



**PREDIKSI KEPUASAN HIDUP PEKERJA KOMUTER INDONESIA  
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)**

**SKRIPSI**

**MUHAMMAD RADITYO**

**1910512037**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI SARJANA SISTEM INFORMASI  
2023**



**PREDIKSI KEPUASAN HIDUP PEKERJA KOMUTER INDONESIA  
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)**

**SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar  
Sarjana Komputer**

**MUHAMMAD RADITYO  
1910512037**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI SARJANA SISTEM INFORMASI  
2023**

## PERNYATAAN ORISINALITAS

Skripsi ini merupakan hasil karya sendiri serta seluruh sumber referensi yang dikutip dan dirujuk telah saya nyatakan dengan benar.

Nama : Muhammad Radityo

NIM : 1910512037

Tanggal : 5 Juni 2023

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan berlaku.

Jakarta, 5 Juni 2023

Yang Menyatakan,



(Muhammad Radityo)

**PERNYATAAN PERSETUJUAN PUBLIKASI  
SKRIPSI UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Radityo  
NIM : 1910512037  
Fakultas : Ilmu Komputer  
Program Studi : S1 Sistem Informasi

Memberikan persetujuan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non Eksklusif (*Non-Exclusive Royalty Free Right*) atas karya ilmiah yang berjudul :

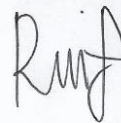
**PREDIKSI KEPUASAN HIDUP PEKERJA KOMUTER INDONESIA  
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)**

Beserta perangkat lain (apabila diperlukan). Hak Bebas Royalti Non Eksklusif ini memberikan hak Universitas Pembangunan Nasional Veteran Jakarta dalam menyimpan, mengalih media, mengelola dalam bentuk pangkalan data, merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis dan pemilik hak cipta. Demikian pernyataan ini dibuat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 5 Juni 2023

Yang Menyatakan,



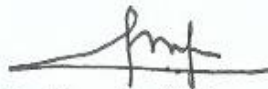
(Muhammad Radityo)

## LEMBAR PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Muhammad Radityo  
NIM : 1910512037  
Program Studi : S1 Sistem Informasi  
Judul Tugas Akhir : PREDIKSI KEPUASAN HIDUP PEKERJA KOMUTER  
INDONESIA MENGGUNAKAN ALGORITMA  
K-NEAREST NEIGHBOR (K-NN).

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Dr. Ermatita, M.Kom.  
Penguji 1



Helena Nurramdhani Irmada, S.Pd., M.Kom.  
Penguji 2



Rio Wirawan, S.Kom., M.M.S.I  
Pembimbing



Helena Nurramdhani Irmada, S.Pd., M.Kom.  
Kepala Program Studi

Ditetapkan di : Jakarta  
Tanggal Ujian : 22 Juni 2023



# **PREDIKSI KEPUASAN HIDUP PEKERJA KOMUTER INDONESIA MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)**

**MUHAMMAD RADITYO**

## **ABSTRAK**

Pekerja komuter adalah seseorang yang bekerja di luar kabupaten/kota tempat tinggal dan secara rutin pergi dan pulang (PP) ke tempat tinggalnya pada hari yang sama. Kegiatan komuter yang menjadi rutinitas sehari-hari seorang pekerja lebih dari sekadar menempuh jarak antara rumah dan bekerja. Kegiatan komuter selain membutuhkan waktu dan biaya yang tidak sedikit, dapat juga menyebabkan stres dan mengganggu hubungan antara pekerjaan dan keluarga. Sehingga bisa berdampak terhadap kepuasan hidup sebagai pekerja komuter. Penelitian ini bertujuan untuk melakukan prediksi kepuasan hidup pekerja komuter Indonesia menggunakan data sekunder yang bersumber dari website *www.kaggle.com* yang diunggah oleh Rezky Yayang Yakhmid pada tahun 2022 dengan judul *Indonesian Commuter Life Satisfaction* yang merupakan hasil survei dari pekerja komuter di Indonesia. *Dataset* ini awalnya memiliki 384 *record* data yang akan dilakukan prediksi dengan pembagian data latih dan data validasi dengan perbandingan 80% berbanding 20%. Prediksi dilakukan dengan menggunakan teknik *data mining* dengan menggunakan algoritma *K-Nearest Neighbor* (K-NN). Algoritma ini bekerja dengan melakukan klasifikasi kelas data baru berdasarkan tetangga terdekatnya. Penelitian ini juga melakukan analisis dan evaluasi performa model dengan nilai akurasi, presisi, *recall*, dan *f1-score* (*f-measure*) untuk setiap nilai K yang diujikan yang berada pada rentang nilai K=3 hingga K=9 dengan tujuan untuk mendapatkan performa teroptimal. Diperoleh bahwa nilai K = 5 memiliki performa teroptimal dengan nilai akurasi 87,04%, nilai presisi sebesar 86,11%, nilai *recall* sebesar 89,01% dan nilai *f1-Score* (*f-measure*) sebesar 87,08%.

**Kata kunci:** prediksi, kepuasan hidup, pekerja komuter, *K-Nearest Neighbor*

# **PREDICTION OF LIFE SATISFACTION OF INDONESIAN COMMUTER WORKERS USING THE K-NEAREST NEIGHBOR (K-NN) ALGORITHM**

**MUHAMMAD RADITYO**

## **ABSTRACT**

*A commuter worker is someone who works outside the regency/city of their residence and routinely goes to and returns to their place of residence on the same day. Commuter activities that become a worker's daily routine are more than just traveling the distance between home and work. Commuter activities not only requiring a lot of time and money, but also can causing stress and disrupting the relationship between work and family. So commuting can have an impact on life satisfaction as a commuter worker. This study aims to predict the life satisfaction of Indonesian commuter workers using secondary data sourced from the website [www.kaggle.com](http://www.kaggle.com) uploaded by Rezky Yayang Yakhamid in 2022 with the title Indonesian Commuter Life Satisfaction which is the result of a survey of commuter workers in Indonesia. This dataset initially has 384 data records that will be predicted by dividing into training data and validation data with a ratio of 80% and 20%. Predictions are made using data mining techniques using the K-Nearest Neighbor (K-NN) algorithm. This algorithm works by classifying new data classes based on their nearest neighbours. This study also analyzes and evaluates the performance of the model with accuracy, precision, recall, and f1-score (f-measure) values for each tested K value which is in the range of values  $K=3$  to  $K=9$  with the aim of obtaining optimal performance. It was found that the value of  $K = 5$  has the optimal performance with an accuracy value of 87.04%, a precision value of 86.11%, a recall value of 89.01% and an f1-Score (f-measure) value of 87.08%.*

**Keywords:** *prediction, life satisfaction, commuter worker, K-Nearest Neighbor*

## KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir ini dengan judul “Prediksi Kepuasan Hidup Pekerja Komuter Indonesia Menggunakan Algoritma K-Nearest Neighbor (K-NN)”.

Adapun tujuan dari penyusunan tugas akhir ini untuk memenuhi salah satu syarat dalam memperoleh gelar Sarjana Komputer pada Program Studi S1 Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta. Pada kesempatan ini, penulis ingin mengucapkan terima kasih kepada:

1. Orang tua dan kakak yang selalu memberi dukungan, doa, dan semangat kepada penulis selama penyusunan tugas akhir ini.
2. Ibu Dr. Ermatita, M.Kom selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Helena Nurramdhani Irmanda, S.Pd., M.Kom selaku Ketua Program Studi S1 Sistem Informasi.
4. Bapak Rio Wirawan, S.Kom., MMSI. selaku dosen pembimbing yang membimbing dan membantu penulis dalam menyelesaikan tugas akhir ini.
5. Bapak I Wayan Widi P., S.Kom., MTI. selaku dosen pembimbing akademik.
6. Bapak dan Ibu dosen Fakultas Ilmu Komputer yang telah memberikan ilmu-ilmu dan pembelajaran yang sangat bermanfaat.
7. Teman-teman program studi S1 Sistem Informasi angkatan 2019 yang memberikan dukungan dan berdiskusi dengan penulis.
8. Dan semua pihak yang telah terlibat namun tidak bisa penulis sebutkan satu per satu tanpa mengurangi rasa hormat penulis.

Jakarta, 5 Juni 2023

Penulis



## DAFTAR ISI

HALAMAN JUDUL.....	i
PERNYATAAN ORISINALITAS .....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS.....	iii
LEMBAR PENGESAHAN.....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI .....	viii
DAFTAR GAMBAR .....	xi
DAFTAR TABEL .....	xiii
DAFTAR LAMPIRAN .....	xiv
BAB I PENDAHULUAN .....	1
1.1    Latar Belakang.....	1
1.2    Rumusan Masalah.....	4
1.3    Tujuan Penelitian .....	4
1.4    Ruang Lingkup Penelitian .....	4
1.5    Manfaat Penelitian .....	5
1.6    Luaran yang diharapkan.....	5
1.7    Sistematika Penulisan .....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1    Pekerja .....	7
2.2    Pekerja Komuter .....	7
2.3    Kepuasan Hidup.....	8
2.4    Data Mining .....	10
2.5    Algoritma <i>K-Nearest Neighbor</i> .....	12
2.6    Sistem Prediksi .....	14
2.6.1    Sistem .....	14

2.6.2	Prediksi .....	15
2.7	<i>Confusion Matrix</i> .....	15
2.8	<i>Python</i> .....	17
2.9	<i>Feature Importance</i> .....	18
2.10	<i>Framework Flask</i> .....	19
2.11	<i>Black Box Testing</i> .....	19
2.12	Penelitian yang Relevan.....	20
BAB III METODOLOGI PENELITIAN.....		23
3.1	Alur Penelitian .....	23
3.2	Identifikasi Masalah.....	24
3.3	Persiapan Data .....	24
3.4	Preprocessing Data .....	26
3.5	Pembagian data .....	27
3.6.	Pemodelan menggunakan Algoritma <i>K-Nearest Neighbor</i> .....	27
3.7	Evaluasi Model .....	28
3.8	Perancangan Sistem Prediksi .....	28
3.9	Pembangunan Sistem Prediksi.....	28
3.10	Pengujian Sistem Prediksi .....	29
3.11	Alat Bantu Penelitian .....	30
3.12	Jadwal Rencana Penelitian.....	30
BAB IV HASIL DAN PEMBAHASAN .....		32
4.1	Tentang <i>Dataset</i> .....	32
4.2	<i>Preprocessing Data</i> .....	37
4.2.1	<i>Data Cleansing</i> .....	37
4.2.2	<i>Data Transformation</i> .....	41
4.2.3	<i>Feature Selection</i> (Seleksi Fitur).....	42
4.3	Implementasi Algoritma <i>K-Nearest Neighbor</i> .....	47

4.3.1	Persiapan Data Latih, Data Validasi dan Data Uji .....	47
4.3.2	Pemodelan menggunakan Algoritma <i>K-Nearest Neighbor</i> ...	50
4.4	Evaluasi Model .....	54
4.5	Perancangan Sistem Prediksi .....	61
4.5.1	<i>Use Case Diagram</i> .....	62
4.5.2	Rancangan Antarmuka Sederhana ( <i>Wireframe</i> ).....	62
4.5.3	Rancangan <i>Database</i> .....	64
4.6	Pembangunan Sistem Prediksi.....	65
4.6.1	Halaman Beranda .....	65
4.6.2	Halaman Tentang Kami.....	66
4.6.3	Halaman Form Prediksi .....	66
4.6.4	Halaman Hasil Prediksi .....	68
4.7	Pengujian Sistem Menggunakan <i>Black Box Testing</i> .....	69
BAB V PENUTUP .....		71
5.1	Kesimpulan .....	71
5.2	Saran .....	71
DAFTAR PUSTAKA .....		73
DAFTAR RIWAYAT HIDUP.....		77
LAMPIRAN .....		78

## DAFTAR GAMBAR

Gambar 1. Tahapan KDD.....	11
Gambar 2. Confusion Matrix .....	15
Gambar 3. Rumus <i>Accuracy</i> .....	16
Gambar 4. Rumus <i>Precision</i> .....	16
Gambar 5. Rumus <i>Recall</i> .....	16
Gambar 6. Rumus <i>F1-Score</i> .....	17
Gambar 7. Alur Penelitian.....	23
Gambar 8. Penghapusan Kolom No.Responden .....	38
Gambar 9. Pengecekan data dengan <i>missing value</i> .....	38
Gambar 10. Pengecekan data duplikat .....	39
Gambar 11. Melakukan penghapusan data duplikat .....	39
Gambar 12. <i>Dataset</i> setelah dilakukan penghapusan data duplikat .....	39
Gambar 13. Boxplot atribut D.4.....	40
Gambar 14. Boxplot atribut D.4 setelah penanganan outlier .....	40
Gambar 15. <i>Dataset</i> setelah dilakukan penanganan outlier .....	41
Gambar 16. Pengecekan nilai Kepuasan Hidup .....	42
Gambar 17. Melakukan transformasi kolom Kepuasan Hidup .....	42
Gambar 18. Pengecekan nilai Kepuasan Hidup setelah transformasi .....	42
Gambar 19. Penghapusan kolom C.1, C.2, C.3, D.5.....	43
Gambar 20. Detail kolom setelah dilakukan penghapusan .....	43
Gambar 21. Perintah kedua penghapusan kolom .....	45
Gambar 22. Detail kolom setelah dilakukan penghapusan kedua.....	45
Gambar 23. Perintah untuk <i>Split</i> data 80% : 20% .....	47
Gambar 24. Normalisasi menggunakan <i>library MinMaxScaler</i> .....	49
Gambar 25. Proses <i>oversampling</i> menggunakan SMOTE.....	50
Gambar 26. Tahapan implementasi algoritma K-NN menggunakan Python.....	54
Gambar 27. Nilai akurasi,presisi, <i>recall</i> , <i>f1-score</i> K=3 menggunakan <i>library sklearn</i> .....	56
Gambar 28. Nilai akurasi,presisi, <i>recall</i> , <i>f1-score</i> K=4 menggunakan <i>library sklearn</i> .....	57

Gambar 29. Nilai akurasi,presisi, <i>recall</i> , <i>f1-score</i> K=5 menggunakan <i>library sklearn</i> .....	57
Gambar 30. Nilai akurasi,presisi, <i>recall</i> , <i>f1-score</i> K=6 menggunakan <i>library sklearn</i> .....	57
Gambar 31. Nilai akurasi,presisi, <i>recall</i> , <i>f1-score</i> K=7 menggunakan <i>library sklearn</i> .....	58
Gambar 32. Nilai akurasi, presisi, <i>recall</i> , <i>f1-score</i> K=8 menggunakan <i>library sklearn</i> .....	58
Gambar 33. Nilai akurasi,presisi, <i>recall</i> , <i>f1-score</i> K=9 menggunakan <i>library sklearn</i> .....	58
Gambar 34. <i>Chart</i> hasil evaluasi setiap nilai K.....	59
Gambar 35. Prediksi Pada Data Latih .....	60
Gambar 36. Prediksi Pada Data Uji .....	60
Gambar 37. <i>Use Case Diagram</i> Sistem .....	62
Gambar 38. Rancangan Antarmuka Halaman Beranda .....	63
Gambar 39. Rancangan Antarmuka Halaman Tentang Kami .....	63
Gambar 40. Rancangan Antarmuka Halaman Form Prediksi .....	63
Gambar 41. Rancangan Antarmuka Halaman Hasil Prediksi .....	64
Gambar 42. Impor hasil pemodelan dengan <i>pickle</i> .....	65
Gambar 43. Tampilan Halaman Beranda.....	66
Gambar 44. Tampilan Halaman Tentang Kami .....	66
Gambar 45. Tampilan Form Prediksi.....	67
Gambar 46. Tampilan Form Prediksi yang telah diisi .....	68
Gambar 47. Tampilan Hasil Prediksi .....	69

## DAFTAR TABEL

Tabel 1. Penelitian yang Relevan .....	20
Tabel 2. Jadwal Rencana Penelitian .....	31
Tabel 3. <i>Dataset</i> Awal.....	32
Tabel 4. Rincian Variabel <i>Dataset</i> .....	33
Tabel 5. <i>Feature Importances</i> .....	44
Tabel 6. Fitur yang Digunakan.....	45
Tabel 7. Data baru yang akan digunakan sebagai data uji .....	48
Tabel 8. Detail data latih untuk perhitungan manual .....	50
Tabel 9. Detail data uji yang akan dilakukan perhitungan manual .....	51
Tabel 10. Perhitungan <i>euclidean distance</i> .....	52
Tabel 11. Pengurutan <i>euclidean distance</i> berdasarkan nilai terkecil .....	53
Tabel 12. <i>Confusion Matrix</i> K=3 .....	54
Tabel 13. Presisi K = 3 .....	55
Tabel 14. <i>Recall</i> K = 3 .....	55
Tabel 15. Nilai akurasi,presisi, <i>recall</i> dan <i>f1 Score(f-measure)</i> untuk setiap nilai K.....	58
Tabel 16. Perbandingan akurasi split data.....	60
Tabel 17. Hasil Prediksi Pada Data Uji.....	61
Tabel 18. Tabel <i>form_response</i> .....	64
Tabel 19. Hasil Pengujian Black Box Testing .....	69

## DAFTAR LAMPIRAN

Lampiran 1 Dataset Awal.....	78
Lampiran 2 Keterangan Dataset Awal .....	94
Lampiran 3 Source Code App.py .....	96
Lampiran 4 Source Code Python Google Colab .....	97
Lampiran 5 Hasil Uji Turnitin.....	100