

**IDENTIFIKASI PENGELOMPOKAN TITIK PENJEMPUTAN
DAN TITIK PENGANTARAN PERJALANAN TAKSI
MENGUNAKAN ALGORITMA DBSCAN**



Disusun Oleh:

N a m a : Anisa Salsabila

NIM : 19523186

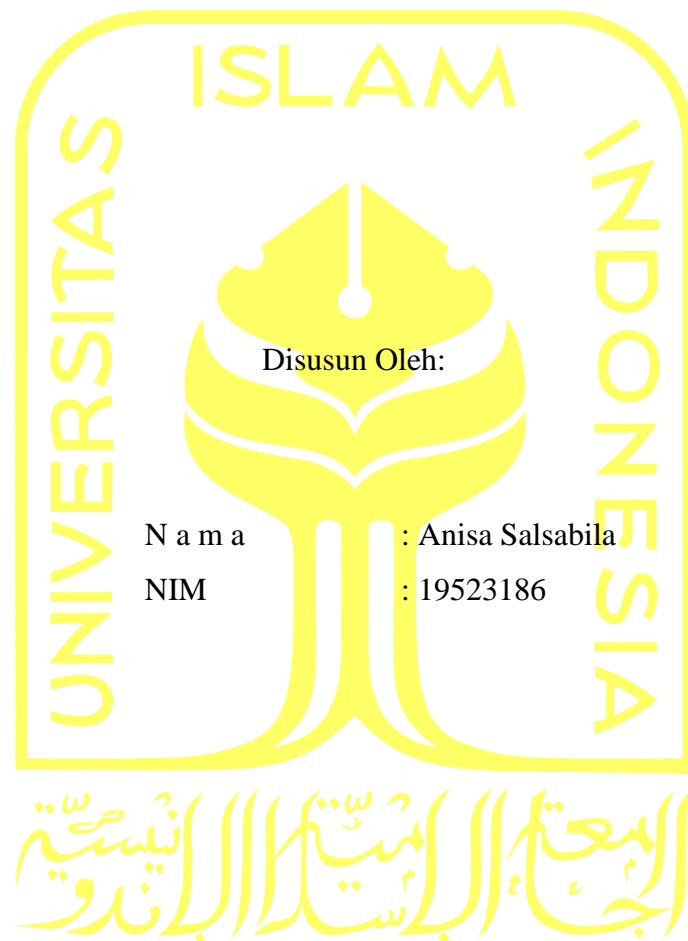
**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA**

2023

HALAMAN PENGESAHAN DOSEN PEMBIMBING

**IDENTIFIKASI PENGELOMPOKAN TITIK PENJEMPUTAN
DAN TITIK PENGANTARAN PERJALANAN TAKSI
MENGUNAKAN ALGORITMA DBSCAN**

TUGAS AKHIR



Yogyakarta, 15 Oktober 2023

Pembimbing,

(Lizda Iswari, S.T, M.Sc.)

HALAMAN PENGESAHAN DOSEN PENGUJI

**IDENTIFIKASI PENGELOMPOKAN TITIK PENJEMPUTAN
DAN TITIK PENGANTARAN PERJALANAN TAKSI
MENGUNAKAN ALGORITMA DBSCAN**

TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 30 Oktober 2023

Tim Peguji

Lizda Iswari, S.T., M.Sc.

Anggota 1

Elyza Gustri Wahyuni, S.T., M.Cs.

Anggota 2

Sheila Nurul Huda, S.Kom., M.Cs.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana
Fakultas Teknologi Industri
Universitas Islam Indonesia



(Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Anisa Salsabila

NIM : 19523186

Tugas akhir dengan judul:

IDENTIFIKASI PENGELOMPOKAN TITIK PENJEMPUTAN DAN TITIK PENGANTARAN PERJALANAN TAKSI MENGUNAKAN ALGORITMA DBSCAN

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 10 Oktober 2023



(Anisa Salsabila)

HALAMAN PERSEMBAHAN

Alhamdulillahirabbil'amin.

Puji dan syukur senantiasa ku ucapkan kepada-Mu, Allah Yang Maha Pemurah. Terima kasih telah memberikanku rezeki, kesehatan, dan kesempatan untuk setiap perjalananku.

Tugas akhir ini kupersembahkan kepada:

Papa Jang Candra dan Bunda Asih

Terima kasih untuk selalu berjuang dalam menafkahi berpuluh tahun lamanya, selalu sabar, menyayangi, dan senantiasa mendoakan tanpa batas dalam mendukung Adek.

Uda Habiib

Teman sejak kecil yang selalu menghibur dan memberikan banyak pelajaran saat tumbuh kembang bersama.

HALAMAN MOTO

“Maka sesungguhnya bersama kesulitan ada kemudahan, sesungguhnya bersama kesulitan ada kemudahan. Maka apabila engkau telah selesai (dari sesuatu urusan), tetaplah bekerja keras (untuk urusan yang lain), dan hanya kepada Tuhanmulah engkau berharap.”

QS. Asy-Syarah: 5 - 8

“The more I read, the more I acquire, the more certain I am that I know nothing.”

- Voltaire

“My youth blossomed when I was stuck with my code.

Someday, it will lead me to a beautiful flower garden for my maturity.”

- Anisa Salsa

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillah. Puji dan syukur senantiasa penulis ucapkan kepada Allah Subhanahu Wa Ta'ala atas berkat rahmat dan nikmat-Nya, sehingga penulisan laporan tugas akhir yang berjudul "Identifikasi Pengelompokan Titik Penjemputan dan Titik Pengantaran Perjalanan Taksi Menggunakan Algoritma DBSCAN" dapat diselesaikan.

Tujuan penulisan laporan tugas akhir ini menjadi salah satu syarat memperoleh gelar Sarjana Informatika di Universitas Islam Indonesia. Laporan tugas akhir ini mengandung keselarasan ilmu teori dan praktik yang telah dipelajari selama menjalani studi. Penulisan laporan tugas akhir tidak lepas dari dukungan dalam bentuk bimbingan dan saran dari banyak pihak. Dengan rendah hati, penulis ingin menyampaikan rasa terima kasih kepada:

1. Kedua orang tua dan kakak penulis yang senantiasa memberikan motivasi, doa, dan kepercayaan kepada penulis.
2. Bapak Dr. Raden Teduh Dirgahayu, S.T., M.Sc., selaku Ketua Jurusan Informatika Fakultas Teknologi Industri Universitas Islam Indonesia.
3. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D., selaku Ketua Program Studi Informatika Program Sarjana Fakultas Teknologi Industri Universitas Islam Indonesia.
4. Ibu Aridhanyati Arifin, S.T., M.Cs., selaku Dosen Pembimbing Akademik.
5. Ibu Lizda Iswari, S.T., M.Sc. selaku Dosen Pembimbing yang telah mengajarkan ilmu pengetahuan serta meluangkan banyak waktu dan kesabaran untuk membimbing penulis sehingga tugas akhir ini dapat diselesaikan.
6. Sahabat-sahabat penulis diantaranya Zola Aghna Vabiola, Kania Alyarani Fortuna, Ratih Putri Nabila, dan Putri Yanda Aulia. Terima kasih telah menemani penulis sejak sekolah menengah hingga saat ini.
7. Teman sejati, Muhammad Fachrezi Munawar dan Timmy The Moon. Terima kasih untuk dukungan, kebaikan hati, dan hiburan tanpa batas waktu.
8. Sahabat baik penulis diantaranya Putri, Rani, Tania, Ara, Verrin, Ici, Tia, dan Trimbi. Semoga perjalanan yang indah mengiringi sahabat di masa sekolah menengah penulis.
9. Teman-teman seperantauan SERGAP 19 dan C4E. Terima kasih atas kesedian waktu dan rasa kekeluargaan yang menemani penulis selama masa perkuliahan.

Penulis menyadari bahwa penelitian tugas akhir ini belum mencapai sempurna. Untuk itu, penulis menerima kritik dan saran yang bermanfaat untuk penelitian selanjutnya di masa

mendatang. Penulis mengucapkan terima kasih dan berharap laporan tugas akhir ini dapat bermanfaat bagi para pembaca.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Yogyakarta, 10 Oktober 2023

A handwritten signature in black ink, appearing to read 'Anisa Salsabila', written in a cursive style.

(Anisa Salsabila)

SARI

Pesatnya perkembangan daerah perkotaan dalam beberapa tahun terakhir, menjadikan rutinitas harian perjalanan penduduk kota dan transportasi umum saling mempengaruhi dan berkaitan erat. Usaha armada taksi sangat berkaitan erat dengan kepuasan pelayanan kepada penumpang mulai dari penjemputan hingga pengantaran ke tempat tujuan. Waktu tunggu taksi menjadi hal yang krusial bagi pengelola taksi untuk mencari lokasi penjemputan yang tepat di lingkungan perkotaan agar tidak membuat penumpang menunggu dengan waktu yang lama. Dengan melakukan identifikasi sejarah permintaan penumpang taksi secara akurat, hal tersebut dapat membantu pengelola taksi mengalokasikan sumber daya taksi pada daerah perkotaan terlebih dahulu. Dengan memanfaatkan algoritma DBSCAN terhadap data perjalanan taksi, maka penelitian ini bertujuan untuk mengidentifikasi *cluster* yang terbentuk dari titik penjemputan dan titik pengantaran penumpang. Data diperoleh melalui Kaggle dan penelitian akan berfokus pada perjalanan taksi di Kota Brooklyn. Data akan dipisahkan menjadi beberapa bagian berdasarkan waktu diantaranya perjalanan taksi yang terjadi pada hari kerja, hari libur, pagi, siang, sore, malam, dan tengah malam. Pada titik penjemputan penumpang, area potensial terdapat di wilayah utara dan barat laut Brooklyn. Sedangkan pada titik pengantaran penumpang, area potensial terjadi di Brooklyn dan Midtown dan Lower Manhattan.

Kata kunci: Area potensial, titik penjemputan, titik pengantaran, *cluster*, DBSCAN.

GLOSARIUM

<i>Cluster</i>	sekumpulan objek yang memiliki kesamaan dalam suatu himpunan.
<i>Clustering</i>	metode yang digunakan untuk memisahkan objek yang tidak serupa dan akan menghasilkan beberapa <i>cluster</i> .
<i>Dataframe</i>	struktur data yang memiliki dua dimensi.
<i>Dataset</i>	kumpulan data yang terletak pada tabel.
<i>Eps</i>	singkatan dari epsilon yaitu nilai yang ditetapkan untuk menentukan radius lingkaran di sekitar titik-titik pada data agar dianggap sebagai titik tetangga.
<i>Library</i>	kumpulan kode atau fungsi dengan prosedur yang mempermudah pemodelan.
<i>MinPts</i>	singkatan dari <i>minimum points</i> yaitu jumlah minimum titik-titik yang harus berada di dalam suatu lingkaran pada radius tertentu agar suatu titik dianggap sebagai anggota <i>cluster</i> .
<i>Missing Value</i>	atribut dari suatu data yang tidak memiliki nilai.
<i>Noise</i>	titik yang tidak termasuk pada anggota suatu <i>cluster</i> .
<i>Outlier</i>	data pencilan yang tidak relevan dan signifikan terhadap kumpulan data.
<i>Output</i>	hasil dari kode program yang telah menjalankan suatu algoritma.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING.....	ii
HALAMAN PENGESAHAN DOSEN PENGUJI.....	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO	vi
KATA PENGANTAR	vii
SARI	ix
GLOSARIUM.....	x
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Metodologi Penelitian	3
1.7 Sistematika Penulisan.....	4
BAB II LANDASAN TEORI.....	5
2.1 <i>Data Preprocessing</i>	5
2.2 <i>Clustering</i>	6
2.3 <i>Density-Based Spatial Clustering of Application with Noise (DBSCAN)</i>	6
2.4 Penelitian Terdahulu.....	9
BAB III METODOLOGI PENELITIAN	11
3.1 Sumber Data	11
3.2 Tahapan Penelitian	11
3.2.1 <i>Data Preprocessing</i>	12
3.2.2 <i>Clustering</i>	15
3.2.3 <i>Pattern Evaluation</i>	15
BAB IV HASIL DAN PEMBAHASAN	16

4.1	Persiapan Data	16
4.2	<i>Data Preprocessing</i>	18
4.3	<i>Clustering</i> Titik Penjemputan.....	22
4.3.1	<i>Clustering</i> Titik Penjemputan Hari Kerja.....	23
4.3.2	<i>Clustering</i> Titik Penjemputan Hari Libur.....	25
4.3.3	<i>Clustering</i> Titik Penjemputan Pagi Hari	26
4.3.4	<i>Clustering</i> Titik Penjemputan Siang Hari	28
4.3.5	<i>Clustering</i> Titik Penjemputan Sore Hari	30
4.3.6	<i>Clustering</i> Titik Penjemputan Malam Hari	31
4.3.7	<i>Clustering</i> Titik Penjemputan Tengah Malam	33
4.4	<i>Clustering</i> Titik Pengantaran.....	35
4.4.1	<i>Clustering</i> Titik Pengantaran Hari Kerja.....	35
4.4.2	<i>Clustering</i> Titik Pengantaran Hari Libur.....	37
4.4.3	<i>Clustering</i> Titik Pengantaran Pagi Hari.....	39
4.4.4	<i>Clustering</i> Titik Pengantaran Siang Hari.....	40
4.4.5	<i>Clustering</i> Titik Pengantaran Sore Hari	42
4.4.6	<i>Clustering</i> Titik Pengantaran Malam Hari	44
4.4.7	<i>Clustering</i> Titik Pengantaran Tengah Malam	46
4.5	<i>Pattern Evaluation</i>	48
4.5.1	<i>Knowledge</i> Titik Penjemputan.....	48
4.5.2	<i>Knowledge</i> Titik Pengantaran.....	52
BAB V KESIMPULAN DAN SARAN		57
5.1	Kesimpulan.....	57
5.2	Saran	59
DAFTAR PUSTAKA		60
LAMPIRAN.....		63

DAFTAR TABEL

Tabel 3. 1 Atribut <i>Dataset</i> “ <i>New York City Taxi Trip Duration</i> ”	12
Tabel 3. 2 Atribut Spasial <i>Clustering</i>	14
Tabel 3. 3 Atribut Analisis Temporal <i>Clustering</i>	14
Tabel 4. 1 Iterasi Parameter Titik Penjemputan Hari Kerja.....	23
Tabel 4. 2 Iterasi Parameter Titik Penjemputan Hari Libur.....	25
Tabel 4. 3 Iterasi Parameter Titik Penjemputan Pagi Hari	27
Tabel 4. 4 Iterasi Parameter Titik Penjemputan Siang Hari	28
Tabel 4. 5 Iterasi Parameter Titik Penjemputan Sore Hari	30
Tabel 4. 6 Iterasi Parameter Titik Penjemputan Malam Hari	31
Tabel 4. 7 Iterasi Parameter Titik Penjemputan Tengah Malam	33
Tabel 4. 8 Iterasi Parameter Titik Pengantaran Hari Kerja.....	35
Tabel 4. 9 Iterasi Parameter Titik Pengantaran Hari Libur.....	37
Tabel 4. 10 Iterasi Parameter Titik Pengantaran Pagi Hari	39
Tabel 4. 11 Iterasi Parameter Titik Pengantaran Siang Hari	41
Tabel 4. 12 Iterasi Parameter Titik Pengantaran Sore Hari	42
Tabel 4. 13 Iterasi Parameter Titik Pengantaran Malam Hari	45
Tabel 4. 14 Iterasi Parameter Titik Pengantaran Tengah Malam	46
Tabel 4. 15 Area Potensial Titik Penjemputan	48
Tabel 4. 16 Area Potensial Titik Pengantaran	52
Tabel 5. 1 Rangkuman Parameter dan Hasil <i>Clustering</i> Titik Penjemputan	57
Tabel 5. 2 Rangkuman Parameter dan Hasil <i>Clustering</i> Titik Pengantaran	58

DAFTAR GAMBAR

Gambar 2. 1 <i>Cluster</i> Sembarang	6
Gambar 2. 2 Pembentukan <i>Cluster</i> Berdasarkan Parameter	7
Gambar 2. 3 <i>Core Point</i> dan <i>Border Point</i>	7
Gambar 2. 4 Titik <i>Density-Reachable</i> dan <i>Density-Connected</i>	8
Gambar 3. 1 Metode Penelitian	12
Gambar 4. 1 <i>Dataset New York City Taxi Trip Duration</i>	16
Gambar 4. 2 Kode Program Mengubah Tipe Data dan Mengecek Tipe Data	16
Gambar 4. 3 <i>Output</i> Pengecekan <i>Missing Value</i>	17
Gambar 4. 4 Kode Program Mengambil Data Bulan April dan Bulan Mei	17
Gambar 4. 5 Dataset Dengan Potongan Kolom <i>sub_urban</i>	18
Gambar 4. 6 Pengecekan Tipe Data	18
Gambar 4. 7 <i>Output</i> Pengecekan <i>Missing Value</i>	19
Gambar 4. 8 Sebaran Titik Penjemputan di Kota New York	20
Gambar 4. 9 Sebaran Titik Penjemputan di Brooklyn	20
Gambar 4. 10 Jumlah Karakteristik Waktu Perjalanan Taksi	21
Gambar 4. 11 Jumlah Titik Penjemputan dan Pengantaran Pada Hari Kerja dan Libur	21
Gambar 4. 12 Hasil <i>Data Preprocessing</i> Titik Penjemputan dan Titik Pengantaran	22
Gambar 4. 13 Visualisasi <i>Cluster</i> Titik Penjemputan Hari Kerja	24
Gambar 4. 14 Jumlah Anggota <i>Cluster</i> Titik Penjemputan Hari Kerja	24
Gambar 4. 15 Visualisasi Peta <i>Cluster</i> Titik Penjemputan Hari Kerja	24
Gambar 4. 16 Visualisasi <i>Cluster</i> Titik Penjemputan Hari Libur	25
Gambar 4. 17 Jumlah Anggota <i>Cluster</i> Titik Penjemputan Hari Libur	26
Gambar 4. 18 Visualisasi Peta <i>Cluster</i> Titik Penjemputan Hari Libur	26
Gambar 4. 19 Visualisasi <i>Cluster</i> Titik Penjemputan Pagi Hari	27
Gambar 4. 20 Jumlah Anggota <i>Cluster</i> Titik Penjemputan Pagi Hari	27
Gambar 4. 21 Visualisasi Peta <i>Cluster</i> Titik Penjemputan Pagi Hari	28
Gambar 4. 22 Visualisasi <i>Cluster</i> Titik Penjemputan Siang Hari	29
Gambar 4. 23 Jumlah Anggota <i>Cluster</i> Titik Penjemputan Siang Hari	29
Gambar 4. 24 Visualisasi Peta <i>Cluster</i> Titik Penjemputan Siang Hari	29
Gambar 4. 25 Visualisasi <i>Cluster</i> Titik Penjemputan Sore Hari	30
Gambar 4. 26 Jumlah Anggota <i>Cluster</i> Titik Penjemputan Sore Hari	31
Gambar 4. 27 Visualisasi Peta <i>Cluster</i> Titik Penjemputan Sore Hari	31

Gambar 4. 28 Visualisasi <i>Cluster</i> Titik Penjemputan Malam Hari	32
Gambar 4. 29 Jumlah Anggota <i>Cluster</i> Titik Penjemputan Malam Hari	32
Gambar 4. 30 Visualisasi Peta <i>Cluster</i> Titik Penjemputan Malam Hari	33
Gambar 4. 31 Visualisasi <i>Cluster</i> Titik Penjemputan Tengah Malam	34
Gambar 4. 32 Jumlah Anggota <i>Cluster</i> Titik Penjemputan Tengah Malam.....	34
Gambar 4. 33 Visualisasi Peta <i>Cluster</i> Titik Penjemputan Tengah Malam	34
Gambar 4. 34 Visualisasi <i>Cluster</i> Titik Pengantaran Hari Kerja.....	36
Gambar 4. 35 Jumlah Anggota <i>Cluster</i> Titik Pengantaran Hari Kerja	36
Gambar 4. 36 Visualisasi Peta <i>Cluster</i> Titik Pengantaran Hari Kerja.....	37
Gambar 4. 37 Visualisasi <i>Cluster</i> Titik Pengantaran Hari Libur.....	38
Gambar 4. 38 Jumlah Anggota <i>Cluster</i> Titik Pengantaran Hari Libur	38
Gambar 4. 39 Visualisasi Peta <i>Cluster</i> Titik Pengantaran Hari Libur.....	38
Gambar 4. 40 Visualisasi <i>Cluster</i> Titik Pengantaran Pagi Hari.....	39
Gambar 4. 41 Jumlah Anggota <i>Cluster</i> Titik Pengantaran Pagi Hari.....	40
Gambar 4. 42 Visualisasi Peta <i>Cluster</i> Titik Pengantaran Pagi Hari.....	40
Gambar 4. 43 Visualisasi <i>Cluster</i> Titik Pengantaran Siang Hari.....	41
Gambar 4. 44 Jumlah Anggota <i>Cluster</i> Titik Pengantaran Siang Hari.....	42
Gambar 4. 45 Visualisasi Peta <i>Cluster</i> Titik Pengantaran Siang Hari.....	42
Gambar 4. 46 Visualisasi <i>Cluster</i> Titik Pengantaran Sore Hari	43
Gambar 4. 47 Jumlah Anggota <i>Cluster</i> Titik Pengantaran Siang Hari.....	44
Gambar 4. 48 Visualisasi Peta <i>Cluster</i> Titik Pengantaran Sore Hari	44
Gambar 4. 49 Visualisasi <i>Cluster</i> Titik Pengantaran Malam Hari	45
Gambar 4. 50 Jumlah Anggota <i>Cluster</i> Titik Pengantaran Malam Hari.....	45
Gambar 4. 51 Visualisasi Peta <i>Cluster</i> Titik Pengantaran Malam Hari	46
Gambar 4. 52 Visualisasi <i>Cluster</i> Titik Pengantaran Tengah Malam	47
Gambar 4. 53 Jumlah Anggota <i>Cluster</i> Titik Pengantaran Tengah Malam.....	47
Gambar 4. 54 Visualisasi Peta <i>Cluster</i> Titik Pengantaran Tengah Malam.....	47
Gambar 4. 55 Visualisasi Peta Titik Penjemputan Hari Kerja.....	50
Gambar 4. 56 Area Potensial Titik Penjemputan Hari Kerja.....	51
Gambar 4. 57 Visualisasi Peta Titik Penjemputan Hari Libur di Flatbush Avenue	51
Gambar 4. 58 Visualisasi Peta Titik Pengantaran di Williamsburg.....	55
Gambar 4. 59 Area Potensial Titik Pengantaran di Flatbush Avenue	55
Gambar 4. 60 Visualisasi Peta Titik Pengantaran.....	56
Gambar 4. 61 Visualisasi Peta Titik Pengantaran di Williamsburg dan Myrtle Avenue	56

BAB I

PENDAHULIAN

1.1 Latar Belakang

Setengah dari 7.6 miliar penduduk dunia tinggal di daerah perkotaan dan setiap tahunnya akan bertambah sekitar 83 juta jiwa yang setara dengan empat kali lipat jumlah penduduk Kota New York (United Nations, Department of Economic and Social Affairs, Population Division, 2017). Pesatnya perkembangan daerah perkotaan dalam beberapa tahun terakhir, menjadikan rutinitas harian perjalanan penduduk kota dan transportasi umum saling mempengaruhi dan berkaitan erat (Jian, Li, & Yu, 2021). Dalam situasi meningkatnya globalisasi dan persaingan global, mobilitas perkotaan berkaitan dengan beragam armada taksi menjadi semakin relevan (Vizuete-Luciano, Guill'en-Pujadas, Alaminos, & Merig'o-Lindahl, 2023). Usaha armada taksi sangat berkaitan erat dengan kepuasan pelayanan kepada penumpang mulai dari penjemputan hingga pengantaran ke tempat tujuan (Riascos & Mateos, 2020). Salah satu kelebihan utama menggunakan layanan taksi adalah menyediakan layanan untuk membawa penumpang dari titik awal penjemputan hingga titik akhir tujuan penumpang. Hal tersebut berbeda dengan transportasi umum lainnya yang beroperasi antar stasiun atau halte dengan rute terbatas (Ulak, Yazici, & Aljarrah, 2020). Taksi menjadi salah satu alat transportasi yang penting untuk melakukan perjalanan sehari-hari bagi penduduk kota (Jian, Li, & Yu, 2021).

Pada masa kini, pemesanan taksi dapat dengan mudah dilakukan melalui platform daring maupun secara konvensional. Setelah penumpang melakukan pemesanan taksi melalui platform daring, penumpang harus menunggu hingga sopir taksi menjemput pada lokasi yang sudah ditentukan. Hal yang serupa terjadi pada pemesanan taksi secara konvensional yang membuat penumpang harus mencari taksi terdekat dan menunggu taksi yang tersedia. Waktu tunggu taksi menjadi hal yang krusial bagi penumpang taksi untuk memilih lokasi penjemputan yang tepat di lingkungan perkotaan (You, et al., 2021). Dengan mengidentifikasi permintaan penumpang taksi secara akurat dapat membantu pengelola taksi mengalokasikan sumber daya taksi pada daerah perkotaan terlebih dahulu. Sehingga membantu sopir taksi menemukan penumpang dengan cepat dan mengurangi lama waktu tunggu penumpang (Zhang et al., 2022).

Pengelola taksi memiliki data sejarah perjalanan taksi berupa waktu penjemputan, waktu pengantaran, jumlah penumpang, serta data geospasial berupa lokasi penjemputan dan pengantaran penumpang. Berdasarkan penelitian yang dilakukan oleh (Yang, Sun, Shang, Wang, & Zhu, 2019), dengan tersedianya kumpulan data geospasial yang besar dapat

memberikan inovasi untuk mengeksplorasi lingkungan spasial perkotaan dan menjadi studi mengenai pola sebaran *cluster* mobilitas penduduk. Salah satu metode pionir yang dapat digunakan untuk memperoleh informasi penting dari data geospasial adalah *Density-Based Spatial Clustering of Application with Noise* (DBSCAN). Algoritma DBSCAN merupakan pembelajaran tanpa pengawasan yang dapat mengidentifikasi kelompok objek yang berbeda di dalam data. Algoritma tersebut akan membentuk objek yang berada pada area dengan kepadatan tinggi ke dalam *cluster* dan menemukan *cluster* bentuk sembarang dalam basis data spasial yang di dalamnya terdapat titik kebisingan atau *noise* (Sugi Almantara, Sri Ariyani, & Alit Swamardika, 2020).

Dengan memanfaatkan algoritma DBSCAN terhadap data perjalanan taksi, maka penelitian ini bertujuan untuk mengidentifikasi *cluster* yang terbentuk dari titik penjemputan dan titik pengantaran penumpang. Hasil identifikasi dan visualisasi *cluster* tersebut dapat membantu pengelola taksi dalam mengalokasikan taksi pada area yang potensial bagi sopir taksi menemukan penumpang. Penelitian ini menggunakan *dataset* dari Platform Kaggle yang berjudul *New York City Taxi Trip Duration* dengan perjalanan taksi yang tercatat pada tahun 2016 dan berfokus pada data spasial di wilayah Kota Brooklyn.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah penelitian ini yaitu:

- a. Bagaimana mengidentifikasi area potensial di perkotaan berupa area titik penjemputan dan titik pengantaran penumpang taksi menggunakan metode *clustering* berbasis densitas?
- b. Bagaimana menganalisis pola penjemputan dan pengantaran penumpang taksi dari penumpang secara temporal?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini yaitu:

- a. *Dataset* yang digunakan adalah *dataset* publik berjudul *New York City Taxi Trip Duration* yang tersedia di *Platform* Kaggle.
- b. Area studi kasus adalah salah satu wilayah New York City yaitu Brooklyn.
- c. Analisis dilakukan pada perjalanan taksi yang terjadi pada Bulan April dan Mei 2016.
- d. Atribut spasial yang digunakan untuk *clustering* adalah *pickup_latitude*, *pickup_longitude*, *dropoff_latitude*, dan *dropoff_longitude*.

- e. Dimensi waktu yang digunakan untuk analisis temporal adalah jam yang dibagi menjadi pagi, siang, malam, tengah malam, serta hari kerja dan hari libur.

1.4 Tujuan Penelitian

Tujuan penelitian ini yaitu:

- a. Mengetahui area potensial titik penjemputan dan titik pengantaran penumpang taksi di perkotaan berdasarkan metode *clustering* berbasis densitas.
- b. Mengetahui pola penjemputan dan pengantaran penumpang taksi berdasarkan analisis secara temporal.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini yaitu:

- a. Pengelola taksi dapat mempelajari karakteristik penumpang taksi melalui hasil identifikasi *cluster* titik penjemputan dan titik pengantaran untuk meningkatkan operasi layanan taksi.
- b. Melalui hasil visualisasi *cluster*, peneliti dan pembaca dapat melihat area potensial berdasarkan mobilitas penduduk yang menggunakan layanan taksi.
- c. Penumpang taksi dapat memilih titik penjemputan yang ramai akan taksi dan tujuan favorit di Kota Brooklyn.
- d. Dapat dijadikan sebagai bahan referensi untuk penelitian lanjut yang menggunakan data spasial dan data temporal.

1.6 Metodologi Penelitian

Tahapan untuk menyelesaikan identifikasi pengelompokan titik penjemputan dan titik pengantaran perjalanan taksi menggunakan algoritma DBSCAN yang mengacu kepada proses *Knowledge Discovery in Database* (KDD) sebagai berikut:

- a. *Data Preprocessing*

Tujuan dari *data preprocessing* adalah menyesuaikan dan meningkatkan kualitas data dengan melakukan *data cleaning* agar meningkatkan akurasi pemodelan.

- b. *Clustering*

Data dengan format yang telah sesuai akan dikelompokkan menggunakan salah satu teknik *clustering* pada yaitu *Density-Based Spatial Clustering of Application with Noise* (DBSCAN).

- c. *Pattern Evaluation and Knowledge*

Pola yang dihasilkan melalui tahapan *clustering* telah menghasilkan informasi implisit serta visualisasi berupa sebaran *cluster* perjalanan taksi di Kota Brooklyn.

1.7 Sistematika Penulisan

Sistematika penulisan dalam penyusunan laporan tugas akhir ini terdiri dari beberapa bab yang mencakup gambaran dari keseluruhan masalah serta penyelesaiannya. Berikut sistematika penulisan yang terbagi dalam 5 bab:

BAB I PENDAHULUAN

Pada bab ini berisi pembahasan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini berisi pembahasan mengenai teori dan kajian pustaka yang dikutip dari beberapa sumber untuk mendukung topik penelitian.

BAB III METODOLOGI PENELITIAN

Pada bab ini berisi uraian tentang identifikasi data mulai dari *data preprocessing*, *clustering*, *pattern evaluation* dan menghasilkan *knowledge*.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini berisi hasil dan pembahasan dari proses metodologi penelitian.

BAB V KESIMPULAN DAN SARAN

Pada bab ini merupakan bab terakhir yang akan membahas kesimpulan dan saran terhadap penelitian yang telah dilakukan pada tugas akhir.

BAB II LANDASAN TEORI

2.1 *Data Preprocessing*

Data preprocessing atau pemrosesan awal data bertujuan untuk membuat data lebih mudah ditangani oleh pemodelan pada *data mining*. Kualitas data memiliki pengaruh yang besar dan signifikan terhadap pemodelan *data mining*. Berbagai teknik pada *data preprocessing* diciptakan untuk membentuk data agar memenuhi persyaratan untuk dimasukkan dalam pemodelan, meningkatkan relevansi target identifikasi, dan mempermudah dalam optimasi pemodelan (Canchen, 2019).

Dalam meningkatkan kualitas data, langkah pertama yang dilakukan adalah *data cleaning*. Pembersihan data atau *data cleaning* merupakan prosedur untuk menangani atau menghilangkan elemen yang tidak standar dari kumpulan data. Elemen tersebut diantaranya seperti *missing value*, *outlier*, *duplicate* dan sejenisnya (Chicco, Oneto, & Tavazzi, 2022). Sebagian besar pemodelan *data mining* tidak bisa mentolerir nilai yang hilang/*missing value* dari data. Nilai ini tidak dapat digunakan untuk perbandingan, tidak tersedia untuk kategorisasi, maupun dioperasikan dengan algoritma aritmatika. Oleh sebab itu, penting untuk menangani *missing value* dengan cara membuang sampel yang memiliki nilai yang hilang. Metode tersebut efektif untuk proporsi data yang memiliki *missing value* tidak mendominasi seluruh data atau jika dihilangkan tidak mengurangi jumlah kumpulan data secara drastis (Canchen, 2019). Namun jika proporsi *missing value* cukup banyak dan jika dihilangkan dapat mengganggu jumlah data yang akan diidentifikasi, *data cleaning* dapat dilakukan dengan cara mengisi *missing value* secara manual maupun menggunakan nilai yang paling mungkin untuk mengisi nilai yang hilang (Han, Kamber, & Pei, 2011).

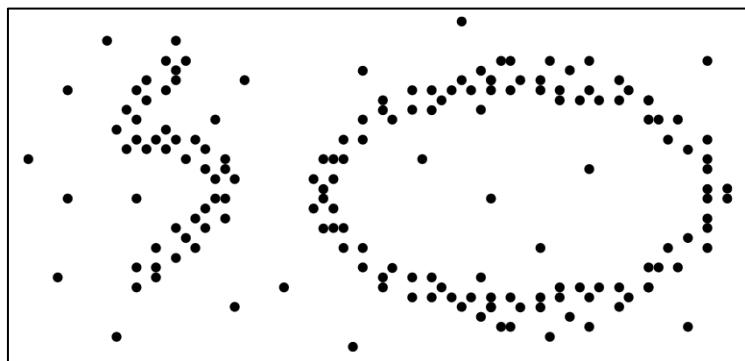
Pada penelitian yang dilakukan oleh (Sari & Primajaya, 2019), *data preprocessing* dilakukan dengan cara memilih 12 atribut yang relevan yang berkaitan dengan *clustering* untuk pertanian di Kabupaten Karawang. Pada salah satu atribut yaitu kecamatan, dilakukan transformasi data untuk memuat informasi titik koordinat garis bujur dan garis lintang kawasan pertanian. Titik koordinat tersebut membantu penelitian dalam melakukan visualisasi titik kecamatan yang tidak terdapat pada atribut awal.

Penelitian pengelompokan titik gempa di Pulau Sulawesi menggunakan *spatio-temporal clustering* yang dilakukan oleh (Manalu, Rahmawati, & Widiharih, 2021), melakukan *data preprocessing* dengan memilih atribut yang akan digunakan pada tahap *clustering*. Hal tersebut

mempermudah *clustering* dengan memilih atribut yang relevan yaitu data spasial berupa *longitude* dan *latitude* serta tanggal terjadinya gempa. *Data preprocessing* dengan mengambil atribut yang relevan juga dilakukan oleh (Johar, Vatesia, & Donny, 2023). Dengan melakukan *spatio-temporal clustering* pada titik api kebakaran hutan, atribut spasial yang digunakan adalah *latitude* dan *longitude* serta atribut temporal berupa *date*.

2.2 Clustering

Clustering ditujukan untuk mendeteksi sekumpulan kelompok objek yang serupa dari kumpulan data sekaligus memisahkan objek yang berbeda dalam kelompok yang berbeda atau disebut dengan kelompok titik *noise* (Xue & Wang, 2021). Alternatif untuk menemukan *cluster* yang berbentuk sembarang adalah dengan memodelkan *cluster* sebagai wilayah padat dalam ruang data dan dipisahkan oleh wilayah yang jarang. Metode pengelompokan berbasis kepadatan atau *density-based method* menjadi strategi utama untuk dapat menemukan kelompok *cluster* dari data berbentuk non-bola. Salah satu bagian dari *density-based method* adalah *Density-Based Spatial Clustering of Application with Noise* (DBSCAN) (Han, Kamber, & Pei, 2011).



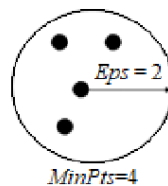
Gambar 2. 1 *Cluster Sembarang*
(Han, Kamber, & Pei, 2011)

2.3 *Density-Based Spatial Clustering of Application with Noise (DBSCAN)*

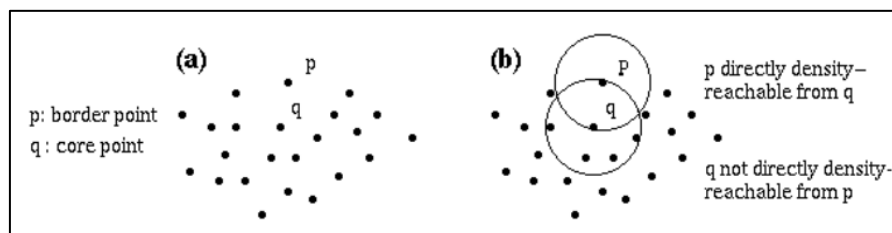
Density-Based Spatial Clustering of Application with Noise (DBSCAN) merupakan salah satu pionir dalam pengembangan teknik pengelompokan berdasarkan kepadatan atau biasa dikenal dengan *density-based clustering* (Sugi Almantara, Sri Ariyani, & Alit Swamardika, 2020). Beragam sektor mengalami peningkatan jumlah data secara eksponensial dan salah satunya sektor transportasi. Sebagian data tersebut mengandung *noise* dan *outlier*. Oleh sebab

itu, DBSCAN menjadi algoritma pengelompokan *cluster* yang populer untuk digunakan pada data yang *noisy*, berbentuk sembarang (*arbitrary shape*), dan *outlier*. DBSCAN mampu mengidentifikasi *cluster* wilayah dengan kepadatan tinggi dengan memisahkannya dari wilayah dengan kepadatan rendah (Hossain, Islam, Rahman Miah, Rony, & Begum, 2021).

DBSCAN memiliki dua parameter yang mempengaruhi jumlah dan kepadatan *cluster* yang dihasilkan yaitu *Eps* dan *MinPts*. Nilai ϵ (*Epsilon*) atau *Eps* adalah jarak maksimum yang digunakan untuk menentukan apakah suatu titik termasuk tetangga dari titik lain. Sedangkan *MinPts* atau *minimum points* merupakan jumlah minimum titik yang harus berada dalam radius *Eps* untuk membentuk suatu *cluster*. Suatu titik data yang berada di dalam radius *Eps* dari setidaknya sejumlah *MinPts* titik tetangga, maka titik tersebut dianggap sebagai bagian dari suatu *cluster*. Kepadatan ini mengacu pada jumlah titik-titik data yang ada dalam radius *Eps* dari suatu titik. Jika kepadatan berada di bawah ambang batas, titik dianggap sebagai *noise* (C. & H., 2022).



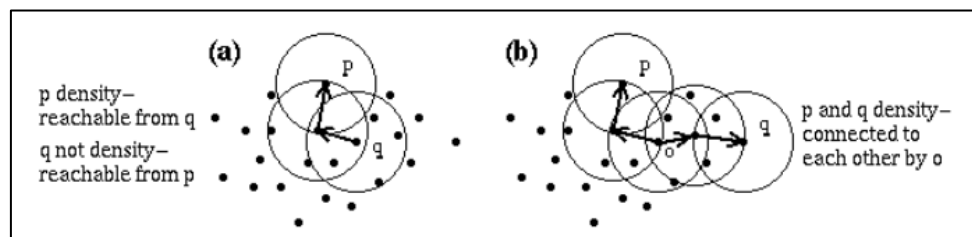
Gambar 2. 2 Pembentukan *Cluster* Berdasarkan Parameter
(Hossain, Islam, Rahman Miah, Rony, & Begum, 2021)



Gambar 2. 3 *Core Point* dan *Border Point*

Berdasarkan penelitian (Ester, Kriegel, Sander, & Xu, 1996), terdapat titik **q** sebagai titik inti atau *core point* dan titik **p** sebagai titik pinggiran atau *border point*. Algoritma DBSCAN memulai dengan titik **q** dan mengambil semua titik yang dapat dijangkau kepadatannya dari titik **q** dengan menggunakan parameter *Eps* dan *MinPts* yang telah ditentukan. Jika titik **p** dideteksi sebagai *border point* (*density-reachable*), DBSCAN akan mengambil titik berikutnya

dalam *database* sehingga membuat titik **p** dan titik **q** berhubungan (*density-connected*) dan terbentuk dalam satu *cluster*.



Gambar 2. 4 Titik *Density-Reachable* dan *Density-Connected*

Berdasarkan (Xin-yan, 2005) dan (Tu, Fu, Huang, Chen, & Ding, DBSCAN Spatial Clustering Analysis of Urban “Production–Living–Ecological” Space Based on POI Data: A Case Study of Central Urban Wuhan, China, 2022), keunggulan algoritma DBSCAN yaitu sangat efektif untuk menyelesaikan masalah dari jumlah data yang besar, dapat memproses titik-titik yang saling bertumpang tindih, dan kecepatan pengelompokan yang tinggi. Hal tersebut sangat efektif dalam menemukan titik-titik *noise* dan membagi area kepadatan tinggi menjadi beberapa *cluster* untuk membentuk *cluster* spasial yang mudah dikenali. Dibandingkan dengan metode pengelompokan K-Means yang hanya tepat digunakan untuk kumpulan sampel data berbentuk cembung, metode pengelompokan DBSCAN dapat diterapkan pada sampel data cembung dan non-cembung.

Pada penelitian (Sari & Primajaya, 2019), penerapan algoritma DBSCAN pada data lahan pertanian padi di Kabupaten Karawang berhasil mengidentifikasi dua *cluster*. Diantaranya perbedaan karakteristik terkait tingkat curah hujan, jumlah produksi yang dipengaruhi oleh luas lahan, serta jenis dan jumlah hama yang menyerang lahan pertanian. Dengan perbedaan karakteristik yang signifikan, *cluster* yang dihasilkan dapat membantu untuk sasaran pemetaan dan bahan analisis rencana penanganan untuk pertanian padi yang lebih baik.

Selain itu, penerapan algoritma DBSCAN (Nurhaliza & Mustakim, 2021) pada pengelompokan kasus Covid-19 berhasil mengelompokkan dua *cluster* yang beranggotakan negara dengan jumlah kenaikan kasus yang signifikan. Terdapat titik *outlier* berupa negara-negara yang berjauhan dari titik-titik yang membentuk *cluster*. Setelah mengidentifikasi *cluster* yang terbentuk, terdapat salah satu *cluster* dengan kenaikan jumlah kasus yang tinggi yang disebabkan oleh kesulitan distribusi alat pelindung diri dan alat-alat medis sehingga mempengaruhi kualitas pelayanan medis. Informasi yang didapatkan dari hasil *clustering* ini

dapat dijadikan acuan untuk penanganan kasus dengan mengamati negara di sekitar yang berada pada satu anggota *cluster*.

Penerapan algoritma DBSCAN juga efisien untuk mengidentifikasi kelompok kabupaten/kota di Indonesia berdasarkan nilai inflasi kelompok pengeluaran. Pada visualisasi pengelompokan *cluster*, terdapat satu *cluster* besar yang memiliki karakteristik yang didominasi dengan nilai inflasi kelompok pengeluaran rendah. Sedangkan terdapat *cluster* kecil dengan nilai inflasi kelompok pengeluaran tinggi (Putri, Permana, Syafriandi, & Zilrahmi, 2023).

Setelah algoritma DBSCAN membentuk *cluster*, salah satu cara yang digunakan untuk mengecek kualitas *cluster* adalah melihat nilai *silhouette coefficient*. Suatu objek *cluster* yang terbentuk secara padat dan terpisah jauh dari tetangga *cluster* lainnya akan menghasilkan nilai *silhouette coefficient* mendekati 1. Sedangkan *cluster* yang terbentuk renggang dan berdekatan dengan tetangga *cluster* lainnya akan menghasilkan nilai *silhouette coefficient* mendekati negatif 1 (Suyanto, 2017).

2.4 Penelitian Terdahulu

Penelitian yang dilakukan oleh (Sari & Primajaya, 2019), algoritma DBSCAN digunakan pada data didapatkan dari Dinas Pertanian Kabupaten Karawang yang direpresentasikan oleh kecamatan, luas panen, luas tanam, hingga organisme pengganggu tanaman. Ditemukan dua *cluster* wilayah kecamatan dari hasil penerapan DBSCAN pada data lahan pertanian padi di Kabupaten Karawang. *Cluster* tersebut memiliki perbedaan karakteristik mengenai tingkat curah hujan, jumlah serangan hama, jenis hama yang menyerang serta luas lahan yang mempengaruhi jumlah produksi pertanian. Nilai yang optimal untuk parameter yaitu *Eps* 325 dan *MinPts* 2. Evaluasi performa dilakukan dengan *average silhouette width* yang menghasilkan skor 0,74.

Metode yang sama dilakukan pada data diperoleh pada situs web resmi *World Health Organization* (WHO). Penelitian menggunakan *clustering* untuk mengelompokkan negara-negara dengan pola kasus Covid-19 serupa. Nantinya hasil dari pengelompokan akan dijadikan rekomendasi sebagai acuan dalam penanganan kasus Covid-19 di suatu negara dengan cara mengamati negara lainnya yang berada pada satu *cluster*. Algoritma DBSCAN diterapkan untuk mendapatkan hasil klasterisasi dan kualitas *cluster* akan diuji menggunakan *silhouette index*. Terdapat 22 percobaan rentang nilai *Eps* dari 0,1 hingga 0,2 dan *MinPts* yaitu 3 dan 4. Penelitian ini memperoleh percobaan terbaik dengan *Eps* 0,2 dan *MinPts* 3 yang menghasilkan

3 *cluster* dengan nilai *silhouette index* sebesar 0,3624. Terdapat *cluster* 1 yang berisi negara-negara kecil atau negara kepulauan yang mulai mengalami peningkatan kasus. Sedangkan *cluster* 2 diisi oleh negara-negara Balkan atau Eropa Tenggara memiliki kenaikan kasus Covid-19 yang signifikan (Nurhaliza & Mustakim, 2021).

Algoritma DBSCAN juga dapat diterapkan selain dari data spasial yang dilakukan oleh (Puspitasari, Cahyana, & Lestari, 2023). Penelitian menggunakan data rekam medis yang didapatkan dari Rumah Sakit As-Sofwan Bekasi. Dari hasil pemrograman *clustering*, terdapat 2 *cluster* pada pengelompokan penyakit pasien. Validitas *cluster* dievaluasi menggunakan *Davies Bouldin Index* yaitu sebesar 1.422086806557305. Dalam menemukan *core point*, dilakukan 3 iterasi secara manual.

Pada penelitian terdahulu, terdapat kesamaan utama setelah menerapkan algoritma DBSCAN terhadap masing-masing *dataset* dengan permasalahan yang berbeda. Setelah mengidentifikasi *cluster* yang terbentuk, kesamaan utama terdapat pada setiap *cluster* yang memiliki karakteristik berbeda secara signifikan. Penerapan algoritma DBSCAN dapat menjawab rumusan masalah pada penelitian terdahulu. Pada penelitian ini, algoritma DBSCAN digunakan pada data perjalanan taksi untuk mengidentifikasi kelompok dari titik penjemputan dan titik pengantaran yang terjadi di Brooklyn. *Cluster* yang terbentuk akan diidentifikasi dan menghasilkan karakteristik yang dapat membantu pengelola taksi dalam mengalokasikan sopir taksi pada area titik penjemputan yang potensial untuk menemukan penumpang.

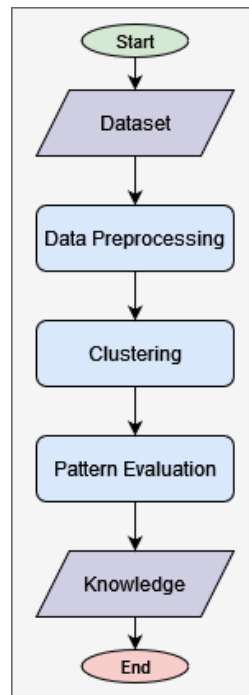
BAB III METODOLOGI PENELITIAN

3.1 Sumber Data

Penelitian ini menggunakan data yang diperoleh melalui *Platform* Kaggle. Kaggle merupakan *platform* bagi komunitas *online* yang diantaranya *data scientist* dan penggemar *machine learning*. Kaggle menyediakan wadah untuk para pengguna berkolaborasi dengan pengguna lain, menemukan dan mempublikasikan *dataset*, menggunakan *notebook* yang terintegrasi dengan GPU, dan menjadi tempat kompetisi para *data scientist* untuk memecahkan tantangan terkait *data science* (Wang, et al., 2021). Data yang digunakan pada penelitian ini berjudul “*New York City Taxi Trip Duration*” yang menjadi salah satu *dataset* untuk kompetisi. Data asli dari catatan perjalanan taksi di Kota New York diterbitkan oleh *The New York City Taxi and Limousine Commission* (TLC). TLC merupakan badan yang bertanggung jawab atas perizinan dan pengaturan Taksi *Medallion* (Kuning) Kota New York. *Dataset* yang diambil merupakan sampel dan sudah dibersihkan oleh Kaggle.

3.2 Tahapan Penelitian

Dalam menjawab rumusan masalah penelitian dan mencapai tujuan penelitian, metode penelitian yang dilakukan terdapat pada Gambar 3. 1. Setelah memperoleh *dataset* dari sumber data, tahap awal yang dilakukan terhadap data adalah *data preprocessing*. Tahap tersebut bertujuan untuk menangani *missing value*, *outlier*, dan *duplicate*. Pada *data preprocessing*, atribut-atribut yang berada di dalam data akan dieksplorasi untuk mengetahui karakteristik data. Selanjutnya, atribut yang diperlukan akan diambil untuk tahap selanjutnya yaitu *clustering*. Algoritma yang digunakan pada *clustering* adalah *Density-Based Spatial Clustering of Application with Noise* (DBSCAN). Setelah algoritma dijalankan dan menghasilkan anggota *cluster*, pola yang terbentuk dari *cluster* tersebut akan diidentifikasi dalam tahap *pattern evaluation* dan menghasilkan *knowledge* yang akan mencapai tujuan penelitian. Beberapa kode program yang digunakan pada penelitian ini diadaptasi melalui *Platform* Kaggle tempat *dataset* asli diunggah. Pada *data preprocessing*, penulis mendapatkan inspirasi melalui *notebook* yang ditulis oleh (Kim, 2023).



Gambar 3. 1 Metode Penelitian

3.2.1 Data Preprocessing

Tahap ini dimulai dengan mengunggah *dataset* yang diperoleh dari sumber data. *Dataset* berisi 1.458.644 baris perjalanan taksi di Kota New York pada tahun 2016. Penelitian ini akan menggunakan catatan perjalanan pada bulan April dan bulan Mei. Terdapat 251.636 catatan perjalanan taksi pada bulan April dan 248.555 catatan pada bulan Mei. Total perjalanan yang tercatat pada dua bulan tersebut adalah 500.191 perjalanan.

Tabel 3. 1 Atribut *Dataset* “New York City Taxi Trip Duration”

No	Atribut	Keterangan	Contoh Atribut
1	<i>id</i>	Kode unik untuk setiap perjalanan	id3504673
2	<i>vendor_id</i>	Kode untuk penyedia taksi terkait dengan catatan perjalanan	1 atau 2
3	<i>pickup_datetime</i>	Tanggal dan waktu saat penjemputan penumpang	2016-04-06 19:32:31
4	<i>dropoff_datetime</i>	Tanggal dan waktu saat tiba di pengantaran penumpang	2016-04-06 19:39:40
5	<i>passenger_count</i>	Jumlah penumpang di dalam kendaraan taksi	4

No	Atribut	Keterangan	Contoh Atribut
6	<i>pickup_longitude</i>	Garis bujur saat penjemputan penumpang	-74.010040
7	<i>pickup_latitude</i>	Garis lintang saat penjemputan penumpang	40.719971
8	<i>dropoff_longitude</i>	Garis bujur saat tiba di pengantaran penumpang	-74.012268
9	<i>dropoff_latitude</i>	Garis lintang saat tiba di pengantaran penumpang	40.706718
10	<i>store_and_fwd_flag</i>	Tanda ini menunjukkan apakah catatan perjalanan disimpan di memori taksi sebelum dikirim ke vendor sebab taksi tidak memiliki koneksi ke server	Y atau N
11	<i>trip_duration</i>	Durasi perjalanan dalam hitungan detik	429

Setiap atribut pada *dataset* akan masuk ke tahap *data cleaning* untuk pengecekan *missing value*, *outlier*, dan *duplicate*. Tahap tersebut juga akan membaca karakteristik dari setiap atribut dan keterkaitan antar atribut. Contohnya, pada atribut *pickup_datetime* dan *dropoff_datetime* apakah terdapat *outlier* seperti tanggal di atas 31 ataupun jam di atas 24. Pada atribut *trip_duration* apakah terdapat catatan perjalanan dengan 0 detik atau waktu perjalanan yang tidak relevan. Selain itu, masing-masing keterkaitan antar atribut diperiksa terhadap *duplicate*. Agar data dapat diidentifikasi dalam pemodelan, setiap nilai atribut harus memiliki tipe data yang sesuai.

Atribut *pickup_datetime* dan *dropoff_datetime* akan disaring pada catatan perjalanan pada bulan April dan bulan Mei. Selanjutnya, dari titik *pickup_longitude* dan *pickup_latitude* akan dikonversi yang menghasilkan atribut baru untuk mencatat wilayah terjadinya titik penjemputan. Konversi titik pada catatan perjalanan bulan April dan bulan Mei menghasilkan 135.000 baris data yang akan digunakan pada penelitian ini. Hasil konversi mencatat titik penjemputan berada di wilayah di Kota New York diantaranya Manhattan, Queens, Brooklyn, Kings, The Bronx, dan Staten Island. Penelitian ini berfokus pada wilayah Brooklyn. Untuk itu, wilayah Brooklyn akan disimpan dalam *data set* baru yang sudah dibersihkan. Setelah *dataset* melewati seluruh tahapan *data cleaning*, maka *dataset* menjadi lebih relevan sesuai target dan tujuan identifikasi serta memenuhi syarat untuk dimasukkan ke dalam pemodelan

clustering. Untuk itu, terdapat 7 atribut yang digunakan untuk membantu identifikasi perjalanan penumpang taksi diantaranya *pickup_datetime*, *dropoff_datetime*, *pickup_longitude*, *pickup_latitude*, *dropoff_longitude*, *dropoff_latitude*, dan *trip_duration*. Terdapat 4 atribut yang diabaikan pada penelitian ini diantaranya *id*, *vendor_id*, *passenger_count*, dan *store_and_fwd_flag*.

Tabel berikut merupakan atribut yang akan digunakan pada metode *clustering*. Pada Tabel 3. 2, atribut spasial berupa *pickup_longitude* dan *pickup_latitude* akan digunakan untuk menghasilkan *cluster* titik penjemputan. Pada *cluster* titik pengantaran, atribut spasial yang digunakan adalah *dropoff_longitude* dan *dropoff_latitude*. Titik spasial *longitude* dan *latitude* akan membentuk titik koordinat yang menjadikan penelitian ini berfokus pada penggunaan metode DBSCAN terhadap data dua dimensi.

Tabel 3. 2 Atribut Spasial *Clustering*

No	Atribut Spasial	Keterangan
1	<i>pickup_longitude</i>	Cluster Titik Penjemputan
2	<i>pickup_latitude</i>	
3	<i>dropoff_longitude</i>	Cluster Titik Pengantaran
4	<i>dropoff_latitude</i>	

Tabel 3. 3 Atribut Analisis Temporal *Clustering*

No	Atribut Temporal	Format Atribut	Kategori Waktu Catatan Perjalanan
1	Hari Kerja	Hari	Senin, Selasa, Rabu, Kamis, Jumat
2	Hari Libur	Hari	Sabtu dan Minggu
3	Pagi	Jam	05:00 – 10:59
4	Siang	Jam	11:01 – 14:59
5	Sore	Jam	15:00 – 17:59
6	Malam	Jam	18:00 – 23:59
7	Tengah Malam	Jam	00:00 – 04:59

Tabel 3. 3 merupakan atribut yang dikembangkan melalui *pickup_datetime* dan *dropoff_datetime*. Sebagai contoh, untuk membentuk atribut temporal pagi, catatan perjalanan akan diambil pada waktu 05:00 – 10:59 sesuai tabel keterangan kategori waktu. Selanjutnya, pada atribut temporal hari kerja dipisahkan antara waktu perjalanan Senin sampai Jumat dan

hari libur pada Sabtu dan Minggu. Hal tersebut dapat membantu untuk mengetahui pola penjemputan dan pengantaran penumpang taksi berdasarkan analisis temporal sebagai bagian dari tujuan penelitian.

3.2.2 *Clustering*

Data perjalanan taksi akan dipisah secara temporal untuk mengetahui pola perjalanan berdasarkan kategori waktu catatan perjalanan yang tertera pada Tabel 3.3. Diperlukan metode *clustering* yang dapat menemukan *cluster* spasial dengan bentuk yang berbeda pada setiap kategori temporal. Penelitian ini akan menggunakan metode *clustering* berbasis densitas yaitu *Density-Based Spatial Clustering of Application with Noise* (DBSCAN). Penelitian ini akan memisahkan *cluster* untuk titik penjemputan dan *cluster* untuk titik pengantaran penumpang taksi. Dengan memasukkan parameter DBSCAN yaitu *Eps* dan *MinPts*, algoritma akan menentukan *core point* dan menghitung titik-titik yang menjadi *border point* dengan jarak yang ditentukan pada *Eps* dan jumlah titik minimum perjalanan pada *MinPts*. Titik-titik yang berada pada nilai ambang batas yang telah ditentukan pada *MinPts* akan termasuk di dalam *cluster*. Jika tidak termasuk pada nilai ambang batas, titik-titik akan dikelompokkan sebagai *noise*. Metode ini sangat efektif dalam menemukan titik *noise* dan membagi area kepadatan tinggi menjadi beberapa *cluster* untuk membentuk *cluster* spasial yang mudah dikenali.

3.2.3 *Pattern Evaluation*

Anggota *cluster* yang terbentuk berdasarkan pemisahan perjalanan pada hari kerja, hari libur, pagi, siang, sore, malam, dan tengah malam akan diidentifikasi secara temporal. Visualisasi *cluster* akan membentuk pola yang memberikan informasi *cluster* titik penjemputan dan *cluster* titik pengantaran di Brooklyn. *Cluster* ini akan menjadi area yang potensial bagi operasi taksi dalam menemukan titik penjemputan dan titik pengantaran berdasarkan sejarah catatan perjalanan. Mobilitas penduduk Brooklyn juga dapat ditemukan dari hasil visualisasi *cluster* yang sudah terbentuk. Penelitian ini mengadaptasi cara untuk memvisualisasikan hasil *cluster* pada peta yang informatif berdasarkan *notebook* milik IBM Developer Skills Network. *Notebook* tersebut menjelaskan langkah-langkah untuk visualisasi data geospasial menggunakan Folium secara gratis.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Persiapan Data

Dataset yang digunakan adalah catatan perjalanan taksi di Kota New York. *Dataset* berbentuk *file* dengan format *csv* (*Comma-Separated Values*) diunggah pada Jupyter Notebook menggunakan bahasa pemrograman Python. Setelah *dataset* diunggah, terlihat tampilan baris dan kolom yang memuat atribut pada *dataset* dan isi dari masing-masing atribut. Terdapat 11 atribut kolom dan 1.458.644 baris data. Gambar berikut merupakan tampilan *dataset* yang berhasil diunggah.

	id	vendor_id	pickup_datetime	dropoff_datetime	passenger_count	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude	store_and_fwd_flag	trip_duration
0	id2875421	2	2016-03-14 17:24:55	2016-03-14 17:32:30	1	-73.982155	40.767937	-73.964630	40.765602	N	455
1	id2377394	1	2016-06-12 00:43:35	2016-06-12 00:54:38	1	-73.980415	40.738564	-73.999481	40.731152	N	663
2	id3858529	2	2016-01-19 11:35:24	2016-01-19 12:10:48	1	-73.979027	40.763939	-74.005333	40.710087	N	2124
3	id3504673	2	2016-04-06 19:32:31	2016-04-06 19:39:40	1	-74.010040	40.719971	-74.012268	40.706718	N	429
4	id2181028	2	2016-03-26 13:30:55	2016-03-26 13:38:10	1	-73.973053	40.793209	-73.972923	40.782520	N	435
...
1458639	id2376096	2	2016-04-08 13:31:04	2016-04-08 13:44:02	4	-73.982201	40.745522	-73.994911	40.740170	N	778
1458640	id1049543	1	2016-01-10 07:35:15	2016-01-10 07:46:10	1	-74.000946	40.747379	-73.970184	40.796547	N	655
1458641	id2304944	2	2016-04-22 06:57:41	2016-04-22 07:10:25	1	-73.959129	40.768799	-74.004433	40.707371	N	764
1458642	id2714485	1	2016-01-05 15:56:26	2016-01-05 16:02:39	1	-73.982079	40.749062	-73.974632	40.757107	N	373
1458643	id1209952	1	2016-04-05 14:44:25	2016-04-05 14:47:43	1	-73.979538	40.781750	-73.972809	40.790585	N	198

Gambar 4. 1 *Dataset New York City Taxi Trip Duration*

Selanjutnya, tipe data akan diperiksa dan disesuaikan terhadap nilai dari kolom agar pemodelan dapat dijalankan. Pada kolom *pickup_datetime* dan *dropoff_datetime*, tipe data merupakan *object*. Tipe data tersebut harus diubah menjadi *datetime* menggunakan *library pandas* agar dapat diproses pada pemodelan. Gambar berikut merupakan kode program untuk mengubah tipe data pada kolom *pickup_datetime* dan *dropoff_datetime*.

```
# mengubah tipe data kolom datetime dari object menjadi datatype
data['pickup_datetime']=pd.to_datetime(data['pickup_datetime'])
data['dropoff_datetime']=pd.to_datetime(data['dropoff_datetime'])

# mengecek update tipe data
print(data['pickup_datetime'].dtypes)
print(data['dropoff_datetime'].dtypes)
```

Gambar 4. 2 Kode Program Mengubah Tipe Data dan Mengecek Tipe Data

Ketika semua tipe data sudah sesuai dengan nilai atribut, *dataset* akan dicek terhadap *missing value*. Setelah dilakukan pengecekan, *dataset* tidak memiliki *missing value* yang ditandai dengan nilai 0. Dataset akan mengandung *missing value* jika *Output* mencetak nilai selain nilai 0.

```

id          0
vendor_id   0
pickup_datetime 0
dropoff_datetime 0
passenger_count 0
pickup_longitude 0
pickup_latitude 0
dropoff_longitude 0
dropoff_latitude 0
store_and_fwd_flag 0
trip_duration 0
dtype: int64

```

Gambar 4. 3 Output Pengecekan Missing Value

Selanjutnya, data akan disortir berdasarkan waktu perjalanan taksi yang tercatat pada bulan April dan bulan Mei. Hasil dari penyortiran data tersebut akan disimpan ke dalam *dataframe* baru. Gambar berikut merupakan kode program untuk menyortir data pada bulan April dan bulan Mei menggunakan manipulasi *library pandas*.

```

# mengambil data pada bulan April
data_april = data.loc[data['pickup_datetime'].dt.month==4]
data_april = data.loc[data['dropoff_datetime'].dt.month==4]

# mengambil data pada bulan Mei
data_mei = data.loc[data['pickup_datetime'].dt.month==5]
data_mei = data.loc[data['dropoff_datetime'].dt.month==5]

```

Gambar 4. 4 Kode Program Mengambil Data Bulan April dan Bulan Mei

Penelitian ini menggunakan data perjalanan taksi yang tercatat di Brooklyn. Untuk itu, diperlukan kolom baru yang berisi wilayah dari catatan perjalanan penumpang taksi. Terdapat fungsi yang digunakan untuk mengalokasikan wilayah berdasarkan kolom *pickup_latitude* dan *pickup_longitude*. Gambar berikut merupakan baru berupa wilayah di Kota New York bernama *sub_urban*. Setelah kolom *sub_urban* terbentuk, isi kolom tersebut merupakan wilayah di Kota New York diantaranya Manhattan, Queens, Brooklyn, Kings, The Bronx, dan Staten Island. Jumlah data yang diperoleh pada persiapan data sebanyak 135.000 baris dan 12 kolom atribut. Setelah data selesai melewati seluruh proses persiapan, dataset siap untuk memasuki tahap *data preprocessing*.

trip_duration	sub_urban
429	Manhattan
2316	Manhattan
731	Manhattan
652	Manhattan
423	Manhattan
...	...
1877	Manhattan
1332	Brooklyn
460	Manhattan
751	Manhattan
570	Manhattan

Gambar 4. 5 Dataset Dengan Potongan Kolom *sub_urban*

4.2 Data Preprocessing

Tahapan *data preprocessing* dilakukan terhadap *dataset* yang sudah diunggah dan telah memiliki atribut *sub_urban*. Berikut merupakan tahap data preprocessing pada penelitian ini, diantaranya:

A. Pengecekan Tipe Data

Pada tahap ini, pengecekan dilakukan terhadap atribut *pickup_datetime* dan *dropoff_datetime* menggunakan *library pandas*. *Output* dari pengecekan tipe data tersebut adalah *object*. Untuk itu, tipe data *object* diubah menjadi *datetime* agar pemodelan dapat dijalankan. Sebab, atribut *datetime* yang berisi nilai seperti 2016-04-06 19:32:31 tidak dapat dilabeli sebagai *object*.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 135000 entries, 0 to 134999
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    135000 non-null object
1   vendor_id            135000 non-null int64
2   pickup_datetime      135000 non-null object
3   dropoff_datetime     135000 non-null object
4   passenger_count      135000 non-null int64
5   pickup_longitude     135000 non-null float64
6   pickup_latitude      135000 non-null float64
7   dropoff_longitude    135000 non-null float64
8   dropoff_latitude     135000 non-null float64
9   store_and_fwd_flag   135000 non-null object
10  trip_duration        135000 non-null int64
11  sub_urban            134874 non-null object
dtypes: float64(4), int64(3), object(5)
memory usage: 12.4+ MB
```

Gambar 4. 6 Pengecekan Tipe Data

B. Menangani *Missing Value*

Setelah tipe data sudah sesuai, pengecekan selanjutnya dilakukan untuk melihat *missing value* setiap atribut. Pada *Output* pengecekan, terdapat bahwa atribut *sub_urban* memiliki 126 baris *missing value*. Penanganan *missing value* dilakukan dengan menghilangkan baris yang mengandung *missing value*. Langkah tersebut dinyatakan efektif sebab tidak membuat *dataset* yang berjumlah 135.000 baris berkurang secara drastis.

```

id          0
vendor_id   0
pickup_datetime  0
dropoff_datetime  0
passenger_count  0
pickup_longitude  0
pickup_latitude  0
dropoff_longitude  0
dropoff_latitude  0
store_and_fwd_flag  0
trip_duration  0
sub_urban   126
dtype: int64

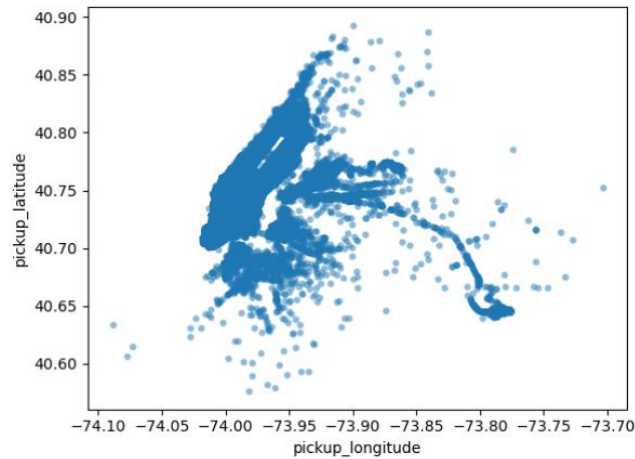
```

Gambar 4. 7 *Output* Pengecekan *Missing Value*

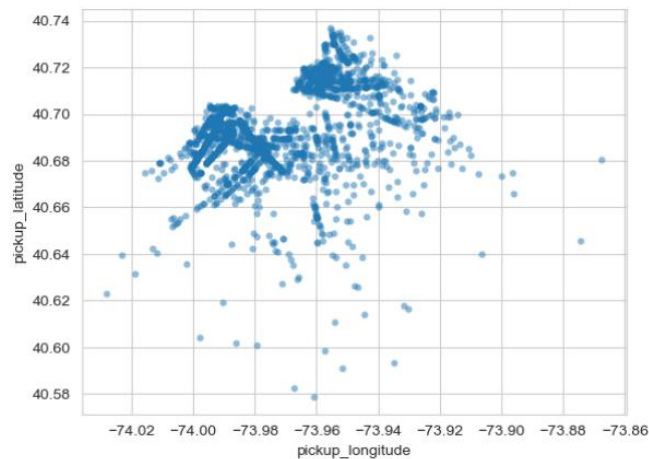
C. Pengecekan *Duplicates*

Setelah menangani *missing value*, dilakukan pengecekan *duplicate*. Hasil dari pengecekan *duplicate* adalah *dataset* tidak memiliki nilai *duplicate*. Selanjutnya, data siap untuk diproses pada tahap eksplorasi untuk mengetahui karakteristik data.

Gambar 4.10 merupakan visualisasi dari atribut *pickup_longitude* dan *pickup_latitude* sehingga membentuk titik koordinat berdasarkan 135.000 baris data perjalanan taksi di Kota New York. Penelitian ini berfokus pada wilayah Brooklyn yang berjumlah 2.610 baris data. Untuk itu, akan dibentuk *dataset* baru yang hanya berisi catatan wilayah perjalanan taksi di Kota Brooklyn seperti visualisasi pada Gambar 4. 11. Sebaran catatan perjalanan taksi untuk titik penjemputan di Kota Brooklyn divisualisasikan menggunakan *scatter plot*. Pada gambar berikut terlihat perbedaan visualisasi data perjalanan pada titik penjemputan di Kota New York dan Kota Brooklyn.



Gambar 4. 8 Sebaran Titik Penjemputan di Kota New York



Gambar 4. 9 Sebaran Titik Penjemputan di Brooklyn

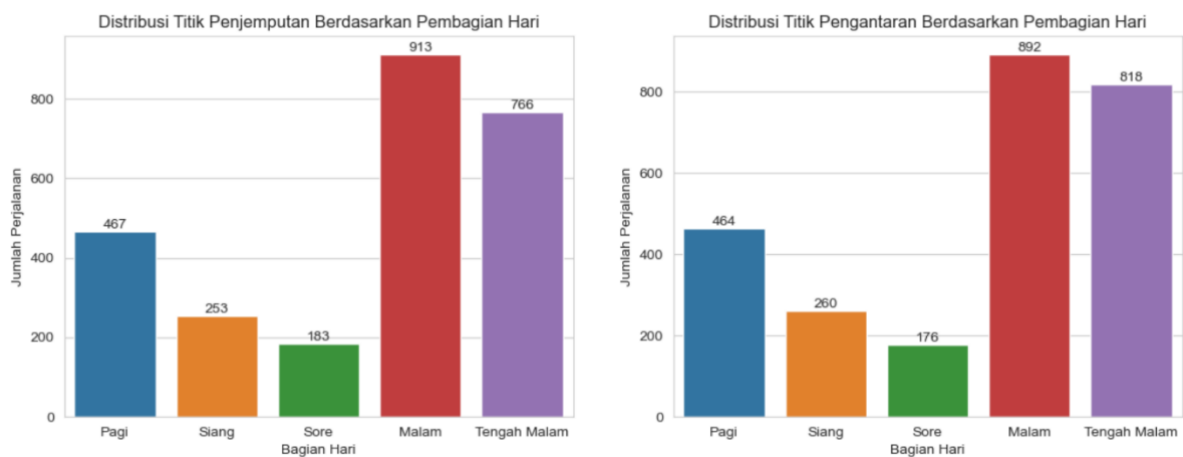
D. Eksplorasi Data

Selanjutnya, dilakukan eksplorasi data untuk memahami karakteristik data. Atribut baru berupa *distance* dapat dikembangkan melalui atribut *pickup_longitude*, *pickup_latitude*, *dropoff_longitude*, dan *dropoff_latitude*. Dengan menggunakan *library numpy*, terdapat fungsi *haversine* yang dapat membuat kalkulasi jarak antara titik tersebut. Sehingga membentuk atribut *distance* yang merupakan jarak antara titik penjemputan dan titik pengantaran penumpang dalam sekali perjalanan.

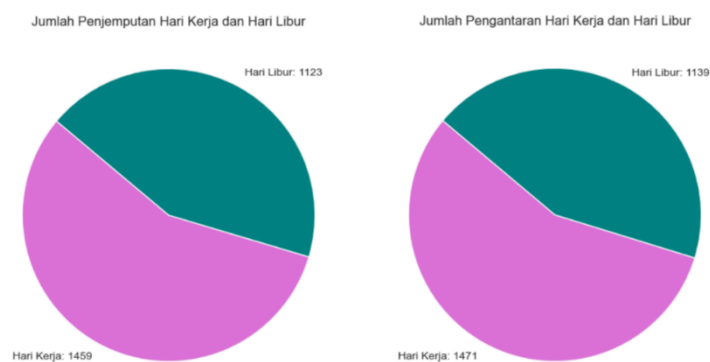
Untuk mengetahui ringkasan statistik berupa nilai minimum, nilai maksimum, kuartil, mean, median, dari suatu *data frame* dapat menggunakan *library pandas*. Terdapat jarak perjalanan yang tercatat sebagai 0 kilometer pada atribut *distance*. Selain itu, 28 baris data

dengan catatan perjalanan 0 kilometer akan dihapus dan menyisakan catatan perjalanan yang relevan dengan jarak yang tercatat pada *dataset* sebanyak 2.582 baris data.

Pada atribut *pickup_datetime* dan *dropoff_datetime*, terdapat catatan berupa waktu penjemputan dan waktu pengantaran pada rentang waktu pukul 00:00 hingga 24:00. Dilakukan pengelompokan jam berdasarkan bagian pada satu hari serta hari kerja dan hari libur yang bertujuan untuk mengetahui catatan perjalanan berdasarkan karakteristik waktu. Pengelompokan tersebut dibagi berdasarkan hari kerja (Senin – Jumat) dan hari libur (Sabtu dan Minggu). Selain itu, catatan perjalanan juga dikelompokkan untuk setiap rentang waktu diantaranya pagi (05:00 – 10:59), siang (11:01 – 14:59), sore (15:00 – 17:59), malam (18:00 – 23:59), dan tengah malam (00:00 – 04:59). Jumlah perjalanan pada penjemputan dan pengantaran melonjak pada malam hari sekitar pukul 18:00 hingga 23:59. Perjalanan taksi di wilayah Brooklyn juga didominasi pada hari kerja.

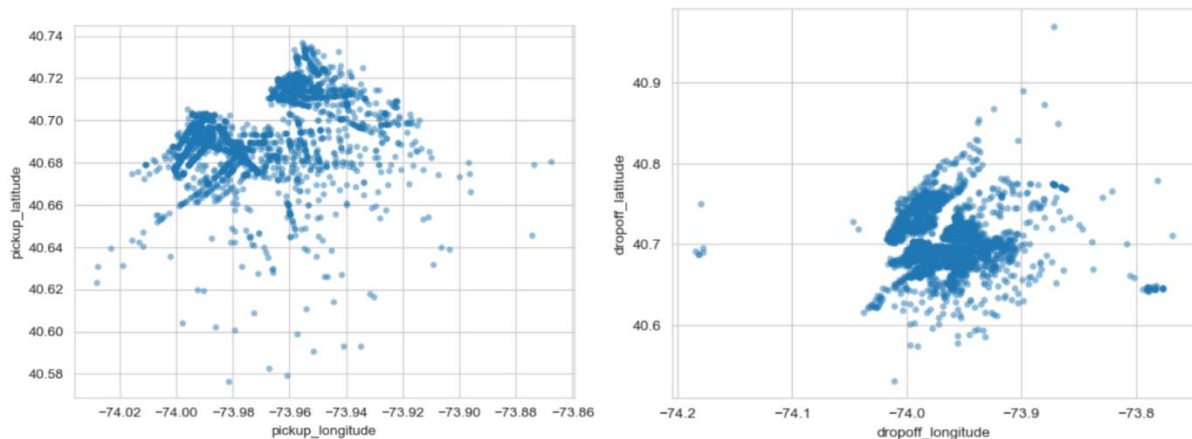


Gambar 4. 10 Jumlah Karakteristik Waktu Perjalanan Taksi



Gambar 4. 11 Jumlah Titik Penjemputan dan Pengantaran Pada Hari Kerja dan Libur

Setelah data melewati seluruh eksplorasi tahapan *data preprocessing*, atribut yang relevan untuk pemodelan *clustering* akan disimpan pada *dataframe* baru. Untuk titik penjemputan, atribut yang digunakan adalah *pickup_longitude* dan *pickup_latitude*. Sedangkan untuk titik pengantaran, atribut yang digunakan adalah *dropoff_longitude* dan *dropoff_latitude*. Gambar berikut merupakan sebaran data pengantaran taksi di Brooklyn.



Gambar 4. 12 Hasil *Data Preprocessing* Titik Penjemputan dan Titik Pengantaran

4.3 *Clustering* Titik Penjemputan

Tahap *clustering* merupakan bagian inti untuk menemukan *cluster* yang terbentuk dari algoritma *Density-Based Spatial Clustering of Application with Noise* (DBSCAN). Tahap ini terbagi menjadi dua yaitu *clustering* untuk titik penjemputan penumpang dan titik pengantaran penumpang di Brooklyn. Atribut spasial menjadi peran penting untuk tahap *clustering*. *Clustering* menggunakan DBSCAN memerlukan dua parameter yaitu *Eps* dan *MinPts*. Penentuan parameter tersebut sangat krusial untuk menghasilkan kualitas *cluster* yang baik. Pada penelitian ini, dilakukan iterasi antara *Eps* dan *MinPts* yang akan menghasilkan *silhouette coefficient* untuk menilai tingkat kepadatan *cluster*. Pasangan parameter yang menghasilkan nilai *silhouette coefficient* mendekati 1 merupakan parameter terbaik dan akan digunakan untuk tahap *clustering*. Algoritma akan mencetak jumlah dan anggota *cluster* yang terbentuk. Terdapat titik *noise* yang dicetak dengan label -1 dan akan dibuang. Titik-titik yang termasuk dalam anggota *cluster* akan disimpan pada *data frame* baru untuk menganalisis area potensial titik penjemputan melalui visualisasi peta.

Untuk mengetahui area potensial tempat terjadinya titik penjemputan berdasarkan rentang waktu, atribut spasial yang sudah dibagi terhadap rentang waktu. Pengelompokan

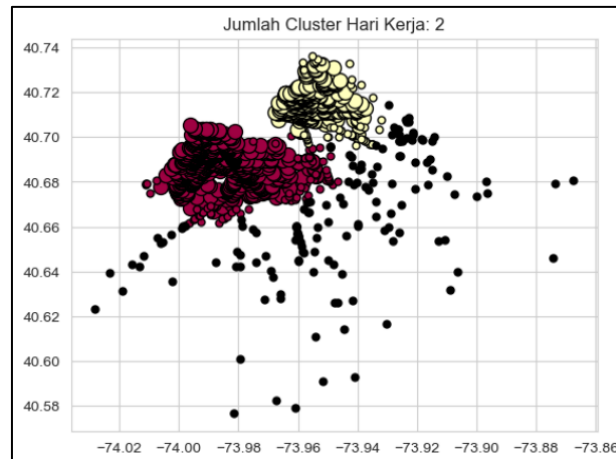
rentang waktu tersebut dibagi berdasarkan hari kerja (Senin – Jumat) dan hari libur (Sabtu dan Minggu). Selain itu, catatan perjalanan juga dikelompokkan untuk setiap rentang jam diantaranya pagi (05:00 – 10:59), siang (11:01 – 14:59), sore (15:00 – 17:59), malam (18:00 – 23:59), dan tengah malam (00:00 – 04:59). Berikut merupakan proses *clustering*, iterasi pencarian parameter terbaik, dan visualisasi *cluster* titik penjemputan berdasarkan karakteristik waktu.

4.3.1 Clustering Titik Penjemputan Hari Kerja

Pada titik penjemputan hari kerja, dilakukan iterasi percobaan *Eps* dengan rentang 0.006 hingga 0.008 dan *MinPts* 21 hingga 30. Setelah mencetak 50 iterasi parameter serta nilai *silhouette coefficient*, ditemukan *Eps* = 0.0075 dan *MinPts* = 29 dengan nilai *silhouette coefficient* tertinggi yaitu 0.546 yang mendekati 1. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 2 *cluster* dan 148 titik *noise*.

Tabel 4. 1 Iterasi Parameter Titik Penjemputan Hari Kerja

<i>Eps</i> \ <i>MinPts</i>	0.006	0.0065	0.007	0.0075	0.008
21	0.533	0.541	0.542	0.493	0.503
22	0.413	0.542	0.541	0.486	0.502
23	0.406	0.538	0.541	0.486	0.495
24	0.402	0.537	0.544	0.485	0.495
25	0.387	0.533	0.543	0.483	0.492
26	0.526	0.527	0.543	0.481	0.490
27	0.525	0.527	0.539	0.544	0.485
28	0.524	0.527	0.539	0.544	0.485
29	0.524	0.533	0.535	0.546	0.485
30	0.520	0.531	0.538	0.540	0.485

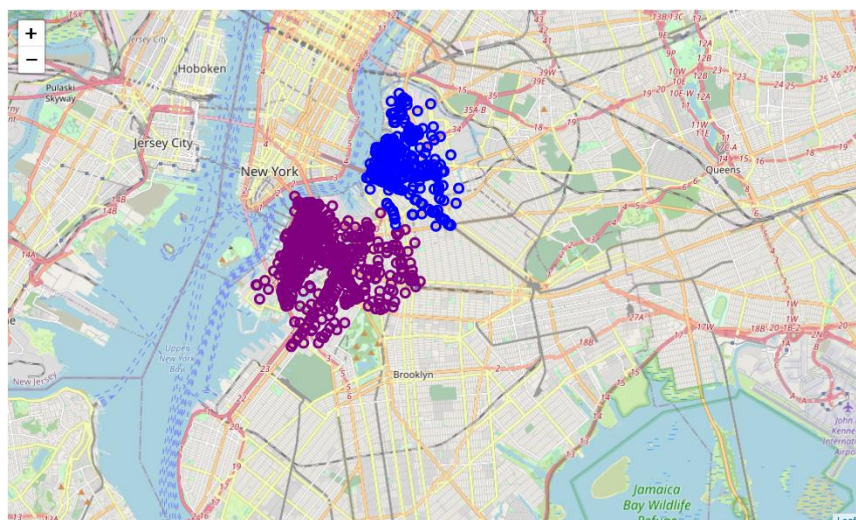


Gambar 4. 13 Visualisasi *Cluster* Titik Penjemputan Hari Kerja

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada hari kerja, terdapat dua *cluster* besar yang berwarna merah dan kuning. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 14. Terdapat 2 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi peta, *cluster* 0 ditandai dengan warna ungu dan *cluster* 1 ditandai dengan warna biru.

cluster	
0	911
1	400
-1	148

Gambar 4. 14 Jumlah Anggota *Cluster* Titik Penjemputan Hari Kerja



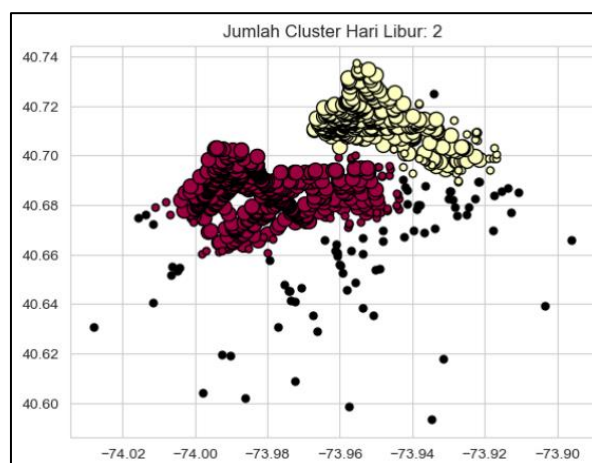
Gambar 4. 15 Visualisasi Peta *Cluster* Titik Penjemputan Hari Kerja

4.3.2 Clustering Titik Penjemputan Hari Libur

Pada titik penjemputan hari libur, dilakukan iterasi percobaan *Eps* dengan rentang 0.006 hingga 0.008 dan *MinPts* 20 hingga 30. Setelah mencetak 55 iterasi parameter serta nilai *silhouette coefficient*, ditemukan *Eps* = 0.008 dan *MinPts* = 25 dengan nilai *silhouette coefficient* tertinggi yaitu 0.524 yang mendekati 1. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 2 *cluster* dan 78 titik *noise*.

Tabel 4. 2 Iterasi Parameter Titik Penjemputan Hari Libur

<i>Eps</i> \ <i>MinPts</i>	0.006	0.0065	0.007	0.0075	0.008
20	0.388	0.470	0.517	0.521	0.443
21	0.452	0.461	0.517	0.521	0.441
22	0.451	0.453	0.517	0.523	0.520
23	0.451	0.449	0.517	0.523	0.521
24	0.449	0.448	0.469	0.523	0.521
25	0.511	0.448	0.450	0.523	0.524
26	0.511	0.455	0.450	0.522	0.523
27	0.513	0.511	0.449	0.519	0.523
28	0.512	0.510	0.455	0.514	0.522
29	0.331	0.511	0.454	0.515	0.520
30	0.312	0.511	0.454	0.518	0.519

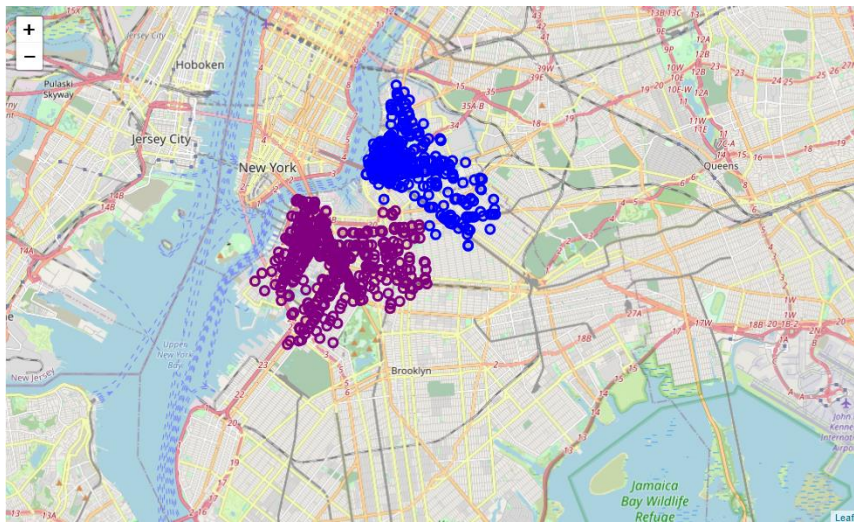


Gambar 4. 16 Visualisasi *Cluster* Titik Penjemputan Hari Libur

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada hari libur, terdapat dua *cluster* besar yang berwarna merah dan kuning. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 17. Terdapat 2 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi peta, *cluster* 0 ditandai dengan warna ungu dan *cluster* 1 ditandai dengan warna biru.

cluster	
0	607
1	438
-1	78

Gambar 4. 17 Jumlah Anggota *Cluster* Titik Penjemputan Hari Libur



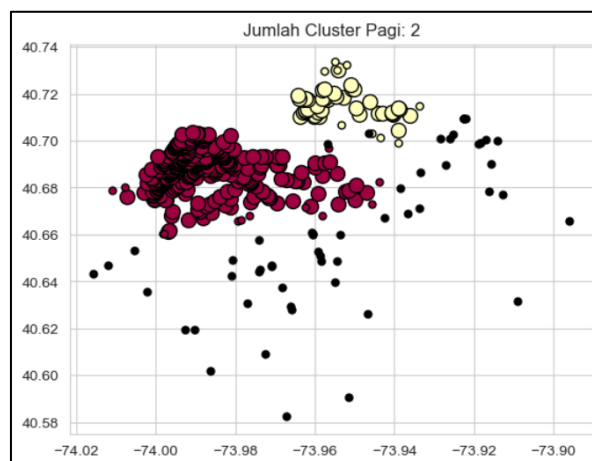
Gambar 4. 18 Visualisasi Peta *Cluster* Titik Penjemputan Hari Libur

4.3.3 Clustering Titik Penjemputan Pagi Hari

Pada titik penjemputan pagi hari, dilakukan iterasi percobaan *Eps* dengan rentang 0.006 hingga 0.008 dan *MinPts* 2 hingga 10. Iterasi pertama dimulai dengan *Eps* = 0.006 terhadap *MinPts* 2 sampai dengan 10, hingga iterasi terakhir yaitu *Eps* = 0.008 terhadap *MinPts* = 10. Setelah mencetak 45 iterasi parameter serta nilai *silhouette coefficient*, ditemukan *Eps* = 0.007 dan *MinPts* = 6 dengan nilai *silhouette coefficient* tertinggi yaitu 0.539 yang mendekati 1. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 2 *cluster* dan 54 titik *noise*.

Tabel 4. 3 Iterasi Parameter Titik Penjemputan Pagi Hari

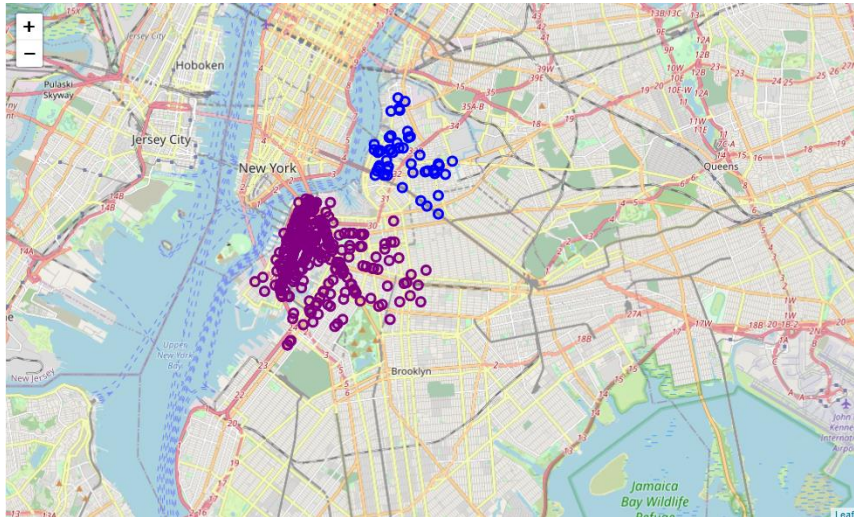
Eps \ $MinPts$	0.006	0.0065	0.007	0.0075	0.008
2	0.303	0.304	0.447	0.456	0.444
3	0.314	0.285	0.432	0.446	0.449
4	0.283	0.268	0.449	0.451	0.453
5	0.365	0.415	0.469	0.445	0.454
6	0.407	0.407	0.539	0.523	0.456
7	0.408	0.407	0.509	0.528	0.530
8	0.460	0.464	0.477	0.537	0.523
9	0.463	0.467	0.473	0.532	0.530
10	0.463	0.457	0.466	0.532	0.528

Gambar 4. 19 Visualisasi *Cluster* Titik Penjemputan Pagi Hari

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada pagi hari, terdapat dua *cluster* besar yang berwarna merah dan kuning. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 20. Terdapat 2 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi peta, *cluster* 0 ditandai dengan warna ungu dan *cluster* 1 ditandai dengan warna biru.

cluster	
0	364
-1	54
1	49

Gambar 4. 20 Jumlah Anggota *Cluster* Titik Penjemputan Pagi Hari



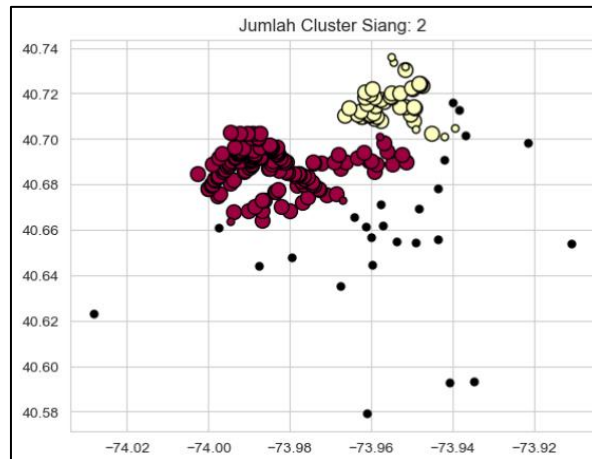
Gambar 4. 21 Visualisasi Peta *Cluster* Titik Penjemputan Pagi Hari

4.3.4 Clustering Titik Penjemputan Siang Hari

Pada titik penjemputan siang hari, dilakukan iterasi percobaan Eps dengan rentang 0.006 hingga 0.008 dan $MinPts$ 2 hingga 10. Setelah mencetak 45 iterasi parameter serta nilai *silhouette coefficient*, ditemukan $Eps = 0.0075$ dan $MinPts = 5$ dengan nilai *silhouette coefficient* tertinggi yaitu 0.546 yang mendekati 1. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 2 *cluster* dan 25 titik *noise*.

Tabel 4. 4 Iterasi Parameter Titik Penjemputan Siang Hari

$Eps \backslash MinPts$	0.006	0.0065	0.007	0.0075	0.008
2	0.369	0.438	0.227	0.227	0.516
3	0.375	0.446	0.389	0.389	0.559
4	0.366	0.412	0.195	0.340	0.366
5	0.382	0.498	0.534	0.546	0.538
6	0.371	0.498	0.534	0.537	0.540
7	0.293	0.267	0.509	0.537	0.541
8	0.228	0.249	0.509	0.512	0.521
9	0.217	0.239	0.280	0.501	0.517
10	0.418	0.457	0.388	0.406	0.501

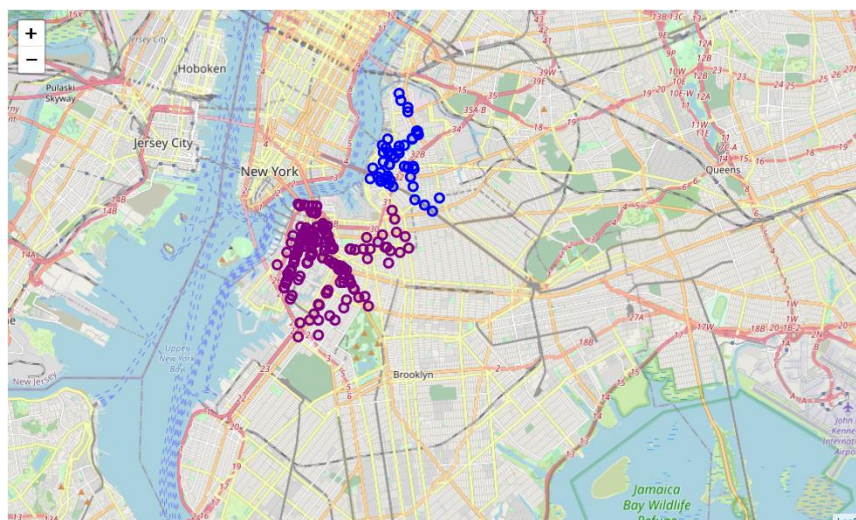


Gambar 4. 22 Visualisasi *Cluster* Titik Penjemputan Siang Hari

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada siang hari, terdapat dua *cluster* besar yang berwarna merah dan kuning. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 23. Terdapat 2 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi peta, *cluster* 0 ditandai dengan warna ungu dan *cluster* 1 ditandai dengan warna biru.

cluster	
0	179
1	49
-1	25

Gambar 4. 23 Jumlah Anggota *Cluster* Titik Penjemputan Siang Hari



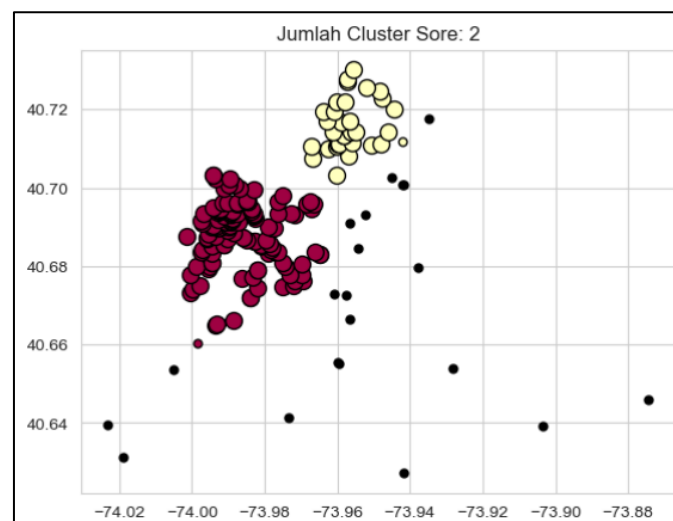
Gambar 4. 24 Visualisasi Peta *Cluster* Titik Penjemputan Siang Hari

4.3.5 Clustering Titik Penjemputan Sore Hari

Pada titik penjemputan sore hari, dilakukan iterasi percobaan *Eps* dengan rentang 0.006 hingga 0.008 dan *MinPts* 2 hingga 10. Setelah mencetak 45 iterasi parameter serta nilai *silhouette coefficient*, ditemukan *Eps* = 0.0075 dan *MinPts* = 4 dengan nilai *silhouette coefficient* tertinggi yaitu 0.551 yang mendekati 1. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 2 *cluster* dan 21 titik *noise*.

Tabel 4. 5 Iterasi Parameter Titik Penjemputan Sore Hari

<i>Eps</i> \ <i>MinPts</i>	0.006	0.0065	0.007	0.0075	0.008
2	0.311	0.311	0.344	0.418	0.418
3	0.309	0.309	0.332	0.408	0.408
4	0.533	0.533	0.404	0.551	0.550
5	0.515	0.533	0.533	0.537	0.546
6	0.436	0.515	0.515	0.520	0.527
7	0.144	0.457	0.473	0.520	0.520
8	0.401	0.441	0.457	0.479	0.488
9	0.383	0.402	0.446	0.475	0.475
10	0.372	0.402	0.424	0.475	0.475



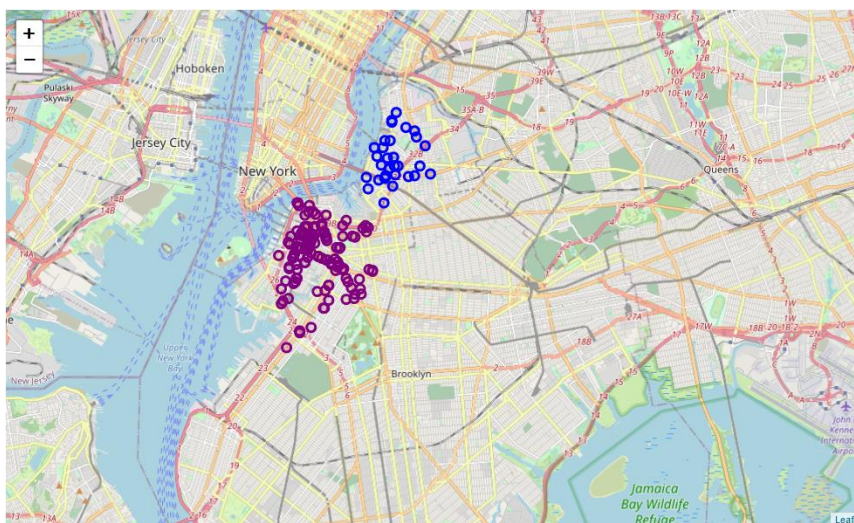
Gambar 4. 25 Visualisasi *Cluster* Titik Penjemputan Sore Hari

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada sore hari, terdapat dua *cluster* yang berwarna merah dan kuning. Untuk jumlah anggota *cluster* dapat dilihat pada

Gambar 4. 26. Terdapat 2 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi peta, *cluster* 0 ditandai dengan warna ungu dan *cluster* 1 ditandai dengan warna biru.

cluster	
0	130
1	32
-1	21

Gambar 4. 26 Jumlah Anggota *Cluster* Titik Penjemputan Sore Hari



Gambar 4. 27 Visualisasi Peta *Cluster* Titik Penjemputan Sore Hari

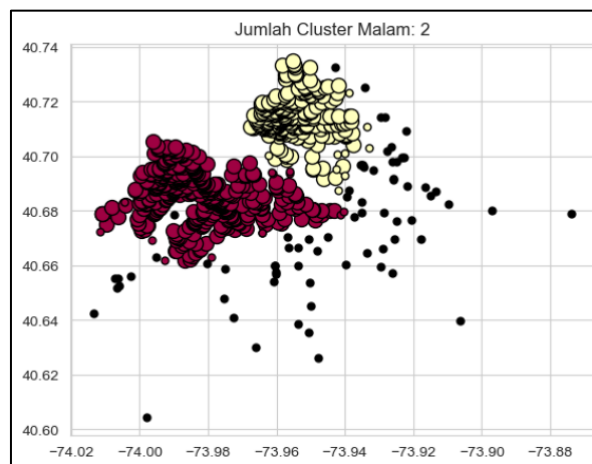
4.3.6 Clustering Titik Penjemputan Malam Hari

Pada titik penjemputan malam hari, dilakukan iterasi percobaan *Eps* dengan rentang 0.005 hingga 0.007 dan *MinPts* 2 hingga 10. Setelah mencetak 45 iterasi parameter serta nilai *silhouette coefficient*, ditemukan *Eps* = 0.005 dan *MinPts* = 6 dengan nilai *silhouette coefficient* tertinggi yaitu 0.546 yang mendekati 1. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 2 *cluster* dan 72 titik *noise*.

Tabel 4. 6 Iterasi Parameter Titik Penjemputan Malam Hari

<i>Eps</i> \ <i>MinPts</i>	0.005	0.0055	0.006	0.0065	0.007
2	0.015	0.156	0.166	0.209	0.221
3	0.031	0.153	0.162	0.208	0.221

<i>Eps</i> \ <i>MinPts</i>	0.005	0.0055	0.006	0.0065	0.007
4	0.044	0.093	0.093	0.209	0.176
5	0.047	0.093	0.090	0.231	0.232
6	0.546	0.283	0.286	0.475	0.482
7	0.266	0.087	0.285	0.456	0.472
8	0.533	0.042	0.117	0.291	0.459
9	0.531	0.039	0.117	0.290	0.291
10	0.527	0.534	0.107	0.127	0.450

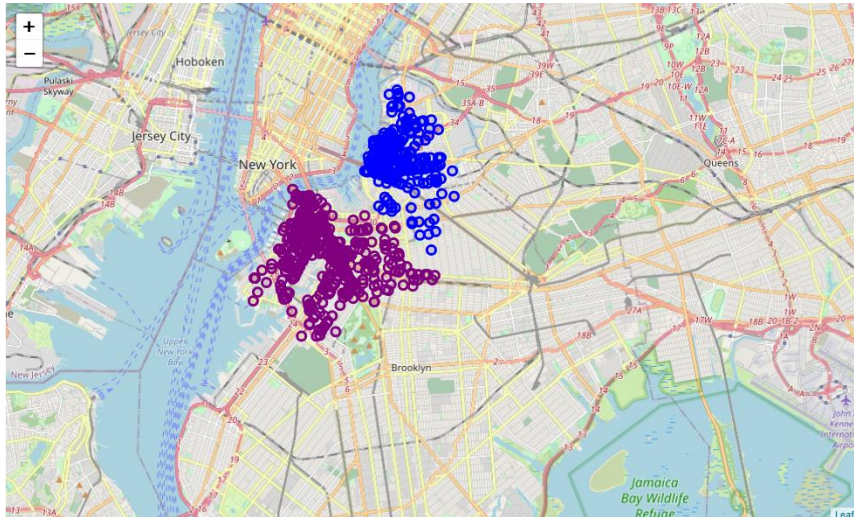


Gambar 4. 28 Visualisasi *Cluster* Titik Penjemputan Malam Hari

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada malam hari, terdapat dua *cluster* besar yang berwarna merah dan kuning. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 29. Terdapat 2 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi peta, *cluster* 0 ditandai dengan warna ungu dan *cluster* 1 ditandai dengan warna biru.

cluster	
0	525
1	316
-1	72

Gambar 4. 29 Jumlah Anggota *Cluster* Titik Penjemputan Malam Hari



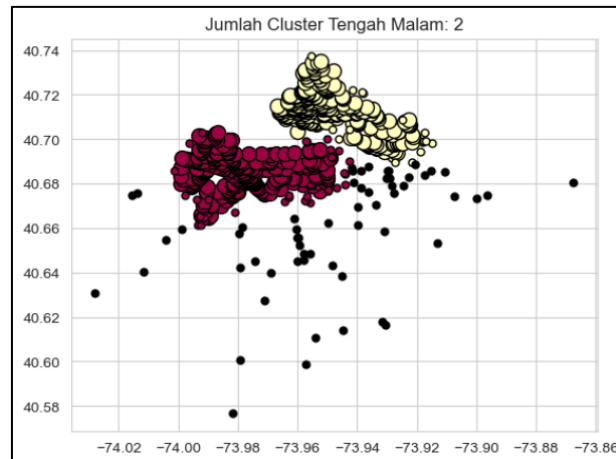
Gambar 4. 30 Visualisasi Peta *Cluster* Titik Penjemputan Malam Hari

4.3.7 Clustering Titik Penjemputan Tengah Malam

Pada titik penjemputan tengah malam, dilakukan iterasi percobaan *Eps* dengan rentang 0.006 hingga 0.008 dan *MinPts* 11 hingga 20. Setelah mencetak 50 iterasi parameter serta nilai *silhouette coefficient*, ditemukan *Eps* = 0.007 dan *MinPts* = 15 dengan nilai *silhouette coefficient* tertinggi yaitu 0.512 yang mendekati 1. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 2 *cluster* dan 60 titik *noise*.

Tabel 4. 7 Iterasi Parameter Titik Penjemputan Tengah Malam

<i>Eps</i> \ <i>MinPts</i>	0.006	0.0065	0.007	0.0075	0.008
11	0.509	0.462	0.469	0.482	0.281
12	0.509	0.507	0.464	0.476	0.477
13	0.509	0.509	0.461	0.476	0.477
14	0.384	0.509	0.508	0.507	0.467
15	0.356	0.510	0.512	0.507	0.467
16	0.459	0.508	0.511	0.509	0.507
17	0.429	0.470	0.511	0.509	0.509
18	0.425	0.447	0.506	0.508	0.508
19	0.405	0.441	0.494	0.508	0.508
20	0.344	0.436	0.494	0.497	0.508

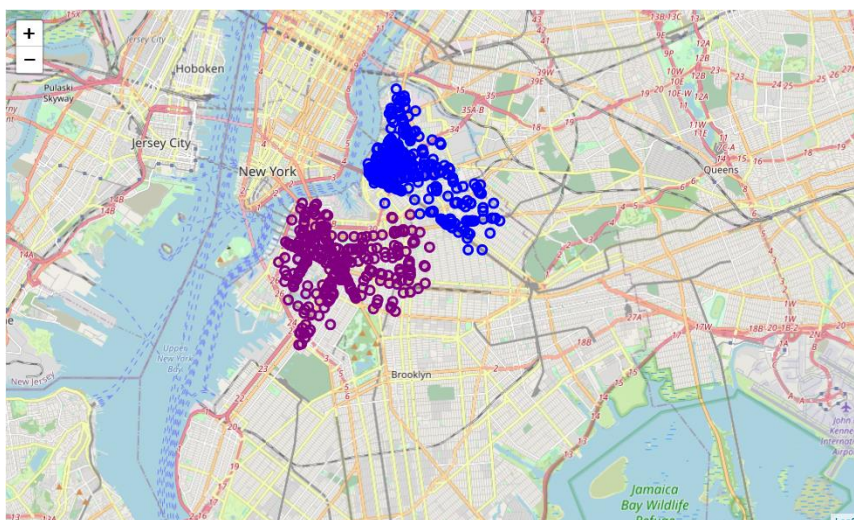


Gambar 4. 31 Visualisasi *Cluster* Titik Penjemputan Tengah Malam

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada tengah malam, terdapat dua *cluster* besar yang berwarna merah dan kuning. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 32. Terdapat 2 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi peta, *cluster* 0 ditandai dengan warna ungu dan *cluster* 1 ditandai dengan warna biru.

cluster	
1	389
0	317
-1	60

Gambar 4. 32 Jumlah Anggota *Cluster* Titik Penjemputan Tengah Malam



Gambar 4. 33 Visualisasi Peta *Cluster* Titik Penjemputan Tengah Malam

4.4 Clustering Titik Pengantaran

Clustering menggunakan DBSCAN memerlukan dua parameter yaitu *Eps* dan *MinPts*. Penentuan parameter tersebut sangat krusial untuk menghasilkan kualitas *cluster* yang baik. Pada penelitian ini, dilakukan iterasi antara *Eps* dan *MinPts* yang akan menghasilkan *silhouette coefficient* untuk menilai tingkat kepadatan *cluster*. Pasangan parameter yang menghasilkan nilai *silhouette coefficient* mendekati 1 merupakan parameter terbaik dan akan digunakan untuk tahap *clustering*. Algoritma akan mencetak jumlah dan anggota *cluster* yang terbentuk. Terdapat titik *noise* yang dicetak dengan label -1 dan akan dibuang. Titik-titik yang termasuk dalam anggota *cluster* akan disimpan pada *data frame* baru untuk mengidentifikasi area potensial titik pengantaran melalui visualisasi peta.

Untuk mengetahui area potensial tempat terjadinya titik pengantaran berdasarkan rentang waktu, atribut spasial yang sudah dibagi terhadap rentang waktu. Pengelompokan rentang waktu tersebut dibagi berdasarkan hari kerja (Senin – Jumat) dan hari libur (Sabtu dan Minggu). Selain itu, catatan perjalanan juga dikelompokkan untuk setiap rentang jam diantaranya pagi (05:00 – 10:59), siang (11:01 – 14:59), sore (15:00 – 17:59), malam (18:00 – 23:59), dan tengah malam (00:00 – 04:59). Berikut merupakan proses *clustering*, iterasi pencarian parameter terbaik, dan visualisasi *cluster* titik pengantaran berdasarkan karakteristik waktu.

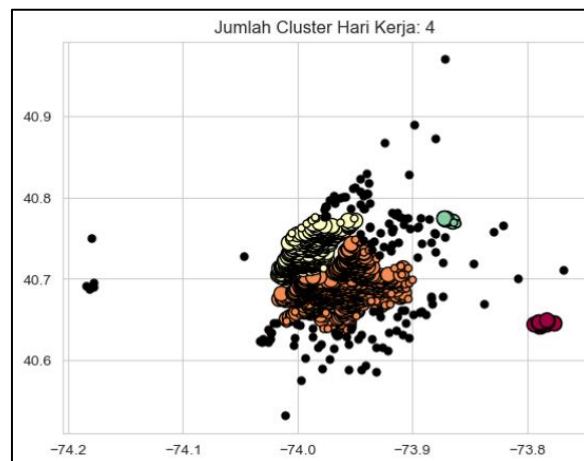
4.4.1 Clustering Titik Pengantaran Hari Kerja

Pada titik pengantaran hari kerja, dilakukan iterasi percobaan *Eps* dengan rentang 0.006 hingga 0.008 dan *MinPts* 2 hingga 10. Setelah mencetak 45 iterasi parameter serta nilai *silhouette coefficient*, ditemukan *Eps* = 0.008 dan *MinPts* = 8 dengan nilai *silhouette coefficient* tertinggi yaitu 0.321. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 4 *cluster* dan 146 titik *noise*.

Tabel 4. 8 Iterasi Parameter Titik Pengantaran Hari Kerja

<i>Eps</i> \ <i>MinPts</i>	0.006	0.0065	0.007	0.0075	0.008
2	0.024	0.047	0.055	0.096	0.139
3	0.052	0.062	0.090	0.122	0.177
4	0.063	0.078	0.133	0.166	0.172
5	0.153	0.147	0.228	0.231	0.203

<i>Eps</i> \ <i>MinPts</i>	0.006	0.0065	0.007	0.0075	0.008
6	0.168	0.179	0.182	0.247	0.235
7	0.004	0.122	0.121	0.306	0.310
8	0.039	0.014	0.106	0.318	0.321
9	0.208	0.089	0.105	0.316	0.318
10	0.206	0.211	0.226	0.098	0.316

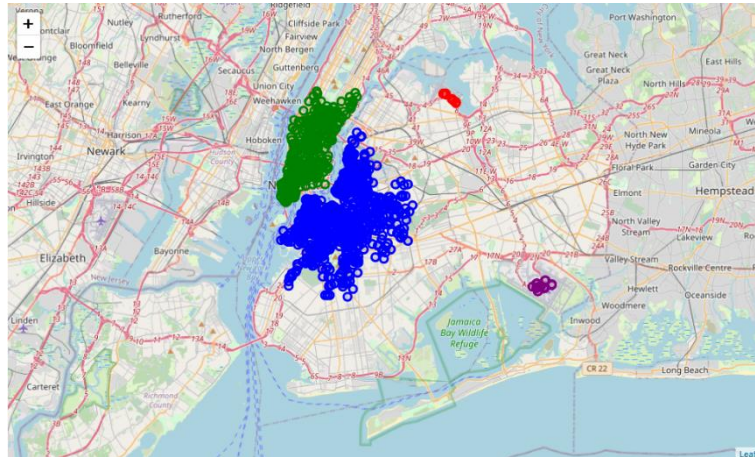


Gambar 4. 34 Visualisasi *Cluster* Titik Pengantaran Hari Kerja

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada hari kerja, terdapat dua *cluster* yang berwarna kuning, oren, hijau, dan merah. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 35. Terdapat 4 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi peta, *cluster* 0 ditandai dengan warna ungu, *cluster* 1 berwarna biru, *cluster* 2 berwarna hijau, dan *cluster* 3 berwarna merah.

cluster	
1	864
2	430
-1	146
0	17
3	14

Gambar 4. 35 Jumlah Anggota *Cluster* Titik Pengantaran Hari Kerja



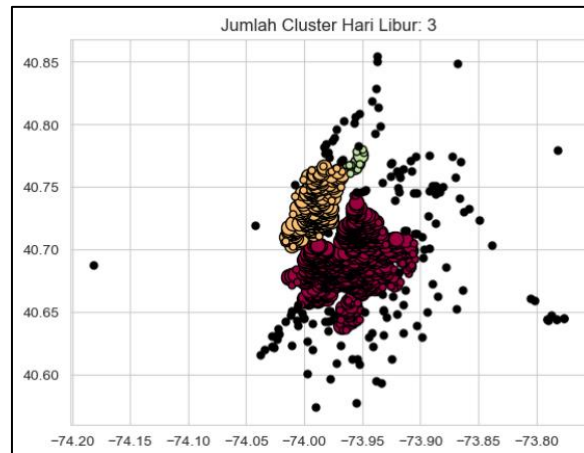
Gambar 4. 36 Visualisasi Peta *Cluster* Titik Pengantaran Hari Kerja

4.4.2 Clustering Titik Pengantaran Hari Libur

Pada titik pengantaran hari libur, dilakukan iterasi percobaan Eps dengan rentang 0.004 hingga 0.0065 dan $MinPts$ 2 hingga 10. Setelah mencetak 45 iterasi parameter serta nilai *silhouette coefficient*, ditemukan $Eps = 0.0065$ dan $MinPts = 7$ dengan nilai *silhouette coefficient* tertinggi yaitu 0.258. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 3 *cluster* dan 151 titik *noise*.

Tabel 4. 9 Iterasi Parameter Titik Pengantaran Hari Libur

$Eps \backslash MinPts$	0.004	0.0045	0.005	0.006	0.0065
2	-0.016	-0.125	-0.036	-0.052	0.090
3	-0.004	-0.123	-0.067	-0.054	0.096
4	-0.027	-0.095	-0.075	-0.005	0.216
5	-0.101	0.014	-0.053	0.098	0.219
6	-0.016	0.023	-0.079	0.100	0.251
7	-0.075	-0.000	-0.024	0.101	0.258
8	-0.094	0.003	-0.155	0.073	0.152
9	-0.126	-0.082	0.017	-0.012	0.151
10	-0.213	-0.078	0.004	-0.049	0.074

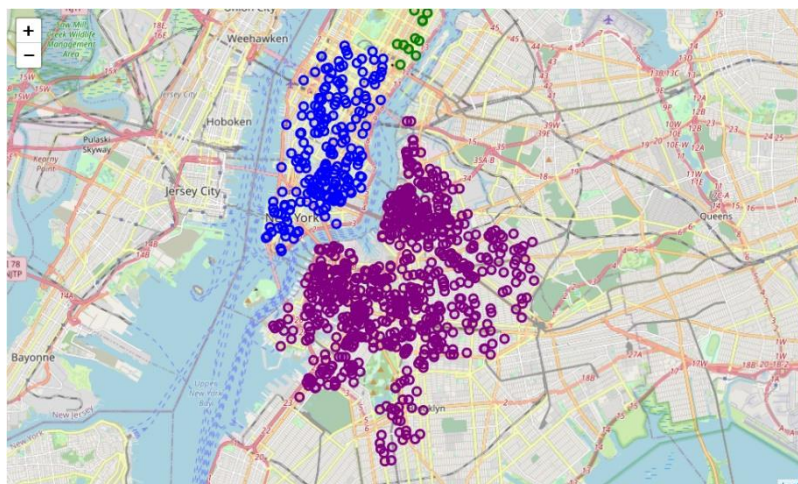


Gambar 4. 37 Visualisasi *Cluster* Titik Pengantaran Hari Libur

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada hari libur, terdapat dua *cluster* yang berwarna kuning, hijau, dan merah. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 38. Terdapat 3 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi peta, *cluster* 0 ditandai dengan warna ungu, *cluster* 1 berwarna biru, dan *cluster* 2 berwarna hijau.

cluster	
0	766
1	210
-1	151
2	12

Gambar 4. 38 Jumlah Anggota *Cluster* Titik Pengantaran Hari Libur



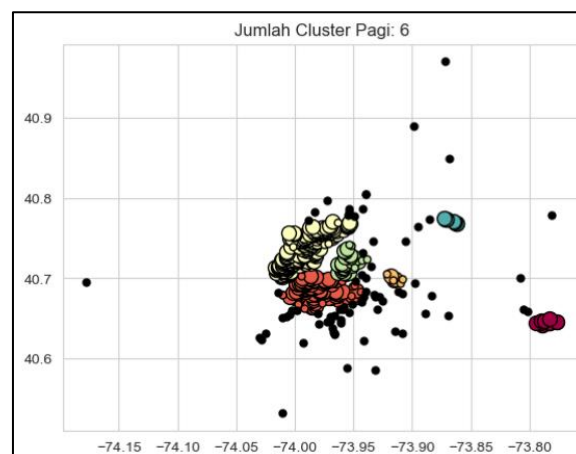
Gambar 4. 39 Visualisasi Peta *Cluster* Titik Pengantaran Hari Libur

4.4.3 Clustering Titik Pengantaran Pagi Hari

Pada titik pengantaran pagi hari, dilakukan iterasi percobaan *Eps* dengan rentang 0.006 hingga 0.008 dan *MinPts* 2 hingga 10. Setelah mencetak 45 iterasi parameter serta nilai *silhouette coefficient*, ditemukan *Eps* = 0.0075 dan *MinPts* = 5 dengan nilai *silhouette coefficient* tertinggi yaitu 0.272. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 6 *cluster* dan 74 titik *noise*.

Tabel 4. 10 Iterasi Parameter Titik Pengantaran Pagi Hari

<i>Eps</i> \ <i>MinPts</i>	0.006	0.0065	0.007	0.0075	0.008
2	-0.008	0.231	0.218	0.234	0.234
3	-0.002	0.211	0.193	0.230	0.244
4	-0.012	0.169	0.216	0.210	0.205
5	-0.045	0.166	0.264	0.272	0.226
6	-0.098	-0.038	0.127	0.199	0.263
7	-0.018	-0.002	0.074	0.156	0.253
8	-0.065	-0.021	0.018	0.067	0.234
9	-0.010	0.039	0.079	0.122	0.217
10	0.018	0.038	0.072	0.082	0.125



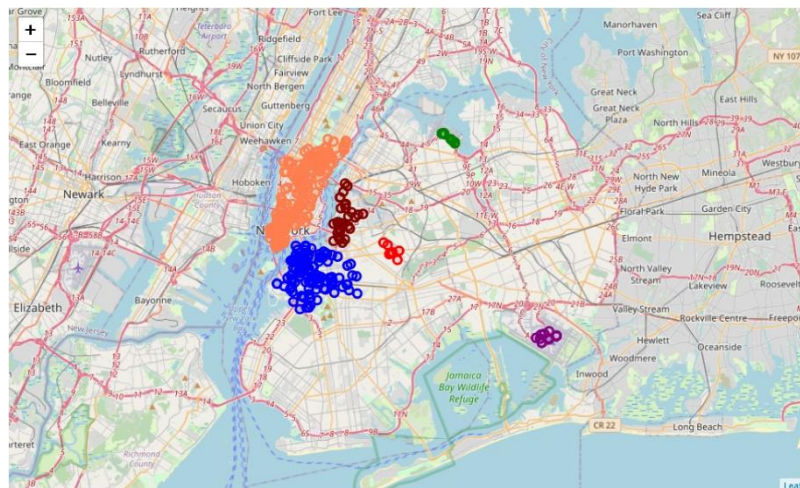
Gambar 4. 40 Visualisasi *Cluster* Titik Pengantaran Pagi Hari

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada pagi hari, terdapat 6 *cluster* yang memiliki beragam warna selain titik berwarna hitam. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 41. Terdapat 6 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi

peta, *cluster* 0 ditandai dengan warna ungu, *cluster* 1 berwarna biru, *cluster* 2 berwarna merah, *cluster* 3 berwarna coral, *cluster* 4 berwarna maroon, dan *cluster* 5 berwarna hijau.

cluster	
3	209
1	120
-1	74
4	32
0	12
5	10
2	7

Gambar 4. 41 Jumlah Anggota *Cluster* Titik Pengantaran Pagi Hari



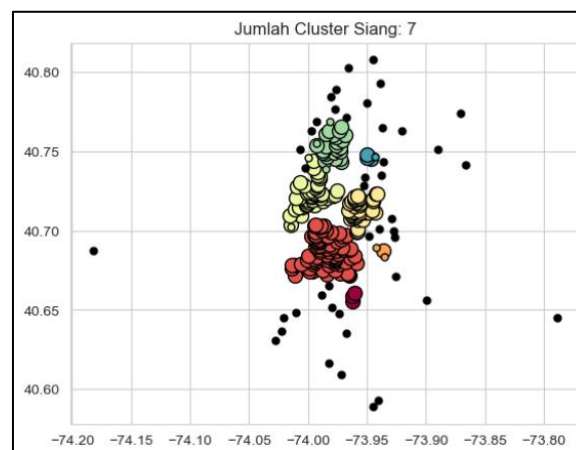
Gambar 4. 42 Visualisasi Peta *Cluster* Titik Pengantaran Pagi Hari

4.4.4 *Clustering* Titik Pengantaran Siang Hari

Pada titik pengantaran siang hari, dilakukan iterasi percobaan *Eps* dengan rentang 0.006 hingga 0.008 dan *MinPts* 2 hingga 10. Setelah mencetak 45 iterasi parameter serta nilai *silhouette coefficient*, ditemukan *Eps* = 0.007 dan *MinPts* = 3 dengan nilai *silhouette coefficient* tertinggi yaitu 0.298. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 7 *cluster* dan 43 titik *noise*.

Tabel 4. 11 Iterasi Parameter Titik Pengantaran Siang Hari

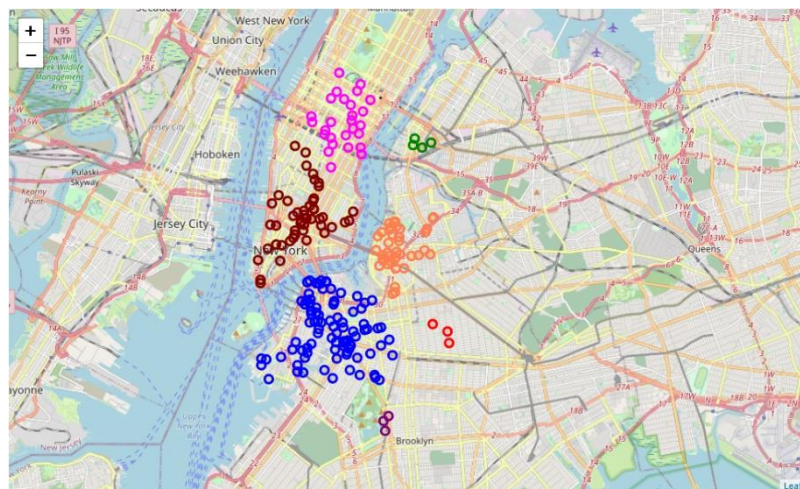
Eps \ $MinPts$	0.006	0.0065	0.007	0.0075	0.008
2	0.118	0.157	0.288	0.213	0.222
3	0.179	0.286	0.298	0.268	0.268
4	0.207	0.260	0.277	0.289	0.294
5	0.202	0.254	0.260	0.280	0.292
6	0.158	0.218	0.253	0.282	0.290
7	0.097	0.165	0.218	0.238	0.285
8	0.026	0.127	0.178	0.222	0.255
9	0.027	0.030	0.152	0.198	0.228
10	0.032	0.030	0.032	0.150	0.209

Gambar 4. 43 Visualisasi *Cluster* Titik Pengantaran Siang Hari

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada siang hari, terdapat 7 *cluster* yang memiliki beragam warna selain titik berwarna hitam. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 44. Terdapat 7 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi peta, *cluster* 0 ditandai dengan warna ungu, *cluster* 1 berwarna biru, *cluster* 2 berwarna merah, *cluster* 3 berwarna coral, *cluster* 4 berwarna maroon, *cluster* 5 berwarna magenta, dan *cluster* 6 berwarna hijau.

cluster	
1	91
4	50
-1	43
3	39
5	27
6	4
0	3
2	3

Gambar 4. 44 Jumlah Anggota *Cluster* Titik Pengantaran Siang Hari



Gambar 4. 45 Visualisasi Peta *Cluster* Titik Pengantaran Siang Hari

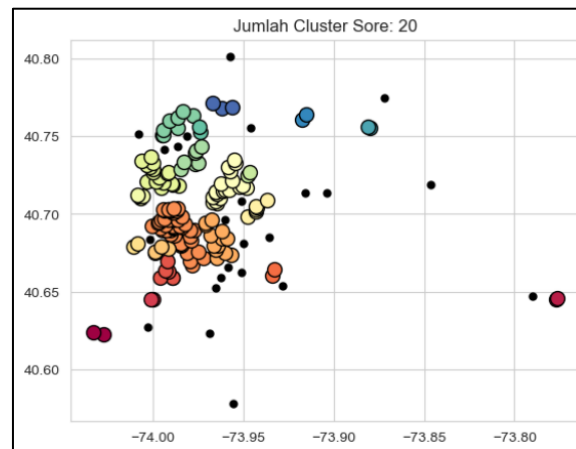
4.4.5 Clustering Titik Pengantaran Sore Hari

Pada titik pengantaran sore hari, dilakukan iterasi percobaan *Eps* dengan rentang 0.0071 hingga 0.0075 dan *MinPts* 2 hingga 10. Setelah mencetak 45 iterasi parameter serta nilai *silhouette coefficient*, ditemukan *Eps* = 0.0072 dan *MinPts* = 2 dengan nilai *silhouette coefficient* tertinggi yaitu 0.256. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 20 *cluster* dan 24 titik *noise*.

Tabel 4. 12 Iterasi Parameter Titik Pengantaran Sore Hari

<i>Eps</i> \ <i>MinPts</i>	0.0071	0.0072	0.0073	0.0074	0.0075
2	0.226	0.256	0.252	0.189	0.194
3	0.137	0.170	0.179	0.133	0.142
4	0.135	0.148	0.158	0.093	0.132

Eps \ $MinPts$	0.0071	0.0072	0.0073	0.0074	0.0075
5	0.083	0.086	0.075	0.020	0.020
6	0.034	0.034	0.036	0.037	0.043
7	-0.004	0.004	0.009	0.012	0.016
8	-0.087	-0.046	-0.046	-0.046	-0.021
9	-0.096	-0.091	-0.091	-0.088	-0.088
10	-0.112	-0.107	-0.107	-0.107	-0.098

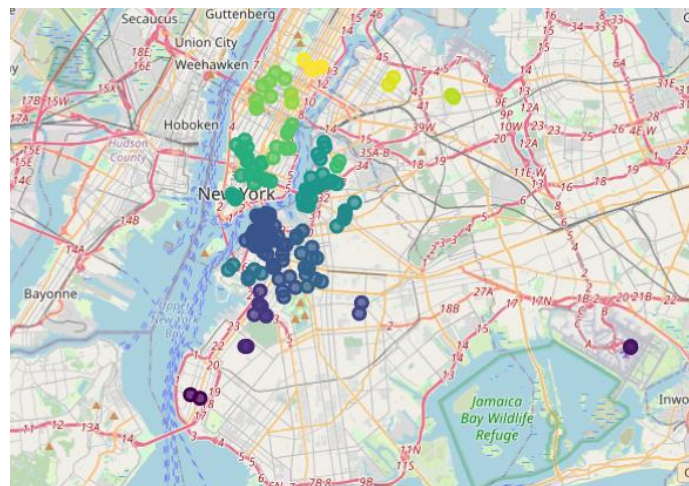


Gambar 4. 46 Visualisasi *Cluster* Titik Pengantaran Sore Hari

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada sore hari, terdapat 20 *cluster* yang memiliki beragam warna selain titik berwarna hitam. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 47. Terdapat 20 *cluster* dengan beragam jumlah titik pengantaran dan 1 data *outlier* yang dilabeli dengan -1. Pada titik ini, yang digunakan untuk analisis temporal adalah *cluster* yang memiliki banyak anggota diantaranya *cluster* 5, 6, 10, 12, 14, dan 15.

cluster		cluster	
5	46	11	4
-1	24	19	3
10	20	8	2
12	17	0	2
6	13	4	2
15	8	13	2
14	7	2	2
9	6	1	2
3	6	16	2
7	4	17	2
		18	2

Gambar 4. 47 Jumlah Anggota *Cluster* Titik Pengantaran Siang Hari



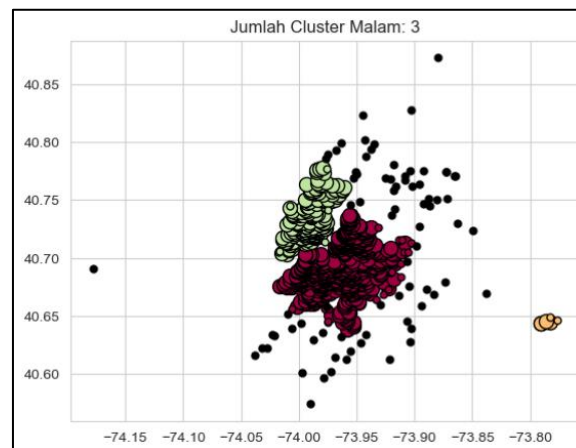
Gambar 4. 48 Visualisasi Peta *Cluster* Titik Pengantaran Sore Hari

4.4.6 Clustering Titik Pengantaran Malam Hari

Pada titik pengantaran malam hari, dilakukan iterasi percobaan *Eps* dengan rentang 0.006 hingga 0.008 dan *MinPts* 2 hingga 10. Setelah mencetak 45 iterasi parameter serta nilai *silhouette coefficient*, ditemukan *Eps* = 0.008 dan *MinPts* = 5 dengan nilai *silhouette coefficient* tertinggi yaitu 0.330. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 3 *cluster* dan 77 titik *noise*.

Tabel 4. 13 Iterasi Parameter Titik Pengantaran Malam Hari

Eps \ $MinPts$	0.006	0.0065	0.007	0.0075	0.008
2	0.010	0.021	0.028	0.057	0.214
3	-0.002	0.006	0.026	0.045	0.287
4	-0.029	0.026	0.025	0.073	0.306
5	0.239	0.172	0.171	0.081	0.330
6	0.217	0.240	0.262	0.320	0.230
7	0.101	0.106	0.237	0.302	0.227
8	0.090	0.104	0.235	0.302	0.306
9	0.167	0.099	0.233	0.236	0.301
10	0.007	0.087	0.098	0.135	0.298

Gambar 4. 49 Visualisasi *Cluster* Titik Pengantaran Malam Hari

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada malam hari, terdapat 3 *cluster* yang berwarna hijau, merah, dan oren. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 50. Terdapat 3 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi peta, *cluster* 0 ditandai dengan warna ungu, *cluster* 1 berwarna biru, dan *cluster* 2 berwarna merah.

cluster	
0	617
2	191
-1	77
1	7

Gambar 4. 50 Jumlah Anggota *Cluster* Titik Pengantaran Malam Hari



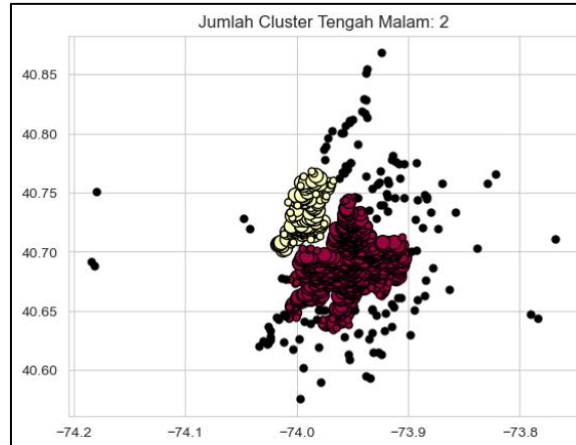
Gambar 4. 51 Visualisasi Peta *Cluster* Titik Pengantaran Malam Hari

4.4.7 Clustering Titik Pengantaran Tengah Malam

Pada titik pengantaran malam hari, dilakukan iterasi percobaan *Eps* dengan rentang 0.006 hingga 0.008 dan *MinPts* 2 hingga 10. Setelah mencetak 45 iterasi parameter serta nilai *silhouette coefficient*, ditemukan *Eps* = 0.0075 dan *MinPts* = 7 dengan nilai *silhouette coefficient* tertinggi yaitu 0.283. Kedua parameter tersebut akan dimasukkan pada algoritma DBSCAN sehingga menghasilkan 2 *cluster* dan 131 titik *noise*.

Tabel 4. 14 Iterasi Parameter Titik Pengantaran Tengah Malam

<i>Eps</i> \ <i>MinPts</i>	0.006	0.0065	0.007	0.0075	0.008
2	-0.097	-0.049	-0.049	-0.009	-0.022
3	-0.053	-0.036	-0.036	-0.035	-0.049
4	0.154	0.203	0.210	0.249	0.274
5	-0.062	0.216	0.193	0.226	0.195
6	-0.137	0.112	0.239	0.243	0.186
7	-0.108	-0.007	0.150	0.283	0.418
8	-0.041	-0.061	0.142	0.278	0.411
9	-0.055	-0.015	0.148	0.267	0.399
10	0.066	0.096	-0.059	0.241	0.259

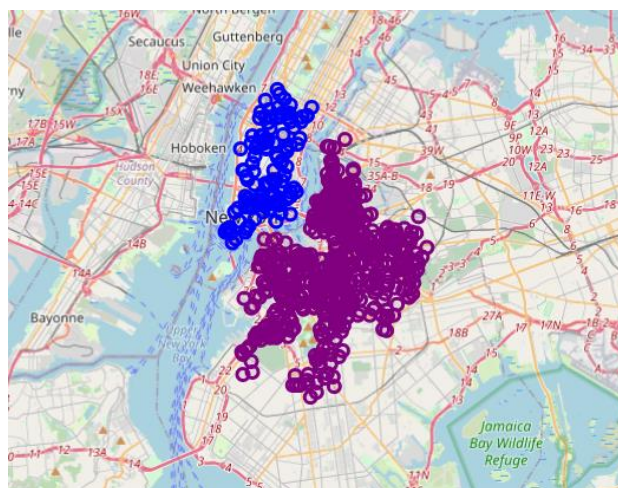


Gambar 4. 52 Visualisasi *Cluster* Titik Pengantaran Tengah Malam

Berdasarkan visualisasi hasil *cluster* titik penjemputan pada tengah malam, terdapat 2 *cluster* padat yang berwarna kuning dan merah. Jumlah anggota *cluster* dapat dilihat pada Gambar 4. 53. Terdapat 2 *cluster* dan 1 *outlier* yang dilabeli dengan -1. Pada visualisasi peta, *cluster* 0 ditandai dengan warna ungu dan *cluster* 1 berwarna biru.

cluster	
0	567
-1	131
1	120

Gambar 4. 53 Jumlah Anggota *Cluster* Titik Pengantaran Tengah Malam



Gambar 4. 54 Visualisasi Peta *Cluster* Titik Pengantaran Tengah Malam

4.5 Pattern Evaluation

Setelah mendapatkan anggota *cluster* pada titik penjemputan dan titik pengantaran, visualisasi peta terhadap anggota *cluster* akan membantu untuk mengetahui nama area dengan kepadatan masing-masing titik. Pengenalan pola ini dapat dijadikan sebagai acuan bagi sopir taksi dan pengelola taksi saat mengoperasikan taksi. Sopir taksi dapat mengetahui area potensial titik penjemputan pada hari kerja dan hari libur, lalu dapat menyesuaikan akomodasi taksi untuk melayani penumpang dengan baik. Sopir taksi juga dapat melihat area potensial yang padat akan penumpang berdasarkan waktu pagi, siang, sore, malam, dan tengah malam.

Melalui pengetahuan area potensial titik penjemputan, sopir taksi dapat menyiapkan bahan bakar taksi secara maksimal untuk menghindari kehabisan bahan bakar. Selain itu, sopir taksi juga dapat mengurangi celah mengendarai taksi tanpa penumpang. Sehingga sopir taksi dapat memaksimalkan pendapatan melalui pengetahuan area potensial tersebut. Sopir taksi juga dapat memilih waktu istirahat pada waktu yang rendah akan permintaan taksi. Berikut merupakan visualisasi peta yang akan memudahkan pengguna dalam mengetahui nama jalan yang menjadi area potensial di Kota Brooklyn.

4.5.1 Knowledge Titik Penjemputan

Berdasarkan visualisasi peta yang diperbesar, terdapat nama area dan jalan yang padat akan permintaan titik penjemputan. Pada setiap pembagian waktu, *cluster* terbagi menjadi dua. *Cluster* 0 terdapat di bagian barat laut Kota Brooklyn, sedangkan *cluster* 1 terdapat pada bagian utara Kota Brooklyn. Saat mengidentifikasi *cluster* yang terbentuk secara padat, titik penjemputan pada *cluster* 0 cenderung terjadi pada daerah yang berdekatan dengan fasilitas umum dan tempat wisata. Sedangkan, *cluster* 1 cenderung memiliki titik penjemputan yang terjadi di persimpangan jalan. Namun, pada *cluster* 1 juga terdapat beberapa titik penjemputan yang berdekatan dengan fasilitas umum seperti tempat untuk bersosialisasi. Berikut merupakan tabel area potensial titik penjemputan.

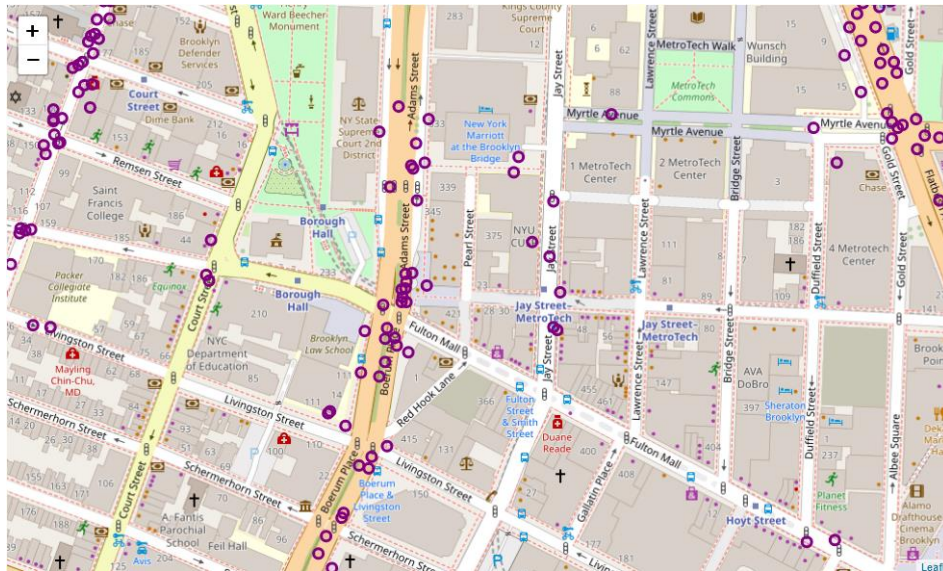
Tabel 4. 15 Area Potensial Titik Penjemputan

Waktu	<i>Cluster</i>	Daerah	Jalan Utama
Hari Kerja	0	Utara Kota Brooklyn	Williamsburg, Bushwick
	1	Barat Laut Kota Brooklyn	Dumbo, Cobble Hill, Carroll Gardens, Prospect Heights, Park

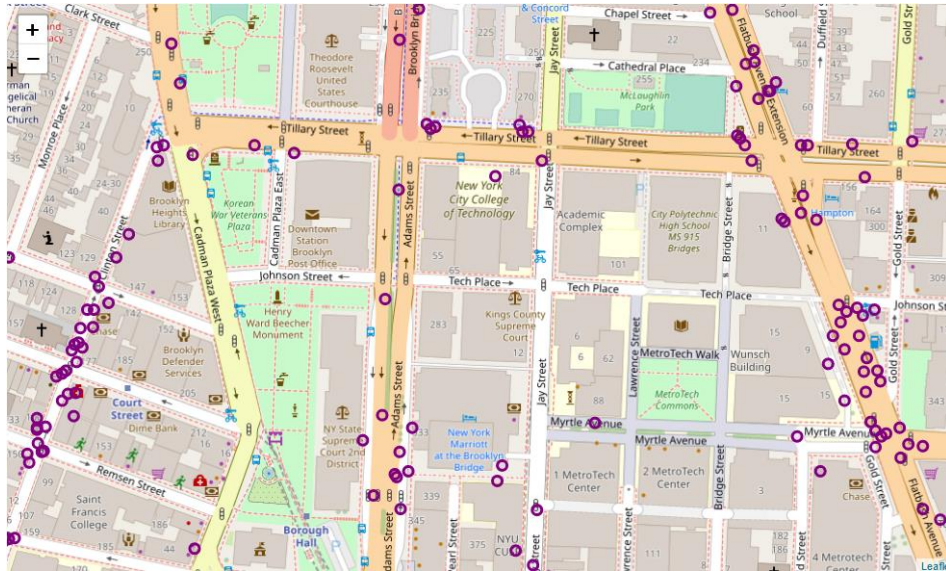
Waktu	Cluster	Daerah	Jalan Utama
			Slope, Clinton Hill, Fort Greene, Triple Cantilever
Hari Libur	0	Utara Kota Brooklyn	Williamsburg, Green Point
	1	Barat Laut Kota Brooklyn	Dumbo, Cobble Hill, Carroll Gardens, Prospect Heights, Park Slope, Clinton Hill, Fort Greene
Pagi	0	Dumbo, Triple Cantilever, Carroll Gardens, South Slope, Prospect Heights, Park Slope, Cobble Hill, Clinton Hill, Fort Greene	Clinton Street, Boerum Place, Flatbush Avenue, Atlantic Avenue
	1	Williamsburg	Bedford Avenue, Grand Street
Siang	0	Dumbo, Triple Cantilever, Carroll Gardens, South Slope, Prospect Heights, Park Slope, Cobble Hill, Clinton Hill, Fort Greene	Flatbush Avenue, Smith Street, Clinton Street, Adams Street, Atlantic Avenue
	1	Williamsburg	Williamsburg Bridge, Bedford Avenue
Sore	0	Dumbo, Triple Cantilever, Carroll Gardens, Prospect Heights, Park Slope, Cobble Hill, Clinton Hill, Fort Greene	Flatbush Avenue, Atlantic Avenue, Boerum Place
	1	Williamsburg	Metropolitan Avenue
Malam	0	Carroll Gardens, South Slope, Park Slope, Gotham Greens, Cobble Hill, Dumbo, Prospect Heights, Clinton Hill	4 th Avenue, Smith Street, Atlantic Avenue, Flatbush Avenue, Court Street, Fulton Street
	1	Williamsburg, Green Point	Metropolitan Avenue, Roebling Street, Wythe Avenue, Borinquen Place, Grand Street, Manhattan Avenue
Tengah Malam	0	Park Slope, Prospect Heights, Gotham Greens, Cobble Hill, Dumbo, Fort Greene, Clinton Hill,	4 th Avenue, Smith Street, Atlantic Avenue, Flatbush Avenue, DeKalb Avenue, Bergen Street, Myrtle Avenue

Waktu	Cluster	Daerah	Jalan Utama
	1	Williamsburg, Bushwick	Broadway, Myrtle Avenue, Jefferson Street, Grand Street, Lorimer Street, Marcy Avenue, Roebling Street, Nassau Avenue, Greenpoint Avenue

Berdasarkan visualisasi peta yang diperbesar pada area Boerum Place dan Adams Street saat hari kerja, diketahui pada bagian barat laut Kota Brooklyn terdapat beberapa titik penjemputan di area perguruan tinggi diantaranya Saint Francis College, Packer Collegiate Institute, dan Brooklyn Law School serta bangunan pemerintahan yaitu Borough Hall. Selain itu, terdapat stasiun transportasi umum di sekitar titik penjemputan Borough Hall dan Jay Street. Simbol-simbol pada visualisasi juga menunjukkan bahwa daerah tersebut terdapat fasilitas umum seperti hotel, pusat berbelanja, dan kantor *bank*. Ketika visualisasi diperbesar pada Flatbush Avenue, terdapat titik penjemputan pada fasilitas umum yaitu William Jay Gaynor Memorial, Cadman Plaza Park, Walt Whitman Park, dan McLaughlin Park. Diantara taman tersebut terdapat apartemen yang bernama Concord Village dan titik penjemputan terjadi di bagian depan apartemen.

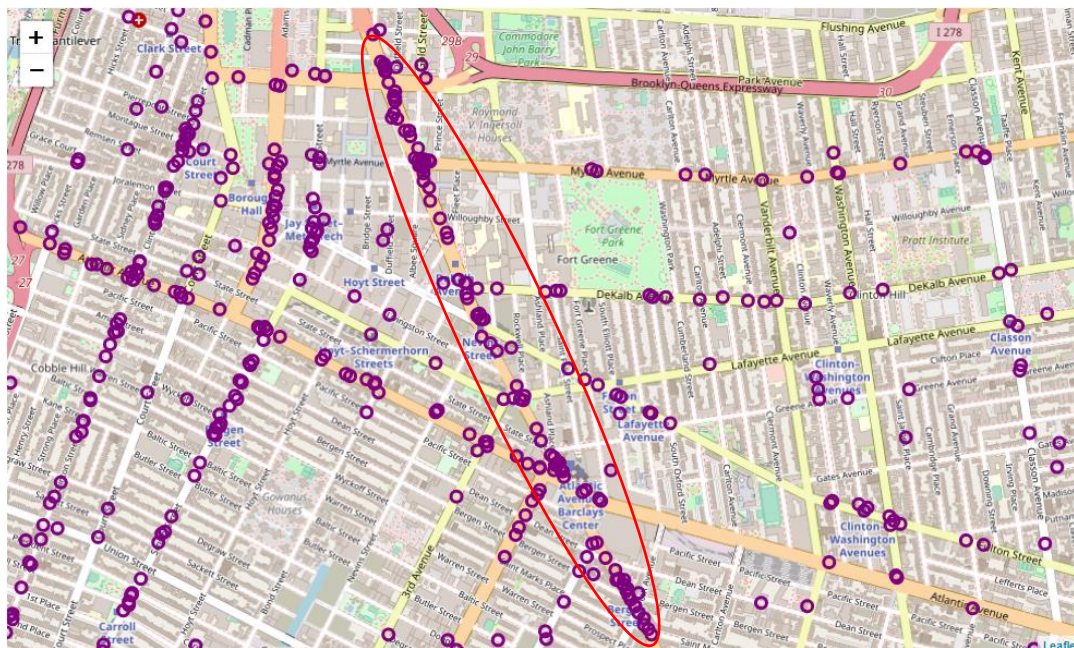


Gambar 4. 55 Visualisasi Peta Titik Penjemputan Hari Kerja



Gambar 4. 56 Area Potensial Titik Penjemputan Hari Kerja

Pada hari libur, titik penjemputan banyak terjadi di sepanjang jalan Flatbush Avenue. Namun, titik penjemputan yang padat terdapat pada persimpangan jalan yang melintasi Flatbush Avenue dan berdekatan dengan pusat perbelanjaan, tempat ibadah, dan tempat hiburan. Flatbush Avenue menjadi jalan raya yang membentang di Kota Brooklyn dan melewati beberapa daerah utama.



Gambar 4. 57 Visualisasi Peta Titik Penjemputan Hari Libur di Flatbush Avenue

4.5.2 Knowledge Titik Pengantaran

Pada visualisasi peta yang diperbesar, terdapat nama area dan jalan yang padat akan permintaan titik pengantaran. Pada setiap pembagian waktu, banyak *cluster* yang terbentuk di Kota Brooklyn hingga ke Manhattan dan Queens. Visualisasi peta akan diperbesar untuk mengetahui daerah titik pengantaran yang terbentuk secara padat melalui pengelompokan sebelumnya. Berikut merupakan tabel area potensial titik pengantaran.

Tabel 4. 16 Area Potensial Titik Pengantaran

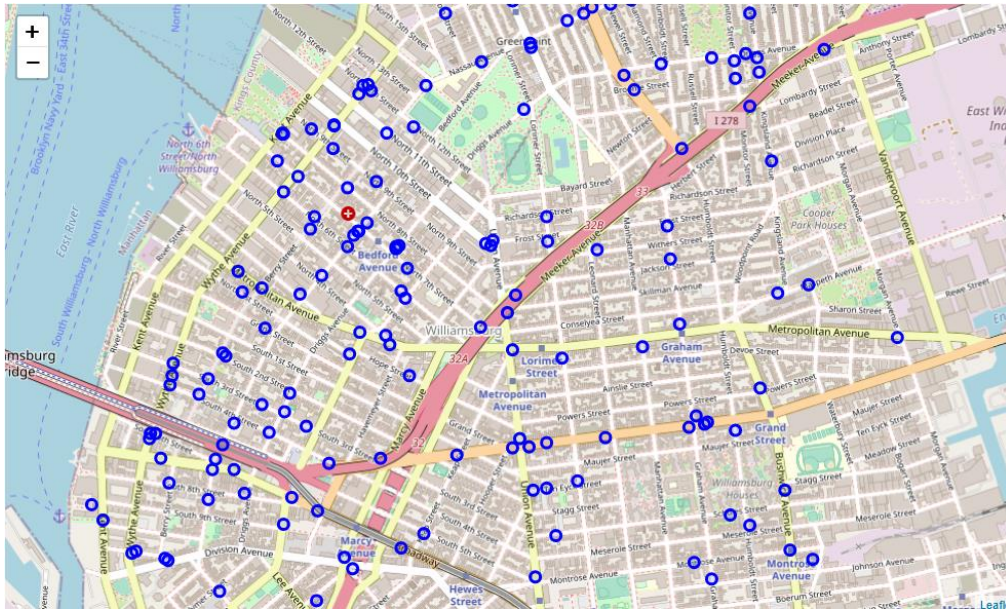
Waktu	Cluster	Daerah	Jalan Utama atau Fasilitas
Hari Kerja	0	Queens	John F. Kennedy International Airport
	1	Barat Laut dan Utara Brooklyn	Barat Laut: Dumbo, Cobble Hill, Carroll Gardens, Prospect Heights, Park Slope, Clinton Hill, Fort Greene, Triple Cantilever, Bedford Stuyvesant, Prospect Park, Boerum Hill, Downtown Brooklyn Utara: Bushwick, Greenpoint, Williamsburg
	2	Manhattan	Lower Manhattan: Delancey Street dan Park Place Midtown Manhattan: Times Square dan Grand Central
	3	Queens	LaGuardia Airport
Hari Libur	0	Barat Laut dan Utara Brooklyn	Barat Laut: Dumbo, Cobble Hill, Carroll Gardens, Prospect Heights, Park Slope, Clinton Hill, Fort Greene, Triple Cantilever, Bedford Stuyvesant, Prospect Park, Boerum Hill, Downtown Brooklyn, Flatbush, Sunset Park Utara: Bushwick, Williamsburg
	1	Manhattan	Midtown Manhattan: Times Square, New York Penn Station Lower Manhattan: Astor Place, Delancey Street, Lafayette Street
	2		Upper East Side Manhattan: Kawasan Timur Central Park
Pagi	0	Queens	John F. Kennedy International Airport
	1	Barat Laut Brooklyn	Cobble Hill, Carroll Gardens, Dumbo, Prospect Heights, Park Slope
	2	Bushwick	Myrtle Avenue

Waktu	Cluster	Daerah	Jalan Utama atau Fasilitas
	3	Manhattan	Midtown Manhattan: Grand Central Lower Manhattan: West Street
	4	Utara Brooklyn	Williamsburg dan Greenpoint
	5	Queens	LaGuardia Airport
Siang	0	Tengah Brooklyn	Prospect Park dan Parkside Avenue
	1	Barat Laut Brooklyn	Atlantic Avenue Barclays Center, Dumbo, Cobble Hill
	2	Brooklyn	Bedford-Stuyvesant
	3	Williamsburg	Bedford Avenue dan Williamsburg Bridge
	4	Manhattan	Lower Manhattan: South of Houston Street
	5		Midtown East Manhattan
	6	Queens	Long Island City dan Court Square
Sore	5	Barat Laut Brooklyn	Dumbo, Downtown Brooklyn, Cobble Hill, Boerum Hill
	6		Prospect Heights dan Clinton Hill
	10	Williamsburg	Bedford Avenue dan Greenpoint Avenue
	12	Manhattan	Chinatown dan Washington Square Village
	14		East Village
	15		Midtown Manhattan: Times Square, Theater District
Malam	0	Barat Laut, Utara Brooklyn, Queens	Barat Laut Brooklyn: Dumbo, Brooklyn Heights, Cobble Hill, Downtown Brooklyn, Fort Greene, Clinton Hill, Carroll Gardens, Gowanus, Park Slope, Prospect Heights, Prospect Park, Crown Heights Utara Brooklyn: Greenpoint, Bushwick, Williamsburg Queens: Ridgewood
	1	Queens	John F. Kennedy International Airport
	2	Manhattan	Lower Manhattan: Delancey Street, 2 nd Avenue, West Village, Washington Square, Alphabet City Midtown Manhattan: Times Square, Garment District, Gramercy
Tengah Malam	0	Brooklyn	Barat Laut: Dumbo, Brooklyn Heights, Cobble Hill, Downtown Brooklyn, Fort Greene, Clinton

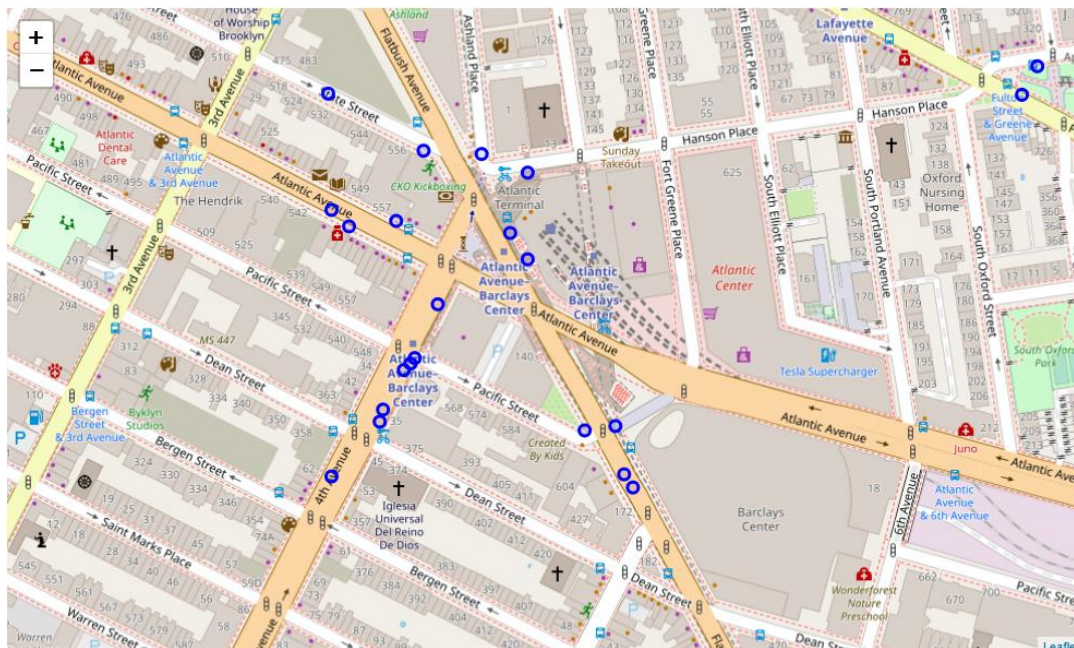
Waktu	Cluster	Daerah	Jalan Utama atau Fasilitas
			Hill, Gowanus, Park Slope, Prospect Heights, Prospect Park, Crown Heights, Gotham Greens Tengah: Flatbush Utara: Williamsburg, Bushwick
	1	Manhattan	Lower Manhattan: Delancey Street, Canal Street, Little Italy, East Village Midtown Manhattan

Titik pengantaran pada hari kerja menyebar di bagian utara Kota Brooklyn yang diantaranya terdapat wilayah Bushwick, Greenpoint, dan Williamsburg. Selain itu, terdapat beberapa jalan utama yang menjadi tujuan penumpang taksi diantaranya Myrtle Avenue, Flatbush Avenue, dan Lafayette Avenue yang berada di wilayah Downtown Brooklyn dan Fort Greene. Gambar berikut menunjukkan area titik pengantaran yang salah satunya banyak terjadi di Williamsburg. Lokasi pada titik pengantaran juga terjadi secara menyebar di bagian utara Kota Brooklyn. Salah satunya adalah Williamsburg yang terkenal dengan kemudahan akses ke Manhattan, kuliner, musik dan hiburan. Daerah tersebut menjadi tujuan yang cenderung dilakukan pada akhir pekan.

Titik pengantaran selanjutnya banyak terjadi di jalan utama seperti Myrtle Avenue dan Lafayette Avenue yang dikelilingi oleh berbagai restoran, cafe, dan butik. Sementara, Flatbush Avenue merupakan salah satu jalan utama yang terpanjang di Kota New York. Jalan tersebut menjadi pusat aktivitas komersial dan tempat bersosialisasi penduduk Kota Brooklyn.



Gambar 4. 58 Visualisasi Peta Titik Pengantaran di Williamsburg



Gambar 4. 59 Area Potensial Titik Pengantaran di Flatbush Avenue

Terdapat sekumpulan titik yang berdekatan pada hari kerja. Tujuan penumpang taksi pada hari kerja diantaranya Atlantic Avenue Barclays Center yang merupakan stasiun kereta bawah tanah. Selain itu, terdapat penumpang menuju Court Street yang didominasi oleh tempat ibadah, perguruan tinggi, dan bangunan pemerintah pemerintah. Pada wilayah Williamsburg, terdapat jalan yang dipenuhi tujuan pada hari kerja yaitu Bedford Avenue. Terdapat beragam restoran, toko baju, dan tempat hiburan yang menjadi kawasan penduduk untuk bersosialisasi.

Berikut merupakan visualisasi peta yang diperbesar pada daerah yang padat akan pengantaran penumpang taksi.

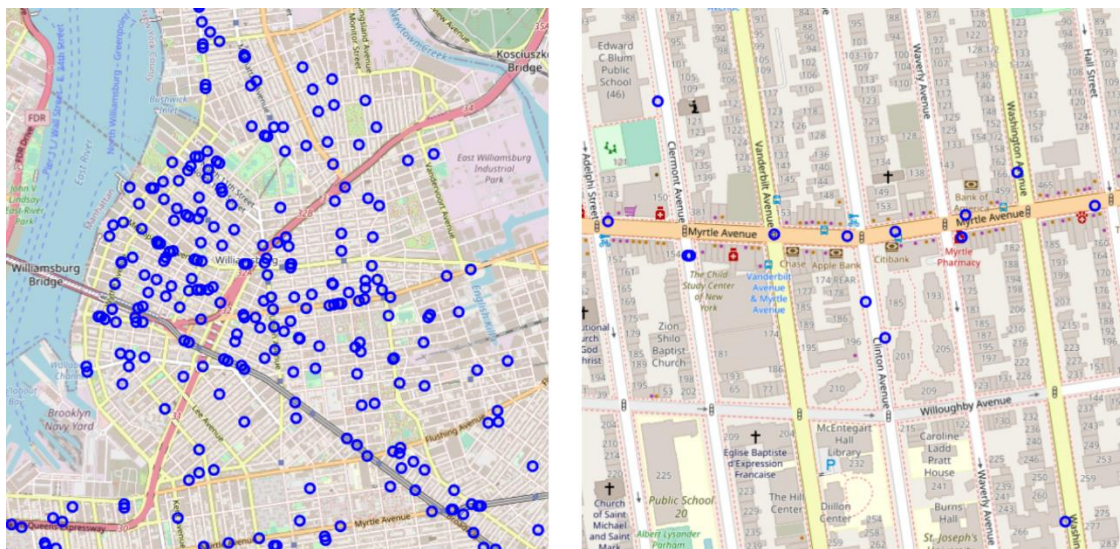


**Atlantic Avenue
Barclays Center**

**Court Street
Brooklyn**

**Bedford Avenue
Williamsburg**

Gambar 4. 60 Visualisasi Peta Titik Pengantaran



Williamsburg

Myrtle Avenue

Gambar 4. 61 Visualisasi Peta Titik Pengantaran di Williamsburg dan Myrtle Avenue

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil identifikasi *clustering* data perjalanan taksi di Kota Brooklyn menggunakan algoritma *Density-Based Spatial Clustering of Application with Noise* (DBSCAN) dapat disimpulkan bahwa:

- a. Penerapan metode *clustering* berbasis densitas terhadap data perjalanan taksi di Kota Brooklyn menggunakan algoritma DBSCAN dapat menghasilkan anggota *cluster* dan titik *noise*. Berdasarkan pemilihan parameter terbaik, *cluster* terbentuk secara padat yang dapat membantu untuk mengidentifikasi area potensial perjalanan taksi. *Clustering* pada titik penjemputan dan pengantaran penumpang dengan hasil algoritma DBSCAN sebagai berikut:
 - 1) Data titik penjemputan dikelompokkan berdasarkan penjemputan yang terjadi pada hari kerja, hari libur, pagi, siang, sore, malam, dan tengah malam. Berikut merupakan tabel rangkuman parameter terbaik, konversi parameter epsilon ke jarak dalam meter, nilai *silhouette coefficient*, jumlah *cluster* dan titik *noise* pada titik penjemputan.

Tabel 5. 1 Rangkuman Parameter dan Hasil *Clustering* Titik Penjemputan

No	Waktu	Parameter		<i>Eps</i> Dalam Satuan Meter	<i>Silhouette Coefficient</i>	Jumlah <i>Cluster</i>	Titik <i>Noise</i>
1	Hari Kerja	<i>Eps</i>	0.0075	834,9 m	0.546	2	148
		<i>MinPts</i>	29				
2	Hari Libur	<i>Eps</i>	0.008	890,56 m	0.524	2	78
		<i>MinPts</i>	25				
3	Pagi	<i>Eps</i>	0.007	779,24 m	0.539	2	54
		<i>MinPts</i>	6				
4	Siang	<i>Eps</i>	0.0075	834,9 m	0.546	2	25
		<i>MinPts</i>	5				
5	Sore	<i>Eps</i>	0.0075	834,9 m	0.551	2	21
		<i>MinPts</i>	4				
6	Malam	<i>Eps</i>	0.005	556,6 m	0.546	2	72

No	Waktu	Parameter		<i>Eps</i> Dalam Satuan Meter	<i>Silhouette Coefficient</i>	Jumlah Cluster	Titik Noise
		<i>MinPts</i>	6				
7	Tengah Malam	<i>Eps</i>	0.007	779,24 m	0.512	2	60
		<i>MinPts</i>	15				

- 2) Data titik pengantaran dikelompokkan berdasarkan pengantaran yang terjadi pada hari kerja, hari libur, pagi, siang, sore, malam, dan tengah malam. Berikut merupakan tabel rangkuman parameter terbaik, konversi parameter epsilon ke jarak dalam meter, nilai *silhouette coefficient*, jumlah *cluster* dan titik *noise* pada titik pengantaran.

Tabel 5. 2 Rangkuman Parameter dan Hasil *Clustering* Titik Pengantaran

No	Waktu	Parameter		<i>Eps</i> Dalam Satuan Meter	<i>Silhouette Coefficient</i>	Jumlah Cluster	Titik Noise
1	Hari Kerja	<i>Eps</i>	0.008	890,56 m	0.321	4	146
		<i>MinPts</i>	8				
2	Hari Libur	<i>Eps</i>	0.0065	723,58 m	0.258	3	151
		<i>MinPts</i>	7				
3	Pagi	<i>Eps</i>	0.0075	834,9 m	0.272	6	74
		<i>MinPts</i>	5				
4	Siang	<i>Eps</i>	0.007	779,24 m	0.298	7	43
		<i>MinPts</i>	3				
5	Sore	<i>Eps</i>	0.0072	801,5 m	0.256	20	24
		<i>MinPts</i>	2				
6	Malam	<i>Eps</i>	0.008	890,56 m	0.330	3	77
		<i>MinPts</i>	5				
7	Tengah Malam	<i>Eps</i>	0.0075	834,9 m	0.283	2	131
		<i>MinPts</i>	7				

- b. Area potensial titik penjemputan dan titik pengantaran taksi berhasil dianalisis secara temporal dengan mengelompokkan waktu kejadian. Pengelola taksi dapat melihat tingkat permintaan yang tinggi berdasarkan hari kerja atau hari libur, lalu melihat tingkat permintaan berdasarkan bagian hari yaitu pagi, siang, sore, malam, dan tengah

malam. Melalui hal tersebut, pengelola taksi dan sopir taksi dapat mengalokasikan armada taksi pada area yang potensial terhadap penumpang taksi. Melalui visualisasi peta, pengelola juga dapat melihat jalan yang padat akan mobilitas sehingga dapat mencari jalan alternatif untuk menghindari kepadatan maupun kemacetan.

- 1) Pada titik penjemputan, pengelola dan sopir taksi dapat melihat kawasan yang padat akan mobilitas. Untuk itu, taksi dapat dialokasikan pada kawasan tersebut untuk memudahkan sopir dalam menemukan penumpang. Selain itu, terdapat titik penjemputan di tempat hiburan dan tempat ibadah. Pengelola taksi dapat memperhatikan jadwal operasional tempat tersebut dengan bantuan visualisasi secara temporal terhadap rentang waktu agar segera menemukan penumpang.
- 2) Pada titik pengantaran, pengelola dan sopir taksi dapat melihat perkiraan tempat tujuan berdasarkan rentang waktu. Pada hari kerja, sopir taksi dapat melihat kawasan dan jalan utama yang padat akan mobilitas. Untuk itu, sopir taksi dapat mempelajari jalan alternatif yang dapat mengantarkan penumpang ke tujuan dan menghindari kemacetan. Mengetahui titik pengantaran penumpang juga dapat membantu pengelola dan sopir taksi untuk mempersiapkan armada taksi secara optimal misalnya dari segi bahan bakar. Selain itu, sopir taksi juga dapat berkendara dengan kecepatan yang aman dan tidak terlalu lambat setelah melihat perkiraan kawasan dari tujuan penumpang taksi. Contohnya, terdapat titik pengantaran menuju bandara yang cukup jauh dari kawasan utara dan barat laut Brooklyn. Sopir taksi dapat mempersiapkan bahan bakar yang cukup dan perkiraan kecepatan berkendara.

5.2 Saran

Adapun saran yang dapat diberikan penulis berdasarkan penelitian ini yang bertujuan untuk meningkatkan kualitas penelitian selanjutnya adalah:

1. Penelitian selanjutnya akan menambah data dengan rentang waktu kejadian dan lokasi pada *dataset* yang digunakan.
2. Menambah metode *clustering* serupa untuk membandingkan hasil *clustering* dari masing-masing metode.

DAFTAR PUSTAKA

- C., M., & H., C. (2022). Parallel DBSCAN *Clustering* Algorithm Using Hadoop Map-reduce Framework for Spatial Data. *I.J. Information Technology and Computer Science*, 3.
- Canchen, L. (2019). Preprocessing Methods and Pipelines of Data Mining: An Overview.
- Chicco, D., Oneto, L., & Tavazzi, E. (2022). Eleven quick tips for data cleaning and feature engineering. *PLoS Comput Biol* 18(12): e1010718.
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering *Clusters* in Large Spatial Databases with Noise. *International Conference on Knowledge Discovery and Data Mining* (hal. 2-6). Portland, Oregon: AAAI Press.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Elsevier.
- Hossain, M. Z., Islam, M. J., Rahman Miah, M. W., Rony, J. H., & Begum, M. (2021). Develop a dynamic DBSCAN algorithm for solving initial parameter selection problem of the DBSCAN algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*, 1.
- Jian, S., Li, D., & Yu, Y. (2021). Research on Taxi Operation Characteristics by Improved DBSCAN Density *Clustering* Algorithm and K-means *Clustering* Algorithm.
- Johar, A., Vatesia, A., & Donny, I. A. (2023). Implementasi Metode Spatio-Temporal *Clustering* Dengan Algoritma ST-DBSCAN Pada Titik Api Kebakaran Hutan Indonesia (2015-2020). *Jurnal Rekursif*, Vol. 11, No. 1.
- Kim, K. (2023, 7 13). *NYC_TAXI_TRIPS*. Diambil kembali dari Kaggle: <https://www.kaggle.com/code/qshick/nyc-taxi-trips>
- Manalu, D. J., Rahmawati, R., & Widiharih, T. (2021). Pengelompokan Titik Gempa di Pulau Sulawesi Menggunakan Algoritma ST-DBSCAN (Spatio Temporal-Density Based Spatial *Clustering* Application with Noise). *JURNAL GAUSSIAN*, Volume 10, Nomor 4, 554 - 561.
- Nurhaliza, N., & Mustakim. (2021). Pengelompokan Data Kasus Covid-19 di Dunia Menggunakan Algoritma DBSCAN. *Indonesian Journal of Informatic Research and Software Engineering*, 1-8.
- Puspitasari, D. A., Cahyana, Y., & Lestari, S. A. (2023). Penerapan Algoritma Density-Based Spastial *Clustering* Algorithm With Noise Untuk Pengelompokkan Penyakit Pasien. *Scientific Student Journal for Information, Technology and Science* , 102-106.

- Putri, M., Permana, D., Syafriandi, & Zilrahmi. (2023). Grouping The Regencies/Cities in Indonesia Based on Expenditure Groups Inflation Value Using DBSCAN Method. *UNP JOURNAL OF STATISTICS AND DATA SCIENCE*, 164 - 171.
- Riascos, A. P., & Mateos, J. L. (2020). *Networks and long-range mobility in cities: A study of more than one billion taxi trips in New York City*.
- Sari, B. N., & Primajaya, A. (2019). Penerapan *Clustering* DBSCAN untuk Pertanian Padi di Kabupaten Karawang. *Jurnal Informatika dan Komputer Vol. 4*.
- Sugi Almantara, I. P., Sri Ariyani, N. W., & Alit Swamardika, I. B. (2020). Spatial Data Analysis using DBSCAN Method and KNN classification. *International Journal of Engineering and Emerging Technology, Vol.5, No.2*.
- Suyanto, I.-(. (2017). *Data mining : untuk klasifikasi dan klusterisasi data / Suyanto*. Bandung: Penerbit Informatika.
- Tu, X., Fu, C., Huang, A., Chen, H., & Ding, X. (2022). DBSCAN Spatial *Clustering* Analysis of Urban “Production–Living–Ecological” Space Based on POI Data: A Case Study of Central Urban Wuhan, China. *Int. J. Environ. Res. Public Health* 2022, 19, 5153.
- Tu, X., Fu, C., Huang, A., Chen, H., & Ding, X. (2022). DBSCAN Spatial *Clustering* Analysis of Urban “Production–Living–Ecological” Space Based on POI Data: A Case Study of Central Urban Wuhan, China. *Int. J. Environ. Res. Public Health* 2022, 19, 5153.
- Ulak, M. B., Yazici, A., & Aljarrah, M. (2020). Value of convenience for taxi trips in New York City.
- United Nations, Department of Economic and Social Affairs, Population Division. (2017). *World Population Ageing 2017 (ST/ESA/SER.A/408)*.
- Vizuete-Luciano, E., Guill'en-Pujadas, M., Alaminos, D., & Merig'o-Lindahl, J. M. (2023). Taxi and urban mobility studies: A bibliometric analysis.
- Xin-yan, L. (2005). DBSCAN spatial *clustering* algorithm and its application in urban planning. *Science of Surveying and Mapping*.
- Xue, Z., & Wang, H. (2021). Effective Density-Based *Clustering* Algorithms for Incomplete Data. *Big Data Mining & Analytics*.
- Yang, J., Sun, Y., Shang, B., Wang, L., & Zhu, J. (2019). Understanding Collective Human Mobility Spatiotemporal Patterns on Weekdays from Taxi Origin-Destination Point Data.

You, L., Guan, Z., Li, N., Zhang, J., Cui, H., Claramunt, C., & Cao, R. (2021). A Spatio-Temporal Schedule-Based Neural Network for Urban Taxi Waiting Time Prediction. *ISPRS International Journal of Geo-Information*.

LAMPIRAN