

**PENERAPAN ALGORITMA *K-NEAREST NEIGHBORS* (KNN) DAN
NAÏVE BAYES UNTUK MENENTUKAN PEMILIHAN PENERIMA
BANTUAN SOSIAL BERDASARKAN EKONOMI MASYARAKAT**

SKRIPSI

DISUSUN OLEH:

DINDA AMELIA

2109010048



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2025

PENERAPAN ALGORITMA *K-NEAREST NEIGHBORS* (KNN) DAN *NAÏVE BAYES* UNTUK MENENTUKAN PEMILIHAN PENERIMA BANTUAN SOSIAL BERDASARKAN EKONOMI MASYARAKAT

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara

DINDA AMELIA

NPM : 2109010048

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2025

LEMBAR PENGESAHAN

Judul Skripsi : PENERAPAN ALGORITMA *K-NEAREST NEIGHBORS* (KNN) DAN *NAÏVE BAYES* UNTUK MENENTUKAN PEMILIHAN PENERIMA BANTUAN SOSIAL BERDASARKAN EKONOMI MASYARAKAT

Nama Mahasiswa : DINDA AMELIA

NPM : 2109010048

Program Studi : SISTEM INFORMASI

Menyetujui
Komisi Pembimbing



(Ferdv Riza ST, M.Kom)
NIDN. 0103068901

Ketua Program Studi



(Dr. Firahmi Rizky, S.Kom., M. Kom.)
NIDN. 0116079201

Dekan



(Dr. Al-Khoyarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

PENERAPAN ALGORITMA *K-NEAREST NEIGHBORS* (KNN) DAN *NAÏVE BAYES* UNTUK MENENTUKAN PEMILIHAN PENERIMA BANTUAN SOSIAL BERDASARKAN EKONOMI MASYARAKAT

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 15 Oktober 2025

Yang membuat pernyataan



DINDA AMELIA
NPM : 2109010048

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : DINDA AMELIA
NPM : 2109010048
Program Studi : SISTEM INFORMASI
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**PENERAPAN ALGORITMA *K-NEAREST NEIGHBORS* (KNN) DAN
NAÏVE BAYES UNTUK MENENTUKAN PEMILIHAN PENERIMA
BANTUAN SOSIAL BERDASARKAN EKONOMI MASYARAKAT**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya

Medan, 15 Oktober 2025

Yang membuat pernyataan



DINDA AMELIA

NPM : 2109010048

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap	: Dinda Amelia
Tempat dan Tanggal Lahir	: Medan, 11 Agustus 2002
Alamat Rumah	: Jl. B Z HAMID Gg. Sado No 71 LK IX
Telepon/Faks/HP	: 082165409208
E-mail	: dameliaa2@gmail.com
Instansi Tempat Kerja	: Kantor Lurah Titi Kuning
Alamat Kantor	: Jl. B Z HAMID NO 38 Titi Kuning

DATA PENDIDIKAN

SD	: SDN 060927	TAMAT: 2014
SMP	: MTs Ex Pga Univa	TAMAT: 2017
SMA	: SMK YPK MEDAN	TAMAT: 2020

KATA PENGANTAR



Pendahuluan

Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Dr. Firahmi Rizky, S.Kom., M.Kom. Ketua Program Studi Sistem Informasi
4. Bapak Mahardika Adbi Prawira Tanjung, S.Kom., M.Kom. Sekretaris Program Studi Sistem Informasi
5. Pembimbing saya yaitu Bapak Ferdy Riza, S.Kom., M.Kom. Terima kasih telah membimbing penulis dari awal sampai akhir skripsian ini.
6. Kepada ibu penulis yang selalu dukung penulis dalam perkuliahan ini, yang selalu jadi penyemangat. Terima kasih atas doa yang tidak pernah putus dan kasih sayang yang sangat teramat tulus. Semua pencapaian ini penulis persembahkan sebagai wujud kecil dari rasa hormat dan cinta yang teramat besar untuk ibu. Hidup lebih lama ya ma.. biar tetap jadi penyemangat penulis selamanya.
7. Ucapan terima kasih penulis sampaikan kepada ayah atas dukungan apapun.
8. Kepada kedua teman penulis Annisa Qomariah dan Aisyah Ambroini, yang selalu support penulis, mendengarkan keluh kesah penulis, dan kasih saran untuk penulis, serta membantu penulis untuk menyiapkan skripsi ini. Penulis berharap pertemanan kita erat selamanya, saling tumbuh, saling peduli tanpa rasa iri,
9. Dan untuk diri penulis sendiri, penulis ucapkan amat teramat terima kasih karna sudah bertahan hidup sampai saat ini atas semua ujian hidup yang telah datang, penulis sangat bangga dengan diri penulis karena bisa menyelesaikan skripsi ini dengan tawa, tangis, ga tidur, stress, masalah hidup tiba tiba ada aja, disakitin 1 orang yang paling penulis sayang, semoga diri ini tidak berhenti disini, belajar tumbuh dan selalu menjadi pribadi yang lebih baik lagi, kedepan nya bisa menuhin

inner child, dapet kerjaan yang bagus, bisa beli rumah dan mobil, dan paling utama mau bahagiain mama. Aamiin.

10. Semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu yang telah membantu penyelesaian skripsi ini.

**PENERAPAN ALGORITMA *K-NEAREST NEIGHBORS* (K-NN)
DAN *NAÏVE BAYES* UNTUK MENENTUKAN PEMILIHAN
PENERIMA BANTUAN SOSIAL BERDASARKAN EKONOMI MASYARAKAT**

ABSTRAK

Salah satu upaya pemerintah untuk meningkatkan kesejahteraan masyarakat, terutama bagi keluarga kurang mampu, adalah program bantuan sosial. Namun, karena proses seleksi masih bersifat subjektif, seringkali sasaran penerima bantuan tidak tepat saat dilaksanakan. Tujuan penelitian ini adalah untuk menggunakan algoritma K-Nearest Neighbor (K-NN) dan Naive Bayes dalam sistem pendukung keputusan untuk menentukan pemilihan penerima bantuan sosial berdasarkan data masyarakat. Algoritma K-NN menghitung jarak kedekatan antara data baru dan data sebelumnya menggunakan metode Euclidean Distance, sedangkan algoritma Naive Bayes menggunakan pendekatan probabilistik yang bergantung pada peluang munculnya atribut pada setiap kelas. Beberapa kriteria digunakan, termasuk pendapatan, pekerjaan, jumlah tanggungan, kondisi rumah, dan status kepemilikan aset. Hasil pengujian menunjukkan bahwa kedua algoritma dapat dengan tepat mengklasifikasikan kelayakan penerima bantuan sosial, tetapi algoritma K-NN sedikit lebih baik daripada Naive Bayes dalam hal ini. Oleh karena itu, penggunaan kedua algoritma ini dapat membantu pihak terkait membuat keputusan yang objektif dan efisien untuk menentukan penerima bantuan sosial yang tepat sasaran

Kata Kunci: Sistem Pendukung Keputusan, K-NN, Naive Bayes, Bantuan Sosial, Data Mining

**APPLICATION OF K-NEAREST NEIGHBORS (K-NN)
AND NAÏVE BAYES ALGORITHMS TO DETERMINE THE SELECTION OF SOCIAL
ASSISTANCE RECIPIENTS BASED ON THE ECONOMY OF THE COMMUNITY**

ABSTRACT

One of the government's efforts to improve public welfare, particularly for low-income families, is through social assistance programs. However, because the selection process is subjective, recipients are often misdirected during implementation. The purpose of this study is to use the K-Nearest Neighbor (KNN) and Naive Bayes algorithms in a decision support system to determine the selection of social assistance recipients based on community data. The KNN algorithm calculates the proximity between new and previous data using the Euclidean distance method, while the Naive Bayes algorithm uses a probabilistic approach that relies on the probability of attributes appearing in each class. Several criteria are used, including income, occupation, number of dependents, housing condition, and asset ownership status. Test results indicate that both algorithms can accurately classify eligibility for social assistance recipients, but the KNN algorithm performs slightly better than Naive Bayes in this regard. Therefore, the use of these two algorithms can help stakeholders make objective and efficient decisions in determining the appropriate social assistance recipients.

Keywords: *Decision Support System, K-NN, Naive Bayes, Social Assistance, Data Mining*

DAFTAR ISI

DAFTAR ISI.....	i
DAFTAR TABEL	ii
DAFTAR GAMBAR.....	iii
BAB I. PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
BAB II LANDASAN TEORI	6
2.1 Bantuan Sosial.....	6
2.2 Kelurahan Titi Kuning	7
2.3 Konsep Bantuan Pangan Non Tunai	8
2.4 Data Mining	9
2.4.1 Pengertian Data Mining	9
2.4.2 Teknik Data Mining	10
2.4.3 Penerapan Data Mining pada Bantuan Sosial	12
2.5 Klasifikasi	14
2.6 Sistem Pendukung Keputusan	15
2.7 <i>Unified Modeling Language</i> (UML).....	18
2.7.1 Algoritma K- Nearest Neighbors	23
2.7.2 Pengertian Algoritma dan Konsep Alogirtma K-NN	23
2.7.3 Rumus Algoritma K-NN	25

2.7.4	Phyton dalam Algoritma K-NN	26
2.8	Algoritma Naïve Bayes	27
2.8.1	Pengertian dan Konsep Algoritma Naïve Bayes	27
2.8.2	Rumus Naïve Bayes	29
2.8.3	Perbandingan Metode Algoritma	30
BAB III METODOLOGI PENELITIAN.....		32
3.1	Pendekatan Penelitian.....	32
3.2	Pengumpulan Data.....	33
3.3	Preprocessing Data	36
3.4	Implementasi Algoritma	36
3.5	Waktu dan tempat penelitian	41
BAB IV HASIL DAN UJI COBA		43
4.1	Hasil	43
4.2	Pembahasan	49
BAB V PENUTUP		72
5.1	Kesimpulan.....	72
5.2	Saran.....	75
DAFTAR PUSTAKA		

DAFTAR TABEL

Table 2.1. Use Case Diagram UML	18
Table 2.2. Activity Diagram UML	20
Table 2.3. Class diagram UML	23
Table 2.4. Perbandingan Metode K-NN dan Naïve Bayes.....	31
Tabel 3.1. Data Latih Kelurahan Titi Kuning.....	33
Tabel 4.3 Blackbox Testing Form Login	67
Tabel 4.4 Blackbox Testing Form Menu.....	68
Tabel 4.5 Blackbox Testing Form Masyarakat	69
Tabel 4.6 Blackbox Testing Form Dataset Latih.....	70
Tabel 4.7 Blackbox Testing Form Prediksi.....	70
Tabel 4.8 Blackbox Testing Form Metode.....	70

DAFTAR GAMBAR

Gambar 2.1. Peta Kelurahan Titi Kuning.....	7
Gambar 3.1. Flowchart Algoritma K-NN	38
Gambar 3.2 Flowchart Algoritma Naïve Bayes	40
Gambar 4.1 Form Login.....	43
Gambar 4.2 Form Menu	44
Gambar 4.3 Form Masyarakat.....	45
Gambar 4.4 Form Dataset Latih.....	46
Gambar 4.5 Form Prediksi	47
Gambar 4.6 Form Login.....	48
Gambar 4.7 Form Metode	48

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Masyarakat dianggap miskin jika mereka tidak punya kemampuan untuk memenuhi kebutuhan dasar seperti tempat tinggal, pendidikan, makanan, pakaian, dan kesehatan secara cukup. Untuk menurunkan angka kemiskinan di tahun berikutnya, pemerintah telah mendorong berbagai program dan upaya untuk menyelesaikan masalah itu. (Fauziyyah & Murnawan, 2024). Bantuan sosial adalah salah satu cara pemerintah yang bertujuan untuk mengurangi kemiskinan di Indonesia. (Fadhli & Nazila, 2023).

Program ini memegang peranan penting dalam mengurangi tingkat kemiskinan dan memperbaiki kesejahteraan sosial masyarakat, terutama di wilayah kelurahan. Sistem seleksi manual yang digunakan saat ini sering kali bergantung pada penilaian subjektif, yang dapat memengaruhi akurasi dan transparansi proses seleksi. Faktor-faktor seperti kurangnya data yang terstruktur, keterbatasan sumber daya, dan ketidakpastian dalam menetapkan kriteria penerima bantuan sosial menjadi kendala signifikan. Akibatnya, terdapat kemungkinan bahwa bantuan tidak tepat sasaran, di mana masyarakat yang sebenarnya layak menerima bantuan justru terlewatkan, sementara masyarakat yang tidak memenuhi kriteria memperoleh bantuan. (Sitanyah et al., 2024). Untuk dapat membantu hal tersebut, perlu didukung suatu penerapan algoritma. Algoritma ini cocok untuk data numerik, misalnya data umum penerima, data ekonomi dan sejenisnya. Tapi algoritma ini harus membandingkan data satu per

satu, kalau jumlah datanya besar, proses prediksi bisa jadi lambat. Metode ini didasarkan pada konsep outranking dengan membandingkan setiap alternatif secara berpasangan berdasarkan kriteria yang relevan. Metode tersebut dijalankan menggunakan Algoritma K-Nearest Neighbors (K-NN) dan Naïve Bayes.(Siringoringo, 2023).

Bantuan ini diberikan agar dapat meringankan masyarakat yang bekerja di sektor pemerintahan. Dengan demikian, sistem yang sudah ada mengalami beberapa masalah terkait penyimpanan data yang jumlahnya banyak.. (Siregar, 2022). Sistem ini diharapkan dapat membantu pihak desa dalam menentukan penerima bantuan yang benar-benar memenuhi kriteria dengan cara yang lebih adil dan efisien. Selain itu, penerapan algoritma ini juga dapat meningkatkan transparansi dan kepercayaan masyarakat terhadap program bantuan sosial di kelurahan tersebut, (Sitanyah et al., 2024).

Algoritma K-NN dan Naïve Bayes dipilih sebagai metode klasifikasi karena keduanya memiliki tingkat akurasi yang cukup tinggi. Nilai prediksi dari query ditentukan berdasarkan klasifikasi dari tetangga terdekat berdasarkan penjelasan sebelumnya (Alfiah, 2021) Dengan hasil penelitian bahwasannya Algoritma K-NN dapat digunakan untuk melakukan penerimaan bantuan sosial. Sedangkan algoritma Naive Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. (Aripin, 2019). Naive Bayes menentukan kelas yang paling optimal dan berfungsi menghitung dan mencari nilai probabilitas paling tinggi untuk mengklasifikasikan sebuah data uji .(Tangkelayuk, 2022).

Tujuan akhir dari pembuatan web bansos dengan algoritma K-NN dan Naive Bayes adalah supaya penyaluran bantuan sosial bisa lebih tepat sasaran dan

adil. Dengan sistem ini, data calon penerima bisa diproses otomatis untuk mengetahui siapa yang layak atau tidak menerima bantuan. K-NN berguna untuk membandingkan data pendaftar baru dengan data lama yang sudah diketahui statusnya, sedangkan Naive Bayes dipakai untuk memprediksi kelayakan berdasarkan data seperti penghasilan, jumlah tanggungan, dan kondisi ekonomi. Lewat web, proses pendaftaran jadi lebih mudah, dan pengelola bisa melihat hasilnya lewat dashboard. Sistem ini membantu mengurangi kesalahan dan memastikan bantuan sampai ke orang yang benar-benar membutuhkan.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka permasalahan dalam penelitian ini adalah sebagai berikut :

1. Bagaimana Algoritma K-Nearest Neighbors (KNN) menggunakan data karakteristik sosial dan ekonomi dalam pemilihan Bantuan Sosial?
2. Seberapa besar kontribusi fitur seperti : pendapatan, pekerjaan, dan kondisi rumah dalam klasifikasi menggunakan Naïve Bayes?
3. Apa saja faktor yang paling mempengaruhi dalam kedua algoritma tersebut?

1.3. Batasan Masalah

Penelitian ini dibatasi pada program BANSOS di Kelurahan Titi Kuning, dengan fokus pada proses penyaluran, ketepatan sasaran, dan persepsi masyarakat,

1. Penelitian ini menggunakan data di Kantor Lurah Titi Kuning. Data yang diambil hanya data sosial ekonomi seperti ; Pendapatan, Kondisi Rumah, Jumlah tanggungan,
2. Data yang digunakan dalam penelitian ini diambil dari kelurahan Titi Kuning.

3. Algoritma yang digunakan dalam metode penelitian ini menggunakan Algoritma K-Nearest Neighbors dan Naïve Bayes.
4. Penelitian ini berfokus untuk menentukan layak atau tidak layak dalam pemilihan bantuan sosial.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisis pelaksanaan program BANSOS di Kelurahan Titi Kuning,

1. Menggunakan algoritma K- Nearest Neighbors dan Naïve Bayes untuk mengklasifikasikan data kandidat penerimaan bantuan sosial berdasarkan karakteristik ekonomi.
2. Menggunakan algoritma K-Nearest Neighbors dan Naïve Bayes untuk mengevaluasi kelayakan kandidat penerima bantuan sosial.
3. Menganalisis kebutuhan system dalam menentukan kelayakan penerima bantuan sosial secara digital dan berbasis data.

1.5. Manfaat Penelitian

Penelitian ini bermanfaat untuk membantu pemerintah mengevaluasi program bantuan sosial dan memberi informasi bagi masyarakat tentang pelaksanaannya di Kelurahan Titi Kuning.

1. Bagi Peneliti: Sebagai pengalaman dalam pengolahan data dan penerapan algoritma K- Nearest Neighbors (KNN) dan Naïve Bayes.
2. Bagi kelurahan Titi Kuning dan Masyarakat sekitar : Membantu dalam menentukan kelayakan penerimaan bantuan sosial secara cepat dan efisien,

serta meningkatkan keadilan bagi masyarakat dalam proses penerimaan bantuan lebih tepat sasaran.

3. Bagi Universitas: Hasil penelitian ini dapat menjadi referensi tambahan di perpustakaan dan perbandingan dalam penelitian lainya.

BAB II

LANDASAN TEORI

2.1. Bantuan Sosial

Bantuan sosial adalah bentuk dukungan yang diberikan oleh pemerintah atau lembaga kepada individu, keluarga, kelompok, atau komunitas. Bantuan ini bisa berupa uang atau barang, tergantung dari kemampuan keuangan pemerintah daerah atau lembaga tersebut. Tujuannya adalah untuk membantu mencapai tujuan program dan kegiatan pemerintah daerah, serta menerapkan prinsip keadilan, martabat, akal sehat, dan manfaat bagi masyarakat.

Pemerintah pusat sudah melakukan banyak upaya untuk membantu masyarakat yang kurang mampu, seperti program bantuan sosial berupa sembako dan uang, program bantuan rumah subsidi, serta berbagai usaha lainnya. Namun, masih ada beberapa masalah yang terjadi ketika program tersebut dijalankan. Salah satu masalah adalah dalam proses penyaluran bantuan, terutama di Kelurahan Titi Kuning. Penyaluran program bantuan sosial di sana belum cukup optimal dan merata. Hal ini terjadi karena proses pemberian bantuan masih menggunakan cara manual dalam memilih penerima manfaat, dan tidak ada sistem yang mendukung dalam mengidentifikasi calon penerima bantuan. Akibatnya, terjadi ketidak tepatan dalam pendistribusian bantuan sosial, dan khawatir program tersebut tidak sampai kepada orang-orang yang benar-benar membutuhkan. (Suprpto et al., 2024).

2.2. Kelurahan Titi Kuning

Kelurahan Titi Kuning adalah salah satu dari 6 (enam) kelurahan di wilayah Kecamatan Medan Johor dengan luas wilayah sebesar 181 hektar. Lingkungan ini memiliki batas-batas sebagai berikut: di sebelah utara berbatasan dengan Kelurahan Kampung Baru Kecamatan Medan Maimun, di sebelah selatan berbatasan dengan Kelurahan Kedai Durian Kecamatan Medan Johor, di sebelah timur berbatasan dengan Kelurahan Suka Maju dan Kelurahan Kedai Durian Kecamatan Medan Johor, sedangkan di sebelah barat berbatasan dengan Kelurahan Pangkalan Masyhur dan Kelurahan Gedung Johor Kecamatan Medan Johor. Berdasarkan data yang dikumpulkan pada akhir bulan April 2017, jumlah penduduk di Kelurahan Titi Kuning sebanyak 2.1532 orang. Secara umum, penduduk di Kelurahan Titi Kuning kebanyakan adalah WNI, dan sekitar 35% adalah WNA keturunan Tionghoa. Kebanyakan wilayah Kelurahan Titi Kuning berupa perkotaan, dengan aktivitas utama penduduk berupa perdagangan.



Gambar 2.1. Peta Kelurahan Titi Kuning.

Sumber: Pemko Medan Kelurahan Titi Kuning

2.3. Konsep Bantuan Pangan Non Tunai

Bantuan Pangan Non Tunai, disingkat BPNT, adalah bantuan sosial yang diberikan dalam bentuk uang elektronik oleh pemerintah kepada keluarga yang berhak menerima. Tujuan dari program ini adalah untuk membantu mengurangi beban pengeluaran keluarga penerima manfaat, memenuhi kebutuhan pangan sehari-hari, memberikan gizi yang lebih seimbang, mencocokkan bantuan dengan kebutuhan dan waktu yang tepat, memberi kebebasan lebih dalam memilih jenis bahan pangan, serta mendorong tercapainya tujuan pembangunan yang berkelanjutan (Sustainable Development Goals/SDGs). (pedoman Umum bpnt,2018:10) (Djaenal., 2021).

Program BPNT bertujuan untuk membantu mengurangi pengeluaran masyarakat terhadap kebutuhan pangan serta memberikan nutrisi yang seimbang kepada keluarga yang menerima manfaat secara tepat sasaran dan tepat waktu. Penyaluran bantuan dilakukan secara non tunai. Untuk mendukung pelaksanaan program ini, Presiden Republik Indonesia telah menerbitkan PERPRES RI Nomor 63 Tahun 2017 tentang Penyaluran BPNT. Program ini membantu mengurangi beban pengeluaran keluarga penerima manfaat dengan memenuhi sebagian kebutuhan pangan dan memberikan gizi yang seimbang kepada peserta KPM (Laurentcia & Yusran, 2021). Adanya Program BPNT diharapkan mampu memperbaiki pelayanan dalam pemberian bantuan kepada masyarakat miskin dengan tujuan penyaluran bantuan subsidi pangan dapat berjalan dengan lebih efektif dan efisien. (Anwar, 2020).

2.4. Data Mining

2.4.1. Pengertian Data Mining

Data mining adalah teknik yang digunakan dalam proses pengolahan data. Proses pengolahan data dalam data mining dilakukan berdasarkan data yang berasal dari masa lalu. Data yang sudah dikumpulkan dan merupakan data masa lalu kemudian diproses kembali untuk mendapatkan informasi baru. Informasi tersebut nantinya digunakan oleh pemilik data dalam mengambil keputusan. Dalam data mining, informasi yang diperoleh bisa dimanfaatkan oleh berbagai bidang ilmu.

Proses pengolahan data dalam data mining menggunakan berbagai teknik seperti klasifikasi, klasterisasi, prediksi, estimasi, dan asosiasi. Teknik-teknik tersebut digunakan sesuai dengan kebutuhan atau hasil yang diinginkan dari proses pengolahan data tersebut (Triayudi, 2023). Data Mining adalah proses menggali pengetahuan dan pola-pola yang tersembunyi dalam data dengan menggunakan metode statistik matematika dan machine learning. Penggunaan sistem informasi yang terdistribusi sangat berpengaruh dalam menjadikan sekumpulan data kecil menjadi data yang sangat

Seiring berjalannya waktu, data akan terus bertambah dari data yang ada saat ini dan data yang akan datang, sehingga terbentuk aliran data yang besar. Algoritma data mining berperan secara efektif dan efisien dalam menganalisis data yang jumlahnya besar. Pada data mining, *clustering* merupakan suatu teknik untuk menganalisis data. Tujuan dari *clustering* adalah untuk mempartisi data menjadi suatu kelompok. Metode clustering telah banyak digunakan di berbagai

bidang seperti bidang penelitian dan pemodelan identifikasi objek, pengolahan citra dan pemetaan zona wilayah. (Sekar Setyaningtyas et al., 2022). Sebagai bagian dari proses klasifikasi, Data Mining menggunakan Algoritma K- Nearest Neighbors dan Naïve Bayes yang merupakan salah satu teknik utama dalam Data Mining untuk memprediksi kategori data yang belum diketahui berdasarkan pola data historis. Data Mining adalah alat yang sangat efektif untuk menggali informasi tersembunyi dari kumpulan data yang sangat besar dan kompleks. Dengan teknik yang tepat, Data Mining dapat membantu organisasi memahami pola dalam data, membuat prediksi, mengelompokkan data, dan menemukan hubungan antar variabel. Dengan penerapan yang tepat, data mining dapat memberikan informasi yang sangat berharga untuk proses pengambilan keputusan yang lebih baik.

2.4.2. Teknik Data Mining

1. *Descriptive Data Mining* (Data Mining Deskriptif)

Descriptive Data Mining atau Data Mining Deskriptif adalah cabang dari data mining yang berfokus pada menemukan pola, struktur, hubungan, dan informasi yang tersembunyi dalam data tanpa membuat prediksi tentang apa yang terjadi di masa depan, tujuan utama dari teknik ini adalah untuk memahami dan merangkum karakteristik utama sehingga dapat digunakan untuk analisis atau pengambilan keputusan berbasis data. Deskriptif Data Mining lebih pada mengeksplorasi data secara menyeluruh untuk menemukan wawasan yang mungkin tidak terlihat secara langsung.

Biasanya, deskriptif data mining digunakan untuk:

- a. menemukan kelompok data yang serupa atau segmentasi data.
- b. mengevaluasi hubungan antara fitur

- c. mengidentifikasi data yang mengandung anomali atau outlier.
- d. menjelaskan demografi kelompok pelanggan.
- e. melihat data kompleks dalam bentuk diagram, grafik, atau peta.

Dibandingkan dengan perhitungan prediktif berbasis formula, deskriptif data mining, berkonsentrasi pada eksplorasi pola, ringkasan data, dan visualisasi dari pada menggunakan satu “rumus tunggal” seperti dalam matematika murni. Adapun beberapa teknik yang umum digunakan dalam *Descriptive Data Mining* yaitu; *Clustering* (pengelompokan), *Association Rule Mining* (Pencarian Aturan Asosiasi), *Summarization* (Perangkuman).

2. *Predictive Data Mining* (Data Mining Prediktif)

Predictive Data Mining atau Data Mining Prediktif adalah jenis data mining untuk memprediksi nilai atau kejadian di masa depan dari data historis yang sudah tersedia. *Predictive Data Mining* berfokus pada membuat model atau yang dapat memprediksi dari hasil data yang belum dilihat dari sebelumnya. Dalam pengambilan keputusan berbasis data. Data mining prediktif adalah metode analisis data yang menggunakan algoritma statistik, pembelajaran mesin, dan teknik pemodelan matematis untuk membuat model prediksi.

Model ini dibuat dengan mempelajari pola dan hubungan dalam data sebelumnya dan kemudian digunakan untuk memprediksi nilai variabel tertentu di masa depan, menggunakan variabel input (juga dikenal sebagai fitur atau atribut) untuk memprediksi variabel target (juga dikenal sebagai label atau output) adalah karakteristik data mining prediktif..

Teknik-teknik yang digunakan dalam *Predictive Data Mining*; *Classification* (Klasifikasi), *Regression* (Regresi). Adapun beberapa proses umum dalam

Predictive Data Mining yaitu; Pengumpulan Data, Praproses Data, Pemisahan Data, Permodelan. Karena pendekatan Data Mining Prediktif sangat bergantung pada metode yang dipilih seperti regresi, klasifikasi, dll, tidak ada rumus yang digunakan. Berdasarkan latar belakang dan beberapa penelitian sebelumnya, penulis akan melakukan pengklasifikasian terhadap proses penerimaan bantuan sosial. Data yang digunakan berasal dari Kantor Lurah Titi Kuning. Penelitian ini bertujuan untuk membandingkan dua algoritma yaitu K-Nearest Neighbor dan Naïve Bayes Classifier, serta mengetahui nilai akurasi masing-masing. Dengan demikian, penelitian ini diharapkan dapat membantu Kelurahan Titi Kuning dalam menentukan penerimaan bantuan sosial secara lebih efektif dan efisien, serta meningkatkan pemahaman tentang metode klasifikasi. (Loka & Marsal, 2023)

2.4.3. Penerapan Data Mining pada Bantuan Sosial

Data mining adalah alat canggih yang dapat membantu dari data bansos yang dapat menemukan siapa yang layak mendapatkan Bantuan Sosial dan mencegah siapapun dalam penyalahgunaan sistem. Data Mining perlu untuk bansos karena penyaluran bansos sering menghadapi tantangan seperti salah sasaran yang mungkin tidak sampai ke tangan yang membutuhkan. Adapun kecurangan seperti manipulasi data, pendaftaran ganda atau data fiktif untuk bantuan secara tidak sah. Data mining hadir untuk mengatasi masalah ini dengan memungkinkan pemerintah atau Lembaga sosial. Berikut ada beberapa contoh penerapan Data mining untuk bantuan sosial :

1. Klasifikasi Penerima Bansos (Prediksi Kelayakan)

Tujuan ini mengidentifikasi calon penerima yang paling membutuhkan berdasarkan kriteria tertentu dan dapat memprediksi kelayakan mereka. Data yang digunakan yaitu :

- a. Data demografi seperti; Usia, jenis kelamin, status perkawinan, jumlah anggota keluarga.
- b. Data Ekonomi : Pendapatan, jenis pekerjaan, kepemilikan asset seperti rumah/tanah/kendaraan, daya listrik terpasang.
- c. Data geografis : Alamat, lokasi, kondisi tempat tinggal.

Penerapan ini dapat memprediksi apakah seseorang cenderung “Layak” atau “Tidak Layak” menerima Bansos berdasarkan pola dari data history penerima yang sudah ada.

2. Klasterisasi calon penerima

Tujuan ini mengelompokkan calon penerima ke dalam segmen-segmen yang berbeda berdasarkan karakteristik dan kebutuhan yang serupa, untuk data yang digunakan sama seperti klasifikasi namun fokus pada variabel yang menggambarkan kebutuhan misalnya (kondisi rumah, jumlah tanggungan, status pekerjaan).

3. Deteksi kecurangan (Fraud Detection)

Mengidentifikasi pola-pola yang tidak biasa atau mencurigakan yang dapat mengindikasikan kecurangan atau salah sasaran dalam penyaluran bansos. Ada beberapa data yang digunakan yaitu; data penerima seperti data ganda, data interaksi atau penyaluran seperti penerimaan bansos yang melebihi batas. Adanya penerapan ini agar terhindar dari satu orang yang terdaftar berkali-kali,

mengidentifikasi penerima yang seharusnya tidak layak (misalnya memiliki asset mewah).

2.5. Klasifikasi

Klasifikasi adalah cara mencari sekelompok model atau fungsi yang bisa menjelaskan dan membedakan berbagai kelas data atau konsep, sehingga model tersebut bisa digunakan untuk memperkirakan kelas dari suatu objek yang kelasnya belum diketahui (Rani, 2016) Klasifikasi adalah sebuah proses untuk menemukan model yang menggambarkan dan membedakan kelas dari konsep data. Model ini diturunkan berdasarkan analisis satu set data pelatihan, model ini digunakan untuk memprediksi label kelas objek yang label kelasnya belum diketahui. (Nikmatun et al., 2019).

Klasifikasi adalah proses yang terbagi menjadi dua tahap, yaitu tahap pembelajaran dan tahap pengklasifikasian. Dalam tahap pembelajaran, algoritma klasifikasi akan membuat model klasifikasi dengan menganalisis data latihan. Tahap ini juga bisa dilihat sebagai proses membentuk fungsi atau pemetaan $Y = F(X)$, di mana Y adalah kelas hasil prediksi dan X adalah data yang ingin diprediksi kelasnya. Selanjutnya, pada tahap pengklasifikasian, model yang sudah dibuat digunakan untuk mengklasifikasikan data. Klasifikasi adalah proses mencari sekelompok model yang mampu menggambarkan dan membedakan kelas data, agar model tersebut bisa digunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya. (Heliyanti Susana, 2022).

Di bidang bantuan sosial, klasifikasi sangat berharga untuk secara otomatis menentukan kelayakan calon penerima, mempercepat proses, dan memastikan

bantuan disalurkan secara tepat sasaran. Mari kita selami bagaimana dua algoritma populer, K-Nearest Neighbors dan Naive Bayes, bekerja dalam konteks ini. K-Nearest Neighbors adalah algoritma yang sangat intuitif, beroperasi berdasarkan prinsip Ketika seorang calon penerima bantuan baru muncul, K-NN tidak membangun model yang kompleks di awal. Sebaliknya, ia langsung mencari sejumlah (K) data historis yang paling mirip dengan calon baru tersebut. Sedangkan Naive Bayes akan menghitung, "Berapa probabilitas pendaftar ini layak, mengingat dia berusia 30 tahun, berpendapatan Rp1.5 juta, dan memiliki 2 anak?" Kemudian akan membandingkannya dengan probabilitas bahwa dia "tidak layak" dengan fitur-fitur yang sama. Kelas dengan probabilitas tertinggi akan dipilih. Pemilihan algoritma yang tepat akan sangat bergantung pada karakteristik spesifik dataset Anda (ukuran, jenis fitur, korelasi antar fitur) dan prioritas Anda (kecepatan, akurasi, interpretasi model). Seringkali, mencoba dan membandingkan performa kedua algoritma ini bersama dengan algoritma lain.

2.6. Sistem Pendukung Keputusan

Sistem Pendukung Keputusan (SPK), juga disebut Sistem Pendukung Keputusan (DSS), adalah sistem informasi canggih yang dimaksudkan untuk membantu manajer, pejabat, dan pengambil keputusan dalam menyelesaikan masalah kompleks dan tidak terstruktur. Ini berbeda dari sistem komputer biasa yang hanya melakukan tugas-tugas biasa. SPK berfungsi sebagai alat bantu yang memberikan wawasan dan analisis mendalam untuk membantu Anda membuat pilihan yang lebih bijaksana yang didasarkan pada data. Sistem pendukung keputusan mempunyai 4 komponen yaitu :

1. *Data base* (Basis Data)

Dalam konteks SPK, database ini dirancang khusus untuk memenuhi kebutuhan informasi dalam proses pengambilan keputusan. Ini mencakup berbagai jenis data, antara lain:

- a. Data internal: data yang berasal dari system internal organisasi seperti data keuangan, data sumber daya manusia, dll.
- b. Data eksternal: Data yang berasal dari luar organisasi, misalkan data *competitor*, data ekonomi, dll.

Database ini dikelola oleh sebuah Sistem Manajemen Basis Data (DBMS), yang merupakan perangkat lunak yang memungkinkan pembuatan, akses, pembaruan, dan pengelolaan data dalam database secara efisien.

2. Model base (Basis Model)

Berisi model-model analitis dan matematis yang dipergunakan dalam memproses data dan membantu analisis keputusan. Contohnya seperti; Model *statistic*, model keuangan, dan model optimasi. Komponen ini membantu menyipulasikan scenario dan menganalisis dampaknya terhadap hasil keputusan.

3. *User Interface/UI*

User interface adalah komponen vital yang memungkinkan interaksi dua arah antara pengguna atau sistem pendukung keputusan. Fungsi utama *interface* menginput pengguna untuk memasukan berbagai jenis informasi kendala sistem seperti;

- a. Data atau parameter spesifik untuk analisis (misalnya: Batasan anggaran, kriteria keputusan)
- b. Pertanyaan atau kueri untuk mengambil database.

- c. Perintah atau instruksi untuk menjalankan fungsi tertentu dalam SPK.

Pentingnya *interface* dirancang sangat baik karena sangat krusial bagi SPK karena interface mudah dipelajari pengguna, mengurangi waktu dan usaha yang diperlukan pengguna untuk berinteraksi dengan sistem, dan akurasi input yang jelas dapat membantu mengurangi kesalahan input dari pengguna. Antarmuka yang efektif dalam SPK bantuan sosial bukan hanya alat teknis, tapi juga instrumen penting untuk memastikan keadilan, akuntabilitas, dan efisiensi dalam penyaluran bantuan kepada mereka yang benar-benar membutuhkan.

2.6.1. Fungsi SPK dalam seleksi penerima bantuan sosial

Salah satu fungsi SPK adalah Meningkatkan objektivitas dan akurasi dalam penetapan kelayakan merupakan tujuan utama SPK. Bayangkan sebuah organisasi yang harus memilih ribuan, jika tidak puluhan ribu, orang yang layak mendapatkan bantuan. Jika tidak ada SPK, proses ini rentan terhadap penilaian pribadi atau "favoritisme", yang dapat menyebabkan gagal mencapai tujuan. SPK mengikuti standar yang ketat, seperti pendapatan, jumlah anggota keluarga, kondisi rumah, status pekerjaan, dan aset yang dimiliki. Setiap kriteria diberi nilai khusus, dan SPK akan secara otomatis menghitung skor kelayakan setiap kandidat. Hasilnya adalah daftar penerima yang diurutkan berdasarkan skor. Ini memastikan bahwa bantuan diberikan hanya kepada mereka yang paling membutuhkan dan memenuhi syarat, bukan karena pertimbangan standar atau emosional.

SPK meningkatkan akurasi dan kecepatan proses seleksi. Proses pengambilan bantuan sosial secara manual seringkali memakan waktu sehari-hari, berminggu-minggu, bahkan berbulan-bulan, karena melibatkan tumpukan dokumen fisik dan perhitungan yang rentan terhadap kesalahan. SPK memiliki

kemampuan untuk mengotomatisasi sebagian besar pekerjaan ini. SPK dapat memproses ribuan data kandidat penerima dalam hitungan menit dengan data yang terorganisir dalam database. Ini menghemat waktu dan sumber daya manusia, tetapi juga.

SPK tidak hanya memberikan rekomendasi kepada penerima, tetapi juga membantu dalam analisis dan evaluasi program yang lebih mendalam. Dengan demikian, SPK bukan hanya alat seleksi, tetapi juga instrumen strategis untuk manajemen program bantuan sosial secara keseluruhan.

2.7. *Unified Modeling Language (UML)*

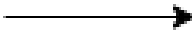
Unified Modeling Language atau lebih sering disingkat UML adalah Bahasa standart yang digunakan dalam membuat model, memvisualisasikan, membangun, dan mendokumentasikan sistem perangkat lunak. Ini adalah Bahasa pemodelan yang sangat agnostic terhadap pemrograman bisa di design menggunakan UML lalu mengimplementasikan nya menggunakan python, c# atau Bahasa lainnya. Tujuan utama dari UML adalah untuk menyederhanakan proses perancangan sistem yang kompleks.

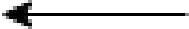
UML memiliki berbagai jenis diagram yang masing-masing berfungsi untuk tujuan tertentu. Contohnya termasuk diagram kelas untuk struktur statis, diagram kasus penggunaan untuk fungsionalitas sistem dari sudut pandang pengguna, dan diagram urutan untuk interaksi objek dari waktu ke waktu. Kombinasi dari jenis diagram ini memberikan gambaran lengkap tentang sistem yang akan atau sedang dikembangkan. Ada beberapa jenis UML yang digunakan yaitu :

- a. Use Case Diagram

Diagram Use Case adalah alat penting untuk UML yang memodelkan fungsionalitas sistem dari sudut pandang pengguna. Dalam konteks bantuan sosial diagram ini membantu kita memahami siapa saja yang berinteraksi dengan sistem dan tujuan interaksi mereka. Berikut adalah simbol-simbol utama yang digunakan dalam diagram use case UML :

Table 2.1. Use Case Diagram UML

Simbol	Nama	Keterangan
	Actor	Simbol ini bisa berupa orang, organisasi atau sistem lain. aktor ini memicu use case atau menerima hasil dari usecase
	Use Case	Simbol use case berfungsi layanan yang disediakan oleh sistem yang bernilai bagi aktor.
	Association	Simbol ini mengindikasikan bahwa seorang aktor berinteraksi dengan sebuah use case, ini adalah hubungan paling umum.
	Generalisasi	Simbol ini digunakan untuk menunjukkan hubungan spesialisasi/generalisasi antara aktor atau use case.

<<include>> 	Include	Simbol Include digunakan Ketika satu use case selalu memerlukan fungsionalitas dari use case lain untuk diselesaikan, use case yang di include adalah bagian yang esensial dari use case dasar.
<<extend>> 	Extend	Simbol extend digunakan Ketika satu use case dapat memperluas fungsionalitas tambahan bersifat opsional.

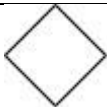
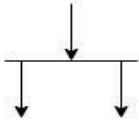
Sumber Nur Adela, 2024

b. Activity Diagram

Activity Diagram adalah salah satu diagram dalam UML yang digunakan untuk memodelkan aliran kerja (workflow) atau urutan aktivitas dalam suatu sistem atau proses dalam aktivitas diagram. Dalam diagram ini Langkah-langkah, keputusan, dan jalur paralel yang terlibat dalam eksekusi suatu tugas dapat digambarkan secara visual. Berikut adalah simbol-simbol yang digunakan dalam Activity Diagram :

Table 2.2. Activity Diagram UML

Simbol	Nama	Keterangan
	Activity	Mempresentasikan satu Langkah yang dilakukan dalam proses bisnis atau alur kerja.
	Initial Node	Simbol ini adalah penanda dimulainya aliran aktivitas dalam

		diagram. Setiap Activity Diagram wajib memiliki satu dan hanya satu Initial Node. Ini seperti "start button" untuk seluruh proses yang digambarkan.
	Final Node	Menandakan akhir dari aliran aktivitas. Sebuah Activity Diagram dapat memiliki satu atau lebih Final Node, yang berarti proses bisa berakhir di beberapa titik yang berbeda, tergantung pada jalur yang diambil.
	Decision Node	Digunakan ketika ada titik di mana proses dapat bercabang menjadi beberapa jalur berdasarkan suatu kondisi. Satu panah masuk ke diamond, dan dua atau lebih panah keluar.
	Fork	Menandakan bahwa aliran proses bercabang menjadi aktivitas-aktivitas paralel yang dapat dieksekusi secara bersamaan. masing-masing menuju aktivitas yang dapat berjalan secara independen.
	Join/Penggabungan	Digunakan untuk menyatukan kembali aktivitas-aktivitas paralel. Ini berarti semua aktivitas yang masuk ke Join Node harus selesai sebelum aliran proses dapat melanjutkan ke aktivitas berikutnya.

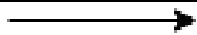
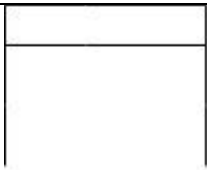
	Control Flow	Menunjukkan aliran kontrol dari satu aktivitas ke aktivitas berikutnya. Ini adalah jalur yang diikuti oleh eksekusi proses. Panah selalu mengarah dari aktivitas sebelumnya ke aktivitas setelahnya
---	--------------	---

Sumber: Nur Adela, 2024

c. Class Diagram

Class Diagram adalah salah satu diagram struktural yang paling fundamental dan banyak digunakan dalam Unified Modeling Language (UML). Diagram ini memvisualisasikan struktur statis dari suatu sistem dengan menunjukkan kelas yang ada di dalam sistem, atribut yang dimiliki setiap kelas metode atau perilaku yang dapat dilakukan oleh setiap kelas Dalam konteks bantuan sosial (bansos), Class Diagram tetap memegang peran yang sama seperti dalam pengembangan perangkat lunak pada umumnya: yaitu untuk memodelkan struktur statis dari sistem yang mengelola data bansos. Namun, penerapannya akan fokus pada entitas-entitas dan hubungan-hubungan spesifik yang relevan dengan domain bansos. Berikut beberapa simbol Class diagram :

Table 2.3. Class diagram UML

Simbol	Nama	Keterangan
	Generalisasi	Menggambarkan relasi antar kelas
	Class	Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama

	Nary Association	Upaya untuk menghindari asosiasi dengan lebih dari 2 objek.
---	------------------	---

Sumber: Nur Adela, 2024

2.8. Algoritma K- Nearest Neighbors

2.7.1. Pengertian Algoritma dan Konsep Algoritma K-NN

Algoritme klasifikasi tetangga terdekat, atau disebut juga K-NN, adalah cara untuk mengklasifikasikan suatu objek berdasarkan data yang sudah dipelajari. Data tersebut direpresentasikan dalam ruang dengan banyak dimensi, di mana setiap dimensi mewakili satu karakteristik atau fitur dari data tersebut. Ketika ada data baru yang ingin diklasifikasikan, algoritma ini mencari k tetangga terdekat dari data tersebut. Hasilnya, data baru akan mendapatkan label yang sama dengan label dari tetangga-tetangga terdekat yang paling banyak muncul. Jika nilai k adalah 1, maka data baru akan memiliki label yang sama dengan tetangga terdekatnya saja (Rahmadini et al., 2023). Algoritma K-Nearest Neighbor (K-NN) adalah salah satu cara paling mudah untuk menyelesaikan masalah klasifikasi dan biasanya menghasilkan hasil yang cukup baik dan berpengaruh. Dari masalah tersebut maka pada penelitian ini akan membuat sebuah website menggunakan metode K- NN sebagai penentuan penerima bantuan sosial dengan beberapa variabel yaitu umur, pekerjaan, penghasilan, jumlah tanggungan, dan kondisi rumah. Variabel tersebut sebagai tolak ukur penentuan penerima bantuan sosial. Maka dirumuskan beberapa permasalahan penelitian ini yaitu untuk mengetahui bagaimana cara membangun sistem yang dapat mengklasifikasikan penerima bantuan sosial menggunakan metode K-NN (Bantuan et al., n.d.).

Konsep K-NN dapat menjadi alat yang sangat berguna dalam konteks bantuan sosial untuk memprediksi dan mengklasifikasikan calon penerima. Bayangkan kita

memiliki data historis tentang setiap orang atau keluarga yang pernah menerima bansos, lengkap dengan informasi tentang status ekonomi, demografi, dan kriteria lainnya. K-NN berusaha untuk menemukan kesamaan antara data historis dan kandidat penerima bansos baru. Pada dasarnya, algoritma K-NN akan melakukan perbandingan ketika seseorang atau keluarga baru mengajukan permohonan bansos. Profil pemohon baru akan dinilai berdasarkan seberapa "mirip" mereka dengan semua profil penerima bansos yang sudah tercatat sebelumnya. Usia, pendapatan, jumlah tanggungan, luas rumah, atau status pekerjaan adalah beberapa cara untuk mengukur kemiripan ini. Algoritma kemudian akan menemukan K profil yang paling mirip atau paling dekat dengan pemohon baru setelah menghitung jarak untuk semua data sebelumnya. Jumlah tetangga terdekat yang kita pilih untuk dipertimbangkan diwakili oleh angka "k" ini, dan memilih k yang tepat sangat penting.

Algoritma K-NN digunakan di Kelurahan Titi Kuning untuk menentukan siapa yang berhak atas bantuan sosial. Cara kerjanya adalah dengan membandingkan data warga baru dengan data sebelumnya tentang status mereka (seperti penghasilan, jumlah keluarga, dan pekerjaan). K-NN mengumpulkan data warga yang paling mirip (tetangga terdekat) dengan warga baru. Jika mayoritas tetangga layak mendapatkan bantuan, warga baru juga dianggap layak; jika tidak, mereka dianggap tidak layak. Metode ini memungkinkan petugas kelurahan memberikan bantuan dengan lebih tepat dan cepat kepada mereka yang membutuhkannya.

2.7.2. Rumus Algoritma K-NN

Rumus K-Nearest Neighbors (K-NN) adalah salah satu algoritma dalam machine learning yang digunakan untuk klasifikasi dan regresi. KNN bekerja dengan cara menghitung jarak antara data baru dan data dalam dataset, lalu memilih k tetangga terdekat untuk menentukan kelas atau nilai. Berikut rumus Algoritma K-NN :

a. jarak *Euclidean*:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \dots \dots \dots (1)$$

Keterangan :

$d(p, q)$: jarak Euclidean antara dua titik p dan q

p_i : Nilai ke-i dari vector p

q_i : nilai ke-i dari vector q

n : jumlah dimensi atau fitur

b. Rumus jarak Manhattan:

$$d(p, q) = \sum_{i=1}^n |p_i - q_i| \dots \dots \dots (2)$$

Penjelasan: Rumus ini untuk mengukur jarak seperti menyusuri jalan di kota (grid), serta lebih tahan terhadap outlier dibanding Euclidean.

c. Rumus jarak Minkowski Distance:

$$d(p, q) = \left(\sum_{i=1}^n |p_i - q_i|^r \right)^{\frac{1}{r}} \dots \dots \dots (3)$$

Penjelasan: Rumus di atas merupakan generalisasi dari Euclidean dan Manhattan.

Dan kita bisa atur parameter r : $r = 1$: Manhattan, $r = 2$: Euclidean.

d. Rumus jarak Chebyshev Distance:

$$d(p, q) = \max_{i=1} |p_i - q_i| \dots \dots \dots (4)$$

Penjelasan: Rumus ini Mengambil selisih fitur terbesar. Selain itu berguna ketika perbedaan besar dalam satu fitur lebih penting dari lainnya.

e. Rumus konseptual:

$$Kelas_prediksi = \text{mode}(\text{Kelas dari } K \text{ tetangga terdekat}) \dots \dots \dots (5)$$

Penjelasan : fungsi statistic yang mengembalikan nilai yang paling sering muncul dalam suatu kumpulan data, hitung frekuensi setiap kelas di antara K tetangga yang terpilih, kelas dengan frekuensi tertinggi adalah prediksi untuk titik data baru.

f. Fungsi regresi (untuk tugas regresi):

$$Nilai_Prediksi = \frac{1}{K} \sum_{j=1}^K Target_Nilai_j \dots \dots \dots (6)$$

a. K : Jumlah tetangga terdekat yang di pertimbangkan.

b. $Target_Nilai_j$: Nilai target (output numerik) dari tetangga ke j .

c. $\sum$: Simbol penjumlahan.

g. Rumus Cosine Similarity didefinisikan persamaan 7

$$\cos(\theta) = \frac{\vec{x} \cdot \vec{y}}{||x|| ||y||} \dots \dots \dots (7)$$

Penjelasan: Rumus ini bisa digunakan untuk mencari kemiripan sudut antara vektor.

a. Jumlah tanggungan keluarga

b. Status pekerjaan

2.7.3. Python dalam Algoritma K-NN

Python ialah software yang biasa digunakan untuk programming tentang image processing. Python ini sering digunakan karena cara mengolah dan menjalankan sebuah program cukup mudah karena library telah disediakan oleh Open CV. Software ini menggunakan Bahasa python sendiri dimana bahasa pemrograman ini lebih mudah daripada bahasa pemrograman C atau C++ (Yaqin et al., 2020). Python bersifat open-source, sehingga siapa pun dapat menggunakan dan mengubahnya secara bebas, termasuk dalam dunia pendidikan dan penelitian, karena kodenya dapat dieksekusi baris demi baris tanpa perlu dikompilasi terlebih

dahulu. Contoh implementasi python pada Bantuan Sosial, misalnya memiliki data sebagai berikut ; Langkah Python :

- a. Konversi data ke angka.
- b. Normalisasi. Terapkan K-NN.
- c. Prediksi untuk warga baru.

Karena strukturnya yang sederhana dan pustaka-pustaka yang terintegrasi dengan baik, Python sangat cocok untuk proyek berbasis K-NN. Kenapa Python Semakin Populer? karena Python telah berkembang menjadi salah satu bahasa pemrograman paling populer di seluruh dunia: Sintaks sederhana mudah dipelajari, Sangat fleksibel dan bisa digunakan untuk berbagai tujuan, seperti pengembangan web, automasi, analisis data, kecerdasan buatan, dan pembelajaran mesin. Selain Algoritma KNN, Naïve Bayes adalah Algoritma klasifikasi yang berbasis probabilitas, berbeda dengan KNN ia menggunakan Teorema Bayes untuk menghitung kemungkinan data akan masuk ke dalam setiap kelas.

2.8. Algoritma Naïve Bayes

2.8.1. Pengertian dan Konsep Algoritma Naïve Bayes

Algoritma Naïve Bayes mudah digunakan dan mudah dihitung dengan komputer. Dasar dari algoritma ini adalah asumsi bahwa setiap nilai atribut bergantung pada nilai atribut lainnya, meskipun asumsi ini tidak selalu benar. Namun, dalam praktiknya, pengklasifikasi Naïve Bayes sering kali bekerja dengan baik. Asumsi tentang ketergantungan yang independen membuat perhitungan untuk membuat model probabilitas Naïve Bayes menjadi lebih sederhana (Hadiani et al., 2024). Naive bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Tujuan klasifikasi adalah untuk secara akurat memprediksi kelas target untuk setiap kasus dalam data. Algoritma Naive Bayes didasarkan pada kondisional probabilitas.

Metode Naive Bayes tidak memerlukan banyak data pelatihan untuk memperkirakan parameter yang dibutuhkan dalam proses klasifikasi, dan ini merupakan salah satu keuntungan dari penggunaan Metode Naive Bayes. Dalam banyak kasus di kehidupan nyata, Naive Bayes seringkali menunjukkan kinerja yang lebih baik daripada yang diperkirakan (Alfiah, 2021). Di sebagian besar waktu, Naive Bayes berperforma jauh lebih baik yang diharapkan di sebagian besar skenario dunia nyata yang kompleks. Asumsi yang sangat kuat (naif) tentang independensi setiap kondisi atau kejadian merupakan ciri pembeda naive Bayes. (Azlil Huriah & Dienwati Nuris, 2023).

Konsep algoritma Naïve Bayes dalam Bansos ini adalah metode klasifikasi probabilitas yang digunakan untuk memprediksi atau mengidentifikasi calon penerima bansos yang paling berhak berdasarkan data historis dan berbagai kriteria yang relevan. Algoritma ini bekerja dengan menerapkan Teorema Bayes bahwa setiap kriteria atau fitur yang digunakan adalah independent satu sama lain. Konsep ini berfungsi sebagai alat bantu keputusan yang cerdas yang dapat membantu Lembaga sosial menyalurkan bantuan secara objektif dan tepat sasaran. Teorema Bayes sebuah formula dalam statistika yang memungkinkan kita untuk menghitung probabilitas berdasarkan pengetahuan sebelumnya. Secara keseluruhan, Naive Bayes menawarkan metode yang efektif dan mudah dipahami untuk menentukan kandidat penerima bansos. Meskipun asumsi independensinya sederhana, algoritma ini terbukti efektif dalam membantu lembaga sosial membuat keputusan berbasis data yang adil dan memberikan bantuan dengan efisien. Untuk mengetahui seberapa

layak seseorang menerima bantuan sosial, algoritma Naive Bayes digunakan untuk menghitung probabilitas berdasarkan data yang dimiliki, seperti penghasilan, jumlah anggota keluarga, dan status pekerjaan. Algoritma ini menganggap bahwa setiap data atau fitur berbeda satu sama lain, sehingga dapat menentukan apakah seorang warga layak atau tidak menerima bantuan berdasarkan kombinasi atribut tersebut. Petugas kelurahan dapat dengan cepat dan akurat menentukan siapa yang berhak mendapatkan bantuan tanpa harus memeriksa semua data secara manual.

2.8.2. Rumus Naïve Bayes

Berikut adalah rumus dasar algoritma Naive Bayes, yang merupakan salah satu algoritma klasifikasi paling populer dalam machine learning berbasis pada Teorema Bayes dengan asumsi independensi antar fitur (atribut).

a. Rumus Teorema Bayes:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \dots\dots\dots (1)$$

Dimana :

$P(C|X)$: Probabilitas kelas C diberikan data X

$P(X|C)$: Probabilitas Data X dalam kelas C

$P(C)$: Probabilitas awal dari kelas C

$P(X)$: Probabilitas Data X secara keseluruhan

b. Rumus Asumsi Naif :

$$P(X|C) = \prod_{i=1}^n P(x_i|C) \dots\dots\dots (2)$$

Penjelasan : Naïve Bayes disebut naif karna mengasumsikan bahwa fitur $x_1, x_2, \dots, x_n$ tidak saling bergantung satu sama lain jika diketahui kelas C.

c. Rumus Final Naïve Bayes :

$$\hat{c} = \arg \max_c p(c) \prod_{i=1}^n p(x_i | c) \dots\dots\dots(3)$$

Dimana : Untuk setiap kelas C kita hitung: Prior kelas P (C), Probabilitas setiap fitur x1 muncul dalam kelas tersebut P (x1 | C). Lalu pilih kelas dengan nilai hasil kali terbesar, itulah prediksi kelas.

d. Rumus Multinomial Naive Bayes

$$P((x_i|ck) = \frac{N_{ik} + a}{N_k + ad} \dots\dots\dots(4)$$

Dimana:

N_{ik} : Jumlah kemunculan Fitur I di kelas k

N_k : Total Jumlah kata pada kelas k

d : Jumlah fitur (ukuran kosakata)

a : Parameter Smoothing Laplace (biasanya $a = 1$)

e. Rumus Bernoulli Naïve Bayes (fitur biner, 0 atau 1)

$$P(x_i | C_k) = P_i^{x_i} \cdot (1 - P_i)^{1-x_i} \dots\dots\dots(5)$$

Dimana : jika fitur hanya menunjukkan keberadaan atau ketiadaan.

2.8.3. Perbandingan Metode Algoritma

Pada akhirnya, pilihan algoritma seringkali tergantung pada karakteristik spesifik dataset Anda dan persyaratan proyek. Seringkali, praktik terbaik adalah mencoba kedua algoritma (dan mungkin algoritma lain juga) dan membandingkan kinerjanya menggunakan metrik evaluasi yang relevan (seperti akurasi, presisi, *recall*, F1-score). Berikut perbandingan metode antara algoritma K-NN dan Naïve Bayes :

Table 2.4. Perbandingan Metode K-NN dan Naïve Bayes

Fitur Utama	K-Neares Neighbors	Naïve Bayes
Konsep Inti	Berbasis jarak, memprediksi kelas berdasarkan mayoritas tetangga terdekat.	Berbasis probabilitas, memprediksi kelas berdasarkan probabilitas tertinggi dari fitur-fitur yang (diasumsikan)
Proses Belajar	Malas (Lazy Learner) hanya menyimpan data belajar terjadi saat prediksi.	Cepat (Eager Learner) menghitung semua probabilitas selama pelatihan.
Kecepatan prediksi	Lambat untuk dataset besar karena perlu menghitung jarak ke semua data Latihan.	Sangat cepat karena probabilitas sudah dihitung di awal
Asumsi Data	Non-parametrik: Tidak ada asumsi tentang distribusi data.	Memiliki asumsi kuat. Semua fitur saling independent
Sensitivitas skala fitur	Sangat sensitive: Perlu normalisasi data	Kurang sensitive, namun normalisasi tetap baik
Efektif untuk Dimensi tinggi	Cendenrung kurang efektif	Cukup Baik
Kebutuhan Memori	Tinggi: membutuhkan seluruh dataset Latihan tersimpan	Rendah: Hanya menyimpan probabilitas yang dihitung.

BAB III

METODOLOGI PENELITIAN

3.1. Pendekatan Penelitian

Pendekatan penelitian ini Menggunakan Pendekatan kualitatif komparatif untuk menilai dan membandingkan kinerja dua algoritma klasifikasi untuk menentukan penerima bantuan sosial. K-NN dan Naïve Bayes. Pendekatan ini berkonsentrasi pada analisis data numerik dan statistik untuk mengetahui secara objektif seberapa efektif masing-masing algoritma. Metode ini memungkinkan evaluasi sistematis dan terukur terhadap efektivitas setiap algoritma yang digunakan untuk mengklasifikasikan kandidat penerima bantuan sosial. Penelitian ini menggunakan metode Data Mining dan Sistem Pendukung Keputusan, yang melibatkan pembuatan model prediktif dan ekstraksi pola dari data yang sangat besar.

Tahapan penelitian meliputi :

1. Pengumpulan Data
2. Pemrosesan Data (Data Processing)
3. Implementasi Algoritma
4. Evaluasi Kinerja Model
5. Analisis Hasil

3.2. Pengumpulan Data

3.2.1. Sumber Data

Penelitian ini mengumpulkan data dari Kantor Lurah Titi Kuning. Data yang akan menjadi focus utama pada penelitian ini adalah data sekunder yang berasal

langsung dari Kantor Lurah Titi Kuning. Dataset ini akan memuat informasi yang komprehensif mengenai atribut atribut kunci yang menjadi pertimbangan dalam penentuan kelayakan penerima bantuan.

Tabel 3.1. Data Latih Kelurahan Titi Kuning

Kelurahan	Penghasilan	Tanggungan	Status_Rumah	Pendidikan	Penerima_Bantuan
Sukamaju	1433912	4	Sendiri	SMP	Tidak
Sukamaju	709805	1	Kontrak	S1	Ya
Sukamaju	2807113	3	Sendiri	S1	Ya
Sukamaju	2554301	5	Kontrak	SD	Tidak
Sukamaju	2372427	1	Sendiri	SMP	Ya
Sukamaju	1670528	4	Sendiri	S1	Ya
Sukamaju	1359791	3	Sendiri	SD	Ya
Sukamaju	1229295	6	Sendiri	SD	Ya
Sukamaju	4039336	6	Kontrak	SMA	Tidak
Sukamaju	766612	6	Sendiri	SMA	Tidak
Sukamaju	749957	3	Sendiri	SMP	Ya
Sukamaju	1285972	5	Sendiri	S1	Ya
Sukamaju	2334068	4	Kontrak	S1	Tidak
Sukamaju	2451701	6	Kontrak	SMA	Ya
Sukamaju	4739227	3	Kontrak	SMP	Ya
Sukamaju	722599	1	Sendiri	S1	Tidak
Sukamaju	2167971	6	Sendiri	SD	Tidak
Sukamaju	4019187	2	Sendiri	SD	Tidak
Sukamaju	2349189	4	Sendiri	SD	Ya

Sukamaju	4268238	1	Kontrak	SD	Ya
Sukamaju	2833632	5	Sendiri	SMA	Ya
Sukamaju	554515	5	Kontrak	SMP	Tidak
Sukamaju	1839319	3	Sendiri	SD	Tidak
Sukamaju	4045146	5	Sendiri	SD	Tidak
Sukamaju	3354228	2	Sendiri	S1	Ya
Sukamaju	2830953	6	Sendiri	SMA	Tidak
Sukamaju	1804256	1	Kontrak	S1	Ya
Sukamaju	2306182	6	Sendiri	S1	Tidak
Sukamaju	3323559	4	Kontrak	S1	Tidak
Sukamaju	1357401	6	Sendiri	SD	Ya
Sukamaju	1278008	3	Kontrak	SMP	Ya
Sukamaju	3687061	6	Sendiri	SMA	Tidak
Sukamaju	1311315	4	Kontrak	SD	Tidak
Sukamaju	3511337	1	Sendiri	SMA	Ya
Sukamaju	3385309	2	Sendiri	SD	Ya
Sukamaju	2718961	1	Kontrak	SMP	Tidak
Sukamaju	864488	1	Kontrak	SMP	Ya
Sukamaju	4353935	3	Kontrak	SMA	Tidak
Sukamaju	4998207	4	Kontrak	S1	Ya
Sukamaju	1547117	1	Sendiri	SD	Ya
Sukamaju	3675376	1	Sendiri	SMA	Tidak
Sukamaju	1161023	2	Sendiri	S1	Ya

Sukamaju	2959357	5	Kontrak	SMP	Tidak
Sukamaju	3533614	2	Kontrak	SD	Tidak
Sukamaju	2113033	4	Sendiri	SMP	Ya
Sukamaju	1083470	4	Sendiri	SD	Ya
Sukamaju	884402	3	Kontrak	S1	Ya
Sukamaju	2411749	6	Kontrak	SD	Ya
Sukamaju	2927562	6	Sendiri	S1	Ya
Sukamaju	1169343	6	Sendiri	SD	Tidak
Sukamaju	2452791	5	Sendiri	SMA	Ya
Sukamaju	1347261	4	Sendiri	SMP	Ya
Sukamaju	3688729	5	Sendiri	SD	Tidak
Sukamaju	2831811	6	Kontrak	SD	Ya
Sukamaju	4303481	6	Kontrak	SMP	Tidak
Sukamaju	3560434	2	Sendiri	SD	Ya
Sukamaju	1864441	4	Sendiri	SMP	Tidak
Sukamaju	3605303	6	Kontrak	S1	Tidak
Sukamaju	3480226	1	Kontrak	SMA	Tidak

Sumber: Data Latih Kelurahan Titi Kuning

3.2.2. Teknik Pengumpulan Data

Data dikumpulkan dari Kantor Lurah Titi Kuning, Data ini mencakup informasi pribadi dan ekonomi warga yaitu :

- a. Jumlah anggota keluarga: Menunjukkan beban tanggungan kepala keluarga, semakin banyak anggota keluarga potensi kebutuhan akan makin tinggi.

- b. Status pekerjaan: Menjelaskan apakah kepala keluarga memiliki pekerjaan tetap, serabutan, atau tidak bekerja sama sekali.
- c. Pendapatan perbulan: Merupakan perkiraan total penghasilan yang diperoleh keluarga dalam satu bulan.
- d. Kepemilikan asset: Meliputi memiliki asset seperti Tanah, mobil, perhiasan, atau tabungan dalam jumlah besar.
- e. Status kelayakan/tidak layak: Ini adalah atribut apakah keluarga tersebut pada akhirnya ditentukan sebagai penerima Bantuan Sosial atau tidak.

3.3. Preprocessing Data

Tujuan utama dari Processing Data ini adalah untuk membersihkan, mengubah, dan mempersiapkan data agar menjadi format yang optimal dan berkualitas tinggi bagi kedua algoritma tersebut. Ada 4 tahapan dalam processing data yaitu:

1. Pembersihan Data/Data Cleaning - Tahapan ini berfokus pada perbaikan atau penghapusan data yang tidak lengkap atau tidak akurat.
2. Transformasi Data – Tahap ini mengubah format data agar lebih cocok untuk pemrosesan algoritma.
3. Normalisasi/Standarisasi Data (Data Scaling) – Tahapan ini penting untuk semua atribut memiliki bobot yang setara dalam perhitungan K-NN yang sensitive dalam skala data.
4. Seleksi fitur – Tahapan ini bertujuan untuk memilih subset atribut yang paling relevan atau informatif untuk membangun model.

3.4. Implementasi Algoritma

3.4.1. Algoritma K-Nearest Neighbors

Metode K-NN diterapkan untuk mengelompokkan data dengan cara mengorganisirnya ke dalam satu atau lebih kategori berdasarkan kesamaan ciri. Algoritma K-NN mengidentifikasi kategori data baru dengan mencari beberapa data yang paling dekat (tetangga) dalam kumpulan data yang telah diberi label sebelumnya, lalu memberikan label pada data baru tersebut berdasarkan kelas terbanyak dari tetangga yang terdekat (Nasution, 2015). Berikut adalah penjelasan lebih lanjut mengenai langkah-langkah implementasi KNN:

1. Menentukan nilai k (jumlah tetangga terdekat)

Penentuan nilai k ; jelaskan bagaimana nilai k dipilih: misal melalui trial dan eror, validasi silang, atau menggunakan nilai umum sebutkan nilai k yang akhir digunakan.

2. Penghitungan jarak

Menjelaskan penggunaan Euclidean Distance (atau jarak lain seperti Manhattan Distance jika dipertimbangkan) untuk menghitung jarak antara setiap data uji dengan seluruh data pelatihan. Berikan rumus matematisnya.

3. Pemilihan tetangga terdekat

Untuk setiap data uji, jelaskan bagaimana k data pelatihan dengan jarak terpendek dipilih.

4. Penentuan kelas

Jelaskan bagaimana kelas data uji ditentukan berdasarkan mayoritas kelas dari k tetangga terdekat. Jika ada kasus seri (tie-breaking), jelaskan bagaimana ditangani (misal: pilih kelas dari tetangga terdekat

Flowchart Algoritma K-NN



Gambar 3.1. Flowchart Algoritma K-NN

Sumber: N Fatahna, S Alifah, S Farisah, 2017

Keterangan :

1. Mulai
 - Proses dimulai dari sini
2. Persiapan data
 - Melakukan preprocessing data (pembersihan, transformasi, encoding)

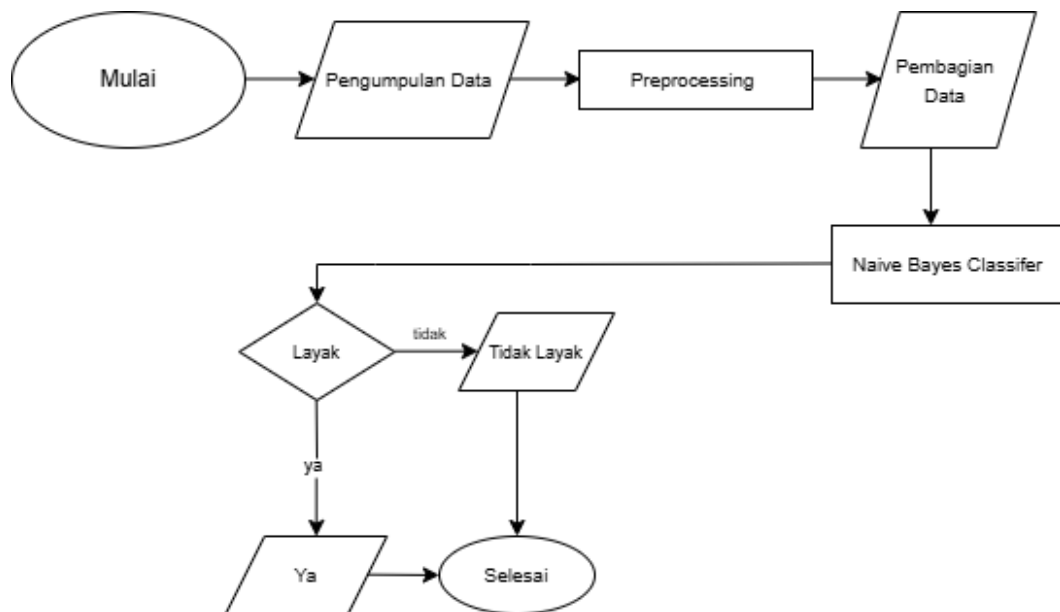
3. Pilih nilai k
 - Menentukan jumlah tetangga terdekat yang akan dipertimbangkan
4. Hitung jarak Euclidean ke setiap data
 - Menghitung jarak antara data uji saat ini dengan setiap data dalam set pelatihan.
5. Urutkan jarak dari terkecil
 - Mengurutkan data pelatihan berdasarkan jarak dari data uji secara ascending.
6. Hitung frekuensi dari kelas k tetangga
 - Menghitung berapa kali setiap kelas (misal: "Layak" atau "Tidak Layak") muncul di antara k tetangga yang dipilih.
7. Tentukan kelas data uji berdasarkan mayoritas
 - Menetapkan kelas untuk data uji berdasarkan kelas yang paling sering muncul di antara k tetangga.
8. Selesai untuk data uji
 - Proses klasifikasi selesai untuk satu data uji.
9. Selesai
 - Proses klasifikasi K—NN selesai.

3.4.2. Algoritma Naïve Bayes

Metode Naive Bayes adalah algoritma yang kuat untuk klasifikasi, terutama dalam aplikasi seperti pengenalan teks dan analisis sentimen. Meskipun memiliki beberapa keterbatasan, kecepatan dan kesederhanaannya menjadikannya pilihan yang baik untuk banyak masalah klasifikasi. Dalam bab selanjutnya, kita akan

membahas implementasi dan evaluasi algoritma Naive Bayes pada dataset tertentu.

Flowchart Algoritma Naïve Bayes:



Gambar 3.2 Flowchart Algoritma Naïve Bayes

Sumber: Jurnal Telkom University, 2014

Keterangan:

1. Mulai
 - Proses dimulai.
2. Pengumpulan Data
 - Kumpulan Data yang relevan tentang calon penerima bansos, termasuk variable independent (misalnya; jumlah anggota keluarga, pendapatan, status pekerjaan, dll).
3. Preprocessing Data
 - Melakukan pembersihan Data/menghapus data yang hilang atau tidak valid.
4. Pembagian Data

- Bagi data menjadi dataset: data latih (training set) dan data uji.
5. Naïve Bayes Classifier
 - Menghitung probabilitas untuk setiap kelas (layak atau tidak layak).
 6. Penentuan kelayakan
 - Menentukan kelas (layak atau tidak layak) berdasarkan probabilitas posterior yang dihitung.
 7. Selesai
 - Proses klasifikasi selesai.

3.5. Waktu dan tempat penelitian

3.5.1. Waktu penelitian

No	Tahapan	Februari 2025	Maret 2025	April 2025	Mei 2025	Juni 2025	Juli 2025
1	Pengajuan judul proposal						
2	Penyusunan proposal						
3	Pengumpulan data						
4	Penelolaan data						
5	Penyusunan skripsi						

3.5.2. Tempat Penelitian

Penelitian ini dilakukan di Kantor Lurah Titi Kuning Kec. Medan Johor, Kota Medan.

3.5.3. Pengumpulan Data

Pengumpulan data merupakan Langkah krusial dalam penelitian ini, karena kualitas dan relevansi data akan mempengaruhi hasil dari penerapan Algoritma K-NN dan Naïve Bayes. Proses pengumpulan ini dilakukan melalui beberapa tahapan sebagai berikut:

- a. Wawancara: Melakukan wawancara dengan calon penerima bantuan sosial untuk mendapatkan informasi yang lebih mendalam mengenai kondisi mereka.
- b. Observasi: Melakukan observasi langsung terhadap kondisi sosial dan ekonomi masyarakat yang menjadi target penerima bantuan sosial

BAB IV

HASIL DAN UJI COBA

4.1. Hasil

Hasil dari Penerapan Algoritma K-Nearest Neighbors (KNN) Dan Naïve Bayes Untuk Menentukan Pemilihan Penerima Bantuan Sosial Berdasarkan Ekonomi Masyarakat dapat dilihat sebagai berikut:

Bansos

Silahkan Login

Silahkan Login :

Username Admin

Password *****

Proses Metode

Bansos

Links

Tema

Contact us

Bansos (Bantuan Sosial) adalah program bantuan yang diberikan oleh pemerintah atau lembaga terkait kepada masyarakat yang membutuhkan, dengan tujuan meningkatkan kesejahteraan.

> Menu

> Masyarakat

> Dataset Latih

> Prediksi

Penerima Bantuan Sosial

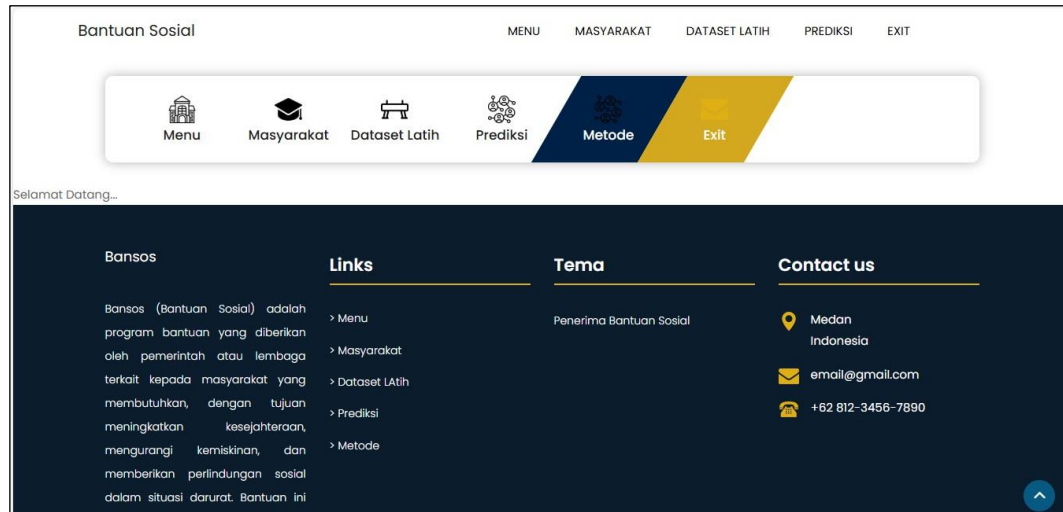
Medan Indonesia

email@gmail.com

+62 812-3456-7890

Gambar 4.1. Form Login

Gambar 4.1 adalah *form* login yang memiliki Tombol Proses dan Tombol Metode. Jika pengguna mengisi *username* dan *password* dengan benar kemudian mengklik Tombol Proses maka aplikasi akan menampilkan *form* Menu Admin. Jika pengguna klik Tombol Metode maka akan tampil *form* Metode. Form Menu Admin dari Penerapan Algoritma K-Nearest Neighbors (KNN) Dan Naïve Bayes Untuk Menentukan Pemilihan Penerima Bantuan Sosial Berdasarkan Ekonomi Masyarakat dapat dilihat pada Gambar 4.2.



Gambar 4.2. Form Menu

Gambar 4.2 *form* menu memiliki beberapa menu yang dapat ditekan dan memiliki fungsi berikut:

1. Menu : Berfungsi menampilkan menu
2. Masyarakat : Berfungsi menampilkan *form* masyarakat
3. Dataset Latih : Berfungsi menampilkan *form* dataset latih
4. Prediksi : Berfungsi menampilkan *form* prediksi
5. Exit : Berfungsi menampilkan *form* login dan menutup menu

Jika pelanggan mengklik menu masyarakat maka akan tampil *form* masyarakat yang disajikan pada Gambar 4.3.

Bantuan Sosial

MENU MASYARAKAT DATASET LATIH PREDIKSI EXIT

Menu Masyarakat Dataset Latih Prediksi Metode Exit

Kelola Masyarakat

Ketik apa saja...

Tambah Masyarakat

Id Masyarakat	Nama	Penghasilan	Tanggungun	Status Rumah	Pendidikan	Aksi
1	Eka	1300000.00	3	Kontrak	SMP	Edit Hapus
2	Fajar	4500000.00	1	Milik Sendiri	Sarjana	Edit Hapus
3	Ani	2000000.00	3	Kontrak	SD	Edit Hapus
4	Andi	200000.00	3	Kontrak	SD	Edit Hapus

Gambar 4.3. Form Masyarakat

Gambar 4.3 *form* masyarakat memiliki beberapa menu yang dapat ditekan dan memiliki fungsi berikut:

1. Menu : Berfungsi menampilkan menu
2. Masyarakat : Berfungsi menampilkan *form* masyarakat
3. Dataset Latih : Berfungsi menampilkan *form* dataset latih
4. Prediksi : Berfungsi menampilkan *form* prediksi
5. Exit : Berfungsi menampilkan *form* login dan menutup menu
6. Kotak Cari : Berfungsi mencari data sesuai yang diketik
7. Tambah : Berfungsi menampilkan *form* pengisian data baru
8. Edit : Berfungsi menampilkan *form* pengedit data lama
9. Hapus : Berfungsi menghapus data yang dipilih

Jika pelanggan mengklik menu dataset latih maka akan tampil *form* dataset latih yang disajikan pada Gambar 4.4.

Bantuan Sosial

MENU MASYARAKAT DATASET LATIH PREDIKSI EXIT

Menu Masyarakat Dataset Latih Prediksi Metode Exit

Kelola Dataset Latih

Ketik apa saja...

Tambah Dataset Latih

Id Latih	Penghasilan	Tanggungan	Status Rumah	Pendidikan	Penerima Bantuan	Aksi
1	1000000.00	4	Kontrak	SD	Layak	Edit Hapus
2	1500000.00	3	Kontrak	SMP	Layak	Edit Hapus
3	3000000.00	2	Milik Sendiri	SMA	Tidak	Edit Hapus
4	5000000.00	1	Milik Sendiri	Sarjana	Tidak	Edit Hapus

Gambar 4.4. Form Dataset Latih

Gambar 4.4 *form* dataset memiliki beberapa menu yang dapat ditekan dan memiliki fungsi berikut:

1. Menu : Berfungsi menampilkan menu
2. Masyarakat : Berfungsi menampilkan *form* masyarakat
3. Dataset Latih : Berfungsi menampilkan *form* dataset latih
4. Prediksi : Berfungsi menampilkan *form* prediksi
5. Exit : Berfungsi menampilkan *form* login dan menutup menu
6. Kotak Cari : Berfungsi mencari data sesuai yang diketik
7. Tambah : Berfungsi menampilkan *form* pengisian data baru
8. Edit : Berfungsi menampilkan *form* pengedit data lama
9. Hapus : Berfungsi menghapus data yang dipilih

Jika pelanggan mengklik menu prediksi maka akan tampil *form* prediksi yang disajikan pada Gambar 4.5.

Bantuan Sosial

MENU MASYARAKAT DATASET LATIH PREDIKSI EXIT

Menu Masyarakat Dataset Latih Prediksi Metode Exit

Kelola Prediksi

Ketik apa saja...

Id Prediksi	Id Masyarakat	Metode	Hasil	Tanggal Prediksi	Aksi
1	1	KNN	Layak	2025-07-28	Hapus
2	1	Naive Bayes	Layak	2025-07-28	Hapus
3	3	KNN	Layak	2025-07-28	Hapus
4	3	Naive Bayes	Layak	2025-07-28	Hapus
5	4	KNN	Layak	2025-07-28	Hapus

localhost/bansos/prediksi.php#

Gambar 4.5. Form Prediksi

Gambar 4.5 *form* prediksi memiliki beberapa menu yang dapat ditekan dan memiliki fungsi berikut:

1. Menu : Berfungsi menampilkan menu
2. Masyarakat : Berfungsi menampilkan *form* masyarakat
3. Dataset Latih : Berfungsi menampilkan *form* dataset latih
4. Prediksi : Berfungsi menampilkan *form* prediksi
5. Exit : Berfungsi menampilkan *form* login dan menutup menu
6. Kotak Cari : Berfungsi mencari data sesuai yang diketik
7. Hapus : Berfungsi menghapus data yang dipilih

Jika pelanggan mengklik menu exit maka akan tampil *form* login yang disajikan pada Gambar 4.6.

Gambar 4.6. Form Login

Gambar 4.6 merupakan form login yang dapat digunakan untuk membuka proses metode dengan mengklik tombol Metode dan hasilnya dapat dilihat pada Gambar 4.7.

Gambar 4.7. Form Metode

Gambar 4.7 merupakan *form* metode yang memiliki beberapa kotak teks untuk dimasukkan data baru sehingga mendapatkan hasil prediksi yaitu Nama, Penghasilan, Tanggungan, Status Rumah, Pendidikan, dan Nilai K=3. Tombol Exit berfungsi untuk keluar dari *form* metode dan kembali ke *form* login.

4.2. Pembahasan

Pembahasan meliputi kebutuhan perangkat, hasil yang digunakan dan pengujian pada penelitian ini.

1. Kebutuhan Perangkat

Kebutuhan perangkat keras dan perangkat lunak untuk membuat aplikasi adalah sebagai berikut:

a. Satu unit laptop dengan spesifikasi sebagai berikut:

- 1) *Processor* minimal *Core 2 Duo*
- 2) RAM minimal 1 Gb
- 3) *Hardisk* minimal 80 Gb

b. Perangkat Lunak dengan spesifikasi sebagai berikut:

- 1) Sistem Operasi *Windows*
- 2) Notepad++
- 3) Xampp/Appserv

2. Metode

Penelitian ini menggunakan banyak metode sebagai proses pencarian hasil.

Berikut adalah tahapannya:

a. K-NN

1) Euclidean Distance

$$\text{Rumus: } d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots (1)$$

Data Latih:

penghasilan : 200000

tanggungan : 3

status_rumah : Kontrak

pendidikan : SD

Data 1:

Penghasilan: $1433912.00 - 200000 = 1233912$

Tanggungan: $4 - 3 = 1$

Status Rumah: Milik Sendiri - Kontrak = 2

Pendidikan: SMP - SD = 1

Data 1 = $1233912^2 + 1^2 + 2^2 + 1^2$

Jarak(d) = $\sqrt{(1522538823750)} = 1233912.0000024$

Data 2:

Penghasilan: $709805.00 - 200000 = 509805$

Tanggungan: $1 - 3 = -2$

Status Rumah: Kontrak - Kontrak = 0

Pendidikan: Sarjana - SD = 4

Data 2 = $509805^2 + (-2)^2 + 0^2 + 4^2$

Jarak(d) = $\sqrt{(259901138045)} = 509805.00001962$

Data 3:

Penghasilan: $2807113.00 - 200000 = 2607113$

Tanggungan: $3 - 3 = 0$

Status Rumah: Milik Sendiri - Kontrak = 2

Pendidikan: Sarjana - SD = 4

$$\text{Data 3} = 2607113^2 + 0^2 + 2^2 + 4^2$$

$$\text{Jarak(d)} = \sqrt{(6797038194789)} = 2607113.0000038$$

Data 4:

$$\text{Penghasilan: } 2554301.00 - 200000 = 2354301$$

$$\text{Tanggungan: } 5 - 3 = 2$$

Status Rumah: Kontrak - Kontrak = 0

Pendidikan: SD - SD = 0

$$\text{Data 4} = 2354301^2 + 2^2 + 0^2 + 0^2$$

$$\text{Jarak(d)} = \sqrt{(5542733198605)} = 2354301.0000008$$

Data 5:

$$\text{Penghasilan: } 2372427.00 - 200000 = 2172427$$

$$\text{Tanggungan: } 1 - 3 = -2$$

Status Rumah: Milik Sendiri - Kontrak = 2

Pendidikan: SMP - SD = 1

$$\text{Data 5} = 2172427^2 + (-2)^2 + 2^2 + 1^2$$

$$\text{Jarak(d)} = \sqrt{(4719439070338)} = 2172427.0000021$$

Data 6:

$$\text{Penghasilan: } 1670528.00 - 200000 = 1470528$$

$$\text{Tanggungan: } 4 - 3 = 1$$

Status Rumah: Milik Sendiri - Kontrak = 2

Pendidikan: Sarjana - SD = 4

$$\text{Data 6} = 1470528^2 + 1^2 + 2^2 + 4^2$$

$$\text{Jarak(d)} = \sqrt{(2162452598805)} = 1470528.0000071$$

Data 7:

$$\text{Penghasilan: } 1359791.00 - 200000 = 1159791$$

$$\text{Tanggungan: } 3 - 3 = 0$$

Status Rumah: Milik Sendiri - Kontrak = 2

Pendidikan: SD - SD = 0

$$\text{Data 7} = 1159791^2 + 0^2 + 2^2 + 0^2$$

$$\text{Jarak(d)} = \sqrt{(1345115163685)} = 1159791.0000017$$

Data 8:

$$\text{Penghasilan: } 1229295.00 - 200000 = 1029295$$

$$\text{Tanggungan: } 6 - 3 = 3$$

Status Rumah: Milik Sendiri - Kontrak = 2

Pendidikan: SD - SD = 0

$$\text{Data 8} = 1029295^2 + 3^2 + 2^2 + 0^2$$

$$\text{Jarak(d)} = \sqrt{(1059448197038)} = 1029295.0000063$$

Data 9:

Penghasilan: $4039336.00 - 200000 = 3839336$

Tanggungan: $6 - 3 = 3$

Status Rumah: Kontrak - Kontrak = 0

Pendidikan: SMA - SD = 2

Data 9 = $3839336^2 + 3^2 + 0^2 + 2^2$

Jarak(d) = $\sqrt{(14740500920909)} = 3839336.0000017$

Data 10:

Penghasilan: $766612.00 - 200000 = 566612$

Tanggungan: $6 - 3 = 3$

Status Rumah: Milik Sendiri - Kontrak = 2

Pendidikan: SMA - SD = 2

Data 10 = $566612^2 + 3^2 + 2^2 + 2^2$

Jarak(d) = $\sqrt{(321049158561)} = 566612.000015$

Data Latih	Penghasilan ²	Tanggungan ²	Status ²	Pendidikan ²	S Kuadrat	d (Jarak)	Kelas
Data 1	1233912 ²	1 ²	2 ²	1 ²	1522538823750	1233912	Tidak
Data 2	509805 ²	-2 ²	0 ²	4 ²	259901138045	509805	Layak
Data 3	2607113 ²	0 ²	2 ²	4 ²	6797038194789	2607113	Layak
Data 4	2354301 ²	2 ²	0 ²	0 ²	5542733198605	2354301	Tidak
Data 5	2172427 ²	-2 ²	2 ²	1 ²	4719439070338	2172427	Layak
Data 6	1470528 ²	1 ²	2 ²	4 ²	2162452598805	1470528	Layak
Data 7	1159791 ²	0 ²	2 ²	0 ²	1345115163685	1159791	Layak
Data 8	1029295 ²	3 ²	2 ²	0 ²	1059448197038	1029295	Layak
Data 9	3839336 ²	3 ²	0 ²	2 ²	14740500920909	3839336	Tidak
Data 10	566612 ²	3 ²	2 ²	2 ²	321049158561	566612	Tidak
...	...	...	...	...	...	...	...

Data 1000	1057515 ²	2 ²	0 ²	2 ²	1118337975233	1057515	Tidak
-----------	----------------------	----------------	----------------	----------------	---------------	---------	-------

2) Manhattan Distance

Rumus: $d = \sum_{i=1}^n |x_i - y_i| \dots \dots \dots (2)$

Data 1:

Penghasilan: $1433912 - 200000 = 1233912$

Tanggungan: $4 - 3 = 1$

Status Rumah: $3 - 1 = 2$

Pendidikan: $2 - 1 = 1$

Jarak = $1233912 + 1 + 2 + 1 = 1233916$

Data 2:

Penghasilan: $709805 - 200000 = 509805$

Tanggungan: $1 - 3 = 2$

Status Rumah: $1 - 1 = 0$

Pendidikan: $5 - 1 = 4$

Jarak = $509805 + 2 + 0 + 4 = 509811$

Data 3:

Penghasilan: $2807113 - 200000 = 2607113$

Tanggungan: $3 - 3 = 0$

Status Rumah: $3 - 1 = 2$

Pendidikan: $5 - 1 = 4$

$$\text{Jarak} = 2607113 + 0 + 2 + 4 = 2607119$$

Data 4:

$$\text{Penghasilan: } 2554301 - 200000 = 2354301$$

$$\text{Tanggungan: } 5 - 3 = 2$$

$$\text{Status Rumah: } 1 - 1 = 0$$

$$\text{Pendidikan: } 1 - 1 = 0$$

$$\text{Jarak} = 2354301 + 2 + 0 + 0 = 2354303$$

Data 5:

$$\text{Penghasilan: } 2372427 - 200000 = 2172427$$

$$\text{Tanggungan: } 1 - 3 = 2$$

$$\text{Status Rumah: } 3 - 1 = 2$$

$$\text{Pendidikan: } 2 - 1 = 1$$

$$\text{Jarak} = 2172427 + 2 + 2 + 1 = 2172432$$

Data 6:

$$\text{Penghasilan: } 1670528 - 200000 = 1470528$$

$$\text{Tanggungan: } 4 - 3 = 1$$

$$\text{Status Rumah: } 3 - 1 = 2$$

$$\text{Pendidikan: } 5 - 1 = 4$$

$$\text{Jarak} = 1470528 + 1 + 2 + 4 = 1470535$$

Data 7:

Penghasilan: $1359791 - 200000 = 1159791$

Tanggungan: $3 - 3 = 0$

Status Rumah: $3 - 1 = 2$

Pendidikan: $1 - 1 = 0$

Jarak = $1159791 + 0 + 2 + 0 = 1159793$

Data 8:

Penghasilan: $1229295 - 200000 = 1029295$

Tanggungan: $6 - 3 = 3$

Status Rumah: $3 - 1 = 2$

Pendidikan: $1 - 1 = 0$

Jarak = $1029295 + 3 + 2 + 0 = 1029300$

Data 9:

Penghasilan: $4039336 - 200000 = 3839336$

Tanggungan: $6 - 3 = 3$

Status Rumah: $1 - 1 = 0$

Pendidikan: $3 - 1 = 2$

Jarak = $3839336 + 3 + 0 + 2 = 3839341$

Data 10:

Penghasilan: $766612 - 200000 = 566612$

Tanggungan: $6 - 3 = 3$

Status Rumah: $3 - 1 = 2$

Pendidikan: $3 - 1 = 2$

Jarak = $566612 + 3 + 2 + 2 = 566619$

Data Latih	Penghasilan	Tanggungan	Status	Pendidikan	d (Jarak)	Kelas
Data 1	1233912	1	2	1	1233916	Tidak
Data 2	509805	2	0	4	509811	Layak
Data 3	2607113	0	2	4	2607119	Layak
Data 4	2354301	2	0	0	2354303	Tidak
Data 5	2172427	2	2	1	2172432	Layak
Data 6	1470528	1	2	4	1470535	Layak
Data 7	1159791	0	2	0	1159793	Layak
Data 8	1029295	3	2	0	1029300	Layak
Data 9	3839336	3	0	2	3839341	Tidak
Data 10	566612	3	2	2	566619	Tidak
...	...	...	...	...	...	...
Data 1000	1057515	2	0	2	1057519	Tidak

3) Minkowski Distance (p=3)

Rumus: $d = (\sum_{i=1}^n |x_i - y_i|^3)^{1/3} \dots \dots \dots (3)$

Ph: Penghasilan

Tg: Tanggungan

St: Status

Pd: Pendidikan

Data 1:

$$\text{Penghasilan: } (1433912 - 200000)^2 = 1.8786789250836\text{E}+18$$

$$\text{Tanggungan: } (4 - 3)^2 = 1$$

$$\text{Status Rumah: } (3 - 1)^2 = 8$$

$$\text{Pendidikan: } (2 - 1)^2 = 1$$

$$\text{Jarak} = (1.8786789250836\text{E}+18 + 1 + 8 + 1)^{1/3} = 1.8786789250836\text{E}+18$$

Data 2:

$$\text{Penghasilan: } (709805 - 200000)^2 = 1.3249889967084\text{E}+17$$

$$\text{Tanggungan: } (1 - 3)^2 = 8$$

$$\text{Status Rumah: } (1 - 1)^2 = 0$$

$$\text{Pendidikan: } (5 - 1)^2 = 64$$

$$\text{Jarak} = (1.3249889967084\text{E}+17 + 8 + 0 + 64)^{1/3} = 1.3249889967084\text{E}+17$$

Data 3:

$$\text{Penghasilan: } (2807113 - 200000)^2 = 1.7720646639079\text{E}+19$$

$$\text{Tanggungan: } (3 - 3)^2 = 0$$

$$\text{Status Rumah: } (3 - 1)^2 = 8$$

$$\text{Pendidikan: } (5 - 1)^2 = 64$$

$$\text{Jarak} = (1.7720646639079\text{E}+19 + 0 + 8 + 64)^{1/3} = 1.7720646639079\text{E}+19$$

Data 4:

$$\text{Penghasilan: } (2554301 - 200000)^2 = 1.30492623122\text{E}+19$$

$$\text{Tanggungan: } (5 - 3)^2 = 8$$

Status Rumah: $(1 - 1)^2 = 0$

Pendidikan: $(1 - 1)^2 = 0$

Jarak = $(1.30492623122E+19 + 8 + 0 + 0)^{1/3} = 1.30492623122E+19$

Data 5:

Penghasilan: $(2372427 - 200000)^2 = 1.0252636861238E+19$

Tanggungan: $(1 - 3)^2 = 8$

Status Rumah: $(3 - 1)^2 = 8$

Pendidikan: $(2 - 1)^2 = 1$

Jarak = $(1.0252636861238E+19 + 8 + 8 + 1)^{1/3} = 1.0252636861238E+19$

Data 6:

Penghasilan: $(1670528 - 200000)^2 = 3.1799470951846E+18$

Tanggungan: $(4 - 3)^2 = 1$

Status Rumah: $(3 - 1)^2 = 8$

Pendidikan: $(5 - 1)^2 = 64$

Jarak = $(3.1799470951846E+18 + 1 + 8 + 64)^{1/3} = 3.1799470951846E+18$

Data 7:

Penghasilan: $(1359791 - 200000)^2 = 1.5600524608008E+18$

Tanggungan: $(3 - 3)^2 = 0$

Status Rumah: $(3 - 1)^2 = 8$

Pendidikan: $(1 - 1)^2 = 0$

Jarak = $(1.5600524608008E+18 + 0 + 8 + 0)^{1/3} = 1.5600524608008E+18$

Data 8:

$$\text{Penghasilan: } (1229295 - 200000)^2 = 1.0904847319568\text{E}+18$$

$$\text{Tanggungan: } (6 - 3)^2 = 27$$

$$\text{Status Rumah: } (3 - 1)^2 = 8$$

$$\text{Pendidikan: } (1 - 1)^2 = 0$$

$$\text{Jarak} = (1.0904847319568\text{E}+18 + 27 + 8 + 0)^{1/3} = 1.0904847319568\text{E}+18$$

Data 9:

$$\text{Penghasilan: } (4039336 - 200000)^2 = 5.6593735843629\text{E}+19$$

$$\text{Tanggungan: } (6 - 3)^2 = 27$$

$$\text{Status Rumah: } (1 - 1)^2 = 0$$

$$\text{Pendidikan: } (3 - 1)^2 = 8$$

$$\text{Jarak} = (5.6593735843629\text{E}+19 + 27 + 0 + 8)^{1/3} = 5.6593735843629\text{E}+19$$

Data 10:

$$\text{Penghasilan: } (766612 - 200000)^2 = 1.8191030582093\text{E}+17$$

$$\text{Tanggungan: } (6 - 3)^2 = 27$$

$$\text{Status Rumah: } (3 - 1)^2 = 8$$

$$\text{Pendidikan: } (3 - 1)^2 = 8$$

$$\text{Jarak} = (1.8191030582093\text{E}+17 + 27 + 8 + 8)^{1/3} = 1.8191030582093\text{E}+17$$

Data Latih	Ph	Tg	St	Pd	S	d	Kelas
Data 1	1.8786789250836E+18	1	8	1	1.8786789250836E+18	1233912	Tidak

Data 2	1.3249889967084E+17	8	0	64	1.3249889967084E+17	509805	Layak
Data 3	1.7720646639079E+19	0	8	64	1.7720646639079E+19	2607113	Layak
Data 4	1.30492623122E+19	8	0	0	1.30492623122E+19	2354301	Tidak
Data 5	1.0252636861238E+19	8	8	1	1.0252636861238E+19	2172427	Layak
Data 6	3.1799470951846E+18	1	8	64	3.1799470951846E+18	1470528	Layak
Data 7	1.5600524608008E+18	0	8	0	1.5600524608008E+18	1159791	Layak
Data 8	1.0904847319568E+18	27	8	0	1.0904847319568E+18	1029295	Layak
Data 9	5.6593735843629E+19	27	0	8	5.6593735843629E+19	3839336	Tidak
Data 10	1.8191030582093E+17	27	8	8	1.8191030582093E+17	566612	Tidak
...	...	...	...	...	...	...	...
Data 1000	1.1826591838701E+18	8	0	8	1.1826591838701E+18	1057515	Tidak

4) Chebyshev Distance

$$\text{Rumus: } d = \max_{i=1}^n (|x_i - y_i|) \dots \dots \dots (4)$$

Data 1:

$$\text{abs}(1433912.00 - 200000) = 1233912$$

$$\text{abs}(4 - 3) = 1$$

$$\text{abs}(3 - 1) = 2$$

$$\text{abs}(2 - 1) = 1$$

Nilai Maximum: 1233912

Data 2:

$\text{abs}(709805.00 - 200000) = 509805$

$\text{abs}(1 - 3) = 2$

$\text{abs}(1 - 1) = 0$

$\text{abs}(5 - 1) = 4$

Nilai Maximum: 509805

Data 3:

$\text{abs}(2807113.00 - 200000) = 2607113$

$\text{abs}(3 - 3) = 0$

$\text{abs}(3 - 1) = 2$

$\text{abs}(5 - 1) = 4$

Nilai Maximum: 2607113

Data 4:

$\text{abs}(2554301.00 - 200000) = 2354301$

$\text{abs}(5 - 3) = 2$

$\text{abs}(1 - 1) = 0$

$\text{abs}(1 - 1) = 0$

Nilai Maximum: 2354301

Data 5:

$\text{abs}(2372427.00 - 200000) = 2172427$

$\text{abs}(1 - 3) = 2$

$\text{abs}(3 - 1) = 2$

$\text{abs}(2 - 1) = 1$

Nilai Maximum: 2172427

Data 6:

$\text{abs}(1670528.00 - 200000) = 1470528$

$\text{abs}(4 - 3) = 1$

$\text{abs}(3 - 1) = 2$

$\text{abs}(5 - 1) = 4$

Nilai Maximum: 1470528

Data 7:

$\text{abs}(1359791.00 - 200000) = 1159791$

$\text{abs}(3 - 3) = 0$

$\text{abs}(3 - 1) = 2$

$\text{abs}(1 - 1) = 0$

Nilai Maximum: 1159791

Data 8:

$\text{abs}(1229295.00 - 200000) = 1029295$

$\text{abs}(6 - 3) = 3$

$\text{abs}(3 - 1) = 2$

$$\text{abs}(1 - 1)=0$$

Nilai Maximum: 1029295

Data 9:

$$\text{abs}(4039336.00 - 200000)=3839336$$

$$\text{abs}(6 - 3)=3$$

$$\text{abs}(1 - 1)=0$$

$$\text{abs}(3 - 1)=2$$

Nilai Maximum: 3839336

Data 10:

$$\text{abs}(766612.00 - 200000)=566612$$

$$\text{abs}(6 - 3)=3$$

$$\text{abs}(3 - 1)=2$$

$$\text{abs}(3 - 1)=2$$

Nilai Maximum: 566612

Data Latih	max(X)	Jarak	Kelas
Data 1	1233912	1233912	Tidak
Data 2	509805	509805	Layak
Data 3	2607113	2607113	Layak
Data 4	2354301	2354301	Tidak
Data 5	2172427	2172427	Layak
Data 6	1470528	1470528	Layak
Data 7	1159791	1159791	Layak
Data 8	1029295	1029295	Layak

Data 9	3839336	3839336	Tidak
Data 10	566612	566612	Tidak
...	...	...	...
Data 1000	1057515	1057515	Tidak

B. Naive Bayes

Langkah: Hitung Prior, Likelihood, Posterior

1) Prior Probability

$$P(\text{Kelas Tidak}) = \frac{394}{1000} = 0.394$$

$$P(\text{Kelas Layak}) = \frac{606}{1000} = 0.606$$

Kelas	Jumlah Data	P(Kelas)
Tidak	394	0.394
Layak	606	0.606

2) Likelihood Probability

Kelas: Tidak

$$\text{Penghasilan} : \frac{(0+1)}{(394+1000)} = 0.00071736011477762$$

$$\text{Tanggungan} : \frac{(75+1)}{(394+6)} = 0.19$$

$$\text{Status Rumah} : \frac{(200+1)}{(394+2)} = 0.50757575757576$$

$$\text{Pendidikan} : \frac{(102+1)}{(394+4)} = 0.25879396984925$$

Atribut	Nilai Uji	Jumlah di Kelas	P(Atribut Tidak)
penghasilan	200000	0	0.0007
tanggungan	3	75	0.19
status_rumah	Kontrak	200	0.5076
pendidikan	SD	102	0.2588

Kelas: Layak

$$\text{Penghasilan} : \frac{(0+1)}{(606+1000)} = 0.00062266500622665$$

$$\text{Tanggungan} : \frac{(101+1)}{(606+6)} = 0.16666666666667$$

$$\text{Status Rumah} : \frac{(277+1)}{(606+2)} = 0.45723684210526$$

$$\text{Pendidikan} : \frac{(162+1)}{(606+4)} = 0.2672131147541$$

Atribut	Nilai Uji	Jumlah di Kelas	P(Atribut Layak)
penghasilan	200000	0	0.0006
tanggungan	3	101	0.1667
status_rumah	Kontrak	277	0.4572
pendidikan	SD	162	0.2672

3) Posterior Probability

$$\text{Kelas Tidak} : 0.394 * 0.000018 = 0.000007$$

$$\text{Kelas Layak} : 0.606 * 0.000013 = 0.000008$$

Kelas	P(Kelas)	P(Atribut Kelas)	Posterior
Tidak	0.394	0.000018	0.000007
Layak	0.606	0.000013	0.000008

3. Uji Coba Program

Uji coba terhadap sistem bertujuan untuk memastikan bahwa sistem sudah berada pada kondisi siap pakai. Instrumen yang digunakan untuk melakukan pengujian ini yaitu dengan menggunakan *Blackbox Testing*:

Tabel 4.3. Blackbox Testing Form Login

No	Form Login	Keterangan	Validitas
1.	Jika pengguna mengisi username dan password dengan benar kemudian melakukan Klik Tombol Submit	Aplikasi menampilkan form Menu	Valid
2	Jika pengguna mengisi username dan password dengan salah kemudian melakukan Klik Tombol Submit	Aplikasi menampilkan pesan kesalahan	Valid

Tabel 4.4. Blackbox Testing Form Menu

No	Form Menu	Keterangan	Validitas
1.	Klik Tombol Menu	Aplikasi menampilkan form Menu	Valid
2.	Klik Tombol Masyarakat	Aplikasi menampilkan form Masyarakat	Valid
3.	Klik Tombol Dataset Latih	Aplikasi menampilkan form Dataset Latih	Valid
4.	Klik Tombol Prediksi	Aplikasi menampilkan form Prediksi	Valid
5.	Klik Tombol Exit	Aplikasi menampilkan form Login	Valid

Tabel 4.5. Blackbox Testing Form Masyarakat

No	Form Masyarakat	Keterangan	Validitas
1.	Klik Tombol Simpan	Aplikasi menyimpan	Valid

		seluruh data di textbox ke dalam table database	
2.	Klik Tombol Ubah	Aplikasi mengubah isi di table database sesuai data yang diubah	Valid
3.	Klik Tombol Hapus	Aplikasi menghapus isi data di database	

Tabel 4.6. Blackbox Testing Form Dataset Latih

No	Form Dataset Latih	Keterangan	Validitas
1.	Klik Tombol Simpan	Aplikasi menyimpan seluruh data di textbox ke dalam table database	Valid
2.	Klik Tombol Ubah	Aplikasi mengubah isi di table database	Valid

		sesuai data yang diubah	
3.	Klik Tombol Hapus	Aplikasi menghapus isi data di database	

Tabel 4.7. Blackbox Testing Form Prediksi

No	Form Prediksi	Keterangan	Validitas
1.	Klik Tombol Hapus	Aplikasi menghapus isi data di database	

Tabel 4.8. Blackbox Testing Form Metode

No	Form Metode	Keterangan	Validitas
1.	Klik Tombol Proses Prediksi	Aplikasi menampilkan hasil prediksi	Valid

3.1 Hasil Uji Coba

Setelah melakukan uji coba terhadap aplikasi, maka dapat disimpulkan hasil yang didapatkan yaitu:

1. *Interface* rancangan telah sesuai dengan *Interface* hasil.
2. Metode K-NN dan Naïve Bayes telah diterapkan pada aplikasi yang dibuat.
3. *Interface* aplikasi bersifat *user friendly*.
4. Aplikasi yang telah dibuat berjalan dengan baik.
5. Aplikasi yang telah dibuat tidak memiliki kesalahan logika.

3.2 Kekurangan Aplikasi

Kekurangan aplikasi pada penelitian ini diantaranya:

1. Aplikasi yang telah dibuat tidak terlalu banyak fitur.
2. Aplikasi yang telah dibuat tidak memiliki petunjuk penggunaan.
3. Aplikasi yang dibuat hanya diterapkan di *web*.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian mengenai “*Penerapan Algoritma K-Nearest Neighbors (KNN) dan Naïve Bayes untuk Menentukan Pemilihan Penerima Bantuan Sosial Berdasarkan Ekonomi Masyarakat*”, dapat diambil beberapa kesimpulan utama yang mencakup aspek metodologi, hasil analisis, akurasi model, serta implikasi penerapan sistem.

1. Efektivitas Penggunaan Algoritma KNN dan Naïve Bayes

Penelitian ini membuktikan bahwa baik KNN maupun Naïve Bayes mampu digunakan untuk melakukan klasifikasi calon penerima bantuan sosial berdasarkan data ekonomi masyarakat. Keduanya bekerja dengan pendekatan yang berbeda:

- a. KNN mengklasifikasikan data berdasarkan kedekatan atribut antara data uji dengan data latih. Semakin kecil nilai jarak (misalnya *Euclidean Distance*), semakin besar kemungkinan data tersebut termasuk dalam kelas yang sama.
- b. Naïve Bayes mengklasifikasikan data berdasarkan probabilitas bersyarat setiap atribut terhadap kelas target dengan asumsi independensi antaratribut.
- c. Kedua algoritma ini memberikan hasil yang cukup baik dalam memprediksi kelayakan penerima bantuan sosial, walaupun karakteristik dataset dan parameter yang digunakan mempengaruhi performa masing-masing algoritma.

2. Perbandingan Akurasi dan Kinerja Model

Berdasarkan pengujian yang dilakukan:

- a. KNN menunjukkan akurasi tinggi pada dataset dengan jumlah atribut numerik yang relevan, terutama saat jumlah tetangga (k) dipilih secara optimal. Algoritma ini relatif sederhana namun efektif ketika data tidak terlalu besar dan distribusi antar kelas cukup jelas. Kelemahan KNN adalah kinerjanya dapat menurun jika terdapat *outlier* atau data latih sangat besar, karena memerlukan komputasi jarak yang intensif.
- b. Naïve Bayes menunjukkan stabilitas akurasi pada berbagai variasi data, terutama jika atribut bersifat kategorikal atau dapat dikonversi ke dalam bentuk probabilitas. Algoritma ini sangat cepat dalam proses pelatihan dan prediksi, namun akurasi dapat sedikit menurun jika asumsi independensi antar atribut tidak terpenuhi sepenuhnya.
- c. Dalam pengujian akhir, KNN menghasilkan akurasi sedikit lebih tinggi dibandingkan Naïve Bayes pada dataset penelitian ini, namun Naïve Bayes memiliki keunggulan dari sisi kecepatan dan efisiensi komputasi.

3. Dampak terhadap Proses Penentuan Penerima Bantuan Sosial

Implementasi sistem berbasis KNN dan Naïve Bayes membawa dampak positif terhadap proses penentuan penerima bantuan sosial:

- a. Lebih objektif: Penilaian kelayakan penerima bantuan tidak hanya berdasarkan pendapat subjektif petugas, tetapi melalui proses klasifikasi berbasis data.
- b. Lebih cepat: Sistem mampu melakukan prediksi secara otomatis dalam waktu singkat.

- c. Transparan: Proses pengambilan keputusan dapat dijelaskan melalui hasil perhitungan jarak (KNN) atau probabilitas (Naïve Bayes).
- d. Mengurangi bias: Dengan adanya model yang berbasis data, kemungkinan diskriminasi atau bias dalam penentuan penerima bantuan dapat diminimalkan.

4. Keterbatasan dan Tantangan

Meskipun kedua algoritma dapat diterapkan, terdapat beberapa keterbatasan yang perlu diperhatikan:

- a. Kualitas data sangat mempengaruhi akurasi model. Data yang tidak lengkap, mengandung kesalahan input, atau tidak representatif dapat menurunkan kinerja algoritma.
- b. Pemilihan parameter K pada KNN sangat krusial. Nilai K yang terlalu kecil membuat model sensitif terhadap *noise*, sedangkan nilai K yang terlalu besar dapat mengaburkan batas antar kelas.
- c. Asumsi independensi Naïve Bayes tidak selalu terpenuhi pada data ekonomi masyarakat, karena beberapa variabel ekonomi sering saling berkorelasi.
- d. Skalabilitas sistem perlu dipertimbangkan jika dataset bertambah besar seiring waktu, khususnya untuk KNN yang memiliki beban komputasi tinggi.

5. Implikasi Penelitian

Berdasarkan hasil yang diperoleh, penerapan KNN dan Naïve Bayes dalam penentuan penerima bantuan sosial memberikan gambaran bahwa:

- a. Teknologi klasifikasi berbasis data mining dapat diintegrasikan dalam sistem informasi pemerintah daerah untuk mendukung pengambilan keputusan.
- b. Hasil model dapat menjadi rekomendasi, bukan keputusan final, sehingga tetap diperlukan validasi lapangan untuk mengonfirmasi kelayakan penerima bantuan.
- c. Pengembangan lebih lanjut dapat dilakukan dengan menggabungkan kedua algoritma (*ensemble method*) atau mengintegrasikan dengan metode lain seperti *Decision Tree* atau *Random Forest* untuk meningkatkan akurasi.

5.2. Saran

Saran dari Penerapan Algoritma K-Nearest Neighbors (KNN) Dan Naïve Bayes Untuk Menentukan Pemilihan Penerima Bantuan Sosial Berdasarkan Ekonomi Masyarakat dapat dilihat sebagai berikut:

1. Peningkatan kualitas dan kuantitas data dengan memperluas cakupan wilayah, menambah jumlah sampel, dan memastikan data bersih dari duplikasi atau kesalahan.
2. Penggunaan normalisasi atau standarisasi data untuk KNN agar perbedaan skala antar atribut tidak mempengaruhi perhitungan jarak.
3. Eksperimen nilai K pada KNN dan penggunaan teknik *cross-validation* untuk memperoleh parameter terbaik.
4. Penggunaan Naïve Bayes dengan modifikasi seperti *Gaussian Naïve Bayes* atau *Multinomial Naïve Bayes* sesuai karakteristik data.

5. Integrasi ke dalam aplikasi berbasis web atau mobile agar dapat digunakan langsung oleh dinas sosial atau perangkat desa.
6. Penggabungan metode (*hybrid model*) seperti KNN-Naïve Bayes atau KNN-Decision Tree untuk memaksimalkan kelebihan masing-masing algoritma.
7. Penerapan visualisasi hasil prediksi agar pihak pengambil kebijakan dapat memahami data dan hasil klasifikasi dengan mudah.

DAFTAR PUSTAKA

- Alfiah, N. (2021). Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes. *Respati*, 16(1), 32. <https://doi.org/10.35842/jtir.v16i1.386>
- Anwar, D. P., Ati, N. U., & Pindahanto, R. (2020). Implementasi Program Bantuan Pangan Non Tunai (Bpnt) Dinas Sosial Dalam Menanggulangi Kemiskinan Di Kelurahan Sisir Kecamatan Batu Kota Batu. *Jurnal Respon Publik*, 14(3), 1–7.
- Aripin, J. J. (2019). *Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi pada BPR Pantura*. <https://repository.nusamandiri.ac.id/index.php/repo/viewitem/13890>
- Azlil Huriah, D., & Dienwati Nuris, N. (2023). Klasifikasi Penerima Bantuan Sosial Umkm Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 360–365. <https://doi.org/10.36040/jati.v7i1.6300>
- Bantuan, P., Studi, S., Desa, K., Jombang, P., Wijaya, A. R., Lazulfa, I., Rizal, M. F., & Andriani, A. (n.d.). *IMPLEMENTASI METODE K-NEAREST NEIGHBOR (KNN) UNTUK MENENTUKAN. 0*, 165–172.
- Djaenal, R., J. E. Kaawoan, & Rachman, I. (2021). Implementasi Kebijakan Program Bantuan Pangan Non Tunai (BPNT) Dinas Sosial Dalam Menanggulangi Kemiskinan di Kelurahan Tosa Kecamatan Tidore Timur Kota Tidore. *Jurnal Governance*, 1(2), 1–8.
- Fadhli, K., & Nazila, L. R. (2023). Pengaruh Implementasi Bantuan Sosial Bpnt Dan Pkh Terhadap Efektivitas Penanggulangan Kemiskinan. *Jurnal Education and Development*, 11(2), 196–202. <https://doi.org/10.37081/ed.v11i2.4654>
- Fauziyyah, H., & Murnawan, M. (2024). Penerapan Metode SWARA-ELECTRE Dalam Pemilihan Penerima Bantuan Sosial Kelompok Usaha Bersama (KUBE). *Jurnal Nasional Teknologi Dan Sistem Informasi*, 9(3), 271–279. <https://doi.org/10.25077/teknosi.v9i3.2023.271-279>
- Hadiani, I., Astuti, R., & Suprpti, T. (2024). Penerapan Algoritma Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Sosial Di Desa Golat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3616–3620. <https://doi.org/10.36040/jati.v7i6.8284>
- Heliyanti Susana. (2022). Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, 4(1), 1–8. <https://doi.org/10.52005/jursistekni.v4i1.96>
- Laurentcia, S., & Yusran, R. (2021). Evaluasi Program Bantuan Pangan Non Tunai

- dalam Penanggulangan Kemiskinan di Kecamatan Nanggalo Kota Padang. *Journal of Civic Education*, 4(1), 7–17. <https://doi.org/10.24036/jce.v4i1.433>
- Loka, S. K. P., & Marsal, A. (2023). Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier untuk Klasifikasi Status Gizi Pada Balita. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 8–14. <https://doi.org/10.57152/malcom.v3i1.474>
- Nasution, A. (2015). Metodologi Penelitian: Metodologi penelitian Skripsi. *Rake Sarasin*, 36.
- Nikmatun, Alvi, I., Waspada, & Indra. (2019). Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2), 421–432.
- Rahmadini, R., Enjel Erika LorencisLubis, Aji Priansyah, Yolanda R.W.N, & Tuti Meutia. (2023). Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Mahasiswa Akuntansi Samudra*, 4(4), 223–235. <https://doi.org/10.33059/jmas.v4i4.7074>
- Rani, L. N. (2016). Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit. *INOVTEK Polbeng - Seri Informatika*, 1(2), 126. <https://doi.org/10.35314/isi.v1i2.131>
- Sekar Setyaningtyas, Indarmawan Nugroho, B., & Arif, Z. (2022). Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 10(2), 52–61. <https://doi.org/10.21063/jtif.2022.v10.2.52-61>
- Siregar, E. T. (2022). Penerapan Metode Multi-Attribute Utility Theory Dalam Pemilihan Penerima Bantuan Sosial Jasa Pelayanan Public Pengabdian Kepada Masyarakat Terdampak Covid-19. *METHODIKA: Jurnal Teknik Informatika Dan Sistem Informasi*, 8(2), 24–28. <https://doi.org/10.46880/mtk.v8i2.1308>
- Siringoringo, F. (2023). Sistem Pendukung Keputusan Pemilihan Penerima Bantuan Sosial Menggunakan Metode Electre & Roc. *Management of Information System Journal*, 1(3), 85–95. <https://doi.org/10.47065/mis.v1i3.746>
- Sitanayah, L., Kansil, R. A. ., & Kumenap, V. D. (2024). Sistem Pendukung Keputusan Seleksi Penerima Bantuan Sosial Menggunakan Metode Simple Additive Weighting. *JOINTER : Journal of Informatics Engineering*, 5(01), 14–19. <https://doi.org/10.53682/jointer.v5i01.313>

- Suprpto, S., Edora, E., & Pasaribu, F. A. (2024). Sistem Pendukung Keputusan Calon Penerima Program Bantuan Sosial (BANSOS) Menggunakan Metode Simple Additive Weighting (SAW). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 188–197. <https://doi.org/10.57152/malcom.v4i1.1057>
- Tangkelayuk, A. (2022). The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 1109–1119. <https://doi.org/10.35957/jatisi.v9i2.2048>
- Triayudi, A. (2023). Penerapan Data Mining Untuk Klasifikasi Penerima Dana Bantuan Sosial Dengan Menggunakan Algoritma K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, 5(2), 532–542. <https://doi.org/10.47065/bits.v5i2.3972>
- Yaqin, M. A., Alfita, R., & Wibisono, K. A. (2020). Alat Identifikasi Jenis Kayu Berbasis Image Processing Dengan Metode K-Nearest Neighbor (KNN). *SinarFe7*, 1–6. <https://garuda.kemdikbud.go.id/documents/detail/2813489>