

# **Laporan Akhir**

## **Analisis Produktivitas Tenaga Kerja Berdasarkan Biaya, Jarak, dan Moda Transportasi (Studi Kasus : Jakarta)**



### **Nama Anggota Kelompok :**

41425063 Luana Breka Manuela Banjarnahor  
41425064 Yosevyn Reginae Sipahutar  
41425065 Albert Rapael Aritonang

**INSTITUT TEKNOLOGI DEL  
FAKULTAS VOKASI  
TEKNOLOGI REKAYASA PERANGKAT LUNAK**

## DAFTAR ISI

<b>BAB I PENDAHULUAN .....</b>	3
1.1    Latar Belakang.....	3
1.2    Rumusan Masalah.....	4
1.3    Tujuan.....	4
1.4    Lingkup.....	4
1.5    Pendekatan .....	5
<b>BAB II TINJAUAN PUSTAKA.....</b>	7
2.1    Produktivitas Tenaga Kerja.....	7
2.2    Biaya Perjalanan dan Pengaruhnya terhadap Produktivitas Tenaga Kerja .....	7
2.3    Jarak Tempuh dan Moda Transportasi dalam Aktivitas Komuter .....	8
2.4    Keterkaitan Biaya, Jarak, dan Moda Transportasi terhadap Produktivitas Tenaga Kerja.....	8
2.5    Referensi Ilmiah Utama.....	9
<b>BAB III DESKRIPSI DATASET .....</b>	10
3.1    Sumber Data.....	10
3.2    Deskripsi Variabel.....	10
3.3    Skala Data.....	11
<b>BAB IV ALGORITMA .....</b>	12
4.1    Logistic Regression .....	12
4.2    Random Forest .....	12
<b>BAB V EXPLORATORY DATA ANALYSIS (EDA) DAN INSIGHT .....</b>	13
5.1    Analisis Statistik Deskriptif.....	13
5.2    Visualisasi Data .....	13
5.3    Insight dari Data .....	17
<b>BAB VI METODOLOGI PENELITIAN (CRISP-DM) .....</b>	19
6.1    Business Understanding .....	19
6.2    Data Understanding .....	20
6.3    Data Preparation .....	20
6.4    Modeling .....	21
6.4    Evaluation .....	21
6.5    Deployment .....	21
<b>BAB VII HASIL DAN PEMBAHASAN .....</b>	22
7.1    Implementasi Persiapan Data .....	22
7.2    Implementasi Preprocessing Data .....	22
7.2.1    Import Data .....	22

<b>7.2.2 Preprocessing.....</b>	<b>23</b>
<b>7.3 Implementasi Pemodelan .....</b>	<b>27</b>
<b>7.3.1 Logistic Regression .....</b>	<b>27</b>
<b>7.3.2 Random Forest .....</b>	<b>27</b>
<b>7.4 Hasil Evaluasi Model .....</b>	<b>28</b>
<b>7.4.1 Evaluasi Kinerja Model Logistic Regression.....</b>	<b>29</b>
<b>7.4.2 Evaluasi Kinerja Model Random Forest .....</b>	<b>30</b>
<b>7.4.3 Perbandingan Performa Model Logistic Regression dan Random Forest .....</b>	<b>30</b>
<b>7.5 Analisis Pengaruh Variabel terhadap Produktivitas Tenaga Kerja .....</b>	<b>30</b>
<b>7.5.1 Feature Importance Menggunakan Random Forest .....</b>	<b>30</b>
<b>7.5.2 Pembahasan Hasil Analisis Feature Importance .....</b>	<b>31</b>
<b>7.6 Insight Penelitian.....</b>	<b>31</b>
<b>7.7 Dashboard dan Visualisasi Komunikasi Data .....</b>	<b>32</b>
<b>7.7.1 Distribusi Moda Transportasi Tenaga Kerja .....</b>	<b>32</b>
<b>7.7.2 Hubungan Jarak Tempuh dan Produktivitas Tenaga Kerja.....</b>	<b>33</b>
<b>7.7.3 Distribusi Biaya Transportasi Tenaga Kerja .....</b>	<b>34</b>
<b>7.7.4 Feature Importance Faktor Transportasi terhadap Produktivitas Tenaga Kerja</b>	<b>34</b>
<b>7.7.5 Dashboard.....</b>	<b>35</b>
<b>7.8 Rekomendasi .....</b>	<b>35</b>
<b>BAB VIII KESIMPULAN DAN SARAN .....</b>	<b>37</b>
<b>8.1 Kesimpulan.....</b>	<b>37</b>
<b>8.2 Saran .....</b>	<b>37</b>
<b>Pembagian Tugas Anggota.....</b>	<b>39</b>
<b>Lampiran .....</b>	<b>40</b>

## **BAB I**

### **PENDAHULUAN**

Pada bab berikut akan dijelaskan terkait latar belakang dari topik project, ruang lingkup, dan pendekatan dalam menyelesaikan permasalahan dalam project, serta penyajian dokumen laporan akhir.

#### **1.1 Latar Belakang**

Transportasi merupakan bagian penting dalam kehidupan masyarakat perkotaan, khususnya bagi tenaga kerja yang setiap hari melakukan mobilitas dari tempat tinggal menuju lokasi kerja. Di kota besar seperti Jakarta, permasalahan transportasi menjadi semakin kompleks akibat kemacetan lalu lintas, jarak tempuh yang jauh, serta biaya transportasi yang relatif tinggi. Kondisi tersebut berpotensi memengaruhi kondisi fisik, psikologis, dan motivasi tenaga kerja. Biaya transportasi yang tinggi dapat menurunkan motivasi kerja karena sebagian pendapatan harus dialokasikan untuk kebutuhan perjalanan. Jarak tempuh yang jauh berisiko menyebabkan kelelahan serta keterlambatan, yang pada akhirnya berdampak pada tingkat kehadiran dan produktivitas tenaga kerja. Selain itu, pemilihan moda transportasi yang kurang efisien juga dapat memperpanjang waktu perjalanan dan menurunkan kenyamanan pekerja.

Seiring dengan perkembangan teknologi, pendekatan data science dan machine learning dapat dimanfaatkan untuk menganalisis pola dan hubungan antara faktor-faktor transportasi dengan produktivitas tenaga kerja berdasarkan data survei. Pendekatan ini memungkinkan pengambilan keputusan yang lebih objektif dan berbasis data. Selain berdampak pada individu, permasalahan transportasi juga memiliki implikasi yang lebih luas terhadap organisasi dan perekonomian. Tingkat kehadiran yang menurun, keterlambatan kerja, serta penurunan motivasi dapat berpengaruh pada kinerja organisasi secara keseluruhan. Oleh karena itu, pemahaman mengenai faktor-faktor transportasi yang memengaruhi produktivitas tenaga kerja menjadi penting sebagai dasar perumusan kebijakan transportasi maupun kebijakan ketenagakerjaan. Dalam penelitian ini, persepsi responden digunakan sebagai pendekatan untuk menilai pengaruh transportasi terhadap produktivitas. Persepsi tersebut diukur melalui beberapa indikator, yaitu pengaruh moda transportasi terhadap produktivitas, pengaruh jarak terhadap kehadiran, dan pengaruh biaya transportasi terhadap motivasi. Ketiga indikator ini kemudian digabungkan untuk membentuk gambaran umum mengenai tingkat pengaruh transportasi terhadap produktivitas tenaga kerja.

Berdasarkan kondisi tersebut, diperlukan suatu analisis yang sistematis dan terstruktur untuk mengetahui sejauh mana biaya transportasi, jarak tempuh, dan moda transportasi berpengaruh terhadap produktivitas tenaga kerja di Jakarta berdasarkan persepsi responden. Pendekatan data science dengan pemodelan machine learning diharapkan dapat memberikan insight yang lebih mendalam dan objektif terhadap permasalahan ini.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, adapun rumusan masalah dalam penggeraan project ini yaitu:

1. Bagaimana gambaran distribusi biaya transportasi, jarak tempuh, dan moda transportasi yang digunakan oleh tenaga kerja di Jakarta berdasarkan data survei?
2. Apakah biaya, jarak, dan moda transportasi dapat digunakan untuk memprediksi pengaruh transportasi terhadap produktivitas tenaga kerja?
3. Bagaimana perbandingan performa algoritma Logistic Regression dan Random Forest dalam memprediksi pengaruh transportasi terhadap produktivitas tenaga kerja?

## 1.3 Tujuan

Tujuan dari penelitian ini adalah:

1. Menganalisis karakteristik biaya, jarak, dan moda transportasi tenaga kerja di Jakarta.
2. Membangun model machine learning untuk memprediksi pengaruh transportasi terhadap produktivitas tenaga kerja.
3. Membandingkan performa algoritma Logistic Regression dan Random Forest.

## 1.4 Lingkup

Penelitian ini dilakukan dengan menggunakan data hasil survei yang diperoleh melalui Google Form dengan jumlah responden sebanyak 2.032 orang tenaga kerja di Jakarta. Lingkup penelitian difokuskan pada analisis hubungan antara faktor transportasi—meliputi biaya, jarak tempuh, dan moda transportasi—with aspek produktivitas tenaga kerja. Variabel yang dianalisis mencakup biaya transportasi (kategori rendah, sedang, tinggi), jarak tempuh (<5 km, 5–10 km, 11–20 km, >20 km), moda transportasi (jalan kaki/sepeda, kendaraan pribadi, ojek online/taksi online, transportasi umum), serta persepsi kenyamanan dan motivasi kerja.

Penelitian ini terbatas pada responden yang berdomisili dan bekerja di wilayah Jakarta, sehingga hasil analisis lebih relevan untuk menggambarkan kondisi transportasi dan produktivitas tenaga kerja di kota tersebut. Data yang digunakan bersifat self-reported dari

survei, sehingga lingkup penelitian tidak mencakup pengukuran langsung produktivitas, melainkan persepsi dan pengalaman responden terkait transportasi.

## **1.5 Pendekatan**

Pendekatan penelitian yang digunakan dalam Tugas Akhir ini disesuaikan dengan permasalahan yang ada, yaitu bagaimana faktor biaya, jarak, dan moda transportasi memengaruhi produktivitas tenaga kerja di Jakarta. Tahapan pendekatan yang dilakukan adalah sebagai berikut:

### **1. Pemahaman Bisnis**

Pada tahap ini dilakukan identifikasi tujuan penelitian, yaitu menganalisis pengaruh faktor transportasi terhadap produktivitas tenaga kerja. Situasi yang ada dinilai dari sudut pandang bisnis dan sosial, khususnya bagaimana biaya, jarak tempuh, dan moda transportasi dapat memengaruhi motivasi, keterlambatan, serta kenyamanan tenaga kerja. Tujuan data mining kemudian didefinisikan untuk menjawab permasalahan tersebut, serta ditentukan strategi analisis yang akan digunakan.

### **2. Pemahaman Data**

Tahap ini berfokus pada pengumpulan dan pemeriksaan data. Data yang digunakan adalah data kuantitatif hasil survei melalui Google Form dengan jumlah responden sebanyak 2.032 orang tenaga kerja di Jakarta. Variabel yang dikumpulkan meliputi biaya transportasi, jarak tempuh, moda transportasi, kenyamanan, motivasi, dan keterlambatan. Data ini akan menjadi acuan utama dalam menjawab permasalahan penelitian.

### **3. Data Preparation / Data Preprocessing**

Data survei yang telah dikumpulkan kemudian diproses agar siap digunakan. Tahapan ini meliputi pembersihan data, pengecekan nilai kosong (missing value), standarisasi kategori, serta pengkodean variabel agar dapat dianalisis lebih lanjut. Jika terdapat data yang tidak lengkap, maka dilakukan penanganan dengan metode imputasi sederhana (misalnya menggunakan rata-rata atau modus), sehingga dataset menjadi lebih efisien dan terstruktur dengan baik.

### **4. Pemodelan**

Tahap ini bertujuan untuk membangun model analisis yang dapat menjelaskan hubungan antara biaya, jarak, moda transportasi, dan produktivitas tenaga kerja. Model yang digunakan berupa analisis statistik dan algoritma data science sederhana seperti regresi linear/logistik untuk melihat pengaruh variabel, serta decision tree atau

clustering untuk mengelompokkan pola responden berdasarkan moda transportasi dan biaya.

#### 5. Evaluasi

Model yang telah dibangun kemudian dievaluasi untuk melihat tingkat akurasi dan relevansinya dalam menjawab permasalahan penelitian. Evaluasi dilakukan dengan membandingkan hasil analisis dengan data survei aktual, serta menggunakan metrik evaluasi seperti akurasi, precision, recall, atau nilai error (misalnya RMSE) sesuai jenis model yang digunakan.

#### 6. Deployment

Tahap akhir adalah penyajian hasil analisis dalam bentuk dashboard interaktif menggunakan Tableau. Dashboard ini menampilkan visualisasi diagram biaya vs motivasi, jarak vs keterlambatan, moda vs kenyamanan, moda vs biaya, dan moda vs jarak. Dengan adanya dashboard, hasil penelitian dapat diakses dan dipahami dengan mudah oleh pihak terkait, baik perusahaan maupun pembuat kebijakan, sehingga dapat digunakan sebagai dasar pengambilan keputusan.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Produktivitas Tenaga Kerja**

Produktivitas tenaga kerja merupakan salah satu indikator penting dalam menilai efisiensi dan kinerja tenaga kerja dalam menghasilkan output. Produktivitas tidak hanya dipengaruhi oleh faktor internal pekerja, seperti keterampilan dan pengalaman, tetapi juga oleh faktor eksternal yang berkaitan dengan lingkungan kerja dan aksesibilitas menuju tempat kerja. Menurut Pratama dan Khoirunurrofik (2023), produktivitas tenaga kerja dapat dipengaruhi secara signifikan oleh kualitas dan ketersediaan infrastruktur transportasi. Infrastruktur transportasi yang memadai mampu memperlancar mobilitas tenaga kerja, menurunkan hambatan perjalanan, serta mengurangi waktu tempuh pekerja menuju lokasi kerja. Kondisi tersebut berdampak pada meningkatnya efisiensi kerja dan kemampuan tenaga kerja dalam mempertahankan performa kerja yang optimal.

Lebih lanjut, Pratama dan Khoirunurrofik (2023) menyatakan bahwa peningkatan akses transportasi berperan dalam mengurangi kelelahan pekerja akibat perjalanan yang panjang, sehingga tenaga kerja memiliki kondisi fisik dan mental yang lebih baik saat menjalankan aktivitas kerja. Dengan demikian, faktor transportasi memiliki kontribusi tidak langsung terhadap peningkatan produktivitas tenaga kerja, khususnya di wilayah dengan tingkat mobilitas tinggi.

#### **2.2 Biaya Perjalanan dan Pengaruhnya terhadap Produktivitas Tenaga Kerja**

Biaya perjalanan merupakan komponen penting dalam aktivitas komuter tenaga kerja, khususnya di wilayah perkotaan. Biaya ini mencakup pengeluaran rutin pekerja untuk transportasi dari tempat tinggal menuju tempat kerja dan sebaliknya. Tingginya biaya perjalanan dapat menjadi beban ekonomi bagi pekerja dan berpotensi memengaruhi tingkat kenyamanan serta kepuasan kerja. Menurut Mafliyanti dan Yudhistira (2022), biaya perjalanan memiliki pengaruh signifikan terhadap perilaku mobilitas pekerja, terutama dalam menentukan pilihan moda transportasi. Pekerja cenderung memilih moda transportasi yang dianggap paling efisien dari sisi biaya, meskipun pilihan tersebut belum tentu optimal dari sisi waktu tempuh dan kenyamanan. Lebih lanjut, Mafliyanti dan Yudhistira (2022) menjelaskan bahwa biaya perjalanan yang tinggi dapat meningkatkan tekanan finansial pada pekerja, yang secara tidak langsung dapat memengaruhi tingkat konsentrasi dan produktivitas kerja. Kondisi ini menjadi semakin relevan di kawasan metropolitan seperti Jakarta, di mana biaya hidup dan biaya transportasi relatif tinggi.

### **2.3 Jarak Tempuh dan Moda Transportasi dalam Aktivitas Komuter**

Jarak antara tempat tinggal dan tempat kerja merupakan faktor utama yang menentukan durasi perjalanan dan tingkat kelelahan tenaga kerja. Semakin jauh jarak tempuh yang harus dilalui, semakin besar pula waktu dan energi yang dikeluarkan oleh pekerja sebelum memulai aktivitas kerja. Dalam konteks transportasi perkotaan, Pratama dan Khoirunurrofik (2023) menegaskan bahwa jarak tempuh yang panjang, apabila tidak didukung oleh infrastruktur transportasi yang baik, dapat menurunkan efisiensi tenaga kerja. Hal ini disebabkan oleh meningkatnya waktu perjalanan dan risiko keterlambatan yang berdampak pada efektivitas jam kerja. Selain jarak, pemilihan moda transportasi juga memegang peranan penting. Mafliyanti dan Yudhistira (2022) menyatakan bahwa pekerja di kawasan Jabodetabek memilih moda transportasi berdasarkan kombinasi antara biaya, jarak, dan ketersediaan layanan. Pemilihan moda transportasi yang tidak efisien dapat meningkatkan waktu perjalanan dan kelelahan, yang pada akhirnya berdampak pada produktivitas tenaga kerja.

### **2.4 Keterkaitan Biaya, Jarak, dan Moda Transportasi terhadap Produktivitas Tenaga Kerja**

Berdasarkan kajian literatur, dapat disimpulkan bahwa biaya perjalanan, jarak tempuh, dan moda transportasi saling berkaitan dan secara bersama-sama memengaruhi produktivitas tenaga kerja. Pratama dan Khoirunurrofik (2023) menekankan bahwa efisiensi sistem transportasi berperan sebagai faktor pendukung utama dalam peningkatan produktivitas tenaga kerja, terutama di wilayah perkotaan dengan tingkat mobilitas tinggi. Sejalan dengan hal tersebut, Mafliyanti dan Yudhistira (2022) menunjukkan bahwa keputusan pekerja dalam memilih moda transportasi didorong oleh pertimbangan biaya dan jarak, yang kemudian berdampak pada kualitas perjalanan kerja. Kualitas perjalanan yang baik berpotensi meningkatkan kesiapan pekerja dalam menjalankan aktivitas kerja, sehingga produktivitas tenaga kerja dapat terjaga. Dengan demikian, penelitian mengenai analisis produktivitas tenaga kerja berdasarkan biaya, jarak, dan moda transportasi menjadi relevan untuk dikaji lebih lanjut, khususnya di Jakarta sebagai kota metropolitan dengan kompleksitas permasalahan transportasi yang tinggi.

## 2.5 Referensi Ilmiah Utama

**Table 1. Referensi**

Penulis & Tahun	Judul Artikel	Jurnal / Sumber	Fokus & Relevansi Penelitian
Pratama, R. & Khoirunurrofik (2023)	Peran Infrastruktur Transportasi terhadap Produktivitas Tenaga Kerja Industri Manufaktur di Indonesia	Jurnal Ekonomi dan Pembangunan (SINTA 2)	Menjelaskan pengaruh infrastruktur dan akses transportasi terhadap produktivitas tenaga kerja. Menjadi dasar teoritis hubungan antara transportasi dan produktivitas pekerja.
Mafliyanti, S. & Yudhistira, M. H. (2022)	Pengaruh Biaya Perjalanan terhadap Perubahan Pilihan Moda Transportasi: Studi Kasus Pekerja Komuter Jabodetabek	Jurnal Perencanaan Wilayah dan Kota (SINTA 2)	Menganalisis pengaruh biaya perjalanan terhadap pemilihan moda transportasi pekerja komuter yang berdampak pada efisiensi perjalanan kerja.

## **BAB III**

### **DESKRIPSI DATASET**

#### **3.1 Sumber Data**

Data yang digunakan dalam penelitian ini diperoleh melalui survei daring menggunakan Google Form. Survei disebarluaskan kepada responden yang merupakan tenaga kerja dengan aktivitas perjalanan harian menuju tempat kerja, khususnya di wilayah Jakarta. Pengumpulan data dilakukan untuk memperoleh gambaran terkait pola transportasi serta persepsi responden terhadap pengaruh transportasi terhadap produktivitas kerja.

Jumlah responden yang berhasil dikumpulkan dalam survei ini adalah sebanyak 2021 responden. Setiap responden mengisi kuesioner yang mencakup informasi mengenai biaya transportasi, jarak tempuh, moda transportasi yang digunakan, serta penilaian persepsi terhadap pengaruh faktor-faktor tersebut terhadap produktivitas, kehadiran, dan motivasi kerja. Data yang terkumpul selanjutnya digunakan sebagai dasar analisis dalam penelitian ini.

#### **3.2 Deskripsi Variabel**

Dataset yang digunakan dalam penelitian ini terdiri dari beberapa variabel yang dikelompokkan berdasarkan perannya dalam analisis. Variabel-variabel tersebut diperoleh dari jawaban responden pada kuesioner survei.

Variabel input (independen) merepresentasikan karakteristik transportasi yang digunakan oleh responden, yaitu

1. Biaya transportasi
2. Jarak tempuh dari tempat tinggal ke tempat kerja
3. Moda transportasi yang digunakan

Variabel-variabel ini digunakan sebagai fitur utama dalam proses pemodelan machine learning.

Selain itu, terdapat variabel persepsi responden yang menggambarkan penilaian subjektif terhadap pengaruh transportasi, meliputi

1. Pengaruh moda transportasi terhadap produktivitas
2. Pengaruh jarak tempuh terhadap kehadiran
3. Pengaruh biaya transportasi terhadap motivasi kerja.

Variabel persepsi ini digunakan untuk membentuk skor pengaruh transportasi.

Berdasarkan skor pengaruh tersebut, dibentuk variabel target yaitu Target\_Pengaruh yang menunjukkan apakah transportasi dianggap berpengaruh atau tidak berpengaruh terhadap produktivitas tenaga kerja.

### **3.3 Skala Data**

Variabel persepsi responden dalam penelitian ini diukur menggunakan skala Likert dengan rentang nilai 1 sampai 5. Nilai 1 menunjukkan bahwa responden menilai faktor transportasi tidak berpengaruh, sedangkan nilai 5 menunjukkan bahwa faktor transportasi sangat berpengaruh.

Penggunaan skala Likert memungkinkan pengukuran persepsi responden secara terstruktur dan kuantitatif. Selanjutnya, nilai-nilai pada skala Likert tersebut diolah dengan menghitung rata-rata untuk membentuk skor pengaruh transportasi, yang kemudian dikonversi menjadi variabel target biner untuk keperluan analisis klasifikasi menggunakan machine learning. Variabel persepsi menggunakan skala Likert 1–5, di mana nilai 1 menunjukkan tidak berpengaruh dan nilai 5 menunjukkan sangat berpengaruh.

## **BAB IV**

### **ALGORITMA**

Bab ini menjelaskan algoritma machine learning yang digunakan dalam penelitian untuk memprediksi pengaruh transportasi terhadap produktivitas tenaga kerja.

#### **4.1 Logistic Regression**

Logistic Regression merupakan algoritma klasifikasi statistik yang digunakan untuk memprediksi probabilitas suatu kejadian dengan hasil keluaran berupa dua kelas atau lebih. Dalam penelitian ini, Logistic Regression digunakan untuk mengklasifikasikan apakah transportasi berpengaruh atau tidak berpengaruh terhadap produktivitas tenaga kerja. Algoritma ini bekerja dengan memodelkan hubungan antara variabel input dan variabel target menggunakan fungsi logistik. Logistic Regression dipilih karena memiliki interpretabilitas yang baik, sehingga hasil model dapat dijelaskan secara logis dan mudah dipahami. Hal ini penting dalam penelitian berbasis survei, di mana interpretasi hasil menjadi salah satu fokus utama. Selain itu, Logistic Regression digunakan sebagai model pembanding awal (baseline) untuk menilai performa algoritma yang lebih kompleks. Dengan menggunakan Logistic Regression, dapat diketahui sejauh mana hubungan linier antara biaya transportasi, jarak tempuh, dan moda transportasi terhadap produktivitas tenaga kerja.

#### **4.2 Random Forest**

Random Forest merupakan algoritma ensemble learning yang mengombinasikan banyak pohon keputusan (decision tree) untuk menghasilkan prediksi yang lebih akurat dan stabil. Algoritma ini bekerja dengan membangun sejumlah pohon keputusan menggunakan data latih yang dipilih secara acak (bootstrap sampling) serta subset fitur yang berbeda pada setiap pohon. Hasil prediksi akhir diperoleh melalui mekanisme voting mayoritas. Penggunaan Random Forest dalam penelitian ini bertujuan untuk menangkap pola hubungan yang lebih kompleks dan bersifat non-linear antara variabel biaya transportasi, jarak tempuh, dan moda transportasi terhadap pengaruh produktivitas tenaga kerja. Hal ini penting karena hubungan antar variabel dalam data survei tidak selalu bersifat linier. Keunggulan lain dari Random Forest adalah kemampuannya dalam mengurangi risiko overfitting dibandingkan dengan model pohon keputusan tunggal. Dengan menggabungkan banyak pohon, model menjadi lebih robust terhadap noise dan variasi data. Selain itu, Random Forest relatif mampu menangani data dengan skala yang berbeda serta distribusi kelas yang tidak seimbang, sehingga cocok digunakan pada data survei yang memiliki variasi karakteristik responden.

## **BAB V**

### **EXPLORATORY DATA ANALYSIS (EDA) DAN INSIGHT**

#### **5.1 Analisis Statistik Deskriptif**

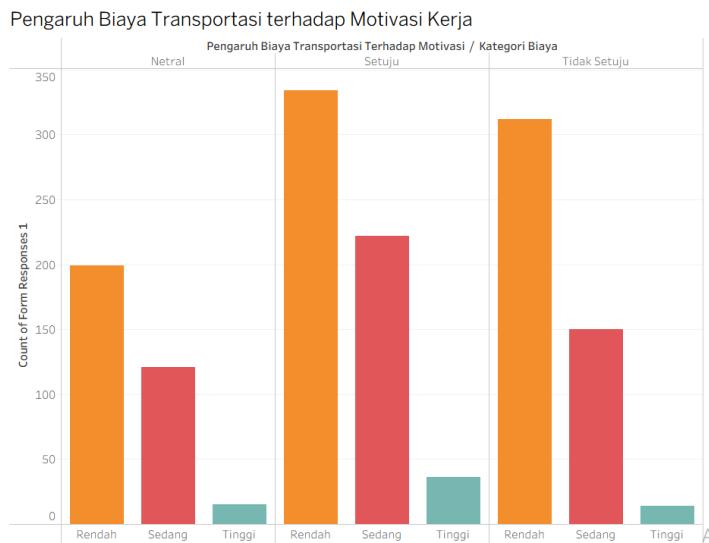
Tahap analisis statistik deskriptif bertujuan untuk memberikan gambaran umum mengenai karakteristik responden dan pola data transportasi yang digunakan dalam penelitian. Analisis ini menjadi dasar awal untuk memahami kondisi data sebelum dilakukan pemodelan machine learning.

Berdasarkan hasil pengolahan data, jumlah responden yang terlibat dalam penelitian ini sebanyak 2021 responden. Jumlah tersebut dinilai cukup representatif untuk menggambarkan kondisi transportasi tenaga kerja di wilayah Jakarta. Dari sisi biaya transportasi, sebagian besar responden berada pada kategori biaya rendah hingga menengah, yang menunjukkan bahwa mayoritas tenaga kerja mengeluarkan biaya transportasi harian yang relatif terjangkau. Namun demikian, masih terdapat responden dengan kategori biaya tinggi, yang mengindikasikan adanya variasi kondisi ekonomi dan jarak tempuh perjalanan kerja. Distribusi jarak tempuh menunjukkan bahwa mayoritas responden menempuh jarak pendek hingga menengah menuju tempat kerja. Hal ini mengindikasikan bahwa sebagian besar tenaga kerja bekerja dalam radius yang relatif dekat dari tempat tinggalnya, meskipun terdapat pula responden dengan jarak tempuh yang cukup jauh, yang berpotensi memengaruhi tingkat kelelahan dan produktivitas kerja. Pada aspek moda transportasi, data menunjukkan bahwa responden menggunakan beragam moda, baik transportasi pribadi maupun transportasi umum. Dominasi moda tertentu mencerminkan preferensi tenaga kerja terhadap faktor kenyamanan, efisiensi waktu, serta ketersediaan sarana transportasi di wilayah perkotaan. Secara keseluruhan, analisis statistik deskriptif ini menunjukkan bahwa data memiliki variasi yang cukup baik pada setiap variabel. Variasi tersebut menjadi indikator bahwa data layak digunakan untuk analisis lebih lanjut, khususnya dalam membangun model machine learning untuk memprediksi pengaruh faktor transportasi terhadap produktivitas tenaga kerja.

#### **5.2 Visualisasi Data**

Setelah analisis statistik deskriptif, tahap selanjutnya adalah visualisasi data untuk menyajikan informasi secara lebih intuitif dan mudah dipahami. Visualisasi digunakan untuk mengidentifikasi pola, tren, serta ketidakseimbangan distribusi data yang tidak selalu terlihat melalui tabel statistik.

## 1. Pengaruh Biaya Transportasi terhadap Motivasi Kerja



**Gambar 1. Pengaruh Biaya Transportasi Terhadap Motivasi Kerja**

Visualisasi ini menunjukkan pola utama bahwa pada semua kategori sikap (Netral, Setuju, dan Tidak Setuju), jumlah responden paling banyak berada pada kategori biaya transportasi rendah, diikuti biaya sedang, dan paling sedikit pada biaya tinggi. Perbandingan penting terlihat jelas pada kelompok Setuju, di mana responden dengan biaya rendah dan sedang jauh lebih dominan dibanding biaya tinggi, menandakan bahwa semakin rendah biaya transportasi, semakin banyak responden yang merasa termotivasi dalam bekerja. Sementara itu, pada kelompok Tidak Setuju, meskipun biaya rendah masih mendominasi, jumlah responden biaya sedang juga cukup besar, mengindikasikan adanya tekanan biaya yang mulai memengaruhi motivasi. Kaitan dengan tujuan penelitian adalah visualisasi ini mendukung analisis bahwa biaya transportasi berperan signifikan terhadap motivasi kerja, khususnya di konteks perkotaan seperti Jakarta, di mana biaya transportasi yang lebih rendah cenderung berkorelasi dengan motivasi kerja yang lebih positif.

## 2. Analisis Keterlambatan Kerja Berdasarkan Jarak Tempuh

Analisis Keterlambatan Kerja Berdasarkan Jarak Tempuh

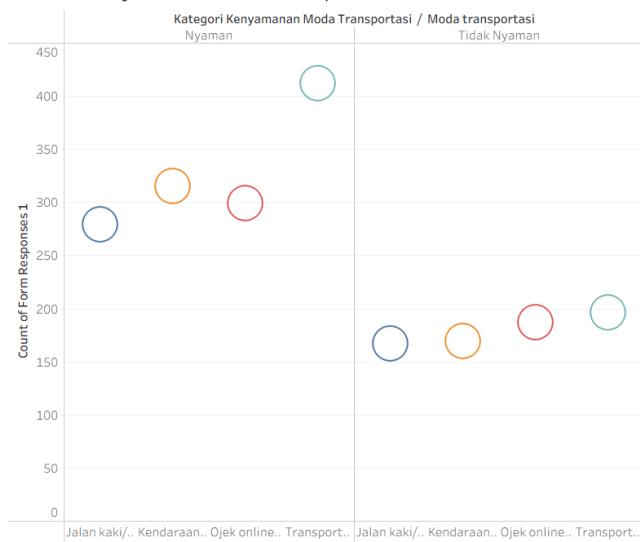
Frekuensi Terlambat akibat Jarak	Jarak			
	5-10 km	11-20 km	< 5 km	> 20 km
Jarang Terlambat	79	79	154	57
Kadang Terlambat	149	105	179	54
Sering Terlambat	96	104	151	63
Tidak Pernah Terlambat	159	138	324	133

**Gambar 2. Analisis Keterlambatan Kerja Berdasarkan Jarak Tempuh**

Visualisasi ini memperlihatkan pola utama bahwa pekerja dengan jarak tempuh < 5 km paling banyak berada pada kategori tidak pernah terlambat, sementara frekuensi keterlambatan cenderung meningkat pada jarak yang lebih jauh. Perbandingan penting terlihat pada jarak > 20 km dan 5–10 km, di mana jumlah responden yang kadang dan sering terlambat relatif lebih tinggi dibanding jarak sangat dekat, menunjukkan bahwa semakin jauh jarak tempuh, semakin besar risiko keterlambatan kerja. Sementara itu, jarak 11–20 km menunjukkan distribusi yang lebih seimbang antara kategori keterlambatan. Kaitan dengan tujuan penelitian adalah temuan ini menegaskan bahwa jarak tempuh merupakan faktor penting yang memengaruhi ketepatan waktu kerja, sehingga relevan dalam analisis produktivitas tenaga kerja di Jakarta, terutama dalam hubungannya dengan efisiensi mobilitas harian.

### 3. Analisis Kenyamanan Moda Transportasi

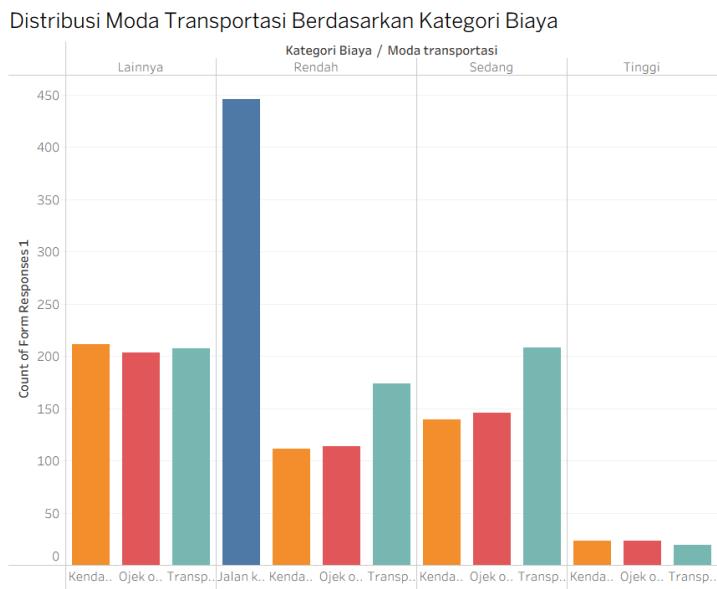
Analisis Kenyamanan Moda Transportasi



**Gambar 3. Analisis Kenyamanan Moda Transportasi**

Visualisasi ini menunjukkan pola utama bahwa sebagian besar responden menilai moda transportasi yang mereka gunakan sebagai nyaman, terutama pada transportasi umum yang memiliki jumlah responden tertinggi dibanding moda lain. Perbandingan penting terlihat antara kategori Nyaman dan Tidak Nyaman, di mana semua moda (jalan kaki/sepeda, kendaraan pribadi, ojek online, dan transportasi umum) konsisten memiliki jumlah penilaian nyaman yang lebih tinggi daripada tidak nyaman, meskipun transportasi umum juga mencatat jumlah tidak nyaman yang relatif lebih besar dibanding moda lain. Kaitan dengan tujuan penelitian adalah hasil ini menegaskan bahwa kenyamanan moda transportasi berperan dalam mendukung mobilitas dan produktivitas tenaga kerja, sehingga pemilihan moda yang lebih nyaman berpotensi meningkatkan efisiensi perjalanan dan kualitas kerja di wilayah Jakarta.

#### 4. Distribusi Moda Transportasi Berdasarkan Kategori Biaya

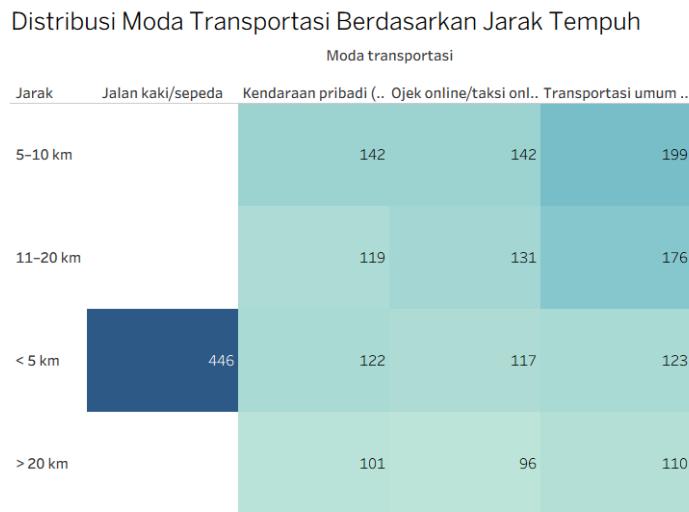


**Gambar 4. Distribusi Moda Transportasi Berdasarkan Kategori Biaya**

Visualisasi ini memperlihatkan pola utama bahwa pada kategori biaya rendah, moda jalan kaki/sepeda mendominasi jumlah responden dibandingkan moda transportasi lainnya, sedangkan pada biaya sedang dan tinggi, distribusi responden cenderung lebih merata meskipun jumlahnya lebih sedikit. Perbandingan penting terlihat pada moda transportasi umum yang memiliki jumlah responden relatif tinggi pada biaya sedang, menunjukkan peran transportasi umum sebagai pilihan utama ketika biaya tidak lagi rendah namun masih terjangkau. Kaitan dengan tujuan penelitian adalah bahwa temuan ini mendukung analisis bahwa biaya transportasi memengaruhi pemilihan moda, yang

pada akhirnya berdampak pada efisiensi perjalanan dan produktivitas tenaga kerja di Jakarta.

## 5. Distribusi Moda Transportasi Berdasarkan Jarak Tempuh



**Gambar 5. Distribusi Moda Transportasi Berdasarkan Jarak Tempuh**

Visualisasi ini menunjukkan pola utama bahwa pada jarak tempuh kurang dari 5 km, moda jalan kaki/sepeda sangat mendominasi dibandingkan moda transportasi lainnya, sedangkan pada jarak yang lebih jauh dominasi tersebut beralih ke transportasi umum. Perbandingan penting terlihat pada jarak 5–10 km dan 11–20 km, di mana transportasi umum memiliki jumlah pengguna tertinggi dibanding kendaraan pribadi dan ojek online, menunjukkan preferensi terhadap moda yang lebih efisien untuk jarak menengah. Kaitan dengan tujuan penelitian adalah bahwa jarak tempuh berpengaruh terhadap pemilihan moda transportasi, yang selanjutnya berdampak pada efisiensi mobilitas harian dan produktivitas tenaga kerja di Jakarta.

### 5.3 Insight dari Data

Berdasarkan hasil Exploratory Data Analysis (EDA) yang telah dilakukan, diperoleh beberapa insight penting terkait karakteristik transportasi tenaga kerja serta potensi pengaruhnya terhadap produktivitas kerja di wilayah Jakarta.

Pertama, dari sisi biaya transportasi, mayoritas responden berada pada kategori biaya rendah hingga menengah. Hal ini menunjukkan bahwa sebagian besar tenaga kerja masih memiliki beban biaya perjalanan yang relatif terjangkau. Namun, responden dengan biaya transportasi tinggi cenderung menunjukkan persepsi yang lebih negatif terhadap motivasi kerja,

mengindikasikan bahwa peningkatan beban biaya dapat berdampak pada penurunan motivasi dan produktivitas kerja.

Kedua, berdasarkan jarak tempuh, sebagian besar responden menempuh jarak dekat hingga menengah menuju tempat kerja. Meskipun demikian, responden dengan jarak tempuh yang lebih jauh menunjukkan kecenderungan frekuensi keterlambatan yang lebih tinggi. Insight ini menunjukkan bahwa jarak perjalanan yang panjang meningkatkan risiko keterlambatan dan kelelahan, yang berpotensi menurunkan kinerja dan produktivitas tenaga kerja.

Ketiga, hasil analisis keterlambatan kerja menunjukkan bahwa tenaga kerja dengan jarak tempuh dekat lebih banyak berada pada kategori tidak pernah terlambat, sedangkan pada jarak menengah hingga jauh, proporsi responden yang kadang atau sering terlambat meningkat. Pola ini memperkuat dugaan bahwa faktor jarak tempuh memiliki peran penting terhadap ketepatan waktu kerja.

Keempat, dari sisi moda transportasi, penggunaan transportasi umum dan kendaraan pribadi mendominasi pola perjalanan responden. Selain itu, distribusi moda berdasarkan jarak tempuh menunjukkan bahwa jalan kaki atau sepeda lebih banyak digunakan pada jarak dekat, sementara transportasi umum lebih dominan pada jarak menengah hingga jauh. Hal ini mengindikasikan bahwa pemilihan moda transportasi sangat dipengaruhi oleh kebutuhan efisiensi perjalanan.

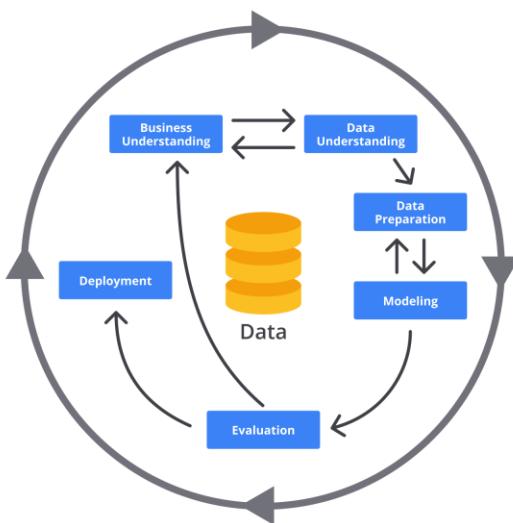
Kelima, hasil visualisasi kenyamanan dan pengaruh biaya terhadap motivasi kerja menunjukkan bahwa responden dengan biaya transportasi rendah dan moda yang lebih nyaman cenderung memiliki persepsi motivasi kerja yang lebih baik. Sebaliknya, biaya tinggi dan tingkat kenyamanan yang lebih rendah berpotensi meningkatkan stres perjalanan, yang berdampak negatif terhadap kondisi psikologis dan produktivitas kerja.

Secara keseluruhan, hasil EDA menunjukkan bahwa biaya transportasi, jarak tempuh, moda transportasi, kenyamanan perjalanan, dan ketepatan waktu kerja saling berkaitan dalam memengaruhi produktivitas tenaga kerja. Insight ini menjadi dasar yang kuat untuk tahap analisis lanjutan menggunakan pendekatan machine learning, khususnya Logistic Regression dan Random Forest, guna memodelkan pengaruh faktor transportasi terhadap produktivitas tenaga kerja secara lebih sistematis dan objektif.

## BAB VI

### METODOLOGI PENELITIAN (CRISP-DM)

Penelitian ini menggunakan metodologi CRISP-DM (Cross Industry Standard Process for Data Mining) karena menyediakan tahapan yang sistematis dan terstruktur dalam proses analisis data dan pemodelan machine learning. Tahapan CRISP-DM yang digunakan meliputi enam tahap sebagai berikut.



**Gambar 6. Metodologi CRISP-DM**

#### 6.1 Business Understanding

Tahap *Business Understanding* bertujuan untuk memahami permasalahan penelitian secara menyeluruh berdasarkan konteks sosial dan ekonomi. Permasalahan utama dalam penelitian ini adalah kondisi transportasi di wilayah Jakarta yang ditandai oleh kemacetan, jarak tempuh yang bervariasi, serta biaya transportasi yang relatif tinggi. Kondisi tersebut berpotensi memengaruhi produktivitas tenaga kerja melalui aspek motivasi kerja, tingkat kehadiran, keterlambatan, dan kenyamanan perjalanan. Berdasarkan permasalahan tersebut, tujuan bisnis dari penelitian ini adalah memperoleh gambaran dan pemahaman mengenai pengaruh faktor transportasi terhadap produktivitas tenaga kerja di Jakarta. Selanjutnya, tujuan data mining ditetapkan untuk membangun model machine learning yang mampu memprediksi apakah faktor transportasi berpengaruh atau tidak berpengaruh terhadap produktivitas tenaga kerja berdasarkan karakteristik biaya, jarak, dan moda transportasi. Keberhasilan penelitian diukur melalui kemampuan model dalam menghasilkan prediksi yang akurat serta menghasilkan insight yang relevan untuk mendukung pengambilan keputusan.

## **6.2 Data Understanding**

Tahap *Data Understanding* berfokus pada pengumpulan, eksplorasi, dan pemahaman terhadap data yang digunakan dalam penelitian. Data diperoleh melalui survei daring menggunakan Google Form yang disebarluaskan kepada tenaga kerja yang berdomisili dan bekerja di wilayah Jakarta. Jumlah responden yang berhasil dikumpulkan sebanyak 2.021 responden, sehingga data dinilai cukup representatif untuk menggambarkan kondisi transportasi tenaga kerja di Jakarta. Variabel yang dikumpulkan meliputi biaya transportasi harian, jarak tempuh perjalanan ke tempat kerja, moda transportasi yang digunakan, serta persepsi responden terhadap pengaruh transportasi terhadap produktivitas, kehadiran, dan motivasi kerja. Pada tahap ini dilakukan pemeriksaan awal terhadap struktur dataset, jenis variabel, distribusi data, serta identifikasi permasalahan kualitas data seperti *missing value*. Pemahaman awal ini menjadi dasar untuk menentukan strategi pengolahan data pada tahap berikutnya.

## **6.3 Data Preparation**

Tahap *Data Preparation* bertujuan untuk menyiapkan data agar dapat digunakan secara optimal dalam proses pemodelan. Tahapan ini meliputi *data cleaning*, *data transformation*, *feature engineering*, *feature selection*, dan *data splitting*. Proses *data cleaning* dilakukan untuk menangani nilai kosong yang terdapat pada beberapa variabel survei. Nilai kosong pada variabel kategorikal diisi menggunakan nilai modus, sedangkan kolom opsional diisi dengan label khusus agar informasi responden tetap terjaga. Setelah proses ini, dataset dinyatakan bersih dan siap untuk diolah lebih lanjut. Tahap *data transformation* dilakukan dengan melakukan pengkodean (*encoding*) terhadap variabel kategorikal seperti biaya transportasi, jarak tempuh, dan moda transportasi ke dalam bentuk numerik. Pengkodean ini diperlukan agar data dapat diproses oleh algoritma Logistic Regression dan Random Forest, serta tetap merepresentasikan tingkat atau urutan kategori. Pada tahap *feature engineering*, variabel persepsi responden mengenai pengaruh moda transportasi terhadap produktivitas, pengaruh jarak terhadap kehadiran, dan pengaruh biaya transportasi terhadap motivasi digabungkan menjadi satu skor pengaruh transportasi. Skor ini kemudian dikonversi menjadi variabel target biner (*Target\_Pengaruh*), yang menunjukkan apakah transportasi dianggap berpengaruh atau tidak berpengaruh terhadap produktivitas tenaga kerja. Tahap *feature selection* dilakukan dengan memilih variabel input utama, yaitu hasil encoding dari biaya transportasi, jarak tempuh, dan moda transportasi. Selanjutnya, dataset dibagi menjadi data latih dan data uji menggunakan metode *train-test split* dengan proporsi 80% data latih dan 20% data uji serta teknik *stratified sampling* untuk menjaga keseimbangan kelas pada variabel target.

#### **6.4 Modeling**

Tahap *Modeling* merupakan tahap pembangunan model machine learning berdasarkan data yang telah dipersiapkan. Dalam penelitian ini digunakan dua algoritma klasifikasi, yaitu Logistic Regression dan Random Forest, sesuai dengan tujuan penelitian yang telah dijelaskan pada bab sebelumnya. Logistic Regression digunakan sebagai model awal (*baseline model*) karena memiliki struktur yang sederhana dan mudah diinterpretasikan. Model ini digunakan untuk melihat hubungan antara variabel biaya transportasi, jarak tempuh, dan moda transportasi terhadap pengaruh transportasi terhadap produktivitas tenaga kerja. Random Forest digunakan sebagai model pembanding yang bersifat non-linier dan lebih kompleks. Algoritma ini dipilih karena kemampuannya dalam menangkap hubungan yang lebih kompleks antar variabel serta mengurangi risiko *overfitting*. Kedua model dilatih menggunakan data latih yang sama agar hasil perbandingan performa dapat dilakukan secara objektif

#### **6.4 Evaluation**

Tahap *Evaluation* dilakukan untuk menilai kinerja model yang telah dibangun dan memastikan kesesuaianya dengan tujuan penelitian. Evaluasi dilakukan menggunakan data uji dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil evaluasi menunjukkan bahwa model Logistic Regression dan Random Forest memiliki tingkat akurasi yang sama dan performa yang relatif seimbang. Hal ini menunjukkan bahwa kedua model mampu memprediksi pengaruh transportasi terhadap produktivitas tenaga kerja berdasarkan faktor biaya, jarak, dan moda transportasi. Evaluasi ini menegaskan bahwa model yang dibangun telah memenuhi tujuan data mining yang ditetapkan pada tahap *Business Understanding*.

#### **6.5 Deployment**

Tahap *Deployment* merupakan tahap akhir dari metodologi CRISP-DM, yaitu penyajian dan pemanfaatan hasil analisis. Pada penelitian ini, hasil analisis disajikan dalam bentuk laporan akhir serta visualisasi data melalui dashboard interaktif menggunakan Tableau. Dashboard yang dibangun menampilkan hubungan antara biaya transportasi dengan motivasi kerja, jarak tempuh dengan keterlambatan, serta moda transportasi dengan kenyamanan perjalanan. Penyajian ini bertujuan untuk memudahkan pemahaman hasil penelitian dan memberikan insight yang dapat dimanfaatkan oleh perusahaan maupun pembuat kebijakan dalam merancang strategi transportasi yang mendukung peningkatan produktivitas tenaga kerja di Jakarta.

## **BAB VII**

### **HASIL DAN PEMBAHASAN**

Bab ini menyajikan hasil penelitian yang diperoleh dari proses implementasi analisis data dan pemodelan machine learning berdasarkan tahapan yang telah dirancang pada bab sebelumnya. Pembahasan difokuskan pada proses persiapan data, preprocessing data, serta implementasi pemodelan menggunakan algoritma Logistic Regression dan Random Forest.

#### **7.1 Implementasi Persiapan Data**

Tahap implementasi persiapan data bertujuan untuk menjelaskan bagaimana data penelitian diperoleh dan disiapkan sebelum dilakukan pengolahan lebih lanjut. Data pada penelitian ini dikumpulkan melalui survei daring menggunakan Google Form yang disebarluaskan kepada tenaga kerja di wilayah Jakarta. Survei ini dirancang untuk memperoleh informasi terkait kondisi transportasi responden, yang meliputi biaya transportasi harian, jarak tempuh perjalanan ke tempat kerja, moda transportasi yang digunakan, serta persepsi responden mengenai pengaruh transportasi terhadap produktivitas kerja. Hasil pengisian Google Form kemudian diekspor ke dalam format CSV, sehingga dapat diolah lebih lanjut menggunakan bahasa pemrograman Python dan lingkungan Jupyter Notebook. Total data yang berhasil dikumpulkan berjumlah 2021 responden, yang selanjutnya digunakan sebagai dataset utama dalam penelitian ini. Pada tahap ini, belum dilakukan pengolahan data secara teknis. Data masih berada dalam bentuk mentah (raw data) dan akan diproses lebih lanjut pada tahap preprocessing.

#### **7.2 Implementasi Preprocessing Data**

Tahap preprocessing data merupakan tahap yang sangat penting karena kualitas data yang digunakan akan memengaruhi performa model machine learning yang dibangun. Tahapan preprocessing yang dilakukan dalam penelitian ini meliputi beberapa langkah berikut.

##### **7.2.1 Import Data**

Proses import data merupakan tahap awal dalam analisis data. Data yang digunakan berasal dari hasil survei mengenai biaya, jarak, dan moda transportasi. Untuk memuat dataset ke dalam lingkungan Python, digunakan library Pandas, yang menyediakan fungsi `read_csv()`. Fungsi ini membaca file CSV dan mengonversinya menjadi DataFrame, struktur data dua dimensi yang memudahkan manipulasi data.

```
#load data
df = pd.read_csv("Data Survei.csv", sep=";")
display(df.head())
```

### 7.2.2 Preprocessing

Preprocessing pada data dilakukan dengan mempersiapkan data mentah yang telah dikumpulkan dari sumbernya untuk diolah lebih lanjut.

#### 1. Data Cleaning

Tahap data cleaning bertujuan untuk memastikan dataset berada dalam kondisi bersih, konsisten, dan siap untuk dianalisis. Pada tahap ini dilakukan identifikasi terhadap nilai hilang (missing value) yang terdapat pada beberapa variabel hasil survei.

```
Tanggal_Waktu          0
Usia                  0
Gender                0
Status_Kerja          1
Lama_Bekerja          0
Biaya_Transport        0
Pengaruh_Biaya_Transportasi_ke_Motivasi 0
Subsidi_Transportasi_Kantor    0
Persetujuan_Subsidi_Jika_Belum_Ada 0
Kepuasan_terhadap_Kebijakan_Subsidi_Transportasi 0
Pengaruh_Biaya_Transportasi_ke_Pengeluaran 0
Jarak                 0
Waktu_Tempuh          0
Pengaruh_Jarak_ke_Kehadiran 0
Frekuensi_Terlambat_karena_Kemacetan 0
Frekuensi_Terlambat_karena_Jarak 0
Moda_transportasi     0
Kenyamanan_Modra_Transportasi 0
Pernah_Ganti_Modra_untuk_Efisiensi 0
Alasan_Ganti_Modra_Transportasi 1960
Frekuensi_Onsite_per_Minggu 0
Frekuensi_Dampak_Fisik_dari_Perjalanan 0
Dampak_Fisik_Perjalanan_Kantor 0
Pengaruh_Transportasi_ke_Konsentrasi_Kerja 0
Pengaruh_Modra_Transportasi_ke_Prodktivitas 0
Faktor_Transportasi_Terhadap_Prodktivitas 0
Persetujuan_Solusi_Transportasi_untuk_Prodktivitas 0
dtype: int64
```

**Gambar 7. Data Cleaning**

Terdapat 1961 missing value. Penanganan missing value dilakukan dengan metode imputasi yang disesuaikan dengan karakteristik data. Untuk variabel kategorikal, nilai kosong diisi menggunakan nilai modus, yaitu kategori yang paling sering muncul, sehingga distribusi data tetap terjaga. Sementara itu, pada kolom tertentu yang bersifat opsional, nilai kosong digantikan dengan label khusus seperti "Tidak diisi" agar data tetap dapat dimanfaatkan tanpa menghilangkan informasi responden.

```
# Isi missing value
df["Status_Kerja"] = df["Status_Kerja"].fillna(
    df["Status_Kerja"].mode()[0]
)
```

```

df["Alasan Ganti Moda Transportasi"] = df["Alasan Ganti
Moda Transportasi"].fillna(
    "Tidak diisi"
)

print("\nCek ulang missing value setelah ditangani:")
print(df.isnull().sum())

```

Setelah imputasi, semua kolom bebas dari missing value. Dataset kini siap untuk langkah selanjutnya, yaitu encoding variabel kategorikal. Penanganan missing value ini penting karena data yang tidak lengkap dapat menurunkan akurasi model atau menghasilkan prediksi yang bias.

## 2. Data Transformation (Encoding)

Setelah data dinyatakan bersih, tahap selanjutnya adalah data transformation berupa encoding terhadap variabel kategorikal. Algoritma Logistic Regression dan Random Forest tidak dapat memproses data dalam bentuk teks, sehingga variabel seperti biaya transportasi, jarak tempuh, dan moda transportasi harus dikonversi ke dalam bentuk numerik.

```

biaya_map = {
    "< Rp20.000": 1,
    "Rp20.000-Rp50.000": 2,
    "Rp50.000-Rp75.000": 3,
    "Rp75.000-Rp100.000": 4,
    "> Rp100.000": 5
}

df["Biaya_Encoded"] = df["Biaya Transport"].map(biaya_map)

```

```

jarak_map = {
    "< 5 km": 1,
    "5-10 km": 2,
    "11-20 km": 3,
    "> 20 km": 4
}

```

```
}
```

```
df["Jarak_Encoded"] = df["Jarak"].map(jarak_map)
```

```
moda_map = {
```

```
    "Kendaraan Pribadi": 1,
```

```
    "Transportasi Umum": 2,
```

```
    "Ojek Online": 3,
```

```
    "Jalan Kaki/Sepeda": 4
```

```
}
```

```
df["Moda_Encoded"] = df["Moda Transportasi"].map(moda_map)
```

Proses encoding algoritma ini memahami urutan biaya, jarak, dan jenis moda transportasi. Misalnya, semakin tinggi nilai Biaya\_Encoded, semakin besar pengeluaran transportasi responden.

### 3. Feature Engineering

Pada tahap feature engineering, dilakukan pengolahan dan pembentukan fitur baru yang merepresentasikan informasi penting dari data mentah. Variabel persepsi responden terkait pengaruh moda transportasi terhadap produktivitas, pengaruh jarak terhadap kehadiran, dan pengaruh biaya transportasi terhadap motivasi digabungkan untuk membentuk satu skor pengaruh transportasi. Skor tersebut dihitung dengan mengambil nilai rata-rata dari seluruh indikator persepsi. Selanjutnya, skor pengaruh ini dikonversi menjadi variabel target biner (Target\_Pengaruh), yang mengklasifikasikan responden ke dalam dua kelas, yaitu transportasi berpengaruh dan tidak berpengaruh terhadap produktivitas tenaga kerja. Pembentukan variabel target ini menjadi dasar dalam penerapan supervised learning pada tahap pemodelan.

```
#feature engineering
```

```
cols_pengaruh = [
```

```
    "Pengaruh Moda Transportasi ke Produktivitas",
```

```
    "Pengaruh Jarak ke Kehadiran",
```

```
    "Pengaruh Biaya Transportasi ke Motivasi"
```

```
]
```

```
df[cols_pengaruh] = df[cols_pengaruh].astype(int)
```

#### 4. Feature Selection

Tahap feature selection dilakukan untuk menentukan variabel input yang paling relevan dan berkontribusi langsung terhadap tujuan penelitian. Fitur yang digunakan dalam pemodelan adalah variabel hasil encoding dari biaya transportasi, jarak tempuh, dan moda transportasi.

```
df["Target_Pengaruh"] = df["Skor_Pengaruh"].apply(  
    lambda x: 1 if x >= 3 else 0  
)
```

Pemilihan fitur ini dilakukan secara terarah berdasarkan rumusan masalah dan tujuan penelitian, sehingga model yang dibangun tetap sederhana, efisien, dan mudah diinterpretasikan. Selain itu, penggunaan fitur yang relevan juga membantu mengurangi kompleksitas model.

#### 5. Data Splitting

Pada penelitian ini, dataset dibagi menjadi dua bagian utama, yaitu data latih (training set) dan data uji (testing set). Data latih digunakan untuk melatih model machine learning, sedangkan data uji digunakan untuk mengukur kemampuan model dalam memprediksi pengaruh transportasi terhadap produktivitas tenaga kerja.

```
X = df[["Biaya_Encoded", "Jarak_Encoded", "Moda_Encoded"]]  
y = df["Target_Pengaruh"]  
  
#data splitting  
  
X_train, X_test, y_train, y_test = train_test_split(  
    X,  
    y,  
    test_size=0.2,  
    random_state=42,  
    stratify=y  
)
```

Pembagian data dilakukan menggunakan metode Train-Test Split dengan proporsi 80% data latih dan 20% data uji. Selain itu, digunakan parameter stratified sampling untuk menjaga proporsi kelas pada variabel target agar tetap seimbang pada kedua subset data. Pendekatan ini bertujuan untuk memastikan bahwa model Logistic Regression dan Random

Forest dapat belajar secara optimal dari data latih serta memiliki kemampuan generalisasi yang baik ketika diuji pada data yang belum pernah dilihat sebelumnya.

### 7.3 Implementasi Pemodelan

Tahap Modeling merupakan tahap implementasi algoritma machine learning menggunakan data yang telah dipersiapkan pada tahap sebelumnya.

#### 7.3.1 Logistic Regression

Logistic Regression digunakan sebagai model awal (baseline model) karena memiliki struktur yang sederhana, efisien, dan mudah diinterpretasikan. Algoritma ini bekerja dengan memodelkan hubungan linier antara variabel input dan variabel target dalam bentuk probabilitas, yang kemudian dikonversi menjadi kelas klasifikasi.

Pada penelitian ini, Logistic Regression dilatih menggunakan data latih dengan variabel input berupa hasil encoding dari biaya transportasi, jarak tempuh, dan moda transportasi, serta variabel target berupa Target\_Pengaruh. Untuk mengatasi kemungkinan ketidakseimbangan kelas pada variabel target, digunakan parameter class\_weight="balanced".

Proses pemodelan Logistic Regression diimplementasikan menggunakan potongan kode berikut:

```
from sklearn.linear_model import LogisticRegression  
lr = LogisticRegression(  
    max_iter=1000,  
    class_weight="balanced"  
)  
lr.fit(X_train, y_train)  
y_pred_lr = lr.predict(X_test)
```

Model yang telah dilatih kemudian digunakan untuk menghasilkan prediksi pada data uji. Hasil prediksi ini selanjutnya digunakan pada tahap evaluasi untuk mengukur performa model.

#### 7.3.2 Random Forest

Random Forest digunakan sebagai model pembanding yang bersifat non-linier dan lebih kompleks. Algoritma ini merupakan metode ensemble learning yang bekerja dengan membangun banyak pohon keputusan (decision tree) dan menggabungkan hasil prediksi dari seluruh pohon tersebut melalui mekanisme voting mayoritas. Keunggulan Random Forest terletak pada kemampuannya dalam menangkap hubungan non-linier antar variabel serta mengurangi risiko overfitting dibandingkan dengan model pohon keputusan tunggal. Oleh

karena itu, Random Forest diharapkan mampu memberikan performa yang lebih baik pada data survei yang memiliki pola kompleks.

Model Random Forest dilatih menggunakan data latih yang sama dengan Logistic Regression agar perbandingan performa dapat dilakukan secara objektif. Proses pemodelannya ditunjukkan pada potongan kode berikut:

```
from sklearn.ensemble import RandomForestClassifier

rf = RandomForestClassifier(
    n_estimators=100,
    random_state=42,
    class_weight="balanced"
)

rf.fit(X_train, y_train)
y_pred_rf = rf.predict(X_test)
```

Hasil prediksi dari Random Forest kemudian dibandingkan dengan hasil Logistic Regression pada tahap evaluasi menggunakan metrik yang sama, sehingga dapat ditentukan model yang paling optimal dalam memprediksi pengaruh transportasi terhadap produktivitas tenaga kerja.

#### 7.4 Hasil Evaluasi Model

Setelah dilakukan proses implementasi algoritma klasifikasi untuk memprediksi moda transportasi responden dalam konteks Analisis Produktivitas Tenaga Kerja Berdasarkan Biaya, Jarak, dan Moda Transportasi di Jakarta, dilakukan evaluasi performa model menggunakan dua metode, yaitu Logistic Regression dan Random Forest. Evaluasi dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score.

**Table 2. Hasil Evaluasi Model**

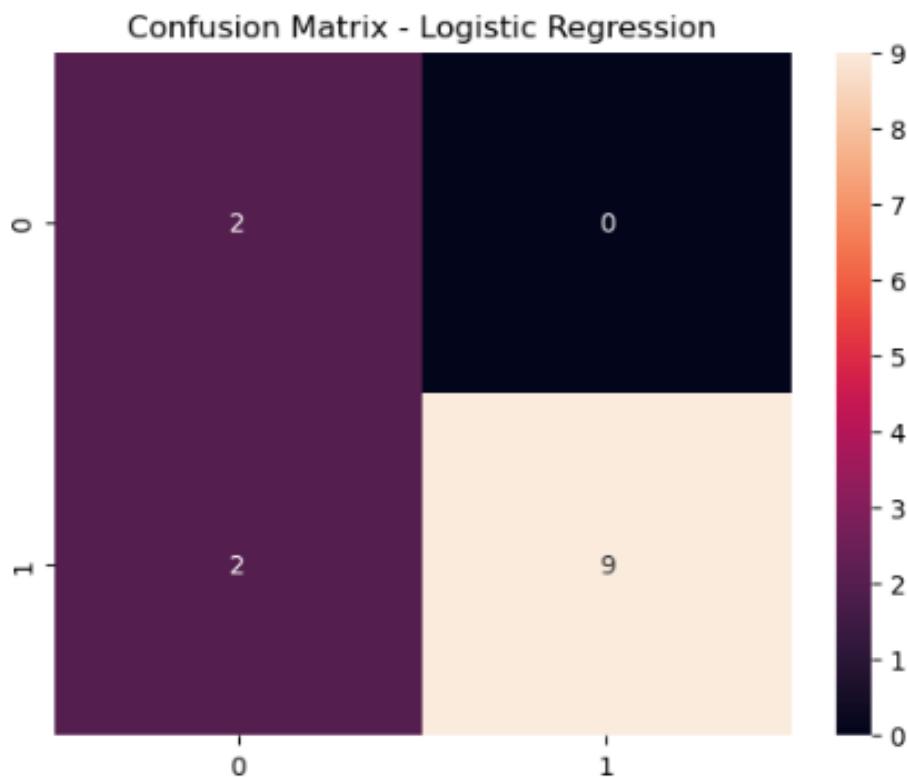
Model	Accuracy	Precision		Recall		F1-score	
		0	1	0	1	0	1
Logistic Regression	0.85	0.50	1.00	1.00	0.82	0.67	0.90
Random Forest	0.85	0.50	1.00	1.00	0.82	0.67	0.90

#### 7.4.1 Evaluasi Kinerja Model Logistic Regression

Evaluasi kinerja model Logistic Regression dilakukan untuk mengetahui kemampuan model dalam memprediksi pilihan moda transportasi responden berdasarkan variabel biaya perjalanan, jarak tempuh, dan moda transportasi. Hasil evaluasi ditunjukkan melalui metrik accuracy, precision, recall, dan F1-score sebagaimana disajikan pada Tabel 2.

Berdasarkan hasil evaluasi, model Logistic Regression memperoleh nilai akurasi sebesar 84,6%, yang menunjukkan bahwa model mampu melakukan klasifikasi dengan tingkat ketepatan yang cukup tinggi. Nilai precision dan recall pada masing-masing kelas menunjukkan bahwa model memiliki performa yang baik dalam mengenali pola data, khususnya pada kelas mayoritas.

Untuk melihat lebih detail terkait kesalahan dan keberhasilan klasifikasi, hasil evaluasi juga ditampilkan dalam bentuk confusion matrix sebagaimana ditunjukkan pada Gambar.



Gambar 8. Confusion Matrix

Berdasarkan confusion matrix pada Gambar, dapat diketahui bahwa model berhasil mengklasifikasikan 9 data kelas positif dengan benar dan 2 data kelas negatif dengan benar. Namun demikian, masih terdapat 2 data kelas positif yang salah diklasifikasikan sebagai kelas negatif. Hasil ini menunjukkan bahwa model memiliki kecenderungan lebih baik dalam mengenali kelas produktivitas tinggi dibandingkan kelas produktivitas rendah.

#### **7.4.2 Evaluasi Kinerja Model Random Forest**

Model Random Forest juga dievaluasi menggunakan metrik yang sama untuk memastikan konsistensi performa antar algoritma. Berdasarkan Tabel 4.X, model Random Forest menghasilkan nilai akurasi sebesar 84,6%, yang sama dengan model Logistic Regression. Nilai precision, recall, dan F1-score yang dihasilkan oleh Random Forest menunjukkan pola yang serupa dengan Logistic Regression. Hal ini mengindikasikan bahwa penggunaan model dengan kompleksitas yang lebih tinggi tidak memberikan peningkatan performa yang signifikan pada dataset penelitian ini.

#### **7.4.3 Perbandingan Performa Model Logistic Regression dan Random Forest**

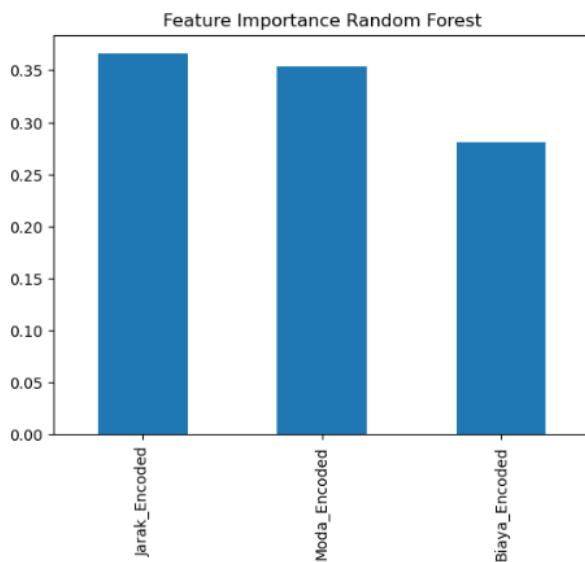
Perbandingan hasil evaluasi antara Logistic Regression dan Random Forest menunjukkan bahwa kedua model memiliki performa yang identik dan konsisten. Tidak terdapat perbedaan signifikan pada nilai accuracy, precision, recall, maupun F1-score. Kesamaan performa ini menunjukkan bahwa pola data yang digunakan dalam penelitian relatif sederhana dan dapat dipelajari dengan baik oleh model linear seperti Logistic Regression. Oleh karena itu, pemilihan model dapat dipertimbangkan berdasarkan efisiensi komputasi dan kemudahan interpretasi hasil.

### **7.5 Analisis Pengaruh Variabel terhadap Produktivitas Tenaga Kerja**

Meskipun hasil evaluasi menunjukkan bahwa kedua model memiliki performa yang baik, evaluasi kinerja saja belum cukup untuk menjelaskan faktor-faktor yang memengaruhi hasil prediksi. Oleh karena itu, dilakukan analisis lanjutan untuk mengetahui kontribusi masing-masing variabel terhadap produktivitas tenaga kerja menggunakan metode feature importance.

#### **7.5.1 Feature Importance Menggunakan Random Forest**

Analisis feature importance dilakukan menggunakan algoritma Random Forest untuk mengetahui tingkat kepentingan masing-masing variabel input dalam memprediksi produktivitas tenaga kerja. Hasil analisis feature importance ditunjukkan pada Gambar.



**Gambar 9. Feature Importance**

Berdasarkan Gambar, variabel jarak tempuh (Jarak\_Encoded) memiliki nilai feature importance tertinggi dibandingkan variabel lainnya. Hal ini menunjukkan bahwa jarak perjalanan menuju tempat kerja merupakan faktor yang paling berpengaruh terhadap produktivitas tenaga kerja. Variabel moda transportasi (Moda\_Encoded) menempati urutan kedua, yang mengindikasikan bahwa jenis moda transportasi yang digunakan pekerja berperan penting dalam menentukan efisiensi perjalanan kerja. Sementara itu, variabel biaya perjalanan (Biaya\_Encoded) memiliki pengaruh yang lebih rendah dibandingkan dua variabel lainnya, meskipun tetap memberikan kontribusi terhadap hasil prediksi model.

### 7.5.2 Pembahasan Hasil Analisis Feature Importance

Hasil analisis feature importance menunjukkan bahwa faktor jarak dan moda transportasi memiliki pengaruh yang lebih dominan terhadap produktivitas tenaga kerja dibandingkan biaya perjalanan. Temuan ini menunjukkan bahwa efisiensi waktu dan kenyamanan perjalanan memiliki peranan yang lebih besar dalam menentukan kesiapan dan performa tenaga kerja di tempat kerja. Dalam konteks Jakarta sebagai wilayah metropolitan dengan tingkat kemacetan yang tinggi, jarak tempuh yang panjang dapat meningkatkan kelelahan dan menurunkan efektivitas kerja. Sementara itu, pemilihan moda transportasi yang efisien dapat membantu mengurangi dampak negatif perjalanan terhadap kondisi fisik dan mental tenaga kerja.

### 7.6 Insight Penelitian

Berdasarkan hasil evaluasi model klasifikasi serta analisis feature importance yang telah dilakukan, diperoleh beberapa insight utama terkait produktivitas tenaga kerja dalam konteks biaya perjalanan, jarak tempuh, dan moda transportasi di Jakarta.

Insight pertama menunjukkan bahwa jarak tempuh merupakan faktor yang paling dominan dalam memengaruhi produktivitas tenaga kerja. Nilai feature importance tertinggi pada variabel jarak mengindikasikan bahwa semakin jauh jarak perjalanan dari tempat tinggal menuju tempat kerja, semakin besar potensi penurunan produktivitas tenaga kerja. Hal ini berkaitan dengan meningkatnya waktu perjalanan dan kelelahan fisik maupun mental sebelum pekerja memulai aktivitas kerja.

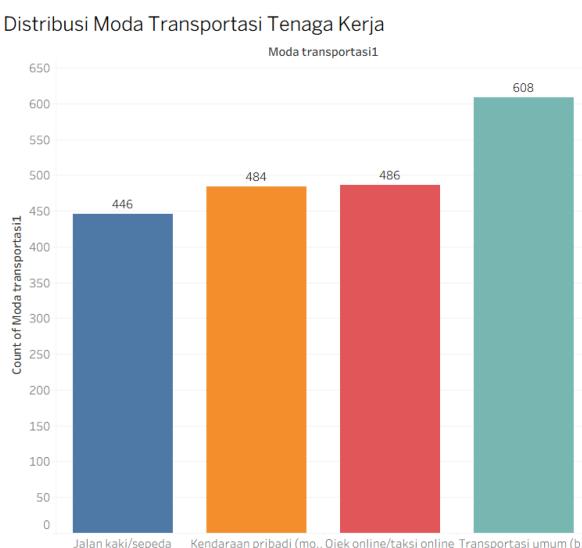
Insight kedua menunjukkan bahwa moda transportasi memiliki peranan penting dalam menentukan tingkat efisiensi perjalanan kerja. Pemilihan moda transportasi yang tepat dapat membantu mengurangi waktu tempuh dan meningkatkan kenyamanan perjalanan, sehingga berdampak positif terhadap kesiapan kerja tenaga kerja. Moda transportasi yang efisien berpotensi meminimalkan dampak negatif perjalanan panjang terhadap produktivitas.

Insight ketiga menunjukkan bahwa biaya perjalanan memiliki pengaruh yang relatif lebih kecil dibandingkan jarak dan moda transportasi. Meskipun demikian, biaya perjalanan tetap berkontribusi dalam menentukan pilihan moda transportasi pekerja. Biaya yang tinggi dapat menjadi beban tambahan bagi tenaga kerja dan berpotensi memengaruhi tingkat kepuasan serta fokus kerja secara tidak langsung.

Secara keseluruhan, insight penelitian ini menunjukkan bahwa faktor non-finansial, khususnya efisiensi waktu dan kenyamanan perjalanan, memiliki peranan yang lebih besar dalam memengaruhi produktivitas tenaga kerja dibandingkan faktor biaya semata.

## 7.7 Dashboard dan Visualisasi Komunikasi Data

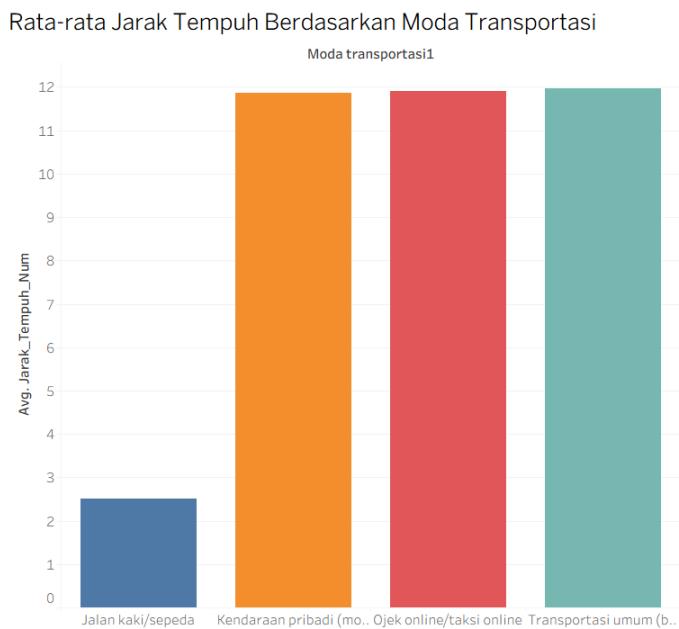
### 7.7.1 Distribusi Moda Transportasi Tenaga Kerja



**Gambar 10. Distribusi Moda Transportasi Tenaga Kerja**

Visualisasi pertama menampilkan distribusi penggunaan moda transportasi oleh tenaga kerja di Jakarta. Hasil visualisasi menunjukkan bahwa mayoritas responden menggunakan kendaraan pribadi dan transportasi umum sebagai moda utama dalam perjalanan menuju tempat kerja. Sementara itu, penggunaan moda jalan kaki atau sepeda memiliki proporsi yang lebih kecil. Temuan ini menggambarkan ketergantungan tenaga kerja terhadap moda transportasi bermotor dalam aktivitas komuter sehari-hari.

### 7.7.2 Hubungan Jarak Tempuh dan Produktivitas Tenaga Kerja

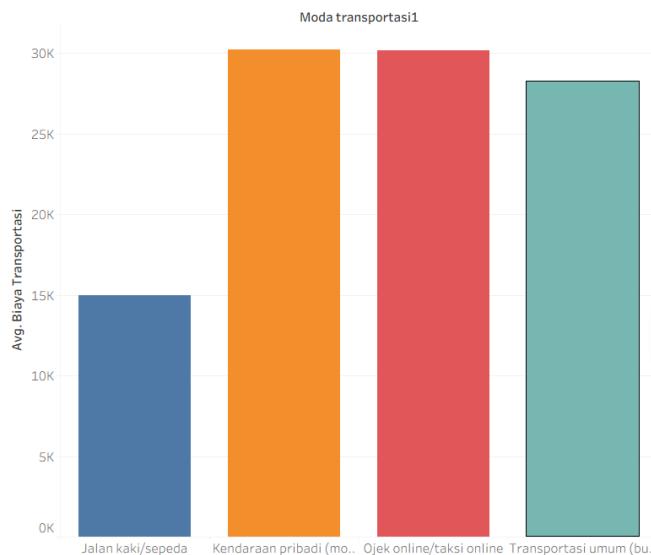


**Gambar 11. Hubungan Jarak Tempuh dan Produktivitas Tenaga Kerja**

Visualisasi kedua menggambarkan hubungan antara jarak tempuh perjalanan kerja dengan tingkat produktivitas tenaga kerja. Hasil visualisasi menunjukkan bahwa semakin jauh jarak tempuh yang harus dilalui, kecenderungan produktivitas tenaga kerja semakin menurun. Hal ini mengindikasikan bahwa jarak perjalanan yang panjang dapat memengaruhi kondisi fisik dan efisiensi kerja tenaga kerja.

### 7.7.3 Distribusi Biaya Transportasi Tenaga Kerja

Rata-rata Biaya Perjalanan Berdasarkan Moda Transportasi

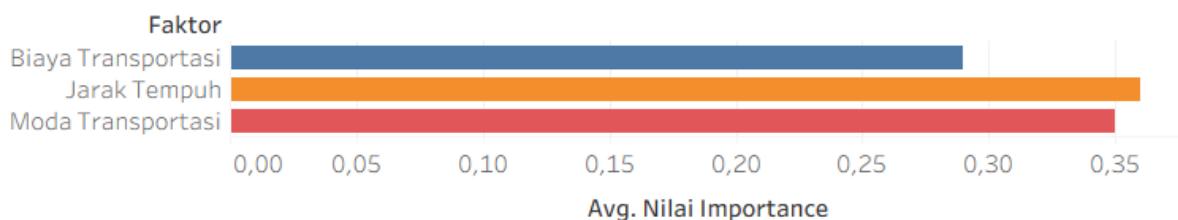


**Gambar 12. Distribusi Biaya Transportasi Tenaga Kerja**

Visualisasi ketiga menyajikan distribusi biaya transportasi yang dikeluarkan tenaga kerja untuk perjalanan menuju tempat kerja. Hasil visualisasi menunjukkan bahwa sebagian besar responden mengeluarkan biaya transportasi pada kategori menengah, sedangkan responden dengan biaya transportasi rendah dan tinggi memiliki proporsi yang lebih kecil. Temuan ini menunjukkan adanya variasi beban biaya perjalanan yang ditanggung oleh tenaga kerja.

### 7.7.4 Feature Importance Faktor Transportasi terhadap Produktivitas Tenaga Kerja

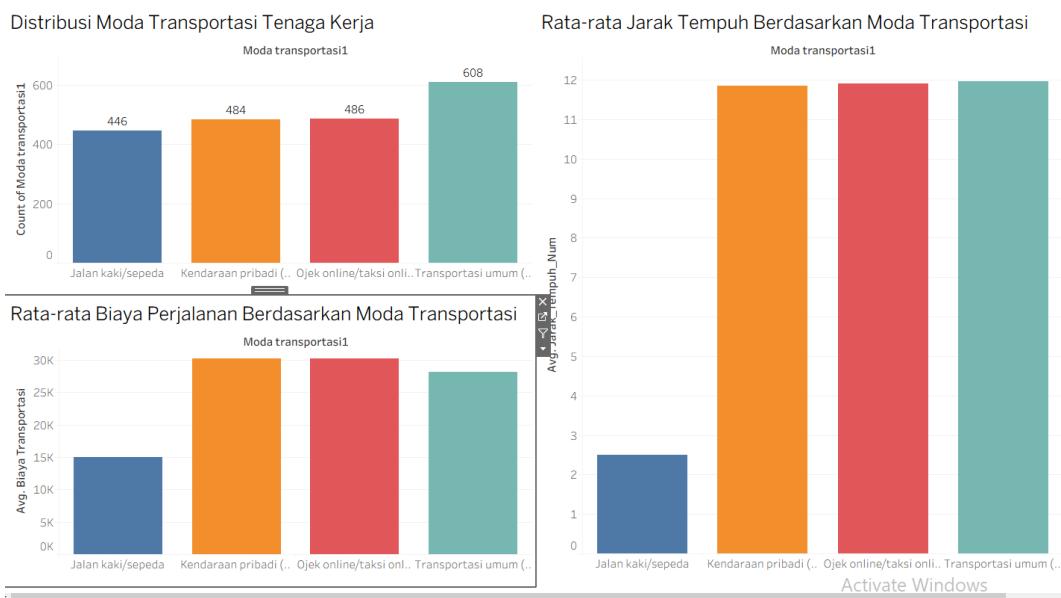
Feature Importance Faktor Transportasi terhadap Produktivitas Tenaga Kerja



**Gambar 13. Feature Importance Faktor Transportasi terhadap Produktivitas Tenaga Kerja**

Visualisasi keempat menampilkan tingkat kepentingan faktor transportasi berdasarkan hasil pemodelan machine learning. Hasil visualisasi menunjukkan bahwa jarak tempuh merupakan faktor yang paling berpengaruh terhadap produktivitas tenaga kerja, diikuti oleh moda transportasi dan biaya transportasi. Temuan ini menegaskan bahwa karakteristik perjalanan kerja, khususnya jarak tempuh, memiliki peran dominan dalam memengaruhi produktivitas tenaga kerja di Jakarta.

### 7.7.5 Dashboard



Gambar 14. Dashboard

Dashboard ini menampilkan ringkasan karakteristik perjalanan tenaga kerja di Jakarta berdasarkan moda transportasi, jarak tempuh, dan biaya perjalanan. Visualisasi 1–3 digunakan untuk memberikan gambaran awal mengenai pola komuter tenaga kerja serta hubungannya dengan produktivitas.

### 7.8 Rekomendasi

Berdasarkan insight yang diperoleh dari hasil penelitian, beberapa rekomendasi dapat diajukan sebagai bahan pertimbangan bagi pemangku kepentingan terkait.

Rekomendasi pertama ditujukan kepada pemerintah dan pembuat kebijakan, yaitu perlunya peningkatan kualitas dan pemerataan infrastruktur transportasi perkotaan. Upaya ini dapat difokuskan pada pengurangan waktu tempuh perjalanan kerja melalui optimalisasi jaringan transportasi umum, pengelolaan kemacetan, serta integrasi antar moda transportasi.

Rekomendasi kedua ditujukan kepada penyedia layanan transportasi, khususnya transportasi umum, untuk meningkatkan kenyamanan, keandalan, dan keterjangkauan layanan. Peningkatan kualitas layanan diharapkan dapat mendorong peralihan penggunaan moda transportasi yang lebih efisien dan ramah lingkungan, sehingga mendukung produktivitas tenaga kerja.

Rekomendasi ketiga ditujukan kepada perusahaan atau pemberi kerja, yaitu dengan mempertimbangkan kebijakan kerja fleksibel seperti pengaturan jam kerja atau sistem kerja jarak jauh (remote working) bagi pekerja dengan jarak tempuh yang jauh. Kebijakan tersebut dapat membantu mengurangi beban perjalanan dan meningkatkan keseimbangan antara kehidupan kerja dan pribadi tenaga kerja.

Selain itu, bagi peneliti selanjutnya, disarankan untuk menambahkan variabel lain seperti waktu tempuh aktual, tingkat kemacetan, dan kondisi lingkungan perjalanan agar analisis produktivitas tenaga kerja dapat dilakukan secara lebih komprehensif.

## **BAB VIII**

### **KESIMPULAN DAN SARAN**

#### **8.1 Kesimpulan**

Berdasarkan hasil implementasi, pengujian, serta analisis yang telah dilakukan dalam penelitian ini, maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Penelitian ini berhasil membangun model klasifikasi untuk memprediksi pilihan moda transportasi tenaga kerja berdasarkan variabel biaya perjalanan, jarak tempuh, dan jenis moda transportasi. Algoritma yang digunakan dalam penelitian ini adalah Logistic Regression dan Random Forest, yang diimplementasikan pada data survei tenaga kerja di Jakarta.
2. Hasil evaluasi kinerja model menunjukkan bahwa Logistic Regression dan Random Forest memiliki performa yang setara, dengan tingkat akurasi sebesar 84,6%. Nilai precision, recall, dan F1-score yang dihasilkan menunjukkan bahwa kedua model mampu mengenali pola data dengan baik, terutama pada kelas mayoritas, meskipun performa pada kelas minoritas masih perlu ditingkatkan.
3. Berdasarkan hasil analisis feature importance, jarak tempuh merupakan variabel yang paling berpengaruh terhadap produktivitas tenaga kerja, diikuti oleh moda transportasi, sedangkan biaya perjalanan memiliki pengaruh yang relatif lebih rendah. Temuan ini menunjukkan bahwa efisiensi waktu dan kenyamanan perjalanan memiliki peranan yang lebih besar dalam memengaruhi produktivitas tenaga kerja dibandingkan faktor biaya semata.
4. Meskipun penelitian ini menggunakan jumlah data uji yang terbatas, hasil analisis menunjukkan bahwa model klasifikasi yang dibangun mampu menangkap pola hubungan antara karakteristik perjalanan kerja dan produktivitas tenaga kerja. Oleh karena itu, hasil penelitian ini dapat dijadikan referensi awal dalam perencanaan transportasi dan perumusan strategi peningkatan efisiensi perjalanan kerja untuk mendukung produktivitas tenaga kerja di Jakarta.

#### **8.2 Saran**

Berdasarkan penelitian yang sudah dilakukan, berikut merupakan saran yang dapat diberikan selama penggeraan

1. Pengembangan Model: Untuk meningkatkan performa prediksi, disarankan menambah jumlah data survei atau menggunakan teknik penyeimbangan data agar model lebih optimal pada semua kategori moda transportasi.
2. Pemanfaatan Hasil: Perusahaan dan pemerintah daerah dapat memanfaatkan prediksi model untuk merancang kebijakan transportasi dan strategi efisiensi perjalanan, sehingga produktifitas tenaga kerja dapat ditingkatkan.
3. Analisis Lanjutan: Disarankan untuk menambahkan variabel lain yang berpotensi mempengaruhi pemilihan moda transportasi, seperti waktu tempuh, kepadatan lalu lintas, atau preferensi pribadi, agar analisis menjadi lebih komprehensif.

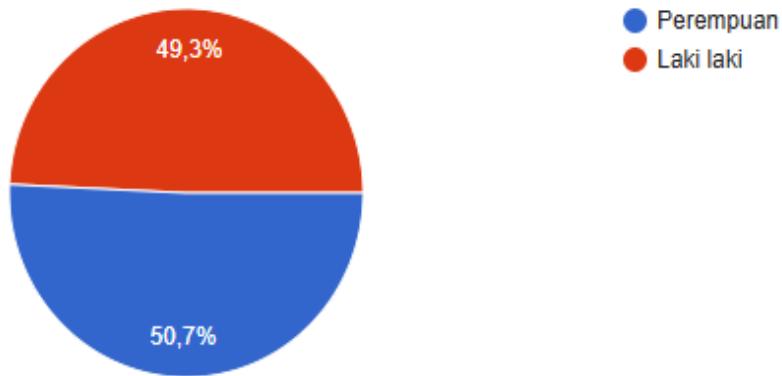
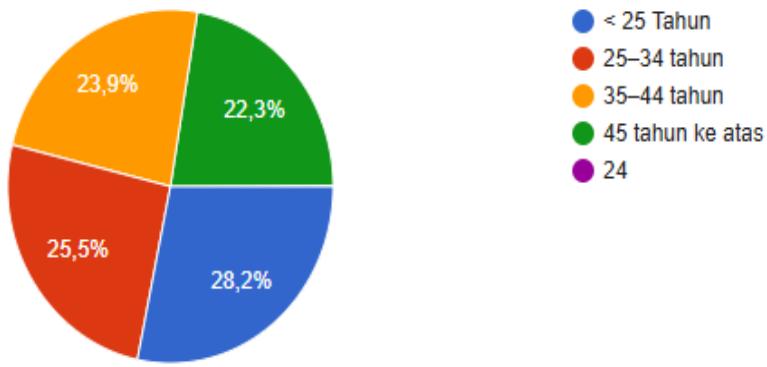
### Pembagian Tugas Anggota

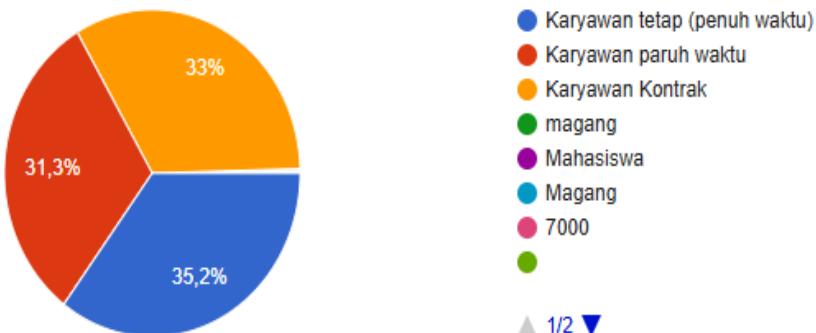
<b>Nama Anggota Kelompok</b>	<b>Peran</b>	<b>Tugas</b>
Yosevyn Sipahutar	Data Collector Data Engineer	<ul style="list-style-type: none"> <li>- Mengumpulkan dataset dari sumber terpercaya (misalnya Kaggle, website resmi, API).</li> <li>- Memastikan data sesuai dengan tujuan proyek.</li> <li>- Melakukan data cleaning (missing value, duplikasi, outlier).</li> <li>- Preprocessing data (normalisasi, encoding, transformasi).</li> </ul>
Luana Banjarnahor	Data Analyst ML Engineer	<ul style="list-style-type: none"> <li>- Melakukan Exploratory Data Analysis (EDA).</li> <li>- Menyusun insight dari hasil analisis.</li> <li>- Membangun model Machine Learning yang sesuai.</li> <li>- Melakukan evaluasi model (akurasi, precision, recall, dll).</li> </ul>
Albert Rapel	Dashboard Developer Document & Presenter	<ul style="list-style-type: none"> <li>- Menyajikan hasil analisis secara informatif dan mudah dipahami.</li> <li>- Menyusun laporan/dokumen proyek (latar belakang, metodologi, hasil, kesimpulan).</li> <li>- Menjadi presenter saat presentasi proyek.</li> </ul>

## Lampiran

## **Lampiran 1 : Jawaban responden terhadap pertanyaan GForm**

	A	B	C	D	E	F	G	H
	Timestamp	Berapa usia Anda saat ini?	Jenis kelamin Anda	Status pekerjaan Anda saat ini	Berapa lama Anda telah bekerja di perusahaan	Berapa rata-rata biaya transportasi harian Anda?	Seberapa besar pengaruh biaya transportasi?	Apakah kantor Anda menyediakan
3	04/12/2025 10:45:51	< 25 Tahun	Laki-laki	Karyawan tetap (penuh waktu)	< 1 tahun	Rp20.000 - Rp50.000		4
4	04/12/2025 10:49:36	< 25 Tahun	Perempuan	Karyawan Kontrak	1-3 tahun	< Rp20.000		4
5	04/12/2025 11:05:14	35-44 tahun	Laki-laki	Karyawan Kontrak	4-6 tahun	Rp20.000 - Rp50.000		4
6	04/12/2025 13:36:34	25-34 tahun	Laki-laki	Karyawan Kontrak	4-6 tahun	< Rp20.000		3
7	04/12/2025 13:42:08	< 25 Tahun	Laki-laki	Karyawan tetap (penuh waktu)	< 1 tahun	Rp20.000 - Rp50.000		4
8	04/12/2025 14:04:32	25-34 tahun	Perempuan	Karyawan tetap (penuh waktu)	1-3 tahun	Rp50.000 - Rp100.000		4
9	05/12/2025 12:36:52	< 25 Tahun	Perempuan	Karyawan tetap (penuh waktu)	4-6 tahun	Rp50.000 - Rp100.000		4
10	05/12/2025 18:07:47	45 tahun ke atas	Laki-laki	Karyawan tetap (penuh waktu)	> 6 tahun	Rp20.000 - Rp50.000		5 Tidak
11	05/12/2025 18:11:21	25-34 tahun	Laki-laki	Karyawan tetap (penuh waktu)	4-6 tahun	Rp20.000 - Rp50.000		5 Tidak
12	05/12/2025 18:12:53	< 25 Tahun	Laki-laki	Karyawan Kontrak	< 1 tahun	Rp20.000 - Rp50.000		5 Tidak
13	05/12/2025 18:14:29	25-34 tahun	Perempuan	Karyawan paruh waktu	< 1 tahun	< Rp20.000		3 Ya
14	05/12/2025 18:16:48	25-34 tahun	Perempuan	Karyawan tetap (penuh waktu)	4-6 tahun	Rp50.000 - Rp100.000		4 Tidak
15	05/12/2025 18:21:58	< 25 Tahun	Laki-laki	Karyawan tetap (penuh waktu)	4-6 tahun	Rp20.000 - Rp50.000		4 Ya
16	05/12/2025 18:23:26	25-34 tahun	Laki-laki	Karyawan Kontrak	< 1 tahun	Rp20.000 - Rp50.000		4 Tidak
17	05/12/2025 18:24:48	< 25 Tahun	Laki-laki	Karyawan Kontrak	1-3 tahun	< Rp20.000		1 Tidak
18	05/12/2025 18:24:54	35-44 tahun	Perempuan	Karyawan tetap (penuh waktu)	> 6 tahun	Rp20.000 - Rp50.000		4 Tidak
19	05/12/2025 18:28:05	25-34 tahun	Perempuan	Karyawan tetap (penuh waktu)	4-6 tahun	Rp50.000 - Rp100.000		3 Ya
20	05/12/2025 18:30:46	35-44 tahun	Laki-laki	Karyawan tetap (penuh waktu)	> 6 tahun	Rp50.000 - Rp100.000		3 Tidak
21	05/12/2025 18:33:28	25-34 tahun	Laki-laki	Karyawan tetap (penuh waktu)	> 6 tahun	< Rp20.000		5 Tidak
22	05/12/2025 18:35:26	25-34 tahun	Perempuan	Karyawan tetap (penuh waktu)	4-6 tahun	Rp20.000 - Rp50.000		4 Ya
23	05/12/2025 18:37:27	< 25 Tahun	Perempuan	Karyawan Kontrak	1-3 tahun	< Rp20.000		2 Ya
24	05/12/2025 18:39:36	25-34 tahun	Laki-laki	Karyawan tetap (penuh waktu)	1-3 tahun	Rp50.000 - Rp100.000		4 Ya





▲ 1/2 ▼

