 Kereta
06:00	06:30	3 Kereta
06:30	07:00	3 Kereta
07:00	07:30	2 Kereta
07:30	08:00	1 Kereta
08:00	08:30	1 Kereta
08:30	09:00	1 Kereta
09:00	09:30	1 Kereta
09:30	10:00	-
10:00	10:30	1 Kereta

Gambar 17 Rekomendasi Penambahan Jadwal

D. Keunggulan

Hasil dari penelitian ini dapat menentukan interval waktu perjalanan yang tepat untuk mengantisipasi kepadatan penumpang pada jam-jam tertentu sesuai dengan pola perjalanan penumpang KRL. Pola perjalanan dan kepadatan penumpang pada stasiun tertentu dapat menjadi dasar dalam menentukan awal keberangkatan KRL. Jika diterapkan pada seluruh lintas layanan KRL dapat menjadi dasar bagi optimalisasi perjalanan untuk mengantisipasi kepadatan penumpang secara keseluruhan

E. Kelemahan

Pada prosesnya, terjadi kesulitan melakukan penggabungan pada dataset awal dengan jadwal keberangkatan kereta sehingga harus dilakukan pengolahan melakukan Microsoft Excel daripada menggunakan Anaconda.

IV. SIMPULAN

Jadwal kereta memiliki korelasi kuat (linier) dan sangat kuat (eksponensial), terhadap kepadatan penumpang pada setiap kereta KRL. Perlunya penelitian lebih lanjut untuk menemukan variabel yang paling mempengaruhi kepadatan kereta seperti demografi wilayah setiap stasiun atau hal lainnya seperti kebiasaan penumpang pada hari-hari tertentu. Mengingat penelitian ini masih berdasarkan data rata-rata harian, maka penelitian lebih lanjut dapat dilakukan untuk mendapatkan gambaran lebih detail pola perjalanan penumpang berdasarkan tren harian untuk mendapatkan rekomendasi yang lebih baik, selain itu jika diterapkan pada seluruh lintas pelayanan KRL dapat digunakan untuk optimalisasi penggunaan sarana KRL

UCAPAN TERIMA KASIH

Ucapan terima kasih kami sampaikan kepada dosen-dosen yang mengajar pada Semester 1: Bapak Hapnes Toba sebagai dosen pengampu mata kuliah Pengelolaan Data Skala Besar, Ibu Mewati Ayub sebagai dosen pengampu mata kuliah Rekayasa Perangkat Lunak, Bapak Setia Budi sebagai pengampu mata kuliah Teknologi *Open Source*, dan Bapak Bernard Renaldy Suteja sebagai dosen pengampu mata kuliah Web Semantik.

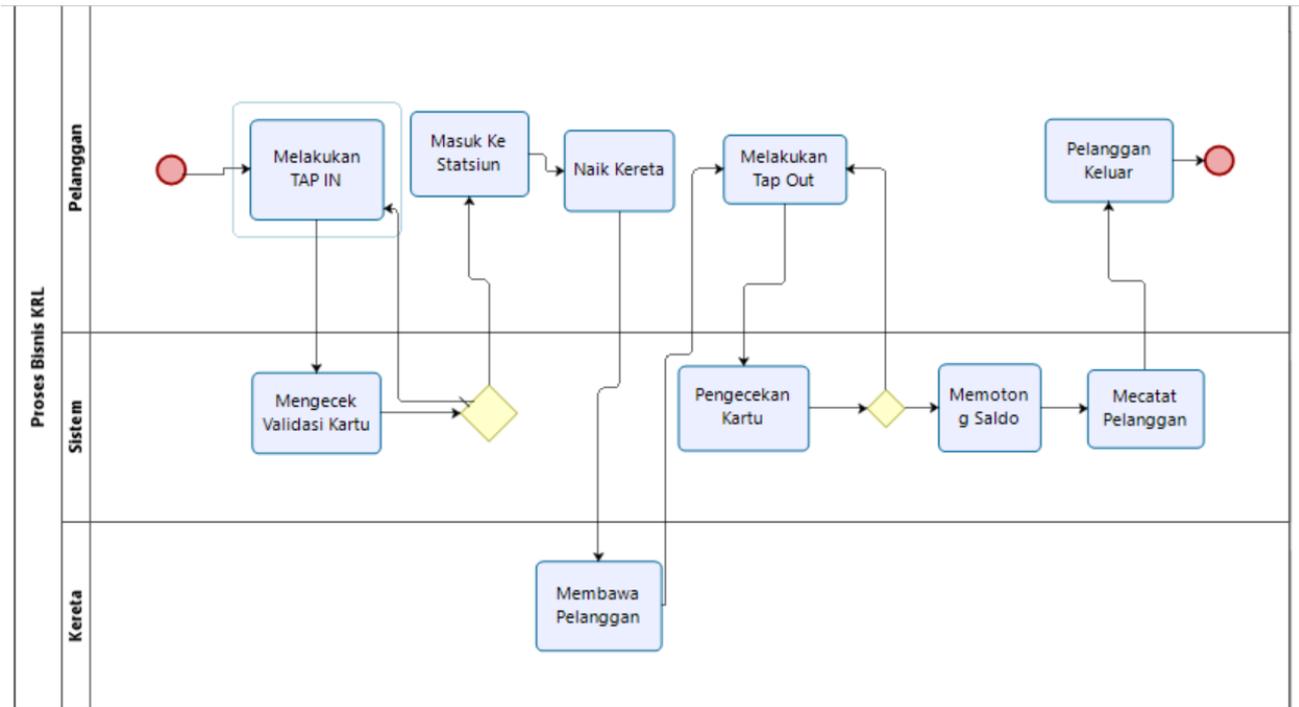
DAFTAR PUSTAKA

- [1] L. Wright and K. Fjellstrom, "Modul 3a," in *Sustainable Urban Transport Sourcebook for Policy-Maker in Developing Cities*, Germany, TZ Verlagsgesellschaft, June 2003.
- [2] P. K. C. Indonesia, "Perjalanan KRL," PT Kereta Commuter Indonesia, [Online]. Available: <http://www.krl.co.id>. [Accessed 4 Jan 2020].
- [3] Kemenhub, "Peraturan Menteri Perhubungan Nomor 18 tahun 2020 tentang Pengendalian Transportasi dalam Rangka Pencegahan Penyebaran Corona Virus Disease 2019 (COVID-19) dan perubahannya yang dituangkan dalam Peraturan Menteri Perhubungan Nomor 41 tahun 2020," Kementerian Perhubungan, Jakarta, Indonesia, 2020.
- [4] G. Leinhardt and S. Leinhardt, "Exploratory Data Analysis: New Tools for the Analysis of Empirical Data," *Review of Research in Education*, vol. 8, p. 85–157, 1980.

- [5] G. L. Taboada and L. Han, "Exploratory Data Analysis and Data Envelopment Analysis of Urban Rail Transit," *Electronics*: MDPI AG, 2020.
- [6] Rodgers and Nicewander, "Thirteen ways to look at the correlation coefficient," *The American Statistician*, vol. 42, no. 1, p. 59–66, 1988.
- [7] G. J. R. Székely and N. K. Bakirov, "Measuring and testing independence by correlation of distances," *Annals of Statistics*, vol. 35, no. 6, p. 2769–2794, 2007.
- [8] Sugiyono, *Statistika untuk Penelitian*, Bandung: Alfabeta, 2000.
- [9] W. McKinney, "Pandas: Powerful Python data analysis toolkit. Release 1.1.4," The Pandas Development Team, 2020.
- [10] J. Sutherland and K. Schwaber, "Scrum Guides," [Online]. Available: <https://scrumguides.org>. [Accessed 4 Jun 2020].
- [11] Eugeneyan, "Data Science and Agile (What works, and what doesn't)," [Online]. Available: <https://eugeneyan.com/writing/data-science-and-agile-what-works-and-what-doesnt>. [Accessed 4 Jan 2020].
- [12] M. Hron and N. Obwegeser, "Scrum in practice: An overview of Scrum adaptations," in *Proceedings of the 2018 51st Hawaii International Conference on System Sciences (HICSS)*, Hawaii, 3-6 January 2018.
- [13] X. Yan, *Linear Regression Analysis: Theory and Computing*, Singapore: World Scientific, 2009.
- [14] D. A. Freedman, *Statistical Models: Theory and Practice*, Cambridge: Cambridge University Press, 2009.
- [15] S. van de Geer, "A New Approach to Least-Squares Estimation, with Applications," *Annals of Statistics*, vol. 15, no. 2, p. 587–602, June 1987.

Lampiran

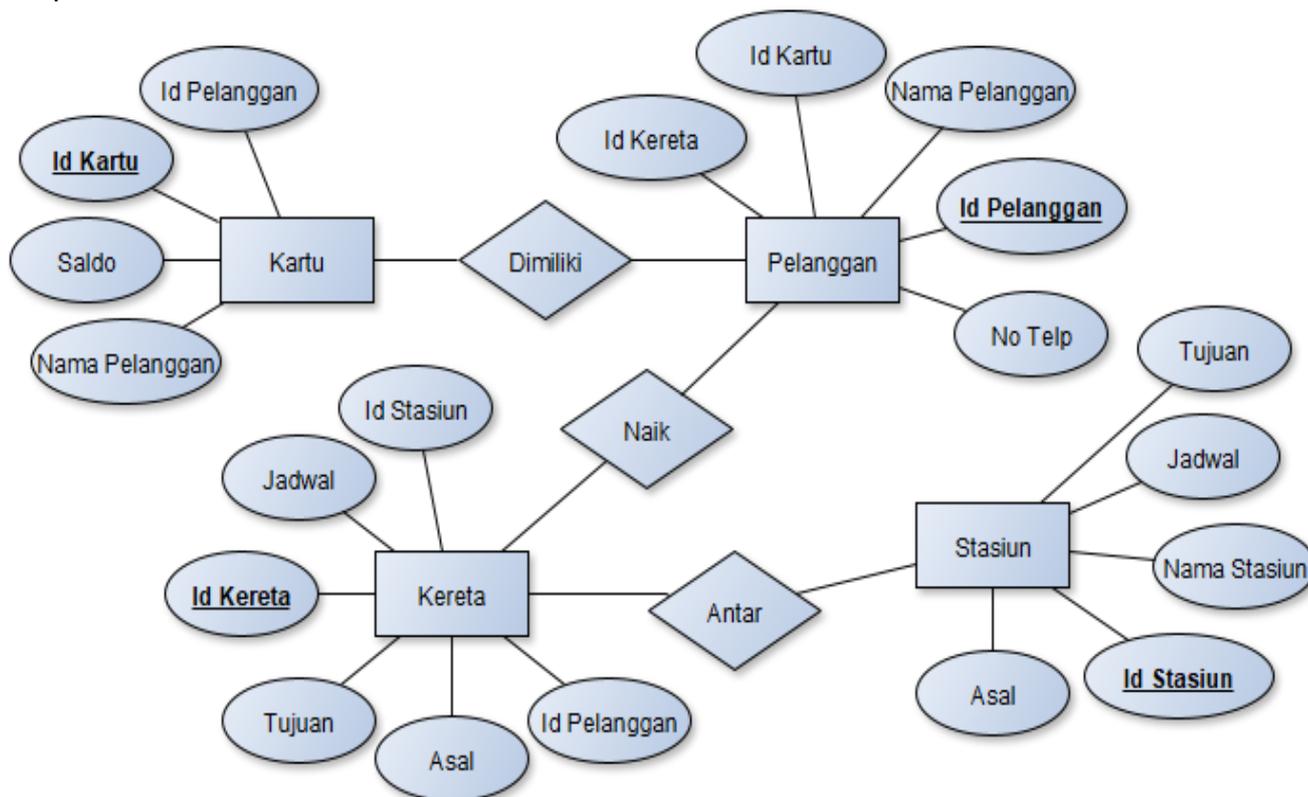
Lampiran 1: Proses Bisnis KRL



Alur Proses Bisnis KRL:

1. Pelanggan melakukan TAP IN
2. Sistem mengecek validasi kartu
3. Jika benar maka pelanggan akan masuk ke stasiun dan jika tidak maka pelanggan tidak bisa masuk
4. Pelanggan naik kereta
5. Pelanggan melakukan TAP OUT
6. Sistem mengecek kartu
7. Sistem memotong saldo jika benar
8. Sistem melakukan pencatatan pelanggan
9. Pelanggan keluar
10. Proses Selesai

Lampiran 2: ERD KRL



Kamus Data :

1. Kartu : (Id Kartu, Id Pelanggan, Saldo, Nama Pelanggan)
2. Kereta : (Id Kereta, Id Stasiun, Jadwal, Tujuan, Asal, Id Pelanggan)
3. Stasiun : (Id Stasiun, Asal, Nama Stasiun, Jadwal, Tujuan)
4. Pelanggan : (Id Pelanggan, Nama Pelanggan, No. Telp, Id Kartu, Id Kereta)

KARTU				
No	Nama Entity	Format Entity	Tipe Data	Key
1	<u>Id Kartu</u>	157988542357	INT	Primary Key
2	Id Pelanggan	197581442604	INT	Key
3	Saldo	50.000	Float	Key
4	Nama Pelanggan	Loudry	Varchar	Key

KERETA				
No	Nama Entity	Format Entity	Tipe Data	Key
1	<u>Id Kereta</u>	01	Float	Primary Key
2	Id Stasiun	001	Float	Key
3	Jadwal	15.00 WIB	TIME	Key
4	Tujuan	Tanah Abang	Varchar	Key
5	Asal	Rangkas Bitung	Varchar	
6	Id Pelanggan	197581442604	INT	Foreign Key

STASIUN				
No	Nama Entity	Format Entity	Tipe Data	Key
1	Id Stasiun	01	Float	Primary Key
2	Asal	001	Float	Key
3	Jadwal	15.00 WIB	TIME	Key
4	Tujuan	Tanah Abang	Varchar	Key
5	Nama Stasiun	Rangkas Bitung	Varchar	Key

PELANGGAN				
No	Nama Entity	Format Entity	Tipe Data	Key
1	Id Pelanggan	157988542357	INT	Primary Key
2	Id Kartu	197581442604	INT	Foreign Key
3	No. Telp	082118156753	Float	Key
4	Nama Pelanggan	Loudry	Varchar	Key
5	Id Kereta	01	Float	Foreign Key