

UNIVERSITAS NASIONAL



**PROGRAM STUDI MAGISTER TEKNOLOGI INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS NASIONAL
2024**

"Analisis Perbandingan Algoritma KNN dan Naïve Bayes dalam klasifikasi data mining menggunakan data sosial ekonomi di Indonesia "

PROPOSAL TESIS

Untuk memenuhi sebagian persyaratan memperoleh gelar
Magister Teknologi Informasi



Di susun oleh:
Kiki Buhori
227064518029

Pembimbing
Dr. Andrianingsih, S.Kom., M.M.S.I

PROGRAM STUDI MAGISTER TEKNOLOGI INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS NASIONAL
2024

HALAMAN PENGESAHAN

TESIS

ANALISIS PERBANDINGAN ALGORITMA KNN DAN NAÏVE BAYES
DALAM KLASIFIKASI DATA MINING MENGGUNAKAN DATA SOSIAL
EKONOMI DI INDONESIA



Kiki Buhori

227064518029

Dosen Pembimbing

A handwritten signature in black ink, appearing to be 'A. Andrianingsih', written over a light blue circular stamp or watermark.

Dr. Andrianingsih, S.Kom., MMSI

NIDN. 0303097902

LEMBAR PERSETUJUAN REVIEW AKHIR

Tesis dengan judul :

**ANALISIS PERBANDINGAN ALGORITMA KNN DAN NAÏVE BAYES
DALAM KLASIFIKASI DATA MINING MENGGUNAKAN DATA SOSIAL
EKONOMI DI INDONESIA**

Dibuat untuk melengkapi salah satu persyaratan menjadi Magister Komputer pada Program Studi Magister Teknolgo Informasi Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tesis ini diujikan pada Sidang Akhir Semester Genap 2023-2024 pada tanggal 24 Agustus Tahun 2024

Dosen Pembimbing

Dr. Andrianingsih, S.Kom., MMSI

NIDN. 0303097902

Ketua Program Studi

Dr. Astul Sani, S.T., M.T., M.Kom., Pd.D

NIDN. 0303067003



LEMBAR PENGESAHAN

JUDUL : ANALISIS PERBANDINGAN ALGORITMA KNN DAN NAÏVE BAYES DALAM KLASIFIKASI DATA MINING MENGGUNAKAN DATA SOSIAL EKONOMI DI INDONESIA

Diajukan untuk memenuhi sebagai persyaratan

Memperoleh gelar Magister Komputer

Disusun Oleh :

Kiki Buhori

227064518029

Tesis ini telah diuji dan dinyatakan lulus pada,

24 Agustus 2024

Telah diperiksa dan disetujui oleh :

Dosen Pembimbing



Dr. Andrianingsih, S.Kom., MMSI
NIDN. 0419068604

Ketua Program Studi



Ir. Asrul Sani, S.T., M.T., M.Kom., Ph.D
NIDN. 0303067003

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah Tesis ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah Tesis ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia Tesis ini digugurkan dan gelar akademik yang telah saya peroleh (Magister) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).



Jakarta, 26 November 2023



Kiki Buhori
227064518029

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT, yang telah memberikan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan tesis ini dengan judul “Analisis Perbandingan Algoritma KNN dan Neavy Bayes dalam klasifikasi data mining menggunakan data sosial ekonomi di Indonesia dengan pemrograman python”. Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister Teknologi Informasi di Universitas Nasional.

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada nama – nama tersebut.

1. Ibu Dr. Andrianingasih, S.Kom., M.M.S.I Selaku pembimbing tesis Program Studi Magister Teknologi Komunikasi dan Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional.
2. Bapak Dr. Agung Triayudi, S.Kom., M.Kom. Selaku Dekan dan Ketua Program Studi S2 Teknologi Informasi, Universitas Nasional beserta jajarannya yang telah mendedikasikan ilmunya kepada penulis sehingga penulis dapat menyelesaikan Thesis ini,
3. Bapak Dr. El Amry Bermawi Putera, M.A, selaku Rektor Universitas Nasional,
4. Bapak dan Ibu Dosen Fakultas Teknologi Komunikasi dan Informatika serta tenaga pendidik yang telah membantu proses perkuliahan,
5. Kepada kedua orang tua yang telah memberikan nasihat, masukan dan semangat untuk menuntut ilmu,
6. Terima kasih tak terhingga untuk rekan Mahasiswa Pascasarjana Magister Teknologi Informasi Angkatan II 2022, atas masukan, semangat dan kebersamaan selama menjalankan perkuliahan. Semoga ilmu yang kita dapatkan menjadi berkah dan bermanfaat.
7. Terima kasih kepada seluruh pihak yang telah membantu penulisan tesis ini sehingga dapat dipaparkan dengan baik.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna dan masih banyak kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari semua pihak yang berkepentingan. Semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi perkembangan ilmu pengetahuan umumnya.

Jakarta, Januari 2024

Penulis

ABSTRAK

Ekonomi global terus pulih seiring dengan membaiknya arus perdagangan, lapangan kerja, serta pendapatan. Namun, pemulihan ekonomi tidak berjalan merata di seluruh negara dan sektor usaha. Perbaikan ekonomi juga berdampak pada perubahan struktural, yang berarti bahwa beberapa sektor, pekerjaan, teknologi, dan perilaku tidak akan kembali ke tren sebelum pandemi. Perkembangan ke depannya bergantung pada kondisi ekonomi setempat. Ekonomi memiliki aspek terpenting didalam suatu negara yang dimana ekonomi menjadikan suatu negara mampu dalam memenuhi kebutuhanya dengan memanfaatkan sumber daya yang terbatas. Penelitian ini bertujuan untuk membandingkan dua algoritma klasifikasi data mining, yaitu Naïve Bayes dan K-Nearest Neighbor, dalam menganalisis data sosial ekonomi di Indonesia. Berdasarkan permasalahan tersebut, metode klasifikasi data mining digunakan dalam menentukan algoritma yang cocok untuk prediksi data sosial ekonomi di Indonesia. Dua algoritma yang digunakan adalah K-NN dan Naive Bayes. Setelah dilakukan pengujian kedua algoritma tersebut menggunakan confusion matrix dan K-Fold Cross Validation didapatkan hasil yang didapatkan dari kedua model memiliki akurasi Naïve Bayes 98,25% dan K-NN 97,78% serta hasil K-Fold Cross Validation Naïve Bayes 98% dan K-NN 96%. Naïve Bayes lebih unggul dibandingkan K-NN dalam konteks klasifikasi data sosial ekonomi di Indonesia ini, terutama dalam hal akurasi. Walaupun K-NN menunjukkan konsistensi yang baik, Naïve Bayes memberikan hasil yang lebih akurat di seluruh skenario Train-Test Split.

Kata Kunci : Klasifikasi, Sosial Ekonomi, Data Mining, K-NN, Naïve Bayes, Python.



ABSTRAK

The global economy continues to recover as trade flows, employment, and incomes improve. However, the economic recovery is uneven across countries and business sectors. The economic recovery has also resulted in structural changes, meaning that some sectors, jobs, technologies and behaviors will not return to pre-pandemic trends. Future developments depend on local economic conditions. The economy has the most important aspect in a country where the economy makes a country capable of meeting its needs by utilizing limited resources. This study aims to compare two data mining classification algorithms, namely Naïve Bayes and K-Nearest Neighbor, in analyzing socio-economic data in Indonesia. Based on this problem, the data mining classification method is used in determining the algorithm that is suitable for predicting socio-economic data in Indonesia. The two algorithms used are K-NN and Naive Bayes. After testing the two algorithms using confusion matrix and K-Fold Cross Validation, the results obtained from the two models have an accuracy of Naïve Bayes 98.25% and K-NN 97.78% and the results of K-Fold Cross Validation Naïve Bayes 98% and K-NN 96%. Naïve Bayes is superior to K-NN in this context of socioeconomic data classification in Indonesia, especially in terms of accuracy. Although K-NN showed good consistency, Naïve Bayes provided more accurate results across the Train-Test Split scenario.

Keywords: Classification, Socio-economic, Data Mining, K-NN, Naïve Bayes, Python.



DAFTAR ISI

PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN ORISINALITAS	Error! Bookmark not defined.
KATA PENGANTAR	Error! Bookmark not defined.
ABSTRAK	Error! Bookmark not defined.
DAFTAR ISI	i
DAFTAR GAMBAR	iii
DAFTAR TABEL	4
BAB I PENDAHULUAN	Error! Bookmark not defined.
1.1. Latar Belakang.....	Error! Bookmark not defined.
1.2. Rumusan Masalah.....	Error! Bookmark not defined.
1.3. Tujuan Penelitian.....	Error! Bookmark not defined.
1.4. Manfaat Penelitian.....	Error! Bookmark not defined.
1.5. Batasan Masalah.....	Error! Bookmark not defined.
1.6. Sistematika Pembahasan.....	Error! Bookmark not defined.
BAB II LANDASAN KEPUSTAKAAN	Error! Bookmark not defined.
2.1. Studi Pustaka.....	Error! Bookmark not defined.
2.1.1. Data Mining.....	Error! Bookmark not defined.
2.1.2. Klasifikasi.....	Error! Bookmark not defined.
2.2. Machine Learning.....	Error! Bookmark not defined.
2.2.1.1. Algoritma Naïve Bayes.....	Error! Bookmark not defined.
2.2.1.2. Algoritma K-Nearest Neighbor.....	Error! Bookmark not defined.
2.3. Perbandingan Studi Literatur.....	Error! Bookmark not defined.
2.4. Pemrograman Python.....	Error! Bookmark not defined.
2.5. Streamlit.....	Error! Bookmark not defined.
2.6. Cross-Industry Standard Process for Data Mining (CRIDP-DM) ...	Error! Bookmark not defined.
2.6.1. Business Understanding.....	Error! Bookmark not defined.
2.6.2. Data Understanding.....	Error! Bookmark not defined.
2.6.3. Data Preparation.....	Error! Bookmark not defined.
2.6.4. Modeling.....	Error! Bookmark not defined.
2.6.5. Evaluation.....	Error! Bookmark not defined.

2.6.6. Deployment.....	Error! Bookmark not defined.
2.7. Validasi dan Evaluasi.....	Error! Bookmark not defined.
2.7.1. Confusion Matrix.....	Error! Bookmark not defined.
2.7.2. K-Fold Cross Validation	Error! Bookmark not defined.
BAB III METODOLOGI PENELITIAN	Error! Bookmark not defined.
3.1. Metode Penelitian	Error! Bookmark not defined.
3.2. Business Understanding.....	Error! Bookmark not defined.
3.3. Data Understanding.....	Error! Bookmark not defined.
3.4. Data Preparation.....	Error! Bookmark not defined.
3.5. Modeling.....	Error! Bookmark not defined.
3.6. Evaluation	Error! Bookmark not defined.
BAB IV HASIL.....	Error! Bookmark not defined.
4.1. Dataset (Sosial Ekonomi).....	Error! Bookmark not defined.
4.2. Business Understanding.....	Error! Bookmark not defined.
4.3. Data Understanding.....	Error! Bookmark not defined.
4.3.1. Variabel Data Penelitian	Error! Bookmark not defined.
4.3.2. Normalisasi Data.....	Error! Bookmark not defined.
4.3.3. Labeling Data.....	Error! Bookmark not defined.
4.3.4. Pembagian Data Train dan Test.....	Error! Bookmark not defined.
4.4. Modeling.....	Error! Bookmark not defined.
4.5. Evaluasi.....	Error! Bookmark not defined.
4.5.1. Evaluasi Confusion Matrix	Error! Bookmark not defined.
4.5.2. Evaluasi Model K-Fold Cross Validation	Error! Bookmark not defined.
BAB V PEMBAHASAN	Error! Bookmark not defined.
BAB V PENUTUP.....	Error! Bookmark not defined.



DAFTAR GAMBAR

Gambar 1. Pertumbuhan Ekonomi Lapangan Usaha	Error! Bookmark not defined.
Gambar 2. Metode CRIPS – DM	Error! Bookmark not defined.
Gambar 3. Alur Penelitian.....	Error! Bookmark not defined.
Gambar 4. Data Sosial Ekonomi	Error! Bookmark not defined.
Gambar 5. Variabel Dataset Penelitian.....	Error! Bookmark not defined.
Gambar 6. Normalisasi Data	Error! Bookmark not defined.
Gambar 7. Proses Labeling	Error! Bookmark not defined.
Gambar 8. Data Train dan Test	Error! Bookmark not defined.
Gambar 9. Modeling Naïve Bayes dan KNN	Error! Bookmark not defined.
Gambar 10. Hasil Confusion Matrix K-Nearest Neighbor (KNN) dan Naïve Bayes (80/20)	Error! Bookmark not defined.
Gambar 11. Hasil Classification Report K-Nearest Neighbor (KNN) dan Naïve Bayes (80/20)	Error! Bookmark not defined.
Gambar 12. Hasil Confusion Matrix K-Nearest Neighbor (KNN) dan Naïve Bayes (70/30)	Error! Bookmark not defined.
Gambar 13. Hasil Classification Report K-Nearest Neighbor (KNN) dan Naïve Bayes (70/30)	Error! Bookmark not defined.
Gambar 14. Hasil Confusion Matrix Naïve Bayes (60/40).....	Error! Bookmark not defined.
Gambar 15. Hasil Classification K-Nearest Neighbor (KNN) dan Naïve Bayes (60/40)....	Error! Bookmark not defined.
Gambar 16. Hasil Confusion K-Nearest Neighbor (KNN) dan Naïve Bayes (50/50).....	Error! Bookmark not defined.
Gambar 17. Hasil Classification K-Nearest Neighbor (KNN) dan Naïve Bayes (50/50)....	Error! Bookmark not defined.
Gambar 18. Hasil K-Fold Cross Validation K-Nearest Neighbor dan Naïve Baye.....	Error! Bookmark not defined.
Gambar 19 Hasil Akurasi.....	Error! Bookmark not defined.
Gambar 20 Hasil Akuras K-Fold Validation	Error! Bookmark not defined.

DAFTAR TABEL

- Table 1. Perbandingan Studi Literatur **Error! Bookmark not defined.**
Table 2. Kelebihan dan Kekurangan algoritma K-Nearest Neighbor dan Naïve Bayes **Error! Bookmark not defined.**
Table 3. Confusion matrix **Error! Bookmark not defined.**

