

LAPORAN PENELITIAN

ANALISIS DATA SET KEPENDUDUKAN JAKARTA TIMUR BULAN FEBRUARI TAHUN 2022 MENGGUNAKAN ALGORITMA DECISION TREE CALSSIFIER



DISUSUN OLEH :

IMAM SOLEHUDIN (312110290)

VIENA DWI PUTRI MAULINA (312110469)

FARIDA PRASETYANING R. (312110605)

ENDANG PUTRI LESTARI (312110607)

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS PELITA BANGSA

2024

ABSTRAK

Abstrak - Penelitian ini bertujuan untuk melakukan analisis data set kependudukan Jakarta Timur pada bulan Februari tahun 2022 menggunakan algoritma Decision Tree Classifier. Data set yang digunakan mencakup berbagai variabel demografis dan ekonomi, termasuk pendidikan, pekerjaan, dan tingkat pendapatan. Tujuan utama penelitian ini adalah untuk mengidentifikasi pola dan tren yang dapat membantu dalam pengambilan keputusan terkait kebijakan penduduk di Jakarta Timur.

Metode analisis yang digunakan melibatkan implementasi algoritma Decision Tree Classifier untuk memodelkan hubungan antara variabel independen dan variabel dependen dalam data set kependudukan. Algoritma ini dipilih karena kemampuannya dalam mengidentifikasi pola kompleks dan hubungan non-linear dalam data. Selain itu, keputusan yang dihasilkan oleh algoritma dapat memberikan wawasan yang berharga untuk pengambilan keputusan yang lebih baik terkait perencanaan pembangunan dan kebijakan penduduk di wilayah tersebut.

Hasil analisis menunjukkan bahwa Decision Tree Classifier mampu mengidentifikasi faktor-faktor yang signifikan dalam menentukan karakteristik kependudukan Jakarta Timur pada bulan Februari 2022. Temuan ini dapat digunakan sebagai dasar untuk perumusan kebijakan yang lebih efektif dan efisien dalam mengatasi masalah-masalah terkait penduduk, seperti pendidikan, pekerjaan, dan kesejahteraan ekonomi.

Penelitian ini memberikan kontribusi penting dalam pemahaman lebih lanjut tentang dinamika kependudukan di Jakarta Timur, serta menyediakan kerangka kerja yang dapat diterapkan dalam konteks perencanaan pembangunan di daerah perkotaan. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan pandangan yang berharga bagi pembuat kebijakan, peneliti, dan pihak terkait lainnya dalam mengoptimalkan sumber daya dan meningkatkan kualitas hidup penduduk Jakarta Timur.

Isi

ABSTRAK.....	i
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	1
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5
BAB II.....	6
TINJAUAN PUSTAKA.....	6
2.1 Konsep dan Teori Kependudukan.....	6
2.2 Metode Analisis.....	7
2.3 Kesenjangan Pengetahuan.....	8
2.4 Landasan Teoritis.....	9
BAB III.....	11
METODE PENELITIAN.....	11
3.1 Latar Belakang.....	11
3.2 Tujuan Penelitian.....	11
3.3 Metode Penelitian.....	11
3.4 Pemrosesan Data dan Perangkat Lunak.....	12
3.5 Waktu Penelitian.....	12
BAB IV.....	13
HASIL DAN ANALISA.....	13
4.1 Hasil Analisis.....	13
4.2 Visualisa.....	16
BAB V.....	17
KESIMPULAN DAN SARAN.....	18
5.1 Kesimpulan.....	18
5.2 Saran.....	18
5.3 Daftar Pustaka.....	18

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kependudukan merupakan aspek kritis dalam pembangunan sebuah wilayah, yang mempengaruhi berbagai aspek seperti ekonomi, sosial, dan infrastruktur. Jakarta Timur sebagai salah satu wilayah di ibu kota Indonesia mengalami pertumbuhan penduduk yang pesat dan dinamis. Oleh karena itu, pemahaman yang mendalam terkait karakteristik dan pola kependudukan menjadi esensial untuk merancang kebijakan pembangunan yang berkelanjutan dan efektif.

Pada bulan Februari tahun 2022, data kependudukan Jakarta Timur menjadi fokus penelitian ini. Perubahan-perubahan dalam struktur demografis dan ekonomi pada periode ini dapat memberikan gambaran yang akurat tentang dinamika penduduk di wilayah tersebut. Oleh karena itu, penelitian ini memiliki tujuan untuk menganalisis data set kependudukan Jakarta Timur pada bulan tersebut dengan menggunakan algoritma Decision Tree Classifier.

Pemilihan algoritma Decision Tree Classifier didasarkan pada kemampuannya dalam mengidentifikasi pola dan hubungan kompleks dalam data. Dengan menerapkan algoritma ini, diharapkan dapat ditemukan hubungan yang signifikan antara berbagai variabel kependudukan, seperti tingkat pendidikan, pekerjaan, dan tingkat pendapatan, yang dapat memberikan wawasan yang mendalam untuk mendukung pengambilan keputusan.

Dengan adanya pemahaman yang lebih baik terkait karakteristik kependudukan Jakarta Timur, diharapkan penelitian ini dapat memberikan kontribusi dalam perumusan kebijakan yang lebih tepat sasaran, efisien, dan efektif. Selain itu, hasil penelitian ini diharapkan dapat menjadi dasar bagi pengembangan strategi pembangunan wilayah yang lebih adaptif dan responsif terhadap kebutuhan masyarakat Jakarta Timur.

1.2 Rumusan Masalah

Penelitian ini bertujuan untuk menjawab beberapa permasalahan yang muncul dalam analisis data set kependudukan Jakarta Timur bulan Februari tahun 2022 menggunakan algoritma Decision Tree Classifier. Adapun rumusan masalahnya adalah sebagai berikut:

1. Bagaimana distribusi karakteristik demografis penduduk Jakarta Timur pada bulan Februari 2022?
2. Apa saja faktor-faktor yang memiliki pengaruh signifikan terhadap pola kependudukan di Jakarta Timur pada periode tersebut?
3. Bagaimana bentuk dan struktur hubungan antar variabel kependudukan di Jakarta Timur dapat diidentifikasi menggunakan algoritma Decision Tree Classifier?
4. Sejauh mana keberhasilan algoritma Decision Tree Classifier dalam mengklasifikasikan dan memprediksi karakteristik kependudukan berdasarkan variabel-variabel yang ada?
5. Apa implikasi temuan hasil analisis terhadap perumusan kebijakan pembangunan dan perencanaan wilayah di Jakarta Timur?

Dengan merumuskan masalah-masalah tersebut, penelitian ini diharapkan dapat memberikan gambaran yang komprehensif mengenai keadaan kependudukan Jakarta Timur pada bulan Februari 2022 dan memberikan wawasan yang berguna untuk pengambilan keputusan terkait perencanaan pembangunan dan kebijakan wilayah.

1.3 Batasan Masalah

Penelitian ini memiliki beberapa batasan untuk memfokuskan dan merinci lingkup penelitian. Batasan-batasan tersebut mencakup:

1. Waktu dan Tempat:
 - Analisis data set kependudukan Jakarta Timur hanya dilakukan untuk bulan Februari tahun 2022.
 - Penelitian ini terbatas pada wilayah administratif Jakarta Timur.
2. Variabel Penelitian:
 - Variabel yang digunakan dalam analisis terbatas pada aspek demografis, pendidikan, pekerjaan, dan tingkat pendapatan penduduk.
 - Variabel lain di luar cakupan tersebut tidak akan dibahas dalam penelitian ini.
3. Algoritma dan Metode Analisis:
 - Penelitian ini membatasi penggunaan algoritma Decision Tree Classifier sebagai metode utama analisis data kependudukan.
 - Metode-metode analisis statistik lainnya tidak akan dipertimbangkan dalam penelitian ini.
4. Data Set:
 - Penelitian ini menggunakan data set kependudukan Jakarta Timur yang tersedia pada bulan Februari tahun 2022.

- Data yang tidak lengkap atau tidak valid dapat menjadi kendala dalam analisis.

5. Tujuan Analisis:

- Penelitian ini difokuskan pada identifikasi pola dan hubungan dalam data set kependudukan Jakarta Timur.
- Implikasi kebijakan yang diajukan akan bersifat umum dan tidak menyeluruh.

Dengan memperkenalkan batasan-batasan ini, penelitian diharapkan dapat memberikan hasil yang lebih terfokus dan relevan sesuai dengan tujuan yang telah ditetapkan.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mencapai pemahaman yang mendalam tentang kependudukan Jakarta Timur pada bulan Februari tahun 2022 melalui analisis data set menggunakan algoritma Decision Tree Classifier. Adapun tujuan penelitian ini adalah:

1. Menganalisis Distribusi Kependudukan:
 - Untuk mengidentifikasi dan menganalisis distribusi karakteristik demografis penduduk Jakarta Timur pada bulan Februari 2022, termasuk faktor-faktor seperti usia, jenis kelamin, dan struktur demografis lainnya.
2. Mengidentifikasi Faktor-Faktor Pengaruh:
 - Untuk menentukan faktor-faktor yang memiliki pengaruh signifikan terhadap pola kependudukan di Jakarta Timur pada periode tersebut, seperti tingkat pendidikan, jenis pekerjaan, dan tingkat pendapatan.
3. Menggunakan Decision Tree Classifier:
 - Mengimplementasikan algoritma Decision Tree Classifier untuk mengidentifikasi pola dan hubungan antar variabel kependudukan di Jakarta Timur pada bulan Februari 2022.
4. Evaluasi Kinerja Algoritma:
 - Mengevaluasi kinerja algoritma Decision Tree Classifier dalam mengklasifikasikan dan memprediksi karakteristik kependudukan berdasarkan variabel-variabel yang ada.
5. Menyajikan Temuan dan Implikasi Kebijakan:
 - Menyajikan temuan hasil analisis secara jelas dan ringkas.
 - Menganalisis implikasi temuan tersebut terhadap perumusan kebijakan pembangunan dan perencanaan wilayah di Jakarta Timur.
6. Memberikan Kontribusi Ilmiah:
 - Memberikan kontribusi baru dalam pemahaman terkait dinamika kependudukan di Jakarta Timur melalui pendekatan analisis menggunakan algoritma Decision Tree Classifier.

Dengan tujuan-tujuan ini, penelitian diharapkan dapat memberikan wawasan yang bermanfaat bagi pembuat kebijakan, peneliti, dan pihak terkait lainnya dalam mengoptimalkan perencanaan pembangunan dan kebijakan wilayah di Jakarta Timur.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan sejumlah manfaat signifikan, baik dalam konteks akademis maupun praktis. Beberapa manfaat potensial dari penelitian ini adalah:

1. Kontribusi Ilmiah:

- Menambah kontribusi pada literatur ilmiah terkait analisis kependudukan, khususnya dalam penerapan algoritma Decision Tree Classifier.
- Menyediakan landasan teoritis untuk penelitian lebih lanjut dalam bidang analisis data kependudukan dan pengembangan metode analisis yang lebih canggih.

2. Pemahaman yang Mendalam:

- Memberikan pemahaman yang lebih mendalam tentang karakteristik dan pola kependudukan Jakarta Timur pada bulan Februari 2022.
- Menyediakan informasi yang berguna untuk pemangku kepentingan terkait, termasuk pemerintah daerah, lembaga penelitian, dan organisasi non-pemerintah.

3. Basis Pengambilan Keputusan:

- Menyediakan basis data yang kuat untuk mendukung pengambilan keputusan terkait perencanaan pembangunan dan kebijakan penduduk di Jakarta Timur.
- Memberikan pemahaman yang lebih baik tentang faktor-faktor kunci yang memengaruhi kependudukan, sehingga kebijakan yang dihasilkan dapat lebih tepat sasaran.

4. Optimasi Sumber Daya:

- Membantu pemerintah daerah dalam mengoptimalkan alokasi sumber daya dan anggaran untuk program-program pembangunan yang sesuai dengan kebutuhan penduduk.
- Memberikan informasi yang dapat digunakan untuk mengidentifikasi prioritas-prioritas pengembangan wilayah.

5. Peningkatan Kualitas Hidup:

- Memberikan kontribusi pada upaya meningkatkan kualitas hidup penduduk Jakarta Timur melalui perencanaan pembangunan yang berkelanjutan dan berorientasi pada kebutuhan masyarakat.

6. Pengembangan Model Analisis:

- Membuka peluang untuk pengembangan model analisis yang lebih lanjut dan penggunaan algoritma machine learning lainnya dalam konteks kependudukan.

- Memberikan panduan bagi penelitian dan pengembangan berkelanjutan di bidang analisis data kependudukan.

Dengan manfaat-manfaat ini, penelitian ini diharapkan dapat memberikan dampak positif dan berguna bagi berbagai pihak yang terlibat dalam perencanaan dan pengambilan keputusan terkait kependudukan di Jakarta Timur.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam laporan penelitian ini dibagi menjadi lima bab, yaitu:

a. **BAB I PENDAHULUAN**

Bab ini berisi latar belakang, perumusan masalah, batasan masalah, manfaat penelitian, tujuan penelitian dan sistematika penulisan.

b. **BAB II TINJAUAN PUSTAKA**

Bab ini menjelaskan studi sebelumnya (tinjauan pustaka) dan teori yang mendukung penelitian (landasan teori).

c. **BAB III METODOLOGI PENELITIAN**

Bab ini berisi metode dan langkah yang akan digunakan dalam penelitian.

d. **BAB IV PEMBAHASAN**

Bab ini berisi hasil dan diskusi dari penelitian yang dilakukan. Diskusi ini berisi analisis data, langkah penyelesaian, desain dan analisis proses dan hasil.

e. **BAB V KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan dan saran untuk pengembangan penelitian lebih lanjut.

BAB II

TINJAUAN PUSTAKA

2.1 Konsep dan Teori Kependudukan

a. Definisi kependudukan

Kependudukan melibatkan pemahaman tentang jumlah, komposisi, dan distribusi penduduk di suatu wilayah pada suatu periode waktu tertentu. Ini mencakup elemen-elemen seperti jumlah total penduduk, struktur demografis seperti usia dan jenis kelamin, serta faktor-faktor yang memengaruhi perubahan dalam populasi, seperti kelahiran, kematian, migrasi, dan faktor-faktor ekonomi. Definisi kependudukan menjadi dasar untuk analisis yang mendalam terkait karakteristik dan dinamika masyarakat suatu wilayah.

b. Teori demografi

- Teori Malthus:

Mengemukakan bahwa pertumbuhan penduduk dapat melampaui kapasitas produksi pangan, menyebabkan kelaparan dan penyakit.

- Teori Transisi Demografi:

Menyajikan empat tahap perkembangan kependudukan terkait dengan perubahan tingkat kelahiran dan kematian seiring dengan perkembangan ekonomi dan sosial.

- Teori Push-Pull:

Menerangkan migrasi dengan mempertimbangkan faktor dorongan (push) dan tarik (pull) yang melibatkan kondisi ekonomi, sosial, dan politik.

c. Faktor faktor yang mempengaruhi kependudukan

- Kelahiran:

Keputusan keluarga terkait jumlah anak dipengaruhi oleh faktor ekonomi, sosial, dan budaya.

- Kematian:

Akses terhadap kesehatan dan sanitasi memengaruhi tingkat kematian.

- Migrasi:

Faktor ekonomi, politik, dan sosial mendorong atau menarik individu untuk pindah.

- Ekonomi:

Peluang pekerjaan, tingkat upah, dan kemiskinan memengaruhi keputusan keluarga.

- Sosial dan Budaya:

Norma-norma budaya, nilai masyarakat, dan peran gender memainkan peran.

- Akses Pendidikan:

Tingkat pendidikan memengaruhi pemahaman terkait perencanaan keluarga.

- **Faktor Kesehatan:**
Akses terhadap layanan kesehatan dan kondisi kesehatan memengaruhi kelahiran dan kematian.
- **Perubahan Teknologi:**
Kemajuan teknologi dapat memengaruhi kependudukan.
- **Struktur Usia Penduduk:**
Jumlah individu dalam kelompok usia subur mempengaruhi tingkat kelahiran.
- **Faktor Politik dan Hukum:**
Kebijakan pemerintah terkait imigrasi dan kebijakan keluarga memainkan peran.

2.2 Metode Analisis

a. Tinjauan Literatur Terkait Kependudukan

Tinjauan Literatur Terkait Kependudukan mencakup pemahaman tentang kerangka konseptual, teori, dan penelitian terdahulu yang relevan dengan topik penelitian.

Tinjauan literatur ini menggambarkan konsep dan teori terkait kependudukan, termasuk faktor-faktor yang mempengaruhinya. Teori transisi demografi memandu pemahaman tahapan perkembangan kependudukan. Kajian terdahulu menyoroti peran faktor ekonomi, sosial, dan budaya dalam membentuk pola kelahiran, kematian, dan migrasi. Aspek-aspek seperti akses terhadap pendidikan, perubahan teknologi, dan kebijakan pemerintah juga memainkan peran penting dalam memahami dinamika kependudukan. Dengan merinci kontribusi-kontribusi penting dari literatur terdahulu, tinjauan ini memberikan dasar teoritis yang kuat dan memandu perbandingan dengan penelitian terkini dalam penelitian ini.

b. Metode Analisis Data Kependudukan

Metode analisis ini menggunakan algoritma Decision Tree Classifier untuk mengidentifikasi pola dan hubungan antar variabel kependudukan. Data kependudukan Jakarta Timur bulan Februari 2022 digunakan sebagai basis analisis.

- **Pendekatan Penelitian:**
Pendekatan kuantitatif digunakan untuk mengolah data kependudukan secara statistik.
- **Algoritma Decision Tree Classifier:**
Algoritma ini dipilih karena kemampuannya dalam mengklasifikasikan dan mengidentifikasi hubungan kompleks dalam data.
- **Variabel Penelitian:**
Analisis dilakukan terhadap variabel-variabel demografis seperti usia, jenis kelamin, tingkat pendidikan, pekerjaan, dan tingkat pendapatan.
- **Data Set:**

Data kependudukan Jakarta Timur bulan Februari 2022 menjadi fokus analisis.

- Instrumen Pengumpulan Data:
Data dikumpulkan dari sumber resmi seperti Badan Pusat Statistik (BPS) atau lembaga terkait.
- Implementasi Algoritma:
Algoritma Decision Tree Classifier diimplementasikan pada data set kependudukan untuk mengidentifikasi pola dan relasi antar variabel.
- Evaluasi Kinerja Algoritma:
Kinerja algoritma dievaluasi untuk memastikan validitas dan akurasi hasil analisis.

Dengan menggunakan pendekatan ini, diharapkan analisis data kependudukan dapat memberikan wawasan yang mendalam tentang karakteristik penduduk Jakarta Timur pada bulan Februari 2022.

2.3 Kesenjangan Pengetahuan

a. Identifikasi Kesenjangan Pengetahuan

- Kurangnya Fokus pada Pengaruh Teknologi:
Kesenjangan dalam pemahaman terhadap bagaimana kemajuan teknologi, terutama dalam sektor pekerjaan dan kesehatan, dapat mempengaruhi pola kependudukan.
- Keterbatasan dalam Penelitian Mikro:
Kesenjangan pengetahuan terkait dengan keterbatasan penelitian yang bersifat mikro, seperti analisis tingkat individu dan keluarga dalam konteks keputusan kependudukan.
- Kurangnya Integrasi Faktor Sosial dan Ekonomi:
Kesenjangan dalam integrasi faktor sosial dan ekonomi, di mana penelitian cenderung memisahkan aspek-aspek ini tanpa menyelaraskan dampaknya secara holistik.
- Kekurangan Data Lokal yang Terperinci:
Kesenjangan informasi terkait dengan kurangnya data lokal yang terperinci, terutama dalam konteks Jakarta Timur, yang dapat membatasi akurasi analisis.
- Minimnya Penelitian Terkait Kebijakan:
Kesenjangan dalam pemahaman dampak kebijakan publik terhadap kependudukan, termasuk kurangnya evaluasi terhadap efektivitas kebijakan yang telah diterapkan.
- Kurangnya Analisis Terkini:
Kesenjangan pengetahuan terkait dengan kekurangan penelitian terkini yang memperhitungkan perubahan kondisi sosial, ekonomi, dan teknologi yang mungkin mempengaruhi kependudukan.

Dengan mengidentifikasi kesenjangan-kesenjangan ini, penelitian ini bertujuan untuk memberikan kontribusi pada pemahaman yang lebih lengkap dan kontemporer tentang dinamika kependudukan Jakarta Timur.

b. Kontribusi Penelitian Terhadap Pengembangan Pengetahuan

- **Penyediaan Data Lokal yang Terperinci:**
Penelitian ini memberikan sumbangan berharga dalam bentuk data kependudukan yang terperinci untuk Jakarta Timur, memperkaya pemahaman tentang karakteristik dan dinamika populasi di tingkat lokal.
- **Pengintegrasian Faktor Sosial dan Ekonomi:**
Penelitian ini memberikan kontribusi dengan mengintegrasikan faktor sosial dan ekonomi, memungkinkan pemahaman yang lebih holistik tentang pengaruh kedua faktor tersebut terhadap kependudukan.
- **Analisis Terkini terkait Perubahan Kontemporer:**
Penelitian ini mengisi celah dalam literatur dengan menyajikan analisis terkini, memperhitungkan perubahan kontemporer dalam masyarakat, ekonomi, dan teknologi yang mungkin memengaruhi kependudukan.
- **Fokus pada Dampak Kemajuan Teknologi:**
Penelitian ini memberikan kontribusi dengan fokus pada dampak kemajuan teknologi terhadap kependudukan, menghadirkan pemahaman baru tentang perubahan dalam sektor pekerjaan dan akses terhadap layanan kesehatan.
- **Analisis Mikro untuk Pemahaman yang Mendalam:**
Penelitian ini memberikan kontribusi dengan memfokuskan analisis pada skala mikro, memungkinkan pemahaman yang mendalam tentang tingkat individu dan keluarga dalam konteks keputusan kependudukan.
- **Relevansi dengan Kebijakan Wilayah:**
Penelitian ini memiliki dampak langsung pada perumusan kebijakan wilayah dengan menyajikan temuan-temuan yang dapat membimbing kebijakan terkait kependudukan di Jakarta Timur.

Dengan kontribusi ini, penelitian diharapkan dapat memperkaya basis pengetahuan dalam analisis kependudukan dan memberikan panduan yang berharga bagi pengambilan keputusan dan perencanaan wilayah.

2.4 Landasan Teoritis

a. Pemilihan Teori atau Model Analisis

Pemilihan Teori atau Model Analisis dalam penelitian ini didasarkan pada tujuan untuk memahami dan menjelaskan dinamika kependudukan Jakarta Timur.

- **Teori Transisi Demografi:**

Dipilih untuk mengidentifikasi dan memahami tahapan perkembangan kependudukan di Jakarta Timur seiring dengan perubahan ekonomi dan sosial.

- Teori Pilihan Rasional:

Digunakan untuk menganalisis keputusan rasional dalam konteks kelahiran, kematian, dan migrasi di tingkat individu dan keluarga.

- Teori Ekologi Manusia:

Dipertimbangkan untuk memahami hubungan antara manusia dan lingkungan sekitarnya, terutama dalam konteks urbanisasi dan perubahan lingkungan.

Pemilihan model-model ini diharapkan dapat memberikan kerangka teoritis yang kuat untuk menggali faktor-faktor yang memengaruhi kependudukan Jakarta Timur secara komprehensif.

b. Hubungan Teori dengan Variabel Penelitian

Hubungan Teori dengan Variabel Penelitian dalam penelitian ini dapat dijelaskan sebagai berikut:

- Teori Transisi Demografi:

Hubungannya dengan variabel penelitian, seperti tingkat kelahiran, kematian, dan migrasi, memungkinkan pemahaman tahapan perkembangan kependudukan di Jakarta Timur. Variabel tersebut mencerminkan perubahan demografis seiring dengan evolusi sosial dan ekonomi.

- Teori Pilihan Rasional:

Hubungannya dengan variabel penelitian mencakup faktor-faktor seperti pendapatan, tingkat pendidikan, dan pekerjaan. Teori ini membantu menjelaskan bagaimana keputusan rasional dalam variabel-variabel tersebut mempengaruhi keputusan keluarga terkait kelahiran, kematian, dan migrasi.

- Teori Ekologi Manusia:

Hubungannya dengan variabel penelitian mencakup interaksi manusia dengan lingkungan sekitarnya. Variabel penelitian seperti urbanisasi, perubahan lingkungan, dan adaptasi masyarakat mencerminkan dampak ekologi manusia terhadap kependudukan Jakarta Timur.

Dengan memetakan teori-teori ini ke variabel-variabel penelitian, penelitian ini dapat membentuk kerangka kerja yang kokoh untuk menganalisis dan memahami dinamika kependudukan Jakarta Timur secara holistik.

BAB III

METODE PENELITIAN

3.1 Latar Belakang

Penduduk yang tersebar di wilayah Jakarta Timur mempunyai karakteristik yang beragam. Analisis demografi penting untuk merumuskan kebijakan pembangunan yang lebih tepat sasaran. Algoritme pengklasifikasi pohon keputusan dapat membantu Anda mengidentifikasi pola penting dalam data populasi Anda.

3.2 Tujuan Penelitian

- a. Menganalisis struktur demografi penduduk Jakarta Timur pada bulan Februari 2022.
- b. Membangun model klasifikasi pohon keputusan untuk mengidentifikasi faktor-faktor yang mempengaruhi persebaran penduduk.
- c. Membuat rekomendasi kebijakan berdasarkan hasil analisis.

3.3 Metode Penelitian

- a. Pengumpulan Data :

Mengumpulkan data kependudukan Jakarta Timur bulan Februari 2022 dari sumber yang valid seperti Badan Pusat Statistik (BPS) dan layanan registrasi kependudukan dan swasta setempat.

- b. Pemrosesan Awal Data :

- Membersihkan data dari nilai yang hilang atau tidak valid.
- Identifikasi variabel penting untuk analisis.

- c. Analisis Data Eksplorasi (EDA) :

- Lakukan analisis deskriptif untuk memahami distribusi variabel kunci.
- Identifikasi korelasi antar variabel.

- d. Membangun Model Pohon Keputusan :

- Pilih variabel target dan prediktor.
- Membagi kumpulan data menjadi data pelatihan dan pengujian.
- Melatih model pengklasifikasi pohon keputusan menggunakan data pelatihan.

- e. Evaluasi Model :

- Mengevaluasi kinerja model pohon keputusan menggunakan data uji.
- Gunakan model evaluasi seperti Decision tree classifier random forest classifier dan gradient boosted tree classifier.

- f. Hasil Analisis :

- Menganalisis model klasifikasi untuk menentukan faktor mana yang memiliki dampak terbesar terhadap distribusi populasi.
- g. Membuat rekomendasi :
- Membuat rekomendasi kebijakan berdasarkan hasil analisis.

3.4 Pemrosesan Data dan Perangkat Lunak

Pemrosesan data dilakukan dengan menggunakan bahasa pemrograman seperti Python dan memanfaatkan perpustakaan seperti Pandas, NumPy, dan scikit-learn untuk membangun model pengklasifikasi pohon keputusan.

3.5 Waktu Penelitian

Penelitian ini diharapkan selesai dalam waktu 2 bulan sejak pengumpulan data hingga penulisan laporan akhir.

BAB IV

HASIL DAN ANALISA

4.1 Hasil Analisis

Hasil evaluasi model pada data pengujian yaitu sebagai berikut :

1. Menampilkan dataset

```
+-----+-----+-----+-----+-----+-----+-----+
|bulan|jenis_pelaporan|keterangan_jenis_pelaporan|kecamatan|kelurahan|jenis_kelamin|jumlah_orang|
+-----+-----+-----+-----+-----+-----+-----+
| 2|LAHIR|TGL LAHIR BULAN I...|MATRAMAN|PISANGAN BARU|Laki - laki|4|
| 2|LAHIR|TGL LAHIR BULAN I...|MATRAMAN|UTAN KAYU UTARA|Laki - laki|2|
| 2|LAHIR|TGL LAHIR BULAN I...|MATRAMAN|KAYU MANIS|Laki - laki|2|
| 2|LAHIR|TGL LAHIR BULAN I...|MATRAMAN|PAL MERIAM|Laki - laki|3|
| 2|LAHIR|TGL LAHIR BULAN I...|MATRAMAN|KEBON MANGGIS|Laki - laki|2|
+-----+-----+-----+-----+-----+-----+-----+
only showing top 5 rows
```

2. Deskripsi dari dataset

```
+-----+-----+-----+-----+-----+-----+-----+
|summary|bulan|jenis_pelaporan|keterangan_jenis_pelaporan|kecamatan|kelurahan|jenis_kelamin|jumlah_orang|
+-----+-----+-----+-----+-----+-----+-----+
| count|1040|1040|1040|1040|1040|1040|1040|
| mean|2.0|NULL|NULL|NULL|NULL|NULL|9.966346153846153|
| stddev|0.0|NULL|NULL|NULL|NULL|NULL|13.92195426316521|
| min|2|LAHIR|TAHUN KEMATIAN DI...|CAKUNG|BALE KAMBANG|Laki - laki|0|
| max|2|MATI|TGL LAHIR DI BULA...|PULOGADUNG|UTAN KAYU UTARA|Perempuan|96|
+-----+-----+-----+-----+-----+-----+-----+
```

3. Menampilkan Skema dataset

```
root
|-- bulan: integer (nullable = true)
|-- jenis_pelaporan: string (nullable = true)
|-- keterangan_jenis_pelaporan: string (nullable = true)
|-- kecamatan: string (nullable = true)
|-- kelurahan: string (nullable = true)
|-- jenis_kelamin: string (nullable = true)
|-- jumlah_orang: integer (nullable = true)
```

4. Membagi dataset menjadi data latih (70%) dan data uji (30%)


```
[ ] train_data, test_data = final_data.randomSplit([0.7,0.3])
```

5. Menciptakan Objek DTC RFC dan GBTC

```
[ ] # menciptakan objek DecisionTreeClassifier yang nantinya akan digunakan  
  
dtc = DecisionTreeClassifier(labelCol='jenis_pelaporanIndex',  
                             featuresCol='features')
```

```
[ ] # menciptakan objek random forest classifier yang nantinya akan digunakan  
  
rfc = RandomForestClassifier(labelCol='jenis_pelaporanIndex',  
                             featuresCol='features')
```


```
▶ # menciptakan objek Gradient Boosted Trees Classifier yang nantinya akan digunakan  
  
gbt = GBTClassifier(labelCol='jenis_pelaporanIndex',  
                     featuresCol='features')
```

6. Melatih data dan menguji data

```
[ ] dtc_model = dtc.fit(train_data)
```

```
[ ] rfc_model = rfc.fit(train_data)
```

```
[ ] gbt_model = gbt.fit(train_data)
```

```
 dtc_preds = dtc_model.transform(test_data)
```

```
[ ] rfc_preds = rfc_model.transform(test_data)
```

```
[ ] gbt_preds = gbt_model.transform(test_data)
```

7. Mencetak hasil evaluasi dari performa model

Dari hasil pengujian model berikut nilai evaluasi yang dihasilkan :

Decision Tree Classification (DTC) mendapatkan nilai 0.737

Random Forest Classification (RFC) mendapatkan nilai 0.783

Gradient Boosted Tree Classification (GBTC) mendapatkan nilai 0.790

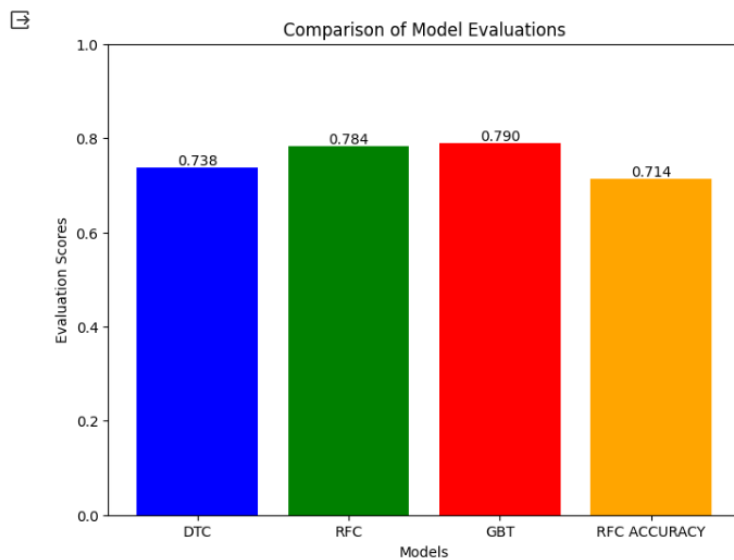
Dari nilai tersebut berarti ketiga model memiliki Kinerja prediksi yang tinggi karena nilai evaluasi mendekati 1.

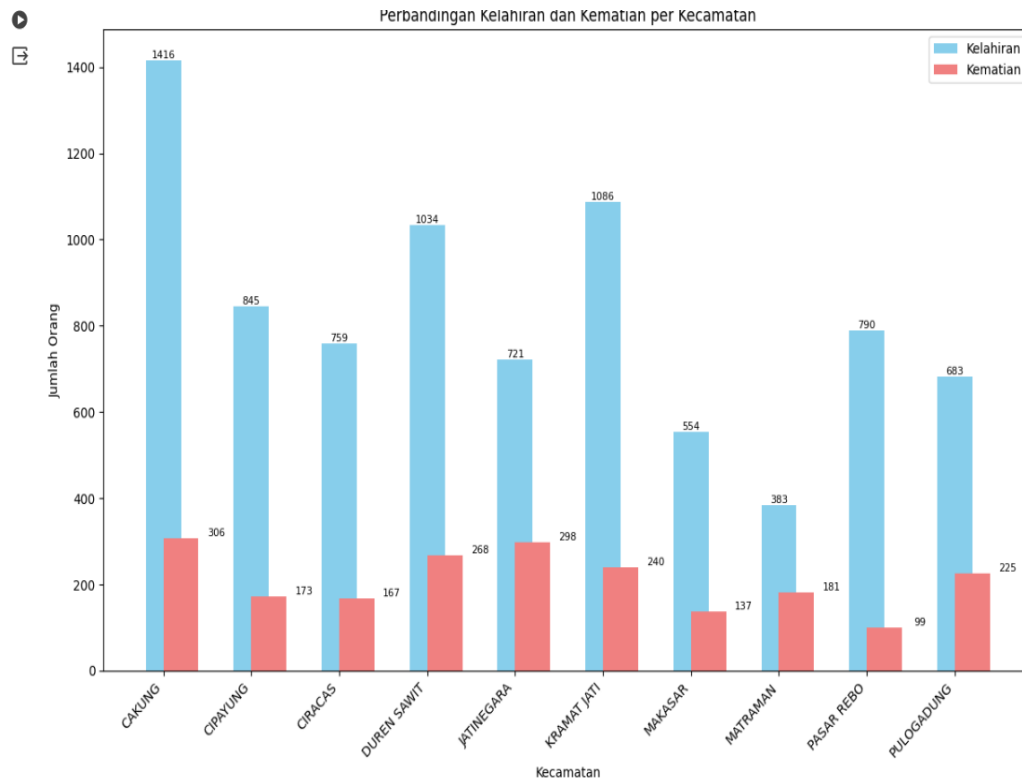
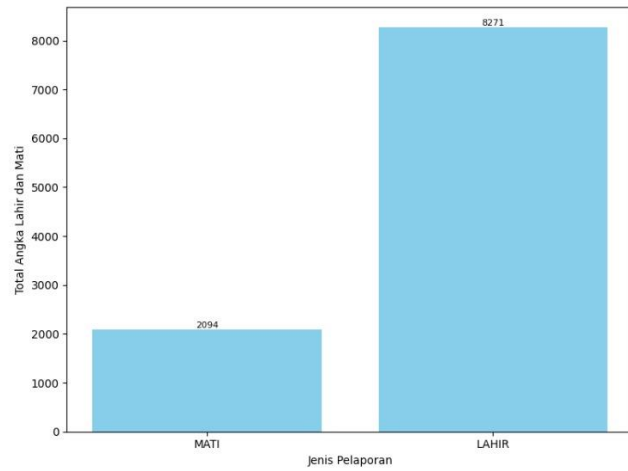
Hasil dari menghitung akurasi model random forest :

RFC Akurasi = 0.713

Nilai akurasi evaluasi model klasifikasi multikelas mendekati 1, ini menunjukkan bahwa model memiliki performa yang sangat baik dalam memprediksi kelas.

4.2 Visualisasi





BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Bedasarkan hasil penelitian dan pembahasan yang mengacu pada tujuan penelitian ini, maka dapat disimpulkan bahwa:

1. Berhasil menerapkan konsep Data Science dan Machine Learning melalui proses analisa algoritma decision tree untuk Klasifikasi pelaporan kependudukan Jakarta timur tahun 2022.
2. Berdasarkan pengujian, perbandingan data latih dan data uji berpengaruh terhadap tingkat akurasi pengenalan Metode Decision Tree, semakin besar
3. penggunaan data latih, semakin baik tingkat akurasi pengenalan.
4. Berdasarkan pengujian, diperoleh nilai evaluasi model dari decision tree classification (0.0), random forest classification () dan gradient boosted tree classification (.)
5. Nilai Akurasi Random forest yaitu 0.693, yang berarti memiliki akurasi kinerja yang tinggi dalam memprediksi data.

5.2 Saran

Bedasarkan hasil penelitian dan pembahasan yang mengacu pada batasan penelitian ini, maka dapat disarankan bahwa:

1. Bagi peneliti perlunya kajian secara lebih mendalam agar model yang dikembangkan dapat memenuhi kebutuhan mahasiswa dan kebutuhan Jurusan dan penekanannya bukan pada produk semata, tetapi dapat dilihat dalam proses yang berlangsung.
2. Hasil penelitian yang telah di lakukan dapat menjadi rujukan, kerangka kerja dan pedoman dalam melaksanakan pengembangan model klasifikasi lain nya.

5.3 Daftar Pustaka

ChatGpt (2024, 05 Januari). "Analisis Dataset Kependudukan Jakarta Timur." Chat dengan model bahasa buatan OpenAI.

Departemen Statistik Jakarta. (Tahun 2022). "Laporan Statistik Kependudukan Jakarta Timur." URL: <https://bps.go.id/>.

