



**PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA  
NAÏVE BAYES, C4.5, DAN K-NEAREST NEIGHBOR UNTUK  
KLASIFIKASI KEMISKINAN DI DKI JAKARTA**

Aisyah Tri Deanita

41820010122

Denis Nila Cahyani

41820010111

Muhammad Anand Rizki Andinta

41820010069

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2023**



**PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA NAÏVE  
BAYES, C4.5, DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI  
KEMISKINAN DI DKI JAKARTA**

*Laporan Tugas Akhir*

Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer

Oleh:

Aisyah Tri Deanita

41820010122

Denis Nila Cahyani

41820010111

Muhammad Anand Rizki Andinta

41820010069

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA**

**2023**

## SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini:

Nama Mahasiswa (1) : Aisyah Tri Deanita

NIM : 41820010122

Nama Mahasiswa (2) : Denis Nila Cahyani

NIM : 41820010111

Nama Mahasiswa (3) : Muhammad Anand Rizki Andinta

NIM : 41820010069

Judul Tugas Akhir : Penerapan Data Mining Menggunakan Algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk Klasifikasi Kemiskinan Di DKI Jakarta

Menyatakan bahwa Tugas Akhir ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Tugas Akhir saya terdapat terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

UNIVERSITAS  
MERCU BUANA

Jakarta, 19 Januari 2024



Aisyah Tri Deanita

## LEMBAR PENGESAHAN

Nama Mahasiswa (1) : Aisyah Tri Deanita  
 NIM : 41820010122  
 Nama Mahasiswa (2) : Denis Nila Cahyani  
 NIM : 41820010111  
 Nama Mahasiswa (3) : Muhammad Anand Rizki Andinta  
 NIM : 41820010069  
 Judul Tugas Akhir : Penerapan Data Mining Menggunakan Algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk Klasifikasi Kemiskinan di DKI Jakarta

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana.


Jakarta, 20 Desember 2023

Menyetujui

Pembimbing : Dr. Bambang Jokonowo, S.Si., M.T.I. (  )  
 NIDN : 0320037002  
 Ketua Penguji : Ratna Mutu Manikam, S.Kom, MT (  )  
 NIDN : 0308017101  
 Penguji 1 : Rinto Priambodo, ST, MTI (  )  
 NIDN : 0327057905  
 Penguji 2 : Abdi Wahab, S.Kom, MT (  )  
 NIDN : 0305068502

Mengetahui,

  
Dr. Bambang Jokonowo, S.Si., M.T.I.  
 Dekan Fakultas Ilmu Komputer

  
Dr. Ruci Meiyanti, M.Kom  
 Ka.Prodi Sistem Informasi

## KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, atas berkat dan rahmat-Nya, saya dapat menyelesaikan Tugas Akhir ini yang berjudul "Penerapan Data Mining Menggunakan Algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk Klasifikasi Kemiskinan di DKI Jakarta". Penulis menyadari bahwa tanpa bantuan dan bimbingan tugas akhir ini mungkin tidak dapat diselesaikan. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Dr. Bambang Jokonowo, S.Si., M.T.I., selaku Dekan Fakultas Ilmu Komputer dan Dosen Pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan Tugas Akhir ini.
2. Dr. Ruci Meiyanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi atas dukungan dan bimbingannya selama saya menjalani perkuliahan.
3. Kepada pihak Keluarga khususnya kedua Orang Tua kami yang tanpa henti memberikan dukungan, semangat, dan doa yang sangat luar biasa kepada penulis.
4. Kepada Rekan Tim yang telah terlibat dalam penyusunan Tugas Akhir ini hingga dapat terselesaikan.
5. Kepada Seluruh pihak yang membantu selama proses pengerjaan Tugas Akhir yang tidak dapat disebutkan satu persatu.

Akhir kata, penulis berharap Tugas Akhir ini bisa bermanfaat untuk semua pihak. Selain itu, kritik dan saran yang membangun sangat penulis harapkan dari para pembaca sekalian agar Tugas Akhir ini bisa lebih baik lagi.

Jakarta, 8 Desember 2023

Penulis

## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan dibawah ini:

Nama Mahasiswa (1) : Aisyah Tri Deanita

NIM : 41820010122

Nama Mahasiswa (2) : Denis Nila Cahyani

NIM : 41820010111

Nama Mahasiswa (3) : Muhammad Anand Rizki Andinta

NIM : 41820010069

Judul Tugas Akhir : Penerapan Data Mining Menggunakan Algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk Klasifikasi Kemiskinan Di DKI Jakarta

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 19 Januari 2024

Yang menyatakan,



Aisyah Tri Deanita

## ABSTRAK

Nama Mahasiswa (1) : Aisyah Tri Deanita  
NIM : 41820010122  
Nama Mahasiswa (2) : Denis Nila Cahyani  
NIM : 41820010111  
Nama Mahasiswa (3) : Muhammad Anand Rizki Andinta  
NIM : 41820010069  
Pembimbing TA : Dr. Bambang Jekonowo, S.Si., M.T.I.  
Judul Tugas Akhir : Penerapan Data Mining Menggunakan Algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk Klasifikasi Kemiskinan di DKI Jakarta

Kemiskinan masih menjadi salah satu permasalahan fundamental yang sulit dihadapi, termasuk di Provinsi DKI Jakarta sebagai Ibu kota negara Indonesia, yang juga tidak luput dari permasalahan kemiskinan. Penelitian ini menggunakan *data mining* untuk melakukan klasifikasi terhadap tingkat kemiskinan di Provinsi DKI Jakarta dengan data yang diperoleh dari Badan Pusat Statistik (BPS). Tujuan penelitian ini adalah untuk secara efektif dan akurat mengklasifikasikan data kemiskinan dan membandingkan hasil dari algoritma Naïve Bayes, C4.5N dan K-NN guna mencari hasil akurasi terbaik. Melalui penerapan algoritma tersebut, didapatkan hasil bahwa performa algoritma dapat dipengaruhi oleh pembagian data, dan peningkatan rasio data uji cenderung meningkatkan akurasi serta konsistensi hasil klasifikasi. Pada rasio pembagian data 70:30, Naïve Bayes mencapai akurasi 81%, C4.5 sebesar 76%, dan K-NN sebesar 71%. Pada rasio 80:20, Naïve Bayes menunjukkan akurasi 93%, C4.5 sebesar 79%, dan K-NN sebesar 86%. Sementara pada rasio 90:10, Naïve Bayes mencapai akurasi 100%, C4.5 sebesar 71%, dan K-NN sebesar 86%. Dengan mempertimbangkan variasi performa ketiga algoritma, dapat disimpulkan bahwa Naïve Bayes lebih unggul sebagai algoritma yang stabil dan dapat diandalkan dalam berbagai skenario pembagian dataset, menunjukkan akurasi tinggi dan kemampuan baik dalam mengidentifikasi kasus positif.

Kata kunci:

kemiskinan, klasifikasi, Naïve Bayes, C4.5, K-Nearest Neighbor

## ABSTRACT

*Name (1)* : Aisyah Tri Deanita  
*Student Number* : 41820010122  
*Name (2)* : Denis Nila Cahyani  
*Student Number* : 41820010111  
*Name (3)* : Muhammad Anand Rizki Andinta  
*Student Number* : 41820010069  
*Counsellor* : Dr. Bambang Jekonowo, S.Si., M.T.I.  
*Title* : Penerapan Data Mining Menggunakan Algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk Klasifikasi Kemiskinan di DKI Jakarta

*Poverty remains one of the fundamental problems that is difficult to overcome, including in DKI Jakarta Province as the capital of Indonesia, which is also not immune to poverty. This research uses data mining to classify the poverty level in DKI Jakarta Province with data obtained from the Central Bureau of Statistics (BPS). The purpose of this study is to effectively and accurately classify poverty data and compare the results of the Naïve Bayes, KNN, and C4.5 algorithms to find the best accuracy results. Through the application of such algorithms, the results were obtained that the performance of the algorithm could be influenced by the division of data, and increased ratio of test data tended to improve the accuracy as well as consistency of classification results. At the data division ratio of 70:30, Naïve Bayes achieved accuracy of 81%, C4.5 of 76%, and K-NN of 71%. At the ratio of 80:20, Naïve Bayes showed accurateness of 93%, C4.5 of 79%, and K-NN of 86%. Whereas at the ratios of 90:10, Naïve Bayes achieves accurateness of 100%, C4.5 of 71%, and K-NN of 86%. By considering the performance variations of the third algorithm, it can be concluded that Naïve Bayes is superior as a stable and reliable algorithm in a variety of data set splitting scenarios, showing high accuracy and good ability in identifying positive cases.*

*Keywords:*

*poverty, classification, Naïve Bayes, C4.5, K-Nearest Neighbor*



## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	i
<b>HALAMAN PERNYATAAN KARYA SENDIRI</b> .....	ii
<b>HALAMAN PENGESAHAN</b> .....	iii
<b>KATA PENGANTAR</b> .....	iv
<b>HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI</b> .....	v
<b>ABSTRAK</b> .....	vi
<b>ABSTRACT</b> .....	vii
<b>DAFTAR ISI</b> .....	viii
<b>DAFTAR TABEL</b> .....	x
<b>DAFTAR GAMBAR</b> .....	xi
<b>DAFTAR LAMPIRAN</b> .....	xiii
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Permasalahan.....	3
1.3 Tujuan Penelitian .....	3
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian .....	4
1.6 Sistematika Penulisan .....	4
<b>BAB II TINJAUAN PUSTAKA</b> .....	6
2.1 Teori / Konsep Terkait.....	6
2.1.1 Dataset.....	6
2.1.2 Data Mining .....	6
2.1.3 Klasifikasi .....	7
2.1.4 Model Algoritma .....	7
2.1.5 Jupyter .....	8
2.1.6 Python .....	8
2.2 Penelitian Terdahulu.....	9

2.3 Analisa Literature Review.....	22
<b>BAB III METODE PENELITIAN .....</b>	<b>23</b>
3.1 Deskripsi Sumber Data .....	23
3.2 Teknik Pengumpulan Data .....	23
3.3 Diagram Alir Penelitian.....	24
3.4 Jadwal Penelitian.....	32
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>33</b>
4.1 HASIL .....	33
4.2 Mendefinisikan ground truth.....	34
4.3 Mengumpulkan data dan pre-processing .....	34
4.4 Mengembangkan model.....	43
4.4.1 Naïve Bayes .....	44
4.4.2 C4.5 .....	45
4.4.3 K-Nearest Neighbor.....	49
4.5 Evaluasi dan Tuning.....	51
4.5.1 Naïve Bayes .....	52
4.5.2 C4.5.....	54
4.5.3 K-Nearest Neighbor.....	56
4.6 PEMBAHASAN .....	58
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>62</b>
5.1 Kesimpulan .....	62
5.2 Saran.....	63
<b>DAFTAR PUSTAKA .....</b>	<b>64</b>
<b>LAMPIRAN.....</b>	<b>69</b>

## DAFTAR TABEL

<b>Tabel 2.1</b> <i>Literature Review</i> .....	21
<b>Tabel 3.1</b> Pengukuran Kinerja <i>Confusion Matrix</i> .....	30
<b>Tabel 3.2</b> Jadwal Penelitian .....	32
<b>Tabel 4.1</b> Hasil predict_proba pada data uji 30%, 20%, dan 10% .....	51
<b>Tabel 4.2</b> Dimensi masing-masing split data latih dan data uji.....	59
<b>Tabel 4.3</b> Hasil evaluasi kinerja ketiga algoritma.....	60



## DAFTAR GAMBAR

<b>Gambar 3.1</b> Tahapan Penelitian .....	24
<b>Gambar 4.1</b> Diagram MDLC .....	33
<b>Gambar 4.2</b> Load data .....	34
<b>Gambar 4.3</b> Proses replace missing value .....	35
<b>Gambar 4.4</b> Menampilkan data sebelum mengubah cell yang kosong .....	36
<b>Gambar 4.5</b> Proses mengubah pada cell yang kosong.....	37
<b>Gambar 4.6</b> Menampilkan data setelah perubahan pada cell yang sebelumnya kosong.....	38
<b>Gambar 4.7</b> Menampilkan jumlah cell yang sudah tidak kosong.....	39
<b>Gambar 4.8</b> Proses transformation data.....	39
<b>Gambar 4.9</b> Menampilkan data pada skala yang sama.....	40
<b>Gambar 4.10</b> Proses pembersihan kolom numerik dari karakter non-numerik...	41
<b>Gambar 4.11</b> Proses memberikan label berdasarkan batas.....	41
<b>Gambar 4.12</b> Menampilkan hasil perubahan pada data label .....	42
<b>Gambar 4.13</b> Proses pengambilan semua kolom kecuali kolom target .....	43
<b>Gambar 4.14</b> Proses split data latih dan data uji.....	43
<b>Gambar 4.15</b> Proses mengetahui dimensi data latih dan data uji .....	43
<b>Gambar 4.16</b> Proses import library pada Jupyter.....	44
<b>Gambar 4.17</b> Proses inisiasi model Naïve Bayes .....	44
<b>Gambar 4.18</b> Menampilkan score Naïve Bayes pada data uji 30%.....	44
<b>Gambar 4.19</b> Menampilkan score Naïve Bayes pada data uji 20%.....	44
<b>Gambar 4.20</b> Menampilkan score Naïve Bayes pada data uji 10%.....	45
<b>Gambar 4.21</b> Proses inisiasi model Decision Tree C4.5.....	45
<b>Gambar 4.22</b> Proses pembuatan pohon keputusan .....	45
<b>Gambar 4.23</b> Pohon keputusan pada data uji 30% .....	46
<b>Gambar 4.24</b> Pohon keputusan pada data uji 20% .....	47
<b>Gambar 4.25</b> Pohon keputusan pada data uji 10% .....	48
<b>Gambar 4.26</b> Proses inisiasi model K-Nearest Neighbor .....	49
<b>Gambar 4.27</b> Hasil <code>y_pred_knn</code> pada data uji 30% .....	50
<b>Gambar 4.28</b> Hasil <code>y_pred_knn</code> pada data uji 20% .....	50

<b>Gambar 4.29</b> Hasil $y_{pred\_knn}$ pada data uji 10%.....	50
<b>Gambar 4.30</b> Proses evaluasi model Naïve Bayes .....	51
<b>Gambar 4.31</b> Proses confusion matrix yang akan ditampilkan dalam bentuk heatmap.....	52
<b>Gambar 4.32</b> Hasil Confusion Matrix model Naïve Bayes pada data uji 30% ...	52
<b>Gambar 4.33</b> Hasil Confusion Matrix model Naïve Bayes pada data uji 20% ...	53
<b>Gambar 4.34</b> Hasil Confusion Matrix model Naïve Bayes pada data uji 10% ...	54
<b>Gambar 4.35</b> Hasil Confusion Matrix model C4.5 pada data uji 30%.....	54
<b>Gambar 4.36</b> Hasil Confusion Matrix model C4.5 pada data uji 20%.....	55
<b>Gambar 4.37</b> Hasil Confusion Matrix model C4.5 pada data uji 10%.....	56
<b>Gambar 4.38</b> Hasil Confusion Matrix model K-Nearest Neighbor pada data uji 30%.....	56
<b>Gambar 4.39</b> Hasil Confusion Matrix model K-Nearest Neighbor pada data uji 20%.....	57
<b>Gambar 4.40</b> Hasil Confusion Matrix model K-Nearest Neighbor pada data uji 10%.....	58

## DAFTAR LAMPIRAN

<b>Lampiran 1</b> Kartu Bimbingan.....	69
<b>Lampiran 2</b> Biodata.....	72

