

Penerapan Algoritma *K-Nearest Neighbor* Menggunakan *RapidMiner* Pada Kepuasan Hidup Pekerja Commuter di Indonesia

¹Bayu Pangestu, ²Bintang Ardian Nugroho, ³Dastin Ramadhani, ⁴Muhammad Fadli Juliana, ⁵Sopyan Hidayat, ⁶Maulana Fansyuri

¹Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

²Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

³Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

⁴Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

⁵Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

⁶Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

¹bayupangestu395@gmail.com, ²bintangardian83@gmail.com, ³ramadhani.dastin0838@gmail.com,
⁴fadliijuliana@gmail.com, ⁵Sopyanhidayat80@gmail.com, ⁶dosen02359@unpam.ac.id

Abstract

The level of life satisfaction of commuter workers in Indonesia is classified using the K-Nearest Neighbor (K-NN) algorithm using the RapidMiner application. This study aims to provide a better understanding of the social and economic conditions of workers who have to travel long distances every day. To collect data, a questionnaire covering various information such as income, number of dependents, location of residence, travel time, and level of life satisfaction was sent. Before being entered into the model, the data is then processed through a cleaning stage, normalizing numeric values, and dividing into test data and training data. One of the reasons for RapidMiner is its visual interface, which allows users to create classification models without writing programming code. The test results show that the K-NN algorithm can accurately classify the level of life satisfaction of commuter workers. Model performance is greatly influenced by the selected variables, namely the K value, and data quality. This study is expected to help related parties, this approach is considered effective in helping data-based decision making.

Keywords: *K-Nearest Neighbor, Life Satisfaction, Commuter, Classification, Rapidminer*

Abstrak

Tingkat kepuasan hidup pekerja komuter di Indonesia diklasifikasikan menggunakan algoritma *K-Nearest Neighbor* (K-NN) menggunakan aplikasi *RapidMiner*. Studi ini bertujuan untuk memberikan pemahaman yang lebih baik tentang kondisi sosial dan ekonomi pekerja yang harus melakukan perjalanan jauh setiap hari. Untuk mengumpulkan data, kuesioner yang mencakup berbagai informasi seperti pendapatan, jumlah tanggungan, lokasi tempat tinggal, waktu tempuh perjalanan, dan tingkat kepuasan hidup dikirim. Sebelum dimasukkan ke dalam model, data kemudian diproses melalui tahap pembersihan, normalisasi nilai numerik, dan pembagian menjadi data uji dan data latih. Salah satu alasan *RapidMiner* adalah antarmuka visualnya, yang memungkinkan pengguna membuat model klasifikasi tanpa menulis kode pemrograman. Hasil pengujian menunjukkan bahwa algoritma K-NN dapat dengan akurat mengklasifikasikan tingkat kepuasan hidup pekerja komuter. Performa model sangat dipengaruhi oleh variabel yang dipilih yaitu nilai K, dan kualitas data. Penelitian ini diharapkan dapat membantu pihak-pihak yang terkait, pendekatan ini dinilai efektif untuk membantu pengambilan keputusan berbasis data.

Kata Kunci: *K-Nearest Neighbor, Kepuasan Hidup, Komuter, Rapidminer, Data Mining.*

A. PENDAHULUAN

Kepuasan hidup sangat penting untuk mengukur kualitas hidup seseorang, terutama bagi pekerja komuter yang menghadapi tekanan perjalanan harian dari rumah ke tempat kerja. Jumlah pekerja komuter di Indonesia terus meningkat seiring dengan pertumbuhan kota dan suburban. Perjalanan yang lama dan berulang ini berdampak pada

kesehatan fisik dan sosial komuter. Oleh karena itu, untuk membuat kebijakan publik yang mendukung kesejahteraan masyarakat, penting untuk mengetahui tingkat kepuasan hidup mereka. Data mining sekarang menjadi alat yang bagus untuk analisis sosial berkat kemajuan teknologi. Algoritma *K-Nearest Neighbor* (K-NN), yang dapat mengelompokkan data berdasarkan kedekatan fitur, adalah salah satu teknik klasifikasi yang paling sederhana dan

umum digunakan. Perangkat lunak data mining berbasis antarmuka visual *RapidMiner* memungkinkan penerapan K-NN tanpa kemampuan pemrograman yang kuat.

Kebutuhan akan pendekatan berbasis data untuk memahami kesejahteraan komuter di Indonesia semakin mendesak, alasan di balik penelitian ini. Data kepuasan hidup komuter masih sedikit dan belum banyak digunakan dalam analisis kebijakan. Penelitian ini diharapkan dapat memberikan klasifikasi tingkat kepuasan hidup yang akurat dengan menggabungkan metode K-NN dan platform *RapidMiner*. Selain itu, penggunaan dataset dari platform *Kaggle* menjadi lebih mudah untuk mendapatkan data yang relevan dan berkualitas. Karena kombinasi metode klasifikasi yang tepat dan visualisasi proses yang ditawarkan oleh *RapidMiner*, ada alasan yang kuat untuk mengangkat penelitian ini sebagai kontribusi dalam bidang data sosial dan pengambilan keputusan berbasis data.

Berdasarkan latar belakang ini, rumusan masalah penelitian ini adalah memprediksi tingkat kepuasan hidup pekerja komuter di Indonesia dengan menggunakan algoritma *K-Nearest Neighbor* (K-NN) yang digunakan oleh *RapidMiner*

Tujuan penelitian ini adalah menggunakan *RapidMiner* dan algoritma K-NN untuk mengklasifikasikan tingkat kepuasan hidup pekerja komuter di Indonesia. Diharapkan bahwa penelitian ini akan memberikan pemahaman yang lebih baik tentang kondisi sosial ekonomi komuter dan pola kepuasan hidup mereka sebagai dasar untuk membuat kebijakan publik yang relevan.

B. METODE

1. Metode yang Digunakan

Metode deskriptif analitis digunakan dalam penelitian ini untuk memberikan gambaran sistematis tentang tingkat kepuasan hidup pekerja komuter di Indonesia. Berdasarkan data yang dikumpulkan dari kuesioner, penelitian ini mencapai tujuan ini. Metode ini menggunakan teknik klasifikasi dari algoritma *K-Nearest Neighbor* (K-NN) yang digunakan oleh perangkat lunak *RapidMiner*.

Penelitian kuantitatif deskriptif dengan pendekatan data mining klasifikasi digunakan dengan tujuannya adalah untuk memetakan atau mengelompokkan data responden sesuai dengan variabel sosial ekonomi dan membaginya ke dalam kelas-kelas yang berisi kepuasan hidup.

Sumber data untuk penelitian ini berasal dari dataset publik "Kepuasan Hidup Komuter" yang tersedia di platform *Kaggle.com*. Kemudian metode pengumpulan data menggunakan kuesioner online untuk mengumpulkan data ini, yang melibatkan beberapa variabel didalamnya

Untuk tahapan penelitian, langkah pertama adalah pengumpulan data. Data ini berasal dari survei kepuasan hidup komuter Indonesia yang dilakukan oleh 384 orang, yang dikumpulkan melalui *Kaggle*. Langkah selanjutnya

adalah pemrosesan pre-data: Sebelum diklasifikasikan, data diproses melalui tahap berikut:

1. Pembersihan data berarti menghapus nilai yang tidak valid atau kosong.
2. Melakukan Normalisasi data numerik dilakukan untuk memastikan bahwa skalanya sama.
3. Pembagian data terdiri dari data latihan (*training data*) dan data uji

Algoritma *K-Nearest Neighbor* (K-NN) digunakan untuk membentuk model klasifikasi. Pemilihan nilai K (jumlah tetangga terdekat) dilakukan secara eksperimental. Metode Euclidean distance digunakan untuk menghitung jarak antar data. Data dimasukkan ke dalam *Excel Studio* lalu diproses secara visual menggunakan operator Read Excel, K-NN, Apply Model, dan Performance (Classification). Setelah itu model diuji pada data uji untuk mengukur performa klasifikasi berdasarkan akurasi dan precision

2. Studi Literatur

Studi literatur dilakukan untuk mencari dan menganalisis penelitian sebelumnya serta teori yang digunakan dalam penelitian. Ini dilakukan untuk membantu pemahaman lebih baik tentang penelitian yang dilakukan dan teori yang digunakan. Selain itu, literatur yang dipelajari berhubungan dengan beberapa teori.

1. Data Mining. Penelitian ini menggunakan data mining untuk mengolah data sosial ekonomi komuter dan mengklasifikasikan tingkat kepuasan hidup berdasarkan pola yang ditemukan. Pengumpulan data, preprocessing, pemodelan, evaluasi, dan implementasi adalah semua bagian dari proses data mining.
2. Klasifikasi. Klasifikasi adalah salah satu teknik data mining yang bertujuan untuk memetakan data ke dalam kategori atau kelas tertentu. Pada penelitian ini, responden dimasukkan ke dalam kategori kepuasan hidup seperti "sangat tidak puas", "biasa", dan "sangat puas". Algoritma *K-Nearest Neighbor* digunakan untuk melakukan proses ini.
3. Algoritma *K-Nearest Neighbor* (K-NN) adalah metode klasifikasi berbasis *instance-based learning* yang menentukan kelas data uji berdasarkan kedekatan jaraknya dengan data latih. Prinsip kerja *Algoritma K-Nearest Neighbor* (K-NN) adalah sebagai berikut: Anda menghitung jarak antara data uji dan data latih, kemudian Anda memilih k tetangga terdekat, dan kemudian Anda menentukan kelas berdasarkan mayoritas dari tetangga tersebut. Nilai k, metode pengukuran jarak (seperti jarak geometri dan Manhattan), dan preprocessing data sangat memengaruhi kinerja K-NN.
4. *RapidMiner*. Perangkat lunak open-source *RapidMiner* adalah perangkat lunak yang digunakan untuk analisis data dan pemodelan statistik secara visual. Aplikasi ini memungkinkan pengguna menerapkan metode data mining seperti klasifikasi tanpa harus menulis kode program. Ada banyak operator di *RapidMiner* yang mendukung proses *end-to-end*, seperti import data, preprocessing, pemodelan dengan K-NN, dan evaluasi performa model klasifikasi.

- Preprocessing Data. Ini mencakup pembagian data menjadi dataset pelatihan (*training*) dan dataset pengujian (*testing*), serta pembersihan data dari nilai yang hilang atau tidak konsisten. Penggunaan preprocessing yang baik akan meningkatkan keandalan dan akurasi model klasifikasi K-NN.
- Evaluasi Model. Evaluasi ini dilakukan untuk mengevaluasi bagaimana model klasifikasi yang dibangun berfungsi optimal. Akurasi, ketepatan, pengembalian, skor F1, dan *matrix confusion* adalah metrik yang digunakan dalam penelitian ini. Evaluasi ini penting untuk mengetahui seberapa baik model dapat memprediksi dan digunakan dalam pengambilan keputusan berbasis data.

C. HASIL DAN PEMBAHASAN

Data utama yang digunakan dalam pengujian ini meliputi dataset yang diperoleh melalui penyebaran kuesioner. Data yang didapatkan akan diproses pada tahap *preprocessing* sebelum diuji. Hal ini mencakup

- Pembersihan data dari nilai yang hilang dan tidak konsisten,
- Normalisasi nilai numerik agar sesuai untuk diproses algoritma,
- Pengelompokan data menjadi dataset (*training*) dan data *testing* (*testing*).

Responis Kelal	Usia	Us Pernikj	Belakang Pendid	Pekerjaan	Janan per Bulan	Upatan per Bulan	(Kepuasan Hidup)	
1	Male	30	Married	Senior High School	Civil Servant	2500000	178.57	Sangat Tidak Pusu
2	Male	40	Married	Senior High School	Others	1500000	85.71	Sangat Tidak Pusu
3	Male	38	Married	Senior High School	Others	1600000	114.29	Sangat Tidak Pusu
4	Male	27	Single	Senior High School	Others	1500000	107.14	Biasa
5	Female	28	Single	Senior High School	Others	1200000	85.71	Sangat Tidak Pusu
6	Male	37	Married	Senior High School	Others	2000000	142.86	Sangat Tidak Pusu
7	Male	25	Single	Senior High School	Others	1500000	107.14	Sangat Tidak Pusu
8	Female	27	Single	Senior High School	Private Employees	2000000	142.86	Sangat Tidak Pusu
9	Male	40	Married	Senior High School	Private Employees	1500000	128.57	Sangat Tidak Pusu
10	Male	27	Single	Senior High School	Others	1500000	107.14	Biasa
11	Male	45	Married	Junior High School	Others	1800000	128.57	Biasa
12	Male	38	Married	Senior High School	Private Employees	1600000	114.29	Sangat Tidak Pusu
13	Male	25	Single	Senior High School	Others	1200000	85.71	Sangat Tidak Pusu
14	Female	30	Married	Senior High School	Others	1500000	107.14	Sangat Tidak Pusu
15	Male	28	Single	Senior High School	Others	1500000	107.14	Sangat Tidak Pusu
16	Male	38	Married	Senior High School	Others	1600000	114.29	Sangat Tidak Pusu
17	Male	33	Married	Senior High School	Others	1800000	128.57	Sangat Tidak Pusu
18	Male	40	Married	Senior High School	Others	2000000	142.86	Sangat Tidak Pusu
19	Male	40	Married	Senior High School	Pilicer/Military	3000000	214.29	Sangat Tidak Pusu
20	Male	38	Married	Senior High School	Others	1600000	114.29	Sangat Tidak Pusu
21	Male	38	Married	Senior High School	Self Employees	2500000	178.57	Sangat Pusu
22	Male	45	Married	Senior High School	Civil Servant	3000000	214.29	Sangat Tidak Pusu
23	Male	36	Married	Senior High School	Private Employees	1200000	85.71	Sangat Tidak Pusu
24	Male	20	Single	Senior High School	Private Employees	1200000	85.71	Sangat Tidak Pusu
25	Male	28	Single	Senior High School	Others	1500000	107.14	Sangat Tidak Pusu
26	Male	23	Married	Senior High School	Private Employees	2500000	178.57	Sangat Tidak Pusu
28	Male	23	Single	Senior High School	Others	1800000	128.57	Sangat Tidak Pusu
29	Male	43	Married	Senior High School	Private Employees	2000000	142.86	Sangat Tidak Pusu
30	Male	36	Widower	Senior High School	Private Employees	3500000	250	Sangat Tidak Pusu

Gambar 1. Dataset

Data awal tentang kepuasan hidup komuter ditunjukkan pada gambar ini. Dalam dataset ini, setiap baris menunjukkan satu responden, dan kolom-kolom berisi informasi seperti pendapatan, tanggungan, tempat tinggal, waktu perjalanan, dan tingkat kepuasan hidup. Sebelum digunakan dalam model, dataset ini dibersihkan dan dinormalisasi, yang merupakan langkah penting untuk proses klasifikasi dengan algoritma K-NN.

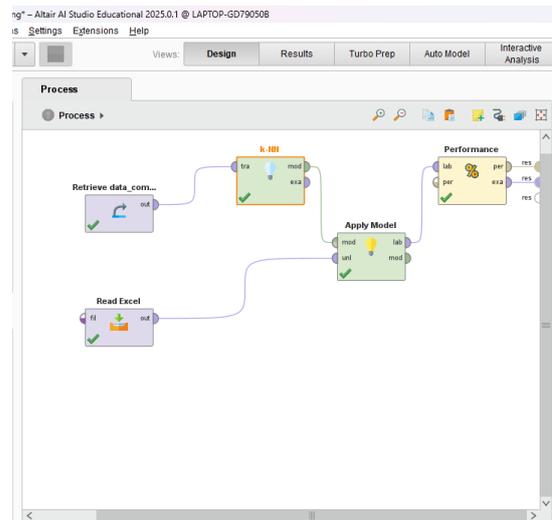
Responis Kelal	Usia	Us Pernikj	Belakang Pendid	Pekerjaan	Janan per Bulan	Upatan per Bulan	(Kepuasan Hidup)	
1	Male	30	Married	Senior High School	Civil Servant	2500000	178.57	Sangat Tidak Pusu
4	Male	27	Single	Senior High School	Others	1500000	107.14	Sangat Tidak Pusu
6	Male	37	Married	Senior High School	Others	2000000	142.86	Sangat Tidak Pusu
10	Male	27	Single	Senior High School	Others	1500000	107.14	Biasa
16	Male	38	Married	Senior High School	Others	1600000	114.29	Sangat Tidak Pusu
19	Male	40	Married	Senior High School	Pilicer/Military	3000000	214.29	Sangat Tidak Pusu
23	Male	36	Married	Senior High School	Private Employees	1200000	85.71	Sangat Tidak Pusu
26	Male	23	Married	Senior High School	Others	1500000	107.14	Sangat Tidak Pusu
31	Male	28	Single	Senior High School	Diploma	1700000	121.43	Sangat Tidak Pusu
32	Female	26	Married	Under Graduate	Civil Servant	5000000	357.14	Sangat Tidak Pusu
34	Male	27	Single	Senior High School	Diploma	3000000	214.29	Sangat Tidak Pusu
40	Male	23	Single	Senior High School	Others	1500000	107.14	Biasa
43	Male	45	Married	Senior High School	Self Employees	4000000	285.71	Sangat Tidak Pusu
46	Male	36	Married	Senior High School	Self Employees	7000000	500	Biasa
47	Male	40	Married	Junior High School	Private Employees	3000000	214.29	Biasa
56	Male	27	Married	Senior High School	Others	2300000	164.29	Sangat Tidak Pusu
57	Female	34	Married	Senior High School	Others	2000000	142.86	Biasa
58	Female	40	Married	Senior High School	Self Employees	4000000	285.71	Sangat Tidak Pusu
64	Female	31	Single	Under Graduate	Private Employees	5000000	357.14	Sangat Tidak Pusu
73	Male	32	Married	Junior High School	Others	1800000	128.57	Biasa
74	Male	24	Single	Senior High School	Others	1500000	107.14	Biasa
77	Male	23	Single	Senior High School	Others	1400000	100	Biasa
78	Male	23	Single	Senior High School	Others	1200000	85.71	Biasa
79	Male	30	Married	Senior High School	Others	1800000	128.57	Biasa
83	Male	23	Single	Senior High School	Others	1600000	114.29	Biasa
85	Male	29	Single	Senior High School	Private Employees	1800000	128.57	Biasa
91	Male	40	Married	Junior High School	Others	2000000	142.86	Biasa
94	Female	34	Married	Under Graduate	Civil Servant	2500000	178.57	Sangat Tidak Pusu
105	Male	30	Married	Senior High School	Private Employees	2000000	142.86	Sangat Tidak Pusu

Gambar 2. Data Testing

Bagian dari dataset yang digunakan untuk data uji ditunjukkan pada gambar ini. Untuk menilai kemampuan model untuk memprediksi data baru yang belum dikenali sebelumnya, data uji ini dipisahkan dari data latih. Setiap baris pada gambar tersebut berisi informasi fitur, seperti pendapatan dan tanggungan, tetapi tanpa label akhir, yang akan diprediksi oleh model K-NN. Pengujian dilakukan menggunakan *RapidMiner Studio* dengan langkah-langkah sebagai berikut:

- Membawa dataset dan data *testing* ke dalam workspace *RapidMiner* melalui proses import.
- Menggunakan operator K-NN (*K-Nearest Neighbor*) untuk klasifikasi.
- Memanfaatkan modul Apply Model untuk menguji prediksi pada data *testing*.
- Menilai performa model menggunakan the Operator Performance (Classification).

Berikut tampilan pengujian dengan *RapidMiner Studio*:



Gambar 3. Tampilan Dasar *RapidMiner*

Gambar ini menunjukkan alur kerja *RapidMiner Studio* untuk klasifikasi. Ada komponen penting seperti:

- Read Excel, untuk mengimpor data latih dan uji dari file eksternal.
- K-NN, operator utama yang menjalankan algoritma *K-Nearest Neighbor*.

- c. Apply Model, menerapkan model K-NN ke data uji.
- d. Performance (*Classification*), menghitung dan menampilkan metrik evaluasi seperti akurasi.

Dengan desain ini, proses klasifikasi bisa ditampilkan secara visual dan terpisah tanpa perlu pemrograman, sehingga memudahkan untuk menguji dan memantau hasilnya. Prediksi tingkat kepuasan hidup untuk setiap data uji digambarkan sebagai hasil dari proses klasifikasi. Output terdiri dari kolom-kolom berikut:

- a. Nilai perkiraan (seperti "sangat tidak puas," "biasa," atau "sangat puas").
- b. Kemungkinan confidence score, yang merupakan tingkat keyakinan model terhadap prediksi tersebut.
- c. Informasi pembanding, seperti label evaluasi kinerja aktual.

Hasil ini digunakan untuk membandingkan prediksi dengan keadaan sebenarnya dan menilai efektivitas model.

Row No.	No. Respond.	Kepuasan H.	Jenis Kelamin	Usia	Status Permt.	Latar Belakang	Agama	Pekerjaan	Kota
1	1	Sangat Tidak...	Male	30	Married	Senior High...	Moslem	Civil Servant	Affor
2	2	Sangat Tidak...	Male	40	Married	Senior High...	Moslem	Others	Affor
3	3	Sangat Tidak...	Male	38	Married	Senior High...	Moslem	Others	Affor
4	4	Sangat Tidak...	Male	27	Single	Senior High...	Moslem	Others	Very
5	5	Sangat Tidak...	Female	28	Single	Senior High...	Moslem	Others	Very
6	6	Sangat Tidak...	Male	37	Married	Junior High S...	Moslem	Others	Affor
7	7	Sangat Tidak...	Male	25	Single	Senior High...	Moslem	Others	Affor
8	8	Sangat Tidak...	Female	27	Single	Senior High...	Moslem	Private Empl.	Neut
9	9	Sangat Tidak...	Male	40	Married	Senior High...	Moslem	Private Empl.	Neut
10	10	Biasa	Male	27	Single	Senior High...	Moslem	Others	Affor
11	11	Biasa	Male	45	Married	Junior High S...	Moslem	Others	Neut
12	12	Sangat Tidak...	Male	38	Married	Senior High...	Moslem	Private Empl.	Neut
13	13	Sangat Tidak...	Male	25	Single	Senior High...	Moslem	Others	Neut

Gambar 4. Hasil Prediksi

Hasil evaluasi performa model ditunjukkan melalui akurasi yang dapat didefinisikan sebagai persentase prediksi yang benar dibandingkan dengan total prediksi. Precision, recall, dan f1-score digunakan untuk mengukur ketepatan dan kelengkapan prediksi untuk masing-masing kelas. Lalu tabel confusion matrix menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas. Nilai akurasi yang ditampilkan menunjukkan bahwa model K-NN memiliki kemampuan klasifikasi yang cukup baik untuk mengukur kepuasan hidup komuter berdasarkan fitur yang tersedia.

	true Biasa	true Sangat Tidak Puas	true Sangat Puas	class precision
pred. Biasa	24	3	2	82.76%
pred. Sangat Tidak Puas	5	34	0	87.18%
pred. Sangat Puas	1	2	6	66.67%
class recall	80.00%	87.18%	75.00%	

Gambar 5. Akurasi Performance

Output dari hasil pengklasifikasian ini adalah kepuasan hidup pekerja komuter (sangat tidak puas, biasa, sangat puas). Dari hasil pengujian, model ini dapat mengklasifikasikan status ekonomi dengan tingkat akurasi yang memadai terhadap data *testing*. Meskipun ada beberapa nilai prediksi yang tidak sesuai dengan label aktual, model tersebut secara umum berfungsi seperti yang diharapkan.

Dari hasil pengujian di atas, algoritma K-NN terbukti efektif dalam mengklasifikasikan status ekonomi berdasarkan data sosial ekonomi keluarga. Kualitas data, pemilihan variabel yang relevan, dan jumlah nilai K yang digunakan sangat mempengaruhi akurasi model. *RapidMiner* menyederhanakan proses pengujian tanpa perlu coding manual, menjadikannya cocok untuk eksperimen cepat dengan visualisasi yang jelas.

D. PENUTUP

1. Kesimpulan

Studi ini menunjukkan bahwa algoritma *K-Nearest Neighbor* (K-NN) dapat dengan baik digunakan untuk mengelompokkan tingkat kepuasan hidup komuter di Indonesia berdasarkan data sosial ekonomi yang diambil dari kuesioner. Dengan menggunakan platform *RapidMiner*, proses klasifikasi menjadi lebih mudah dan teratur karena antarmuka visualnya yang mendukung berbagai langkah dalam pengolahan data, mulai dari preprocessing hingga evaluasi model.

Faktor-faktor seperti pemilihan variabel yang tepat, kualitas data, dan nilai parameter K yang digunakan sangat mempengaruhi seberapa baik model K-NN bekerja; meskipun begitu, model K-NN dapat memberikan prediksi tentang tingkat kepuasan hidup yang cukup akurat.

Secara keseluruhan, metode ini bisa menjadi alat awal untuk memahami pandangan dan kondisi sosial ekonomi orang-orang yang berkomuter serta membantu dalam pengambilan keputusan yang berdasarkan data. Penelitian ini bisa diperluas dengan menambah jumlah data, menganalisis lebih banyak variabel, dan menggabungkan algoritma K-NN dengan teknik lain untuk meningkatkan ketepatan dan akurasi klasifikasi.

2. Ucapan Terima Kasih

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua orang yang telah membantu, membantu, dan mendorong penulis untuk menyelesaikan makalah ini. Secara khusus, ucapan terima kasih disampaikan kepada:

- a. Universitas Pamulang, khususnya Fakultas Ilmu Komputer, yang telah memberikan kesempatan dan fasilitas untuk menyelesaikan tugas dan ujian akhir semester ini;
- b. Dosen pengampu mata kuliah, yang telah memberikan bimbingan, arahan, dan masukan yang

sangat berarti selama proses penyusunan laporan dan pengembangan sistem; dan

- c. Rekan-rekan mahasiswa, yang telah memberikan dukungan, kerja sama, dan semangat dalam menyelamatkan laporan.

E. DAFTAR PUSTAKA

- Aisyah Fatmawati, A. I. (2024). IMPLEMENTASI NAÏVE BAYES UNTUK KLASIFIKASI KELAYAKAN PENERIMA BANTUAN SOSIAL . *JATI (Jurnal Mahasiswa Teknik Informatika)* , 745-750.
- Argina, A. M. (2020). Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*, 1(2), 29-33.
- Bayu Susilo, N. A. (2024). Penerapan Algoritma *K-Nearest Neigh* bor untuk Prediksi. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1466-1476.
- Dewi, A. O. (2020). Big Data di Perpustakaan dengan Memanfaatkan Data Mining. *Anuva: Jurnal Kajian Budaya, Perpustakaan, dan Informasi*, 4(2), 223-230.
- Dewi, M. A. (2021). Penerapan Framework Scrum pada Pengembangan Sistem Informasi Keuangan Berbasis Web. *Jurnal Teknologi Informasi dan Komunikasi*.
- Harun, R. P. (2020). Penerapan Data Mining untuk Menentukan Potensi Hujan Harian dengan Menggunakan Algoritma K Nearest Neighbor (KNN). *Jurnal Manajemen Informatika dan Sistem Informasi.*, 3(1), 8-15.
- Irawati, R. A. (2024). Sistem Klasifikasi Status Sosial Ekonomi Keluarga di Desa Popalia Menggunakan Metode K-NN . *Jurnal Sains dan Informatika*, 11-20.
- Karyono, G. (2016). Analisis Teknik Data Mining Algoritma C4. 5 Dan *K-Nearest Neigh* bor Untuk Mendiagnosa Penyakit Diabetes Mellitus. *STMIK–Politeknik PalComTech.*, 12.
- Khormarudin, A. N. (2016). Teknik data mining: Algoritma K-Means clustering. *J. Ilmu Komput.*, 1-12.
- Lorena Br Ginting, S. Z. (2015). Teknik Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma *K-Nearest Neigh* borhood. *KOMPUTIKA-Jurnal Sistem Komputer UNIKOM*, 3(2).
- Mahulae, I. G., Harli, E., & Dja'far, H. I. (2022). PERANCANGAN SISTEM INFORMASI PEMBAYARAN DAN PENDATAAN PADA BENGKEL KARINA MOTOR MENGGUNAKAN NETBEANS. *Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI)*, 3(4), 691-698.
- Maulidina Azizah, N. R. (2024). ANALISIS KLASTERISASI WILAYAH PENYANDANG DISABILITAS DI PROVINSI JAWA BARAT. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 638-644.
- Mukhamad Said Riza Zudi, A. M. (2024). IMPLEMENTASI DATAMINING CLUSTERING DALAM MENGELOMPOKKAN PENDUDUK PENYANDANG DISABILITAS PROVINSI JAWA BARAT MENGGUNAKAN ALGORITMA K-MEANS. 1-15.
- Mulyati, S. H. (2020). Rancang bangun aplikasi data mining prediksi kelulusan ujian nasional menggunakan Algoritma (Knn) *K-Nearest Neigh* bor dengan metode Euclidean Distance pada SMPN 2 Pagedangan. *JIKA (Jurnal Informatika)*, 4(1), 65-73.
- Nikmatun, I. A. (2019). Implementasi data mining untuk klasifikasi masa studi mahasiswa menggunakan algoritma *K-Nearest Neigh* bor. *Jurnal Simetris.*, 10(2), 421-432.
- Purwandari, N., & Fauzi, A. (2020). PERANCANGAN SISTEM INFORMASI MANAJEMEN. *JUNSIBI: Jurnal Sistem Informasi Bisnis*, 1(2), 54-64.
- Ridlwaniul Fitra, I. R. (2022). Penerapan Metode Algoritma *K-Nearest Neigh* bor Menggunakan *RapidMiner* Studio pada Klasifikasi Status Sosial Ekonomi Studi Kasus : Kelurahan Kapuk Muara Rt 010 Rw 04 . *Informatika, Universitas Nusa Mandiri*, 653-660.

- Santoso, A. &. (2022). Perancangan Aplikasi Digital Arsip Kampus Dalam Meningkatkan. *Jurnal Makhumatika*, 9(1), 37.
- Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4, 5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, , 13(1), 16-25.
- Winda Lidysari, H. S. (2022). Penerapan Data Mining Dalam Menentukan Kelayakan Penerima Bantuan Sosial Pemko Dengan Algoritma C4.5 (Kasus Kantor Kelurahan Martoba) . *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, 53-61.