

**PERBANDINGAN SISTEM PREDIKSI MENGGUNAKAN METODE
MONTE CARLO DENGAN METODE K-NN PADA NILAI PESERTA
DIDIK UJI KOMPETENSI KEJURUAN
(STUDI KASUS PADA SMK NEGERI 1 STABAT)**

SKRIPSI

DISUSUN OLEH

AZZAHRAH
NPM. 2009020072



UMSU
Unggul | Cerdas | Terpercaya

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

MEDAN

2024

**PERBANDINGAN SISTEM PREDIKSI MENGGUNAKAN METODE
MONTE CARLO DENGAN METODE K-NN PADA NILAI PESERTA
DIDIK UJI KOMPETENSI KEJURUAN
(STUDI KASUS PADA SMK NEGERI 1 STABAT)**

SKRIPSI

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

**AZZAHRAH
NPM. 2009020072**

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2024**

LEMBAR PENGESAHAN

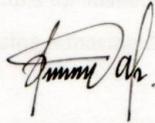
Judul Skripsi : **Perbandingan Sistem Prediksi Menggunakan Metode Monte Carlo Dengan Metode K-NN Pada Nilai Peserta Didik Uji Kompetensi Kejuruan (Studi Kasus Pada SMK Negeri 1 Stabat)**

Nama Mahasiswa : Azzahrah

NPM : 2009020072

Program Studi : Teknologi Informasi

Menyetujui
Komisi Pembimbing



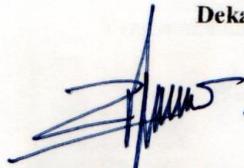
(Indah Purnama Sari, S.T., M.Kom)
NIDN. 0116049001

Ketua Program Studi



(Fatma Sari Hutagalung, S.Kom., M.Kom)
NIDN. 0117019301

Dekan



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

PERNYATAAN ORISINALITAS

**PERBANDINGAN SISTEM PREDIKSI MENGGUNAKAN METODE
MONTE CARLO DENGAN METODE K-NN PADA NILAI PESERTA
DIDIK UJI KOMPETENSI KEJURUAN
(STUDI KASUS PADA SMK NEGERI 1 STABAT)**

SKRIPSI

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

Medan, 09 Agustus 2024

Yang membuat pernyataan



AZZAHRAH

NPM. 2009020072

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

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

Nama : AZZAHRAH
NPM : 2009020072
Program Studi : Teknologi Informasi
Karya Ilmiah : Skripsi

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

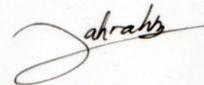
**PERBANDINGAN SISTEM PREDIKSI MENGGUNAKAN METODE
MONTE CARLO DENGAN METODE K-NN PADA NILAI PESERTA
DIDIK UJI KOMPETENSI KEJURUAN
(STUDI KASUS PADA SMK NEGERI 1 STABAT)**

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, 09 Agustus 2024

Yang membuat pernyataan



AZZAHRAH

NPM. 2009020072

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : AZZAHRAH
Tempat dan Tanggal Lahir : Medan, 11 Desember 2001
Alamat Rumah : Jl. Satu No B33 Pulo Brayan Bengkel Baru
Telepon/Faks/HP : 081328430772
E-mail : azzahrahzahrah11@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : SD Negeri 060873 Medan TAMAT : 2013
SMP : SMP Swasta Laksamana Martadinata TAMAT : 2016
SMA : SMA Swasta Laksamana Martadinata TAMAT : 2019

KATA PENGANTAR



Dengan segala puji syukur kepada Allah *Subhanahu wa ta'ala* dan atas dukungan dan do'a dari orang tercinta, akhirnya Skripsi ini dapat di selesaikan dengan baik dan tepat waktu. Dalam penyusunan skripsi ini, penulis banyak mendapatkan arahan, nasihat dan bimbingan sehingga skripsi ini dapat terselesaikan dengan baik. Penulis ingin menyampaikan terima kasih kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU).
2. Bapak Prof. Dr. Muhammad Arifin, S.H., M.Hum, selaku Wakil Rektor I Universitas Muhammadiyah Sumatera Utara.
3. Bapak Prof. Dr. Akrim, M.Pd, selaku Wakil Rektor II Universitas Muhammadiyah Sumatera Utara.
4. Bapak Assoc. Prof. Dr. Rudianto, S.Sos., M.Si, selaku Wakil Rektor III Universitas Muhammadiyah Sumatera Utara.
5. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom, selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
6. Bapak Halim Maulana, S.Kom, M.Kom dan Bapak Lutfi Basit, S.Sos, M.I.Kom, selaku Wakil Dekan I dan III Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatrea Utara.
7. Ibu Fatma Sari Hutagalung, S.Kom, M.Kom, selaku Ketua Program Studi Teknologi Informasi.

8. Bapak Mhd. Basri, S.Si., M.Kom, selaku Sekretaris Program Studi Teknologi Informasi
9. Ibu Indah Purnama Sari, S.T., M.Kom, selaku Dosen Pembimbing dalam penelitian ini. Terima kasih telah banyak membantu, meluangkan waktu, memberikan bimbingan dan arahan dengan sangat baik kepada penulis sehingga dapat menyelesaikan penelitian ini. Semoga Ibu selalu diberikan kesehatan, keberkahan, dan kesuksesan dalam segala aspek kehidupan.
10. Kedua orang tua penulis tersayang, Ayahanda Zul dan Ibunda Rida selaku orang tua penulis yang selalu memberikan doa, semangat, kasih sayang serta pengorbanan tanpa henti demi memastikan saya mendapatkan pendidikan terbaik.
11. Kepada Aulia Jannah, Isnaini Faiz Qathrunada, Thamita Anggraini, dan Adila Mawadda Meuraxa, terima kasih atas dukungan tanpa henti kalian. Kehadiran kalian membuat perjalanan ini terasa lebih ringan dan menyenangkan. Terima kasih telah menjadi teman yang keren dan luar biasa. Semoga pertemanan kita terus terjalin dan sukses selalu menyertai langkah kita.
12. Teman-teman satu kelompok KKN yaitu Aulia, Aini, Mita, Dila, Mala, Ayu, Dimas, Zharfan, Ibal, Ija, Haikal, Royhan, Rahmad, dan Naufal. Terima kasih untuk canda, tawa, dan momen-momen berharga yang kita lalui bersama.
13. Semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu yang telah membantu penyelesaian penelitian ini.

14. Tidak lupa untuk berterima kasih kepada diri saya yang telah berjuang melewati berbagai tantangan dan rintangan, serta tidak pernah menyerah meskipun dalam keadaan sulit. Setiap langkah yang saya ambil, baik itu keberhasilan maupun kegagalan, telah menjadi bagian penting dari perjalanan ini. Semua pengalaman ini telah membentuk saya menjadi pribadi yang lebih baik dan lebih kuat.

Penulis menyadari bahwa masih banyak kekurangan dalam penyusunan skripsi ini. Oleh karena itu, penulis akan sangat menghargai dan menerima kritikan dan saran untuk membuat skripsi ini menjadi lebih baik lagi. Semoga skripsi ini dapat memberikan manfaat bagi kita semua dan semoga bantuan serta partisipasi yang diberikan oleh semua pihak bernilai ibadah disisi Allah Subhanahu Wata'ala.

Wassalamualaikum Warahmatullahi Wabarakatuh

Medan, 09 Agustus 2024

Penulis



AZZAHRAH

**PERBANDINGAN SISTEM PREDIKSI MENGGUNAKAN METODE
MONTE CARLO DENGAN METODE K-NN PADA NILAI PESERTA
DIDIK UJI KOMPTENSI KEJURUAN
(STUDI KASUS PADA SMK NEGERI 1 STABAT)**

ABSTRAK

Penelitian ini bertujuan untuk membandingkan kinerja dua metode prediksi, yaitu *Monte Carlo* dan *K-Nearest Neighbors* (K-NN), dalam memprediksi nilai peserta didik pada Uji Kompetensi Kejuruan (UKK). *Monte Carlo* dikenal sebagai metode statistik yang menggunakan simulasi acak untuk menghasilkan hasil prediktif, sementara K-NN adalah metode pembelajaran mesin yang menggunakan tetangga terdekat untuk klasifikasi dan regresi. Dalam penelitian ini, data nilai peserta didik dikumpulkan dan dibagi menjadi data latih dan data uji. Akurasi prediksi dari kedua metode tersebut dihitung dan dibandingkan. Hasil penelitian menunjukkan bahwa metode *Monte Carlo* memberikan akurasi prediksi sebesar 85,26%, sedangkan metode K-NN memberikan akurasi prediksi sebesar 85,37%. Dengan demikian, K-NN sedikit lebih unggul dalam hal akurasi prediksi dibandingkan *Monte Carlo*. Penelitian ini memberikan wawasan berharga dalam memilih metode prediksi yang lebih efektif untuk evaluasi nilai peserta didik pada UKK.

Kata Kunci: Akurasi; *K-Nearest Neighbors*; *Monte Carlo*; Prediksi; Uji Kompetensi Kejuruan.

**COMPARISON OF PREDICTION SYSTEMS USING THE MONTE CARLO
METHOD WITH THE K-NN METHOD ON THE SCORES OF
VOCATIONAL COMPETENCY TEST STUDENTS
(CASE STUDY AT SMK NEGERI 1 STABAT)**

ABSTRACT

This study aims to compare the performance of two prediction methods, Monte Carlo and K-Nearest Neighbors (K-NN), in predicting student scores in the Vocational Competency Test (UKK). Monte Carlo is known as a statistical method that uses random simulation to generate predictive results, while K-NN is a machine learning method that uses nearest neighbors for classification and regression. In this study, student score data were collected and divided into training and testing data. The prediction accuracy of both methods was calculated and compared. The results showed that the Monte Carlo method achieved a prediction accuracy of 85,26%, whereas the K-NN method achieved a prediction accuracy of 85,37%. Thus, Monte Carlo is slightly superior in terms of prediction accuracy compared to K-NN. This research provides valuable insights into selecting a more effective prediction method for evaluating student scores in UKK.

Keywords: Accuracy; K-Nearest Neighbor; Monte Carlo; Prediction; Vocational Competency Test.

DAFTAR ISI

LEMBAR PENGESAHAN	ii
PENYATAAN ORISINALITAS.....	iii
PENYATAAN PERSETUJUAN PUBLIKASI.....	iii
RIWAYAT HIDUP	iv
KATA PENGANTAR.....	v
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	4
1.5.1 Bagi Pendidikan.....	4
1.5.2 Bagi Penulis	4
1.5.3 Bagi Institusi.....	4
1.5.4 Bagi Masyarakat	5
BAB II LANDASAN TEORI	6
2.1. Sistem Prediksi	6
2.2. Uji Kompetensi Kejuruan (UKK)	7
2.3. <i>Monte Carlo</i>	7
2.4. <i>Data Mining</i>	8
2.4.1. Pengelompokkan <i>Data Mining</i>	9
2.4.2. Tahapan <i>Data Mining</i>	11
2.5. <i>K-Nearest Neighbor</i> (K-NN).....	12
2.6. <i>Python</i>	13

2.7. Streamlit	14
2.8. Visual Studio Code.....	15
2.9. Flowchart.....	16
2.10. Penelitian Terkait.....	17
2.11. Penelitian Terdahulu.....	18
BAB III METODOLOGI PENELITIAN	21
3.1. Jenis Penelitian	21
3.2. Tahapan Penelitian	21
3.3. Tempat dan Waktu Penelitian	22
3.4. Perangkat Penelitian	23
3.5. Flowchart Penelitian.....	24
BAB IV HASIL DAN PEMBAHASAN.....	27
4.1. Analisis Data	27
4.2. Prediksi Menggunakan <i>Monte Carlo</i>	28
4.2.1. Import Library	28
4.2.2. Membaca dataset	28
4.2.3. Menentukan dan memisahkan kolom target dan fitur	29
4.2.4. Implementasi Simulasi <i>Monte Carlo</i>	29
4.2.5. Menampilkan Hasil Distribusi.....	30
4.3. Prediksi Menggunakan <i>K-Nearest Neighbors</i> (K-NN)	30
4.3.1. Import Library	30
4.3.2. Import Dataset	31
4.3.3. Data Cleaning	31
4.3.4. Exploratory Data Analysis (EDA).....	31
4.3.5. Split Data	32
4.3.6. Modeling.....	32
4.3.7. Prediksi	33
4.3.8. Evaluasi Performa <i>K-Nearest Neighbors</i>	33
4.4. Pengujian <i>Monte Carlo</i>	34
4.4.1. Analisis Hasil Prediksi.....	34
4.5. Pengujian <i>K-Nearest Neighbors</i> (K-NN)	35
4.5.1. Hasil Pengujian <i>K-Nearest Neighbors</i> (K-NN).....	36

4.6. Perbandingan <i>K-Nearest Neighbors</i> (K-NN) dan <i>Monte Carlo</i>	44
4.7. Keunggulan dan Kelemahan Metode <i>Monte Carlo</i> dan K-NN.....	47
BAB V KESIMPULAN DAN SARAN	49
5.1. Kesimpulan.....	49
5.2. Saran.....	50
DAFTAR PUSTAKA	51
LAMPIRAN.....	55

DAFTAR TABEL

Tabel	Halaman
Tabel 2.1 Simbol Flowchart beserta artinya	16
Tabel 2.2 Penelitian terdahulu mengenai <i>Monte Carlo</i> dan K-NN.....	18
Tabel 3.1 Tabel Pelaksanaan Penelitian.....	23
Tabel 4.1 Data Siswa Teknik Pemesinan SMK Negeri 1 Stabat	27
Tabel 4.2. Skenario Pengujian <i>K-Nearest Neighbors</i>	35
Tabel 4.3. Confusion Matrix K-NN Pengujian 1 Dengan Variasi k=2.....	36
Tabel 4.4. Confusion Matrix K-NN Pengujian 1 Dengan Variasi k=7.....	38
Tabel 4.5. Confusion Matrix K-NN Pengujian 2 dengan Variasi k=5.....	40
Tabel 4.6. Confusion Matrix K-NN Pengujian 2 dengan Variasi k=5.....	41
Tabel 4.7. Hasil Pengujian <i>K-Nearest Neighbors</i> Variasi k=2 dan k=5	43
Tabel 4.8 Perbandingan antara Metode <i>Monte Carlo</i> dan K-NN	45

DAFTAR GAMBAR

Gambar	Halaman
Gambar 2.1 Tahapan Data Mining.....	11
Gambar 2.2 Logo bahasa pemrograman python	14
Gambar 2.3 Contoh tampilan sederhana streamlit	15
Gambar 2.4 Visual Studio Code.....	16
Gambar 3.1 Tahapan Penelitian	21
Gambar 3. 2 Flowchart Penelitian.....	24
Gambar 3.3 Flowchart Metode <i>Monte Carlo</i>	25
Gambar 3.4 Flowchart Metode <i>K-Nearest Neighbors</i>	26
Gambar 4.1 Import library pada <i>Monte Carlo</i>	28
Gambar 4.2 Membaca dataset csv	28
Gambar 4.3 Memisahkan antara fitur dan target.....	29
Gambar 4.4 Simulasi <i>Monte Carlo</i>	29
Gambar 4.5 Menampilkan hasil distribusi.....	30
Gambar 4.6 Import library pada <i>K-Nearest Neighbors</i>	30
Gambar 4.7 Import dataset pada proses <i>K-Nearest Neighbors</i>	31
Gambar 4.8 Memeriksa missing value pada <i>K-Nearest Neighbors</i>	31
Gambar 4. 9 Visualisasi data nilai_ket.....	31
Gambar 4.10 Split data pada proses <i>K-Nearest Neighbors</i>	32
Gambar 4.11 Model <i>K-Nearest Neighbors</i>	32
Gambar 4.12 Prediksi <i>K-Nearest Neighbors</i>	33
Gambar 4.13 Confusion Matrix <i>K-Nearest Neighbors</i>	33
Gambar 4.14 Tampilan Confusion Matrix	34
Gambar 4. 15 Tampilan 10 simulasi <i>Monte Carlo</i>	34
Gambar 4.16 Heatmap pengujian 1 dengan k=2	36
Gambar 4.17 Hasil performa pengujian 1 dengan variasi k=2.....	37
Gambar 4.18 Heatmap pengujian 2 dengan variasi k=2	38
Gambar 4.19 Hasil performa pengujian 2 dengan variasi k=2.....	39
Gambar 4.20 Heatmap pengujian 1 dengan variasi k=5	39

Gambar 4.21 Hasil pengujian 1 dengan variasi $k=5$	40
Gambar 4.22 Heatmap pengujian 2 dengan variasi $k=5$	41
Gambar 4.23 Hasil pengujian 2 dengan variasi $k=5$	42
Gambar 4.24 Tampilan streamlit.....	47

DAFTAR LAMPIRAN

Lampiran	Halaman
Lampiran 1. SK – 1 Persetujuan Topik/Judul Penelitian	55
Lampiran 2. SK-2 Penetapan Dosen Pembimbing.....	56
Lampiran 3. SK-3 Formulir Berita Acara Pembimbing.....	57
Lampiran 4. SK-4 Surat Permohonan Sempro.....	59
Lampiran 5. SK-5 Surat Plagiasi.....	60
Lampiran 6. Surat Izin Riset	61
Lampiran 7. Letter of Acceptance (LoA).....	62
Lampiran 8. Data Siswa Teknik Pemesinan SMK Negeri 1 Stabat.....	63
Lampiran 9. Kode Program.....	70
Lampiran 10. Dokumentasi Penelitian.....	73

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Pendidikan kejuruan, terutama di tingkat Sekolah Menengah Kejuruan (SMK), memainkan peran sentral dalam membekali peserta didik dengan keterampilan praktis untuk memasuki dunia kerja. Dalam konteks ini, prediksi nilai uji kompetensi kejuruan menjadi kritis untuk memberikan gambaran yang akurat tentang kesiapan peserta didik. Prediksi ini bukan hanya sekedar evaluasi, tetapi juga merupakan landasan bagi pengambilan keputusan strategis dalam mengarahkan pembelajaran dan menyusun kurikulum. Meskipun sudah ada upaya untuk mengembangkan sistem prediksi nilai uji kompetensi kejuruan, tetap ada tantangan yang perlu diatasi. Variabilitas dalam kemampuan peserta didik, perubahan dinamis dalam kurikulum, dan faktor-faktor eksternal yang memengaruhi hasil uji kompetensi menjadi beberapa faktor yang mempersulit proses prediksi yang akurat.

SMK Negeri 1 Stabat dipilih sebagai studi kasus dalam penelitian ini, mengingat peran signifikan sekolah ini dalam pendidikan kejuruan di wilayah tersebut. Fokus penelitian ini tidak hanya pada prediksi nilai uji kompetensi, tetapi juga pada perbandingan kinerja dua metode prediksi, yakni Metode *Monte Carlo* dan Metode K-NN. Dengan merinci ruang lingkup penelitian, diharapkan hasil penelitian dapat memberikan kontribusi yang lebih spesifik terhadap pengembangan sistem prediksi. Tujuan utama penelitian ini adalah untuk membandingkan keefektifan Metode *Monte Carlo* dengan Metode K-NN dalam memprediksi nilai peserta didik uji kompetensi kejuruan di SMK Negeri 1 Stabat.

Melalui perbandingan ini, diharapkan dapat diidentifikasi metode yang lebih unggul dalam memberikan prediksi yang akurat, sehingga dapat menjadi dasar bagi pengembangan sistem prediksi yang lebih efisien di lingkungan pendidikan kejuruan.

Dalam era big data dan revolusi industri 4.0, kemampuan untuk memanfaatkan teknologi dalam analisis data pendidikan menjadi sangat krusial. Oleh karena itu, penelitian ini tidak hanya relevan bagi SMK Negeri 1 Stabat, tetapi juga bagi lembaga pendidikan lainnya yang menghadapi tantangan serupa dalam mengevaluasi dan meningkatkan kualitas pendidikan mereka. Melalui studi ini, diharapkan dapat memberikan kontribusi signifikan dalam pengembangan metode analisis data yang lebih baik untuk mendukung pendidikan kejuruan di Indonesia.

Berdasarkan latar belakang yang telah diuraikan diatas maka penulis akan mencoba melakukan hal perbandingan prediksi nilai uji kompetensi kejuruan dengan metode *Monte Carlo* dan K-NN serta bahasa pemrograman Python sebagai visualisasi data. Maka penulis mengangkat sebuah tema pada penelitian ini dengan judul “Perbandingan Sistem Prediksi Menggunakan Metode *Monte Carlo* Dengan Metode K-NN Pada Nilai Peserta Didik Uji Kompetensi Kejuruan (Studi Kasus SMK Negeri 1 Stabat)”.

1.2. Rumusan Masalah

Berdasarkan uraian yang menjadi latar belakang penelitian ini, maka didapatkan rumusan masalah sebagai berikut:

1. Bagaimana memperoleh hasil yang efektif dalam melakukan perbandingan sistem prediksi menggunakan Metode *Monte Carlo* dan Metode K-NN dalam memprediksi nilai peserta didik pada uji kompetensi kejuruan.

2. Identifikasi Kelebihan pada masing-masing metode prediksi.
3. Penerapan streamlit untuk menampilkan data prediksi dengan menggunakan simulasi *Monte Carlo* dan K-NN.

1.3. Batasan Masalah

Untuk memberikan fokus yang lebih jelas pada penelitian ini dan mencegah masalah yang dihadapi menjadi terlalu meluas, penting untuk menetapkan batasan permasalahan yang akan dibahas, yaitu:

1. Penelitian ini dilakukan di SMKN 1 Stabat yang beralamat di jalan KH Wahid Hasyim, Stabat.
2. Penelitian dengan menggunakan *Simulasi Monte Carlo* dan *K-Nearest Neighbor* dan membandingkan hasil mana yang lebih akurat antara *Monte Carlo* dan *K-Nearest Neighbor*.
3. Penelitian yang dilakukan menggunakan data 5(lima) Tahun terakhir untuk disimulasikan dengan metode *Monte Carlo* dan K-NN.
4. Menggunakan streamlit untuk menampilkan hasil dari metode *Monte Carlo* dan K-NN.

1.4. Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini sebagai berikut:

1. Mengetahui prediksi simulasi *Monte Carlo* dan metode K-NN pada nilai peserta didik uji kompetensi kejuruan.
2. Mengembangkan streamlit untuk memuat hasil dari simulasi *Monte Carlo* dan metode K-NN.

3. Membandingkan keefektifan Metode *Monte Carlo* dengan Metode K-NN dalam memprediksi nilai peserta didik uji kompetensi kejuruan.
4. Mengidentifikasi kelebihan dan kekurangan masing-masing metode prediksi.

1.5. Manfaat Penelitian

Adapun manfaat yang diharapkan penulis adalah:

1.5.1 Bagi Pendidikan

- a. Pengembangan pengetahuan mahasiswa pada penerapan metode prediksi dalam konteks pendidikan.
- b. Memberikan pandangan mengenai perbandingan antara dua algoritma yaitu *Monte Carlo* dan *K-Nearest Neighbor*. Hasil dari penelitian ini dapat memberikan kesimpulan dari kedua algoritma tersebut. Sehingga dapat dimanfaatkan sebagai sumber literasi yang kedepannya bisa dikembangkan oleh para peneliti lain di masa depan.

1.5.2 Bagi Penulis

- a. Menambah wawasan mengenai metode *Monte Carlo* dan *K-Nearest Neighbor* dalam memprediksi.
- b. Mengetahui metode prediksi mana yang lebih efektif dapat membantu dalam evaluasi peserta didik pada uji kompetensi kejuruan.
- c. Pengembangan keterampilan dengan memperluas pengetahuan dan keterampilan dalam pengembangan visualisasi data dengan *framework* streamlit.

1.5.3 Bagi Institusi

- a. Menambah informasi mengenai perbandingan kedua metode yang menghasilkan keakuratan dalam memprediksi.

- b. Pengembangan *framework* python dengan menggunakan streamlit dalam memvisualisasi data.

1.5.4 Bagi Masyarakat

- a. Memberikan pengetahuan mengenai metode prediksi *Monte Carlo* dan *K-Nearest Neighbor*, serta penerapan *framework* streamlit sebagai visualisasi data.
- b. Meningkatkan minat dan keterampilan di bidang Informatika, khususnya *Data Mining* yang merupakan bagian dari kecerdasan buatan.

BAB II

LANDASAN TEORI

2.1. Sistem Prediksi

Meskipun definisi sistem bervariasi di berbagai bidang, semua sistem memiliki beberapa elemen umum yang penting. Sistem harus terdiri dari elemen-elemen, lingkungan, interaksi antar elemen dan dengan lingkungannya, serta tujuan yang ingin dicapai. Berdasarkan hal ini, sistem dapat didefinisikan sebagai kumpulan komponen yang saling terhubung untuk mencapai tujuan bersama. Elemen-elemen sistem ini bisa berupa manusia, mesin, prosedur, dokumen, data, atau komponen lainnya yang terorganisir. Selain berinteraksi satu sama lain, elemen-elemen tersebut juga berinteraksi dengan lingkungan sekitarnya dalam upaya mencapai tujuan yang telah ditetapkan sebelumnya.

Prediksi merupakan proses estimasi sistematis mengenai kemungkinan peristiwa di masa mendatang, berdasarkan data historis dan terkini. Tujuannya adalah meminimalkan perbedaan antara perkiraan dan kejadian aktual. Penting untuk dipahami bahwa prediksi tidak selalu memberikan kepastian mutlak, melainkan berupaya memberikan perkiraan yang seakurat mungkin tentang apa yang akan terjadi. Proses ini tidak ditujukan untuk memberikan jawaban yang pasti tentang kejadian di masa depan, tetapi lebih kepada upaya mendekati kemungkinan yang paling akurat. Dengan demikian, prediksi berfungsi sebagai alat untuk mempersiapkan diri menghadapi berbagai skenario yang mungkin terjadi, berdasarkan analisis informasi yang tersedia saat ini. (Herdianto, et. al. 2013).

Menurut Kamus Besar Bahasa Indonesia (KBBI), prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang

akan datang dengan menggunakan data masa lalu. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan merupakan input bagi proses perencanaan dan pengambilan keputusan.

2.2. Uji Kompetensi Kejuruan (UKK)

Uji Kompetensi Kejuruan (UKK) merupakan upaya untuk mengetahui tingkat kompetensi dan kompetensi yang telah dicapai mahasiswa melalui pengembangan produk dan manufaktur. UKK dilaksanakan pada akhir masa studi, karena hasil UKK Anda tersedia dalam bentuk sertifikat untuk membantu Anda dalam mencari pekerjaan (Setiawati, 2015).

2.3. Monte Carlo

Metode Monte Carlo adalah "kelas algoritma komputasi yang luas yang mengandalkan pengambilan sampel acak berulang untuk memperoleh hasil numerik". Pada prinsipnya pendekatan ini dapat digunakan untuk menyelesaikan masalah komputasi dengan variabel acak. Ketika suatu program menggunakan metode Monte Carlo digunakan dalam program untuk memperkirakan indikator kinerja tertentu dari sistem nyata (Luke Bertot, et. al. 2018). Metode *Monte Carlo* membantu menganalisis, memecahkan, dan mengoptimalkan berbagai masalah matematika atau fisika dengan menggunakan sejumlah besar sampel statistic acak untuk mensimulasikan kejadian acak (Ang Li, et. al. 2018).

Monte Carlo adalah teknik statistik yang menggunakan sampel acak untuk menyelesaikan masalah matematika dan statistik. Teknik ini menggunakan bilangan acak untuk mensimulasikan proses yang kompleks, dan hasilnya digunakan untuk memperkirakan solusi dari masalah tersebut (Ghozali, 2018).

Rumus membangkitkan angka acak dari *Monte Carlo* dapat dilihat pada persamaan (2.1):

$$Z_i = (a * Z_{i-1} + c) \text{ mod } m \quad \dots\dots\dots (2.1)$$

Keterangan :

Z_i = Nilai angka acak ke- i

a = Konstanta Penggali ($a < m$)

Z_{i-1} = nilai acak sebelumnya (Z_0 adalah nilai awal yang mewakili kunci yang dihasilkan, disebut juga nilai seed, jadi nilai Z_0 adalah bilangan bulat dengan kondisi $Z_0 \geq 0$ dan $Z_0 < m$)

c = Konstanta Pergeseran ($c < m$)

m = Konstanta Modulus ($m > 0$)

2.4. Data Mining (Penambangan Data)

Data mining adalah proses yang mencakup pengumpulan dan penggunaan data historis untuk menemukan pola, keteraturan, atau hubungan dalam dataset yang besar. Hasil dari proses ini dapat digunakan untuk meningkatkan kualitas pengambilan keputusan di masa mendatang (Suyanto, 2017).

Penambangan data hanyalah sebuah teknik untuk menemukan pola yang berguna dalam data, penambangan data memiliki definisi dan standar kriteria untuk data mining. Data mining juga dikenal sebagai penemuan pengetahuan, pembelajaran mesin, dan analisis prediktif (Kotu et al, 2014).

Data mining merupakan proses yang memanfaatkan statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi serta mengidentifikasi informasi yang berguna dan pengetahuan terkait dari berbagai database berukuran besar (Wahyudin, 2019). Data mining dibagi menjadi beberapa

kelompok yaitu deskripsi, estimasi, prediksi, klasifikasi, pengklusteran, dan asosiasi.

Data Mining dikatakan salah satu bagian dari *Knowledge Discovery in Databases* (Sari et al., 2023) karena digunakan untuk mencari pengetahuan dari suatu basis data (Tarigan et al., 2022).

Dari penjelasan diatas dapat disimpulkan bahwa Data mining merupakan Teknik pengumpulan, pemakaian data dengan menggunakan statistik, machine learning dalam mengidentifikasi suatu informasi yang bermanfaat.

2.4.1. Pengelompokkan *Data Mining*

Data mining dibagi menjadi 6 kelompok (Gunadi et al, 2016), Adapun penjabarannya yaitu :

a. Deskripsi

Deskripsi pola dan kecendrungan sering memberikan penjelasan tentang pola atau kebiasaan dalam data. Misalnya, analisis suara mungkin tidak dapat mengetahui informasi tentang siapa yang tidak cukup profesional dan mendapatkan dukungan sedikit dalam pemilihan presiden.

b. Estimasi

Estimasi merupakan teknik yang memiliki kemiripan dengan klasifikasi, namun dengan fokus utama pada nilai numerik daripada kategori. Dalam proses ini, data lengkap digunakan sebagai representasi, di mana nilai variabel target berfungsi sebagai acuan prediksi. Tahap berikutnya melibatkan analisis mendalam untuk memperkirakan nilai dari variabel-variabel prediktif.

Berbeda dengan klasifikasi yang mengelompokkan data ke dalam kategori-kategori tertentu, estimasi bertujuan untuk menghasilkan perkiraan angka yang akurat. Proses ini memanfaatkan informasi historis yang lengkap untuk membentuk dasar prediksi. Setelah model estimasi terbentuk, dilakukan evaluasi menyeluruh guna menghasilkan perkiraan yang tepat untuk variabel-variabel yang ingin diprediksi.

c. Prediksi

Prediksi juga mirip dengan klasifikasi dan estimasi, tetapi dalam prediksi, nilai hasil akan berlaku di masa mendatang. Contohnya, dalam bisnis, prediksi dapat mencakup harga beras dalam tiga bulan mendatang atau tingkat pengangguran 5 tahun yang akan datang.

d. Pengklusteran

Clustering adalah proses pengelompokan rekaman berdasarkan pengamatan dan pembentukan kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan rekaman yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan rekaman-rekaman dalam kluster lain. Pengklusteran berbeda dengan klasifikasi karena tidak melibatkan variabel target. Pengklusteran tidak bertujuan untuk melakukan klasifikasi, estimasi, atau prediksi nilai dari variabel target.

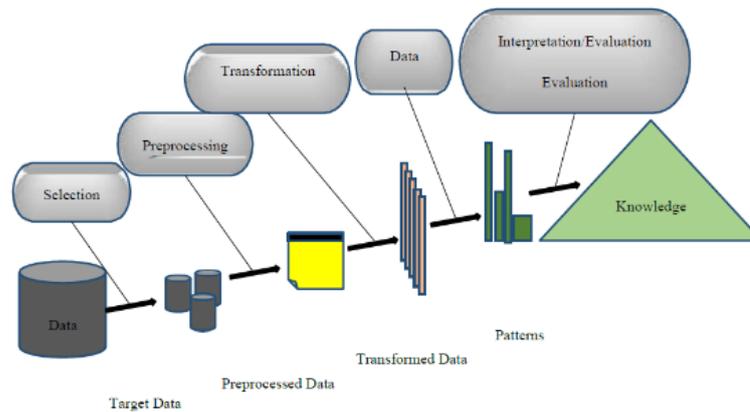
e. Asosiasi

Asosiasi bertujuan untuk menemukan atribut yang muncul bersamaan. Dalam konteks bisnis, asosiasi lebih dikenal sebagai analisis keranjang belanja. Contoh asosiasi dalam bisnis adalah menemukan barang yang

biasanya dibeli bersamaan di supermarket dan barang yang jarang atau tidak pernah dibeli bersamaan.

2.4.2. Tahapan *Data Mining*

Berikut ini 5 tahapan *data mining* yang dapat dilihat pada Gambar 2.1 dibawah ini:



Gambar 2.1 Tahapan Data Mining

a. Pemilihan Data (*Data Selection*)

Tahap awal dalam proses data mining melibatkan seleksi data dan atribut dari kumpulan data operasional yang tersedia. Data terpilih kemudian dipisahkan dari data asli untuk digunakan dalam analisis selanjutnya.

b. Pembersihan (*Pre-processing*)

Langkah penting sebelum analisis data adalah pembersihan data yang telah diseleksi. Proses ini bertujuan untuk menghilangkan duplikasi dan inkonsistensi, sehingga menghasilkan dataset yang lebih berkualitas dan dapat diandalkan untuk tahap berikutnya.

c. Transformation

Tahap transformasi adalah langkah di mana data yang sudah terpilih dimodifikasi ke dalam format yang lebih sesuai untuk pengolahan

selanjutnya. Proses ini bertujuan untuk mengubah struktur atau bentuk data agar lebih optimal dan siap digunakan dalam tahapan analisis berikutnya.

d. *Data Mining*

Data mining yaitu proses untuk menemukan wawasan atau pola yang bernilai dalam data yang telah diseleksi dengan menerapkan berbagai teknik atau metodologi khusus. Menurut (Maulana & Al-Khowarizmi, 2022) *Data Mining* pada dasarnya dilakukan untuk proses penemuan pengetahuan dengan mengekstraksi informasi yang tersembunyi dan informasi yang berguna dari kumpulan banyak data.

e. *Interpretation (evaluation)*

Pada fase ini, pola-pola yang dihasilkan dari data mining diidentifikasi. Pola informasi yang dihasilkan harus disajikan sedemikian rupa sehingga mudah dipahami oleh pihak yang berkepentingan.

2.5. *K-Nearest Neighbors (K-NN)*

K-Nearest Neighbors merupakan pendekatan untuk menemukan kasus dengan menghitung kedekatan antara kasus baru dan kasus lama berdasarkan pencocokan tertimbang dari rangkaian fitur yang ada. Misalnya, mungkin diinginkan untuk menggunakan solusi dari pasien sebelumnya untuk menemukan solusi bagi pasien baru. Untuk mengetahui kasus pasien mana yang digunakan, maka dihitung kedekatan kasus pasien baru dengan seluruh kasus pasien lama. Kasus pasien lama terdekat digunakan sebagai solusi untuk digunakan pada kasus pasien baru. Algoritma KNN merupakan algoritma yang efektifitasnya terhadap

data training dan ketahanan pada data noise (Imron & Kusumah, 2018; Salim et al., 2020).

Secara umum, K-NN merupakan metode yang fleksibel dan dapat diterapkan dalam berbagai konteks, mulai dari prediksi prestasi belajar siswa hingga analisis sentimen terhadap isu-isu sosial. Penerapannya juga dapat disesuaikan dengan kebutuhan spesifik dari masing-masing studi atau kasus.

Rumus umum K-NN yang digunakan untuk melakukan perhitungan yaitu rumus *Euclidean Distance* (Tangkelayuk, 2022) yang terlihat pada persamaan (2.2):

$$D(x, y) = \sqrt{\sum_{k=1}^n (X_{\text{training}} - Y_{\text{testing}})^2} \dots\dots\dots(2.2)$$

Keterangan:

X_{training} : data training ke-i

Y_{testing} : data testing

i: record (baris) ke-i dari tabel

n: jumlah data training

2.6. Python

Python adalah bahasa pemrograman yang umum digunakan untuk membangun pembelajaran mesin. Namun, sekarang banyak pengembang yang menggunakan *Python* untuk membuat *web*, sistem informasi seluler, atau *game*. *Python* merupakan bahasa pemrograman tingkat tinggi yang biasa digunakan oleh pengembang saat membuat aplikasi (Graciela Fausten Novindri & Ocsa Nugraha Saian, 2022). *Python* merupakan bahasa pemrograman yang mudah dipelajari karena memiliki *sintaks* yang mudah dibaca dan dipahami sehingga cocok untuk

programmer pemula, serta terdapat banyak perpustakaan dan *framework* yang memudahkan pengembangan.

Karena sifatnya yang serba guna dan mudah digunakan, bahasa pemrograman python menjadi bahasa pemrograman yang paling banyak digunakan. Terutama untuk mereka yang masih pemula dalam memakai bahasa pemrograman ini.



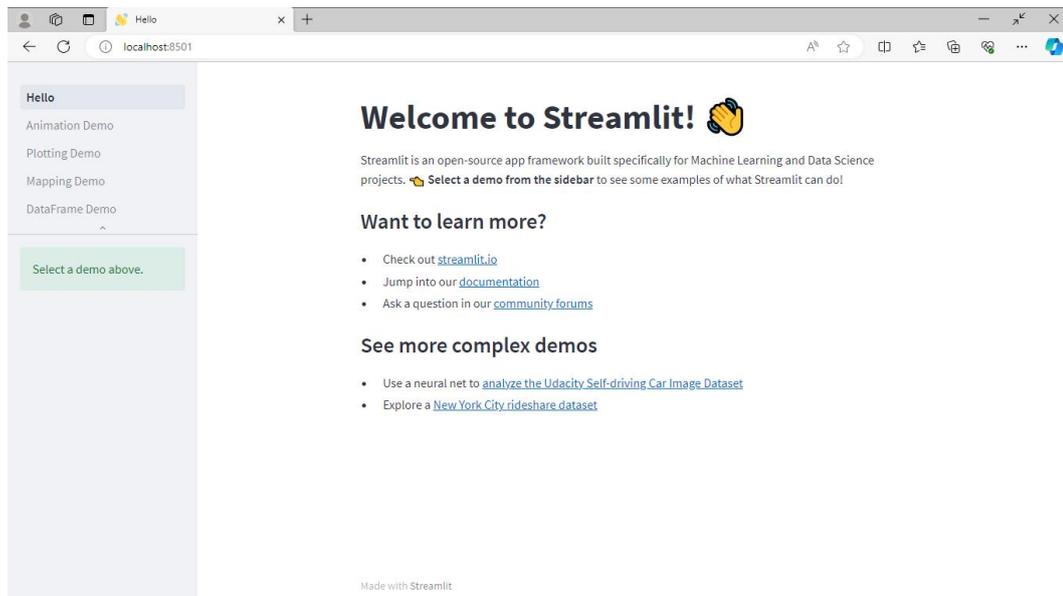
Gambar 2.2 Logo bahasa pemrograman python

2.7. Streamlit

Streamlit adalah kerangka kerja sumber terbuka yang berbasis *Python* yang dirancang untuk dengan mudah membuat aplikasi web interaktif di bidang ilmu data dan pembelajaran mesin.

Streamlit menyediakan berbagai pustaka dan fitur yang dapat membantu para pengembang dalam membangun aplikasi *web* dengan cepat dan mudah, seperti *tools* untuk memvisualisasikan data, fitur interaktif seperti *dropdown* dan *slider*, serta dukungan untuk integrasi dengan berbagai pustaka *Python* populer seperti Pandas, Numpy, dan Matplotlib. Selain itu, Streamlit juga menyediakan fitur-fitur yang mempermudah pengembangan aplikasi *web*, seperti *autoreload*, dimana aplikasi akan secara otomatis reload saat terjadi perubahan pada kode. Streamlit juga menyediakan dukungan untuk deployment ke berbagai platform cloud, seperti AWS dan Google Cloud.

Pada Gambar 2.3 berupa contoh tampilan sederhana yang menggunakan *framework* streamlit:



Gambar 2.3 Contoh tampilan sederhana streamlit

2.8. Visual Studio Code

Visual Studio Code adalah editor kode sumber yang dikembangkan oleh Microsoft untuk Windows, Linux dan macOS. Ini termasuk dukungan untuk debugging, kontrol git dan GitHub yang tertanam, penyorotan sintaksis, penyelesaian kode cerdas, cuplikan, dan pemfaktoran ulang kode. Sangat dapat disesuaikan, pengguna dapat mengubah tema, pintasan keyboard, pengaturan, dan memasang ekstensi untuk fungsionalitas tambahan.

Visual Studio Code (VS Code) ini adalah editor teks ringan dan handal yang dikembangkan oleh Microsoft untuk sistem operasi multiplatform, sehingga tersedia juga untuk versi Linux, Mac, dan Windows.

Teks editor ini mendukung langsung bahasa pemrograman JavaScript, TypeScript, dan Node.js, serta bahasa pemrograman lainnya dengan bantuan penggunaan plugin yang dapat dipasang via marketplace Visual Studio Code (seperti C++, C#, Python, Go, Java, dst). Visual Studio Code menawarkan banyak fitur, termasuk diantaranya IntelliSense, Git Integration, Debugging, dan fitur ekstensi yang menambah kemampuan teks editor.



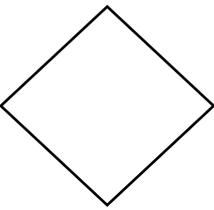
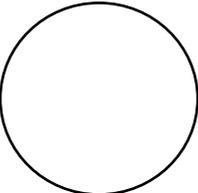
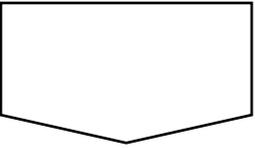
Gambar 2.4 Visual Studio Code

2.9. Flowchart

Flowchart adalah bagan arus yang menggambarkan langkah penyelesaian suatu masalah yang memperhatikan urutan dan hubungan antar proses beserta instruksinya. Simbol flowchart dapat dilihat pada Tabel 2.1 sebagai berikut:

Tabel 2.1 Simbol Flowchart beserta artinya

No.	Simbol Flowchart	Nama	Arti Simbol Flowcart
1.		Terminator	Awal atau akhir konsep (prosedur)
2.		Process	Process operasional
3.		Document	Dokumen atau laporan

4.		Decision	Keputusan atau sub-point. Garis yang terhubung dengan bentuk decision merujuk pada situasi-situasi yang berbeda sesuai dengan keputusan yang digambarkan.
5.		Data	Input dan Output (Contohnya, Input: feedback dari pelanggan, Output: desain produk baru)
6.		On-Page Reference/Connector	Penghubung alur dalam halaman yang sama
7.		Off-Page Reference/Off-Page Connector	Penghubung alur dalam halaman yang berbeda
8.		Flow	Arah alur dalam konsep (procedur)

2.10. Penelitian Terkait

Penelitian yang dilakukan oleh Yovi Apridiansyah mengenai metode *Monte Carlo* menjelaskan hasil prediksi calon mahasiswa baru pada tahun 2020

didapatkan tingkat akurasi sebesar 92.49%, sehingga dapat diambil kesimpulan metode *Monte Carlo* mampu membantu dalam pengambilan keputusan untuk memprediksi calon mahasiswa baru di masa datang (Yovi, et. al. 2022).

Penelitian terkait selanjutnya yang dilakukan oleh penulis lain, seperti penelitian mengenai ketersediaan produk herbal. Simulasi *Monte Carlo* digunakan untuk memprediksi penjualan produk herbal dan meningkatkan volume penjualan produk. Hasil penelitian ini memperoleh akurasi sebesar 97% (Elvin Syahrin, et.al. 2019).

Penelitian selanjutnya untuk metode *K-Nearest Neighbor* yang dilakukan oleh Umaidah dengan judul Penerapan Algoritma *K-Nearest Neighbor* (K-NN) Dengan Pencarian Optimal Untuk Prediksi Prestasi Siswa, dengan nilai *accuracy*=93.63%.

2.11. Penelitian Terdahulu

Penelitian terdahulu mengenai *Monte Carlo* dan *K-Nearest Neighbor* (K-NN) terdapat pada Tabel 2.2 sebagai berikut:

Tabel 2.2 Penelitian terdahulu mengenai *Monte Carlo* dan K-NN

No	Nama Peneliti	Judul Penelitian	Metode	Tahun
1.	Lalu Abd Rahman Hakim, Ahmad Ashril Rizal, Dwi Ratnasari	Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN)	K-Nearest Neighbor	2019

2.	Panoto dan Agus	Penerapan Algoritma <i>K-Nearest Neighbors</i> Untuk Prediksi Kelulusan Mahasiswa Pada Stmik Sinar Nusantara Surakarta	K-Nearest Neighbor	2015
3.	Yulia Irfayanti dan M. Ibnu Satria	Implementasi Metode <i>K-Nearest Neighbor</i> Untuk Prediksi Penjualan Beton Instan Pada PT. Decon Multi Industri	K-Nearest Neighbor	2020
4	Cita Lestari dan Hani	Klasifikasi Citra Daun Tumbuhan Obat Menggunakan Deteksi Tepi Canny Dan Metode <i>K-Nearest Neighbor (KNN)</i>	K-Nearest Neighbor	2023
5	Hanna Willa Dhany	Performa Algoritma <i>K-Nearest Neighbour</i> dalam Memprediksi Penyakit Jantung	K-Nearest Neighbor	2021
6	Okta Veza dan Andi Lewis Pratama	Dashboard Simulasi Perhitungan Persediaan Oli Menggunakan Metode <i>Monte Carlo</i> pada PT. Laras Era Perdana	<i>Monte Carlo</i>	2021
7	Muhammad Irsan Mulana dan Edy Victor Haryanto	Penerapan Metode <i>Monte Carlo</i> Untuk Peramalan Pembelian Aksesoris Laptop Pada Cv Gaharu Berbasis Android	<i>Monte Carlo</i>	2022

8	Okta Veza, Nofri Yudi Arifin ,Dan Sari Setyaning Tyas	Analisis dan Perancangan Arsitektur Pemodelan Simulasi Dalam Menentukan Calon Mahasiswa Non Aktif	<i>Monte Carlo</i>	2021
9	Muhammad Habib dan Yuhandri	Penerapan Metode <i>Monte Carlo</i> Dalam Memprediksi Produksi Es Balok Terhadap Optimalisasi Kebutuhan	<i>Monte Carlo</i>	2022
10	Yoksan dan Mart	Simulasi Prediksi Jumlah Peserta Didik Baru di SMKS YPPI TUALANG Menggunakan Metode <i>Monte Carlo</i>	<i>Monte Carlo</i>	2022

BAB III

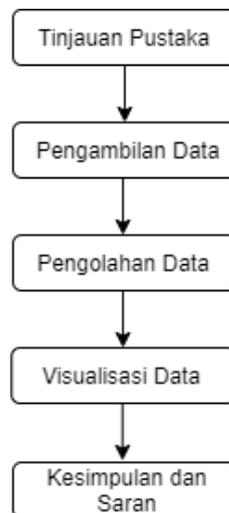
METODOLOGI PENELITIAN

3.1. Jenis Penelitian

Jenis penelitian ini adalah dengan menggunakan penelitian kuantitatif dengan pendekatan deskriptif karena menggunakan angka, mulai dari pengumpulan data, penafsiran terhadap data tersebut, serta hasil yang sudah di analisis akan dijabarkan secara deskriptif.

3.2. Tahapan Penelitian

Tahapan untuk metodologi penelitian dapat dilihat pada Gambar 3.1 merupakan bagan alur penelitian.



Gambar 3.1 Tahapan Penelitian

1. Tinjauan Pustaka

Tinjauan pustaka dilakukan mempelajari dan mengkaji teori yang dijadikan referensi yang berhubungan dengan penelitian dan sebagai pendukung tercapainya solusi dari masalah yang ada. Tinjauan pustaka yang dilakukan

adalah dengan mempelajari kembali referensi yang berupa buku, jurnal, serta informasi yang berkaitan dengan topik permasalahan dalam penelitian ini.

2. Pengambilan Data

Pengambilan data yang dilakukan yaitu data sekunder yang mana data diperoleh dari tata usaha SMK Negeri 1 Stabat. Data yang digunakan berupa data tahun ajaran, nilai rapor, nilai praktek, dan nilai presentasi. Data-data tersebut digunakan untuk menentukan metode mana yang akurat pada kedua metode tersebut.

3. Pengolahan Data

Pengolahan data yang dilakukan yaitu perhitungan data yang telah dikumpulkan sebelumnya berupa data nilai uji kompetensi kejuruan menggunakan metode *Monte Carlo* dan metode K-NN.

4. Visualisasi Data

Memvisualisasikan data dengan menggunakan *framework* streamlit, yang dimana menampilkan hasil dari kedua metode tersebut.

5. Kesimpulan dan Saran

Tahap akhir dari penelitian adalah memberikan kesimpulan dan saran dari penelitian yang telah dilakukan untuk studi kasus perbandingan sistem prediksi metode *Monte Carlo* dengan *K-Nearest Neighbors* pada nilai uji kompetensi kejuruan.

3.3. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di SMK Negeri 1 Stabat. Penelitian ini difokuskan pada data nilai Uji Kompetensi Kejuruan jurusan Teknik Pemesinan. Waktu penelitian dilakukan pada bulan Januari 2024 hingga Juni 2024.

Berikut ini merupakan tabel pelaksanaan penelitian dapat dilihat pada Tabel

3.1 berikut:

Tabel 3.1 Tabel Pelaksanaan Penelitian

No.	Kegiatan	2024					
		Jan	Feb	Mar	Apr	Mei	Jun
1.	Tahap Persiapan Penelitian						
	a. Penyusunan dan Pengajuan Judul						
	b. Pengajuan Proposal						
	c. Perizinan Penelitian						
2.	Tahap Pelaksanaan						
	a. Pengumpulan Data						
	b. Pengolahan Data						
	c. Visualisasi Data						
3.	Tahap Penyusunan Proposal						
4.	Pengujian Penelitian						

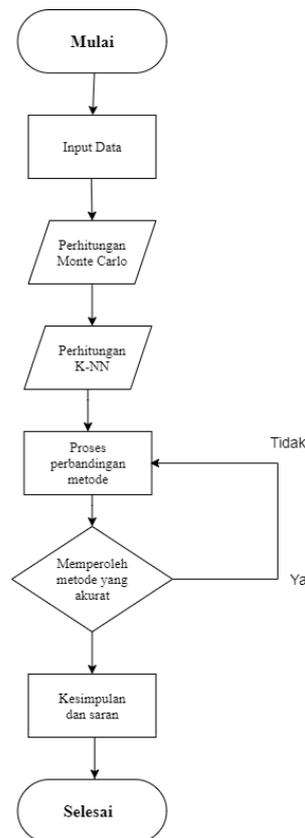
3.4. Perangkat Penelitian

Kebutuhan bahan dan alat yang digunakan pada penelitian ini antara lain:

1. Hardware: laptop dengan spesifikasi intel core i5, Ram 8 GB, dan SSD 120 GB.
2. Tools: Windows 10, Microsoft Excel, Visual Studio Code, dan Google Colab.

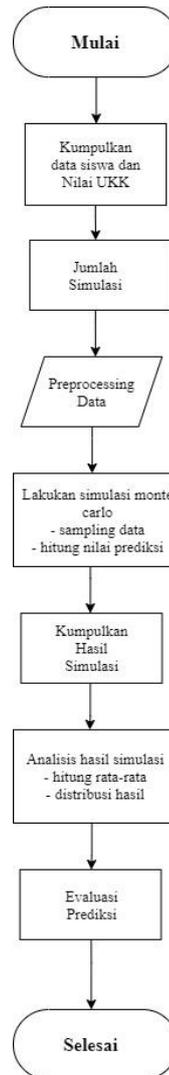
3.5. Flowchart Penelitian

Berikut ini merupakan flowchart pengerjaan penelitian sebagai berikut:



Gambar 3.2 Flowchart Penelitian

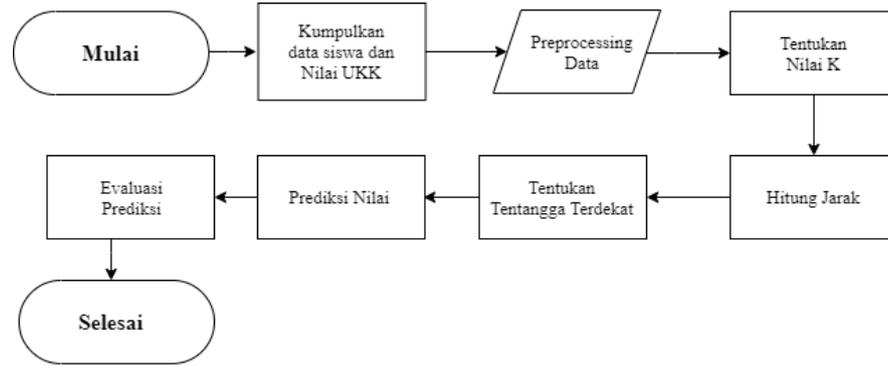
Berdasarkan pada Gambar 3.2 merupakan flowchart tahapan penelitian yang dimulai dengan tinjauan pustaka kemudian pengambilan sample data di tempat penelitian, selanjutnya melakukan perhitungan data menggunakan metode *Monte Carlo* setelah mendapatkan hasil dari perhitungan tersebut. Langkah selanjutnya, menghitung data dengan menggunakan *K-Nearest Neighbors*. Setelah melakukan perhitungan dari kedua metode tersebut akan dibandingkan untuk mendapatkan hasil prediksi mana yang lebih efektif dari kedua metode tersebut. Lalu, akan di visualisasikan dengan bahasa pemrograman python dengan *framework* streamlit. Kemudian pada tahap terakhir diperoleh kesimpulan dan saran dari penelitian tersebut.



Gambar 3.3 Flowchart Metode *Monte Carlo*

Pada Gambar 3.3 merupakan flowchart mengerjakan metode *Monte Carlo* yang diawali dengan mengumpulkan data siswa. Kemudian, mempreprocessing data serta melakukan metode *Monte Carlo*. Dan terakhir, melakukan analisis hasil simulasi dan evaluasi prediksi.

Selanjutnya, berikut ini merupakan flowchart metode *K-Nearest Neighbors*:



Gambar 3.4 Flowchart Metode K-NN

Pada Gambar 3.4 Langkah pertama dalam alur ini adalah mengumpulkan data siswa dan nilai UKK mereka. Data yang terkumpul kemudian masuk ke tahap preprocessing data, di mana data tersebut mungkin akan dibersihkan atau diolah agar siap untuk digunakan dalam model prediksi. Setelah preprocessing, langkah selanjutnya adalah menentukan nilai K, yaitu jumlah tetangga terdekat yang akan digunakan dalam algoritma K-NN untuk membuat prediksi. Selanjutnya, dilakukan perhitungan jarak antara data yang ingin diprediksi dengan data lain dalam dataset untuk menemukan tetangga terdekat. Setelah jarak dihitung, proses dilanjutkan dengan menentukan tetangga terdekat, yaitu K data yang paling mirip dengan data yang sedang dianalisis. Berdasarkan tetangga terdekat ini, sistem kemudian melakukan prediksi nilai untuk data baru tersebut. Langkah terakhir dalam alur ini adalah evaluasi prediksi, di mana hasil prediksi dibandingkan dengan data aktual untuk menilai akurasi dan efektivitas model K-NN. Setelah evaluasi selesai, proses ini mencapai tahap "Selesai," yang menandakan bahwa seluruh proses telah diselesaikan.

BAB IV
HASIL DAN PEMBAHASAN

4.1. Analisis Data

Data yang digunakan pada penelitian ini yaitu data seluruh siswa SMK Negeri 1 Stabat Tahun 2020 sampai 2024. Data tersebut akan dianalisis yang kemudian dilakukan prediksi nilai Uji Kompetensi Kejuruan menggunakan metode *Monte Carlo* dan *K-Nearest Neighbors* (K-NN) menggunakan Google colab.

Adapun data siswa SMK Negeri 1 Stabat yang ditampilkan pada tabel 4.1 dibawah ini.

Tabel 4.1 Data Siswa Teknik Pemesinan SMK Negeri 1 Stabat

NO	NIS	NAMA SISWA	NILAI AKHIR	NILAI SIKAP	NILAI PENGE TAHUAN	NILAI KETERAMPILAN
1	0002017217	Abdul Hakim	85	Sangat Baik	Baik	Baik
2	0002017218	Aldi Hardiansyah	92	Sangat Baik	Baik	Baik
3	0002017219	Aldi Irwanda	81	Baik	Baik	Baik
4	0002017221	Andreana Hasibuan	79	Baik	Baik	Baik
5	0002017222	Andrian Pratama	88	Sangat Baik	Baik	Baik
...
159	0002021175	Tama Irawan	90	Sangat Baik	Sangat Baik	Baik
160	0002021176	Taufik Hidayat Nst	90	Sangat Baik	Baik	Baik
161	0002021177	Wanda Syahputra	90	Sangat Baik	Baik	Baik
162	0002021178	Wendy Jay Prastyo	88	Sangat Baik	Baik	Baik
163	0002021179	Yuda Firmansyah	87	Sangat Baik	Baik	Sangat Baik

4.2. Prediksi Menggunakan *Monte Carlo*

4.2.1. Import Library

Dalam membangun model *Monte Carlo*, ada beberapa pustaka (*library*) yang digunakan pada penelitian ini untuk dilakukan pemrosesan dataset, antara lain:

1. Numpy, digunakan untuk mengolah dan memanipulasi data yang disimpan dalam bentuk array.
2. Pandas, berfungsi untuk mengubah file dataset yang berbentuk CSV menjadi tabel virtual.
3. Matplotlib, digunakan untuk membuat plot histogram yang menggambarkan distribusi prediksi.

Gambar 4.1. menampilkan proses import library untuk memprediksi nilai dengan menggunakan model *Monte Carlo*.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Gambar 4.1 Import library pada *Monte Carlo*

4.2.2. Membaca dataset

Dataset yang digunakan merupakan data berformat file CSV (*Comma Separated Values*). Dataset diimport ke Google Colab menggunakan library pandas dan disimpan dalam dataframe dengan nama “ukkk.csv” yang dapat dilihat pada gambar 4.2.

```
# Membaca dataset
file_path = 'ukkk.csv'
data = pd.read_csv(file_path)
```

Gambar 4.2 Membaca dataset csv

4.2.3. Menentukan dan memisahkan kolom target dan fitur

Selanjutnya menentukan dan memisahkan kolom target dengan fitur. Kolom target yang dipilih yaitu kolom 'nilai_akhir', kemudian memisahkan kolom dengan fitur seperti pada Gambar 4.3.

```
# Misalkan kita ingin memprediksi nilai kolom 'target' berdasarkan kolom-kolom lain
target_column = 'nilai_akhir' # Gantilah sesuai dengan nama kolom target di dataset Anda
features = data.columns.difference([target_column])

# Memisahkan fitur dan target
X = data[features]
y = data[target_column]
```

Gambar 4.3 Memisahkan antara fitur dan target

4.2.4. Implementasi Simulasi *Monte Carlo*

Setelah melakukan pemisahan antara fitur dan target dilanjutkan dengan simulasi *Monte Carlo* seperti pada Gambar 4.4.

```
# Fungsi Monte Carlo untuk prediksi
def monte_carlo_simulation(X, y, num_simulations=10):
    predictions = []
    for i in range(num_simulations):
        # Random sampling with replacement
        sample_indices = np.random.choice(len(X), size=len(X), replace=True)
        sample_X = X.iloc[sample_indices]
        sample_y = y.iloc[sample_indices]

        # Menghitung rata-rata dari sample_y sebagai prediksi (sederhana)
        prediction = sample_y.mean()
        predictions.append(prediction)
        print(f'Simulasi {i+1}: Prediksi = {prediction}')

    return predictions

# Melakukan simulasi Monte Carlo
predictions = monte_carlo_simulation(X, y)
```

Gambar 4.4 Simulasi *Monte Carlo*

Fungsi `monte_carlo_simulation` melakukan prediksi menggunakan metode *Monte Carlo*. Fungsi ini menerima dua parameter: `X`, yaitu data fitur, dan `y`, yaitu data target, serta jumlah simulasi yang ingin dilakukan (`num_simulations`, defaultnya 1000). Dalam setiap simulasi, dilakukan pengambilan sampel acak dengan penggantian dari data fitur dan target menggunakan `numpy`. Sampel-sampel ini digunakan untuk menghitung rata-rata dari data target (`sample_y.mean()`) sebagai prediksi. Hasil prediksi dari setiap simulasi disimpan dalam daftar `predictions`, dan

setiap prediksi juga dicetak di layar. Setelah semua simulasi selesai, fungsi ini mengembalikan daftar berisi semua hasil prediksi. Fungsi ini kemudian dipanggil dengan `predictions = monte_carlo_simulation(X, y)` untuk melakukan simulasi dan menyimpan hasilnya dalam variabel `predictions`.

4.2.5. Menampilkan Hasil Distribusi

Pada Gambar 4.5 untuk memvisualisasikan distribusi hasil prediksi dari model *Monte Carlo* dalam bentuk histogram. Dengan menampilkan frekuensi prediksi dalam interval tertentu, pengguna dapat dengan mudah memahami pola dan kecenderungan dari hasil prediksi yang dihasilkan oleh model.

```
# Menampilkan distribusi hasil prediksi
plt.hist(predictions, bins=30, edgecolor='k', alpha=0.7)
plt.xlabel('Prediksi')
plt.ylabel('Frekuensi')
plt.title('Distribusi Prediksi Monte Carlo')
plt.show()
```

Gambar 4.5 Menampilkan hasil distribusi

4.3. Prediksi Menggunakan *K-Nearest Neighbors* (K-NN)

4.3.1. Import Library

Untuk membangun model *K-Nearest Neighbors*, ada beberapa library yang digunakan pada penelitian ini untuk dilakukan pemrosesan dataset, antara lain:

1. Numpy, mengolah dan memanipulasi data yang telah disimpan ke dalam bentuk array.
2. Pandas, berfungsi untuk mengubah file dataset yang berbentuk CSV menjadi tabel virtual.

Gambar 4.6 menampilkan proses import library untuk memprediksi nilai dengan menggunakan model K-NN.

```
[1] # Import Library
import pandas as pd
import numpy as np
```

Gambar 4.6 Import library pada *K-Nearest Neighbors*

4.3.2. Import Dataset

Dataset yang digunakan merupakan data berformat file CSV (*Comma Separated Values*). Dataset diimport ke Google Colab menggunakan library pandas dan disimpan dalam dataframe dengan nama “df_ukk” yang dapat dilihat pada gambar 4.7.

```
# Load dataset
df_ukk = pd.read_csv('test.csv')
```

Gambar 4.7 Import dataset pada proses *K-Nearest Neighbors*

4.3.3. Data Cleaning

Proses pembersihan data (data cleaning) dilakukan untuk mempersiapkan data agar dapat diproses lebih lanjut. Tahapan ini meliputi penghapusan data yang memiliki nilai kosong (missing value) dan data duplikat (double) dengan nilai yang sama. Tujuannya adalah untuk mengoptimalkan kinerja model K-NN, yang dapat dilihat pada gambar 4.8.

```
# Data Cleaning missing value
df_ukk.isnull().sum

# Checkin if the data double
df_ukk.duplicated().sum
```

Gambar 4.8 Memeriksa missing value pada *K-Nearest Neighbors*

4.3.4. Exploratory Data Analysis (EDA)

Setelah proses pembersihan data selesai dilakukan, tahap selanjutnya adalah melakukan Exploratory Data Analysis (EDA). Tahap ini melibatkan visualisasi data dari beberapa atribut yang terdapat pada dataset Siswa SMK. Salah satu visualisasi yang dilakukan adalah pada atribut nilai_ket, yang ditampilkan pada gambar 4.9.

```
#eda
sns.set_theme(style="darkgrid")
sns.countplot(x='nilai_ket', data=df_ukk, palette='rocket')
plt.ylabel("Jumlah")
plt.xlabel('0=Baik 1=Sangat Baik')
plt.title("Kompeten")
plt.show()
```

Gambar 4.9 Visualisasi data nilai_ket

4.3.5. Split Data

Pada proses split data, yang pertama dengan memisahkan nilai x dan y menggunakan `drop ()` pada atribut `nilai_ket` yang dapat dilihat pada gambar 4.10.

```
X = df_ukk.drop(['nilai_ket'], axis=1)
y = df_ukk['nilai_ket']

# Split data menjadi training dan testing set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=0)
```

Gambar 4.10 Split data pada proses *K-Nearest Neighbors*

Proses pembagian data yang kedua melibatkan pemisahan dataset menjadi dua bagian, yaitu data training dan data testing, dengan proporsi masing-masing 80% untuk data training dan 20% untuk data testing. Pembagian dataset ini menggunakan fungsi `train_test_split ()` dari modul `sklearn.model_selection`, seperti yang ditunjukkan pada gambar 4.10.

4.3.6. Modeling

Sebelum melakukan modeling yakni menentukan nilai k. Nilai k yang digunakan adalah k=2 dan k=5 yang berarti jumlah tetangga yang akan digunakan untuk memprediksi label suatu data baru sebanyak 2 tetangga terdekat dan 5 tetangga terdekat.

Setelah menentukan nilai k, langkah berikutnya adalah membangun model K-NN dengan menggunakan metrik jarak "euclidean" seperti yang ditunjukkan pada gambar 4.11.

```
# Buat model KNN
knn = KNeighborsClassifier(n_neighbors=5, metric='euclidean', p=2)

# Train model
knn.fit(X_train, y_train)
```

Gambar 4.11 Model *K-Nearest Neighbors*

4.3.7. Prediksi

Setelah model K-NN selesai dibangun, kemudian prediksi pada data baru dapat dilakukan menggunakan metode `predict` dengan model K-NN yang dapat dilihat pada gambar 4.12.

```
# Prediksi pada testing set
y_pred = knn.predict(X_test)
```

Gambar 4.12 Prediksi *K-Nearest Neighbors*

4.3.8. Evaluasi Performa *K-Nearest Neighbors*

Setelah model K-NN dibangun menggunakan training set, langkah berikutnya adalah mengevaluasi performa K-NN dengan menggunakan testing set. Performa K-NN diperoleh dengan menghitung nilai confusion matrix dari model K-NN, seperti yang ditunjukkan pada gambar 4.13.

```
#Confusion Matrix
from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, recall_score, f1_score, classification_report
cm = confusion_matrix(y_test, y_pred)
print(cm)
```

Gambar 4.13 Confusion Matrix *K-Nearest Neighbors*

Confusion matrix menyediakan informasi komprehensif tentang bagaimana model klasifikasi memprediksi kelas data. Ini menunjukkan jumlah prediksi benar (true positive dan true negative) dan prediksi salah (false positive dan false negative) yang dibuat oleh model. Nilai dari confusion matrix digunakan untuk menghitung akurasi, presisi, recall, F1 Score, dan error rate berdasarkan rata-rata macro, seperti yang ditunjukkan pada gambar 4.14. Akurasi menunjukkan seberapa tepat model dalam memprediksi keseluruhan data, sedangkan presisi, recall, dan F1 Score mengevaluasi kualitas prediksi untuk masing-masing kelas. Error rate digunakan untuk mengukur tingkat kesalahan yang terjadi dalam proses K-NN.

```

# Buat model KNN
knn = KNeighborsClassifier(n_neighbors=5, metric='euclidean', p=2)

# Train model
knn.fit(X_train, y_train)

# Prediksi pada testing set
y_pred = knn.predict(X_test)

# Evaluasi model
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='macro')
recall = recall_score(y_test, y_pred, average='macro')
f1 = f1_score(y_test, y_pred, average='macro')
error_rate = 1 - accuracy

# Print evaluasi model
print("Akurasi:", accuracy)
print("Presisi:", precision)
print("Recall:", recall)
print("F1 Score:", f1)
print("Error Rate:", error_rate)

```

Gambar 4.14 Tampilan Confusion Matrix

4.4. Pengujian *Monte Carlo*

4.4.1. Analisis Hasil Prediksi

Hasil prediksi memperlihatkan hasil dari 10 simulasi *Monte Carlo* yang saya lakukan untuk memprediksi suatu nilai berdasarkan dataset yang diberikan. Setiap baris dalam tabel menunjukkan hasil prediksi yang dihasilkan dari setiap simulasi, yang dinyatakan dalam bentuk persentase.

```

Simulasi 1: Prediksi = 85.0925925925926
Simulasi 2: Prediksi = 85.36585365853658
Simulasi 3: Prediksi = 85.30864197530865
Simulasi 4: Prediksi = 85.38993710691824
Simulasi 5: Prediksi = 84.29878048780488
Simulasi 6: Prediksi = 85.39877300613497
Simulasi 7: Prediksi = 85.64596273291926
Simulasi 8: Prediksi = 85.7560975609756
Simulasi 9: Prediksi = 85.39024390243902
Simulasi 10: Prediksi = 85.09876543209876

```

Gambar 4.15 Tampilan 10 simulasi *Monte Carlo*

Pada Gambar 4.15 setiap simulasi melakukan pengambilan sampel acak dengan penggantian dari dataset asli. Proses ini dikenal sebagai bootstrapping, di mana setiap sampel yang diambil dapat dipilih kembali dalam pengambilan sampel berikutnya. Hal ini memungkinkan untuk menghasilkan berbagai kemungkinan hasil prediksi yang berbeda setiap kali simulasi dijalankan.

Dari tabel, kita dapat melihat bahwa hasil prediksi berkisar antara 85,09% hingga 85,75%. Prediksi ini bervariasi karena sifat acak dari proses pengambilan sampel. Misalnya, prediksi pertama adalah 85,09%, sementara prediksi tertinggi adalah 85,75% pada simulasi kedelapan. Didapatkan nilai rata-rata dari simulasi tersebut 85,27%, rata-rata dari prediksi ini memberikan gambaran tentang nilai yang diharapkan, sementara variasi di antara prediksi menunjukkan tingkat ketidakpastian dan variabilitas dalam data.

4.5. Pengujian *K-Nearest Neighbors* (K-NN)

Pada penelitian ini terdapat 163 jumlah data yang mana dataset ini dibagi menjadi 2 yaitu data training dan data testing. Pembagian data training dan data testing harus berdasarkan variasi pembagian data (test size) karena test size akan mempengaruhi performa model *K-Nearest Neighbors*. Pada model *K-Nearest Neighbors* ada 4 kali pengujian berdasarkan test size yang digunakan yaitu 20% dan 25% yang dapat dilihat pada tabel 4.2. Hasil dari pengujian ini nantinya akan dievaluasi menggunakan *confusion matrix*, lalu akan mendapatkan performa model *K-Nearest Neighbors*.

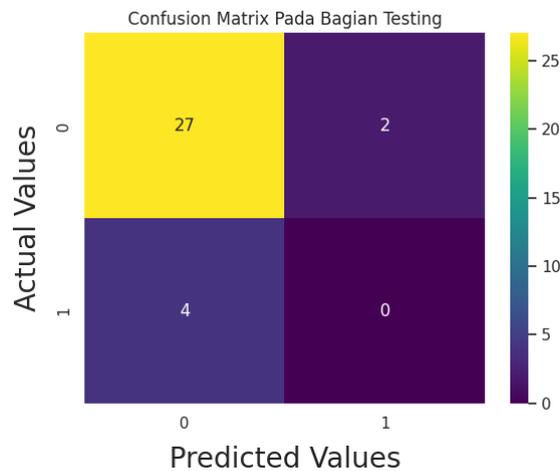
Tabel 4.2. Skenario Pengujian *K-Nearest Neighbors*

NO.	Pengujian Variasi k = 2 dan k = 5	Test Size (Data Testing)	Data Training	Data Testing
1		20%	130	33
2		25%	122	41

4.5.1. Hasil Pengujian *K-Nearest Neighbors* (K-NN)

1) Pengujian 1 dengan Variasi k=2

Berikut ini adalah hasil *confusion matrix* pada pengujian pertama dengan variasi k=2 dari text size 20% dimana data training berjumlah 130 data dan data testing berjumlah 33 data yang dapat dilihat pada gambar 4.16.



Gambar 4.16 Heatmap pengujian 1 dengan k=2

Untuk keterangan selanjutnya dari hasil *confusion matrix* pada pengujian pertama dengan variasi k=2 yang telah digambarkan diatas dalam bentuk heatmap dapat dilihat pada tabel 4.3. dibawah ini:

Tabel 4.3. Confusion Matrix K-NN Pengujian 1 Dengan Variasi k=2

K-NN Dengan Variasi k=2 Test Size 20%		
Actual Values \ Predicted Values	0 (Baik) Negative	1 (Sangat Baik) Positive
	0 (Baik) Negative	27 (TN)
1 (Sangat Baik) Positive	4 (FN)	0 (TP)

Dapat dilihat dari tabel 4.3 terdapat 165 data testing yang digunakan dalam *confusion matrix* pengujian 1 dengan variasi k=2 yang dijelaskan sebagai berikut:

- Ada 27 data siswa aktual yang tergolong dalam kelas “Baik” dan model berhasil memprediksi sebagai “Baik” (TN = True Negative).
- Tidak ada data siswa aktual yang tergolong dalam kelas “Sangat Baik” dan model berhasil memprediksi sebagai “Sangat Baik” (TP = True Positive).
- Ada 4 data siswa aktual yang tergolong ke dalam kelas “Sangat Baik”, tetapi model memprediksi sebagai “Baik” (FN = False Negative).
- Ada 2 data siswa aktual yang tergolong ke dalam kelas “Baik”, tetapi model memprediksi sebagai “Sangat Baik” (FP = False Positive).

Berikut ini adalah nilai akurasi (*accuracy*), presisi (*precision*), *recall*, *F1* Score, dan *Error Rate* pada pengujian 1 dengan variasi k=2:

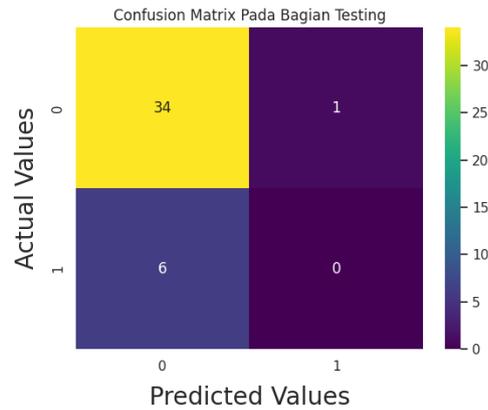
```
Akurasi: 81.82%
Presisi: 43.55%
Recall: 46.55%
F1 Score: 45.00%
Error Rate: 18.18%
```

Gambar 4.17 Hasil performa pengujian 1 dengan variasi k=2

Dari hasil performa pada gambar 4.17 tampak bahwa hasil performa Model K-NN pengujian 1 dengan variasi k=2 yaitu didapatkan nilai akurasi 81,82%, presisi 43,55%, *recall* 46,55%, *F1 Score* 45,00%, dan *Error Rate* 18,18%.

2) Pengujian 2 dengan Variasi k=2

Berikut ini adalah hasil *confusion matrix* pada pengujian pertama dengan variasi k=2 dari text size 25% dimana data training berjumlah 122 data dan data testing berjumlah 41 data yang dapat dilihat pada gambar 4.18.



Gambar 4.18 Heatmap pengujian 2 dengan variasi k=2

Untuk keterangan lebih lanjut mengenai hasil *confusion matrix* dari pengujian kedua dengan variasi k=2 yang telah digambarkan diatas dalam bentuk heatmap dapat dilihat pada Tabel 4.4. dibawah ini:

Tabel 4.4. Confusion Matrix K-NN Pengujian 1 Dengan Variasi k=7

K-NN Dengan Variasi k=2 Test Size 25%		
Predicted Values \ Actual Values	0 (Baik) Negative	1 (Sangat Baik) Positive
0 (Baik) Negative	34 (TN)	1 (FP)
1 (Sangat Baik) Positive	6 (FN)	0 (TP)

Dapat dilihat dari Tabel 4.4. terdapat 33 data testing yang digunakan dalam *confusion matrix* pengujian 1 dengan variasi k=2 yang dijelaskan sebagai berikut:

- Ada 34 data siswa aktual yang tergolong dalam kelas “Baik” dan model berhasil memprediksi sebagai “Tidak” (TN = True Negative).
- Tidak ada data siswa aktual yang tergolong dalam kelas “Sangat Baik” dan model berhasil memprediksi sebagai “Sangat Baik” (TP = True Positive).
- Ada 6 data siswa aktual yang tergolong ke dalam kelas “Sangat Baik”, tetapi model memprediksi sebagai “Baik” (FN = False Negative).

- Tidak ada data siswa aktual yang tergolong ke dalam kelas “Baik”, tetapi model memprediksi sebagai “Sangat Baik” (FP = False Positive).

Hasil dari *confusion matrix* digunakan untuk mencari performa model *K-Nearest Neighbors* yaitu akurasi (*accuracy*), presisi (*precision*), *recall*, *F1 Score*, dan *Error Rate*. Berikut ini adalah nilai akurasi (*accuracy*), presisi (*precision*), *recall*, *F1 Score*, dan *Error Rate* pada pengujian 2 dengan variasi k=2:

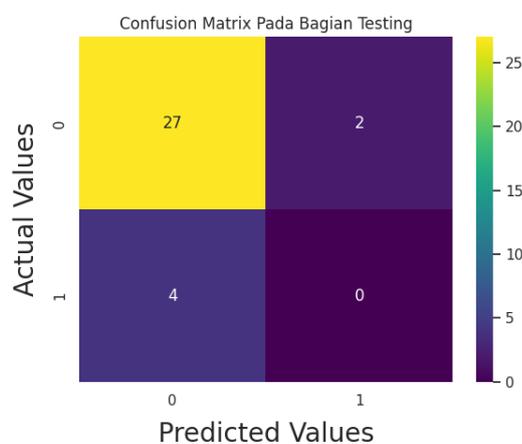
```
Akurasi: 82.93%
Presisi: 42.50%
Recall: 48.57%
F1 Score: 45.33%
Error Rate: 17.07%
```

Gambar 4.19 Hasil performa pengujian 2 dengan variasi k=2

Pada gambar 4.19 tampak bahwa hasil performa Model K-NN pengujian 2 dengan variasi k=2 yaitu didapatkan nilai akurasi 82,93%, presisi 42,50%, *recall* 48,57%, *F1 Score* 45,33%, dan *Error Rate* 17,07%.

3) Pengujian 1 dengan Variasi k=5

Berikut ini adalah hasil *confusion matrix* pada pengujian kedua dengan variasi k=5 dari text size 20% dimana data training berjumlah 122 data dan data testing berjumlah 41 data yang dapat dilihat pada gambar 4.20.



Gambar 4.20 Heatmap pengujian 1 dengan variasi k=5

Untuk informasi lebih lanjut mengenai hasil confusion matrix dari pengujian pertama dengan variasi $k=5$ yang telah digambarkan sebelumnya dalam bentuk heatmap, dapat dilihat pada tabel 4.5 di bawah ini.

Tabel 4.5. Confusion Matrix K-NN Pengujian 2 dengan Variasi $k=5$

K-NN dengan Variasi $k=5$ Test Size 20%		
Predicted Values \ Actual Values	0 (Baik) Negative	1 (Sangat Baik) Positive
0 (Baik) Negative	27 (TN)	2 (FP)
1 (Sangat Baik) Positive	4 (FN)	0 (TP)

Dapat dilihat dari tabel 4.5. terdapat 41 data testing yang digunakan dalam *confusion matrix* pengujian 1 dengan variasi $k=5$ yang dijelaskan sebagai berikut:

- Ada 27 data siswa aktual yang tergolong dalam kelas “Baik” dan model berhasil memprediksi sebagai “Baik” (TN = True Negative).
- Tidak ada data siswa aktual yang tergolong dalam kelas “Sangat Baik” dan model berhasil memprediksi sebagai “Sangat Baik” (TP = True Positive).
- Ada 4 data siswa aktual yang tergolong ke dalam kelas “Sangat Baik”, tetapi model memprediksi sebagai “Baik” (FN = False Negative).
- Ada 2 data siswa aktual yang tergolong ke dalam kelas “Baik”, tetapi model memprediksi sebagai “Sangat Baik” (FP = False Positive).

Berikut ini adalah nilai akurasi (*accuracy*), presisi (*precision*), *recall*, *F1* Score, dan *Error Rate* pada pengujian 1 dengan variasi $k=5$:

```

Akurasi: 81.82%
Presisi: 43.55%
Recall: 46.55%
F1 Score: 45.00%
Error Rate: 18.18%

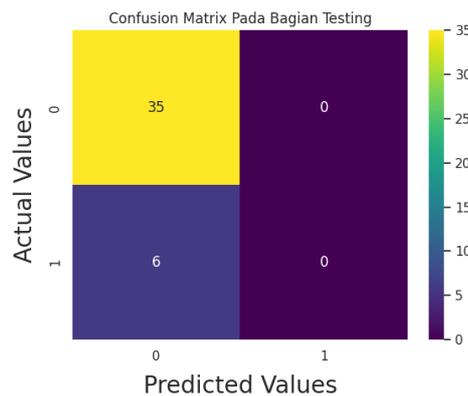
```

Gambar 4.21 Hasil pengujian 1 dengan variasi $k=5$

Dari gambar 4.21. tampak bahwa hasil performa Model K-NN pengujian 2 dengan variasi k=5 yaitu didapatkan nilai akurasi 81,82%, presisi 43,55%, *recall* 46,55%, *F1 Score* 45,00%, dan *Error Rate* 18,18%.

4) Pengujian 2 dengan Variasi k=5

Berikut ini adalah hasil *confusion matrix* pada pengujian kedua dengan variasi k=5 dari text size 25% dimana data training berjumlah 122 data dan data testing berjumlah 41 data yang dapat dilihat pada gambar 4.22.



Gambar 4.22 Heatmap pengujian 2 dengan variasi k=5

Untuk keterangan selanjutnya dari hasil *confusion matrix* pada pengujian kedua dengan variasi k=5 yang telah digambarkan diatas dalam bentuk heatmap dapat dilihat pada tabel 4.6. dibawah ini:

Tabel 4.6. Confusion Matrix K-NN Pengujian 2 dengan Variasi k=5

K-NN Dengan Variasi k=5 Test Size 25%		
Actual Values \ Predicted Values	0 (Baik) Negative	1 (Sangat Baik) Positive
	0 (Baik) Negative	35 (TN)
1 (Sangat Baik) Positive	6 (FN)	0 (TP)

Dapat dilihat dari tabel 4.6. terdapat 41 data testing yang digunakan dalam *confusion matrix* pengujian 2 dengan variasi $k=5$ yang dijelaskan sebagai berikut:

- Ada 35 data siswa aktual yang tergolong dalam kelas “Baik” dan model berhasil memprediksi sebagai “Baik” (TN = True Negative).
- Tidak ada data siswa aktual yang tergolong dalam kelas “Sangat Baik” dan model berhasil memprediksi sebagai “Sangat Baik” (TP = True Positive).
- Ada 6 data siswa aktual yang tergolong ke dalam kelas “Sangat Baik”, tetapi model memprediksi sebagai “Baik” (FN = False Negative).
- Tidak ada data siswa aktual yang tergolong ke dalam kelas “Baik”, tetapi model memprediksi sebagai “Sangat Baik” (FP = False Positive).

Hasil dari *confusion matrix* digunakan untuk mencari performa model *K-Nearest Neighbors* yaitu akurasi (*accuracy*), presisi (*precision*), *recall*, *F1 Score*, dan *Error Rate*. Berikut ini adalah nilai akurasi (*accuracy*), presisi (*precision*), *recall*, *F1 Score*, dan *Error Rate* pada pengujian 2 dengan variasi $k=5$:

```

r Akurasi: 85.37%
  Presisi: 42.68%
  Recall: 50.00%
  F1 Score: 46.05%
  Error Rate: 14.63%

```

Gambar 4. 23 Hasil pengujian 2 dengan variasi $k=5$

Dari gambar 4.23. tampak bahwa hasil performa Model *K-NN* pengujian 2 dengan variasi $k=5$ yaitu didapatkan nilai akurasi 85,37%, presisi 42%, *recall* 50,00%, *F1 Score* 46,05%, dan *Error Rate* 14,63%.

Analisis Performa dari Pengujian *K-Nearest Neighbors* (K-NN) Dengan Variasi k=2 dan k=5

Berdasarkan hasil pengujian yang telah dilakukan, nilai k terbaik untuk model K-NN dilihat berdasarkan rata – rata nilai akurasi, presisi, recall, f1 score, dan error rate. Dari tabel yang ditunjukkan, tampak bahwa rata – rata nilai akurasi K-NN variasi k=5 lebih besar yakni 82,59% dibandingkan dengan nilai akurasi K-NN variasi k=2 yakni 82,37%. Begitu juga dengan rata – rata nilai presisi K-NN variasi k=5 lebih besar yakni 43,11% dibandingkan dengan nilai presisi K-NN variasi k=2 yakni 43,02%. Untuk rata – rata nilai recall K-NN variasi k=5 juga mendapatkan nilai lebih besar yakni 48,27% dibandingkan dengan nilai recall K-NN variasi k=2 yakni 47,56%. Selanjutnya juga nilai rata – rata nilai F1 Score K-NN variasi k=5 lebih besar yakni 45,52% dibandingkan dengan nilai F1 Score K-NN variasi k=2 yakni 45,16%. Dan untuk rata – rata nilai error rate K-NN variasi k=5 lebih kecil yakni 16,40% dibandingkan dengan nilai error rate K-NN variasi k=2 yakni 17,62%. Semakin kecil nilai error rate maka semakin bagus karena persentase terjadinya kesalahan kecil pada saat model K-NN digunakan untuk mengklasifikasi data.

Tabel 4.7. Hasil Pengujian *K-Nearest Neighbors* Variasi k=2 dan k=5

Test Size	K-NN dengan Variasi k=2					K-NN dengan Variasi k=5				
	Akurasi	Presisi	Recall	F1 Score	Error Rate	Akurasi	Presisi	Recall	F1 Score	Error Rate
20%	81,82%	43,55%	46,55%	45,00%	18,18%	81,82%	43,55%	46,55%	45,00%	18,18%
25%	82,93%	42,50%	48,57%	45,33%	17,07%	85,37%	42,68%	50,00%	46,05%	14,63%
Rata-Rata	82,37%	43,02%	47,56%	45,16%	17,62%	82,59%	43,11%	48,27%	45,52%	16,40%

Dari tabel hasil pengujian diatas, maka nilai k terbaik yang dipilih adalah nilai $k=5$ karena memiliki akurasi paling tinggi (85,37%) dan error rate paling rendah (14,63%).

4.6. Perbandingan *K-Nearest Neighbors* (K-NN) dan *Monte Carlo*

Dari hasil penelitian, dapat dilihat bahwa metode K-NN memberikan akurasi prediksi yang lebih tinggi dibandingkan dengan metode *Monte Carlo*. Metode K-NN berhasil mencapai akurasi sebesar 85,37%, sedangkan metode *Monte Carlo* hanya mencapai rata-rata akurasi sebesar 85,26%. Metode *Monte Carlo* menggunakan sampel acak setiap kali program dijalankan, yang berarti bahwa hasil prediksi dapat sedikit bervariasi pada setiap pengujian. Hal ini dapat menyebabkan variasi kecil dalam akurasi prediksi yang dihasilkan oleh *Monte Carlo*. Di sisi lain, metode K-NN biasanya memberikan hasil yang lebih konsisten karena tidak bergantung pada pengambilan sampel acak, hasilnya akan tetap sama jika parameter dan data yang digunakan tidak berubah. Namun, jika kita mempertimbangkan ketidakpastian yang terkait dengan sifat acak dari *Monte Carlo*, maka performa metode K-NN yang sedikit lebih unggul bisa menjadi lebih signifikan dalam konteks ini. Konsistensi hasil yang dihasilkan oleh K-NN memberikan keunggulan dalam situasi di mana kestabilan dan prediktabilitas hasil sangat penting. Oleh karena itu, meskipun *Monte Carlo* memiliki kelebihan dalam fleksibilitas dan kemampuan untuk menangani ketidakpastian, K-NN tetap unggul dalam hal konsistensi performa, sehingga dalam kasus ini, metode K-NN masih dapat dianggap lebih andal.

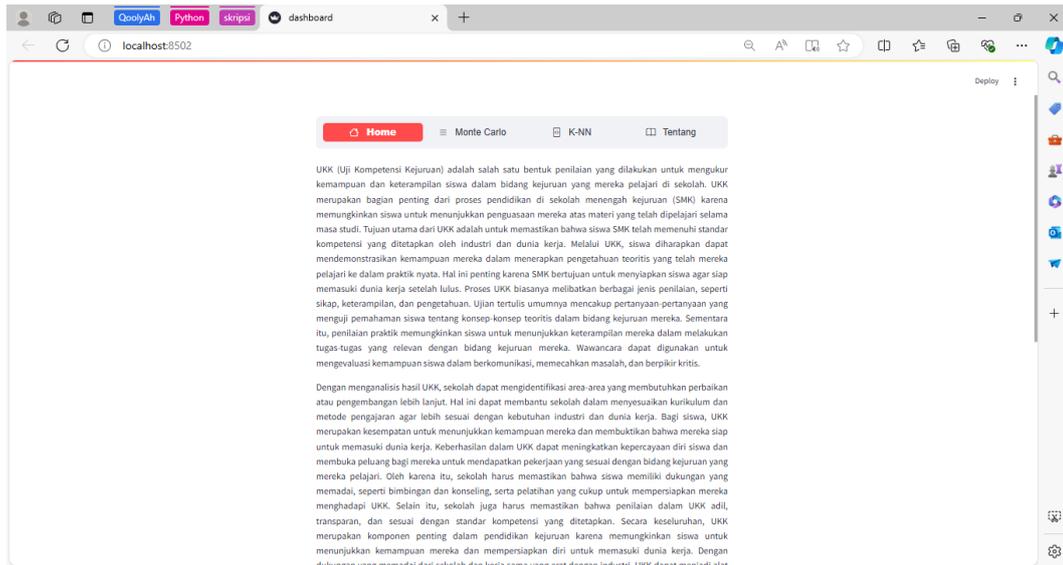
Tabel 4.8 Perbandingan antara Metode *Monte Carlo* dan K-NN

Aspek	Monte Carlo	K-NN
Akurasi Prediksi	85,26%	85,37%
Konsistensi Hasil	Tidak konsisten, hasil bervariasi karena sampel acak	Konsisten, hasil tetap sama jika parameter dan data tidak berubah
Pengaruh Sifat Acak	Menggunakan sampel acak, menyebabkan variasi kecil dalam akurasi prediksi	Tidak bergantung pada pengambilan sampel acak
Waktu Pemrosesan	Bisa lebih cepat untuk beberapa skenario karena tidak memerlukan perhitungan jarak antar data	Cenderung lebih lambat pada dataset besar karena membutuhkan perhitungan jarak antar titik data dalam ruang multidimensi
Fleksibilitas dalam Penanganan Data	Mampu menangani ketidakpastian dan variasi dalam data	Lebih sensitif terhadap perubahan dalam data; membutuhkan data yang bersih dan terstruktur untuk hasil yang optimal
Ketergantungan pada Parameter	Kurang bergantung pada parameter spesifik, lebih fokus pada simulasi	Sangat bergantung pada parameter K, jarak, dan metrik yang digunakan
Skalabilitas	Skalabilitas baik pada dataset besar, terutama dengan optimasi penggunaan sampel acak	Skalabilitas bisa menjadi isu pada dataset besar karena kompleksitas perhitungan jarak
Implementasi	Lebih mudah diterapkan dalam situasi dengan data yang memiliki banyak ketidakpastian	Memerlukan pemahaman yang baik tentang pemilihan parameter dan normalisasi data
Fleksibilitas Model	Dapat digunakan pada berbagai jenis masalah prediksi tanpa banyak modifikasi model	Cocok untuk masalah klasifikasi dan regresi, tetapi perlu disesuaikan dengan baik dengan struktur data
Penanganan Outlier	Dapat lebih toleran terhadap outlier karena hasilnya tergantung pada rata-rata dari banyak sampel	Sensitif terhadap outlier, terutama jika K kecil
Ketepatan dan Presisi	Bisa lebih bervariasi karena sifat acak dari sampel yang digunakan	Cenderung lebih tepat dan presisi jika data terstruktur dengan baik dan parameter diatur dengan benar

Penggunaan Sumber Daya Komputasi	Cenderung membutuhkan sumber daya komputasi yang lebih rendah pada dataset besar	Bisa sangat membutuhkan sumber daya komputasi, terutama memori, saat menangani dataset yang sangat besar
Kesesuaian dengan Berbagai Domain	Sangat cocok untuk masalah dengan ketidakpastian tinggi, seperti simulasi fisika atau keuangan	Sangat cocok untuk masalah dengan struktur data yang jelas, seperti pengenalan pola atau analisis data terstruktur
Kompleksitas Pemahaman	Bisa lebih rumit untuk dipahami karena melibatkan konsep probabilitas dan statistik	Lebih mudah dipahami dan diimplementasikan dalam konteks klasifikasi sederhana
Keandalan dalam Konteks Tertentu	Kurang andal dalam situasi yang memerlukan prediksi yang sangat konsisten	Sangat andal dalam situasi di mana konsistensi hasil sangat penting
Kesimpulan	Meskipun fleksibel dan cocok untuk situasi dengan ketidakpastian tinggi, kurang andal dalam konsistensi performa	Lebih andal dan konsisten, unggul dalam kestabilan hasil dan situasi yang memerlukan prediksi yang stabil

Visualisasi Sistem Prediksi menggunakan metode *Monte Carlo* dan *K-Nearest Neighbors* dengan Streamlit

Antarmuka ini dirancang untuk mendukung penelitian yang dilakukan mengenai perbandingan sistem prediksi menggunakan metode *Monte Carlo* dan K-NN pada nilai peserta didik dalam UKK. Desain visualnya yang sederhana menunjukkan bahwa aplikasi ini lebih difokuskan untuk menyampaikan informasi secara rinci mengenai kedua metode tersebut dapat dilihat pada gambar 4.24.



Gambar 4.24 Tampilan streamlit

4.7. Keunggulan dan Kelemahan Metode *Monte Carlo* dan *K-Nearest*

Neighbors

Monte Carlo:

- Keunggulan: Metode ini sangat fleksibel dan dapat digunakan dalam berbagai jenis distribusi data. Selain itu, metode ini memungkinkan untuk melakukan simulasi yang berulang sehingga memberikan gambaran yang lebih komprehensif mengenai hasil prediksi.
- Kelemahan: Proses simulasi yang dilakukan secara berulang membutuhkan waktu komputasi yang lebih lama dan hasil prediksi cenderung kurang konsisten dibandingkan dengan metode K-NN.

K-Nearest Neighbors:

- Keunggulan: Metode ini memberikan hasil prediksi yang lebih akurat dan konsisten. Proses prediksi dengan metode K-NN juga lebih cepat dibandingkan dengan metode *Monte Carlo*.

- Kelemahan: Metode ini sangat bergantung pada pemilihan nilai k yang optimal dan metrik jarak yang digunakan. Selain itu, metode ini kurang efektif apabila terdapat data yang sangat besar dan kompleks.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka dapat disimpulkan dari tujuan penelitian ini sebagai berikut:

1. Algoritma *Monte Carlo* dan *K-Nearest Neighbors* (K-NN) berhasil diterapkan dalam nilai uji kompetensi kejuruan pada data siswa SMK Negeri 1 Stabat.
2. Aplikasi berbasis web menggunakan Streamlit berhasil dikembangkan untuk memuat dan menampilkan hasil prediksi dari simulasi *Monte Carlo* dan metode K-NN. Aplikasi ini memudahkan pengguna dalam melihat, menganalisis, dan membandingkan hasil prediksi dari kedua metode secara interaktif.
3. Hasil analisis perbandingan menunjukkan bahwa metode *Monte Carlo* dan K-NN memiliki keefektifan yang berbeda dalam memprediksi nilai peserta didik uji kompetensi kejuruan. Metode K-NN menunjukkan performa yang lebih konsisten dengan rata-rata akurasi prediksi yang lebih tinggi dibandingkan dengan metode *Monte Carlo*.
4. Identifikasi kelebihan dan kekurangan masing-masing metode menunjukkan bahwa metode *Monte Carlo* memiliki keunggulan dalam fleksibilitas dan kemampuan menangani variabilitas data yang tinggi, sementara metode K-NN unggul dalam kecepatan dan akurasi prediksi pada dataset yang lebih terstruktur dan homogen.
5. Hasil akhir yang didapatkan dari perbandingan algoritma *Monte Carlo* dan *K-Nearest Neighbors* (K-NN) adalah algoritma *K-Nearest Neighbors* (K-NN) di

tetapkan sebagai algoritma terbaik dengan nilai akurasi 85,37% dalam memprediksi nilai UKK pada data siswa SMK Negeri 1 Stabat.

5.2. Saran

Adapun beberapa saran untuk peneliti selanjutnya, antara lain:

1. Mengeksplorasi penggunaan metode lain untuk meningkatkan akurasi prediksi.
2. Mengkaji dampak penggunaan data yang lebih besar dan kompleks terhadap akurasi dan konsistensi hasil prediksi.
3. Mengembangkan sistem prediksi yang terintegrasi dengan sistem informasi sekolah untuk mempermudah proses evaluasi dan pemantauan hasil belajar peserta didik.

DAFTAR PUSTAKA

- Abdar, M., Kalhori, S. R. N., Sutikno, T., Subroto, I. M. I., & Arji, G. (2015). Comparing performance of *Data Mining* algorithms in prediction heart diseases. *International Journal of Electrical and Computer Engineering*, 5(6), 1569–1576. <https://doi.org/10.11591/ijece.v5i6.pp1569-1576>
- Agustini, S. (2022). Pemodelan dan Simulasi Monte Carlo dalam Identifikasi Kebutuhan Bahan Bakar Minyak (BBM). *Jurnal Informatika Ekonomi Bisnis*, 4(3), 90–95. <https://doi.org/10.37034/infeb.v4i3.149>
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, 10(2), 160–165. <https://doi.org/10.33096/ilkom.v10i2.303.160-165>
- Bertot, L., Genaud, S., & Gossa, J. (2018). An overview of cloud simulation enhancement using the monte-carlo method. *Proceedings - 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGRID 2018*, 386–387. <https://doi.org/10.1109/CCGRID.2018.00064>
- Chala Beyene, M. (2020). *Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques*. November. <http://www.ijpam.eu>
- Ghozali, I. (2018). Aplikasi statistika dengan program SPSS. Badan Penerbit Universitas Diponegoro.
- Graciela Fausten Novindri, & Ocsa Nugraha Saian, P. (2022). Implementasi Flask Pada Sistem Penentuan Minimal Order Untuk Tiap Item Barang Di Distribution Center Pada Pt Xyz Berbasis Website. *Jurnal Mnemonic*, 5(2), 81–85. <https://doi.org/10.36040/mnemonic.v5i2.4670>
- Gunadi, G., & Sensuse, D. I. (2016). Penerapan metode data mining market basket analysis terhadap data penjualan produk buku dengan menggunakan algoritma apriori dan frequent pattern growth (fp-growth): studi kasus percetakan pt. Gramedia. *Telematika MKOM*, 4(1), 118-132.
- Imron, M., & Kusumah, S. A. (2018). Application of Data Mining Classification Method for Student Graduation Prediction Using K-Nearest Neighbor (K-NN)

- Algorithm. *IJIS: International Journal of Informatics and Information Systems*, 1(1), 1–8.
- Kotu, V., & Deshpande, B. (2014). *Predictive analytics and data mining: concepts and practice with rapidminer*: Morgan Kaufmann.
- Kurniawan, Y. I., & Barokah, T. I. (2020). Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan *K-Nearest Neighbor*. *Jurnal Ilmiah Matrik*, 22(1), 73–82. <https://doi.org/10.33557/jurnalmatrik.v22i1.843>
- Kotu, V., & Deshpande, B. (2014). *Predictive analytics and data mining: concepts and practice with rapidminer*: Morgan Kaufmann.
- Li, A., Ren, Y., Yang, D., & Li, Z. (2018). A Monte Carlo simulation-based algorithm for a repairable system in GO methodology. *Proceedings - 2018 5th International Conference on Dependable Systems and Their Applications, DSA 2018*, 119–125. <https://doi.org/10.1109/DSA.2018.00029>
- Lubis, A. R., Lubis, M., & Al-Khowarizmi. (2020). Optimization of Distance Formula in *K-Nearest Neighbor Method*. *Bulletin of Electrical Engineering and Informatics*, 9(1), 326–338. <https://doi.org/10.11591/eei.v9i1.1464>
- Lubis, A. R., Lubis, M., Al-Khowarizmi, & Listriani, D. (2019). Big Data Forecasting Applied Nearest Neighbor Method. *International Conference on Sustainable Engineering and Creative Computing (ICSECC)*, 116–120. <https://doi.org/10.1109/ICSECC.2019.8907010>
- Macuácuá, J. C., Centeno, J. A. S., & Amisse, C. (2023). *Data Mining approach for dry bean seeds classification*. *Smart Agricultural Technology*, 5(April). <https://doi.org/10.1016/j.atech.2023.100240>
- Maulana, H., & Al-Khowarizmi, A.-K. (2022). Analysis of the Effectiveness of Online Learning Using Eda Data Science and Machine Learning. *Sinkron*, 7(1), 222–231. <https://doi.org/10.33395/sinkron.v7i1.11206>
- Mawarti, L., Sugiman, & Kharis, M. (2018). Perbandingan Uji Hasil Simulasi Monte Carlo dan Simulasi Bootstrap dalam Analisis Saham untuk Menghitung Nilai VaR Data. *UNNES Journal of Mathematics*, 7(2), 253–261.
- Muhathir, Sibarani, T. T. S., & Al-Khowarizmi. (2020). Analysis *K-Nearest Neighbors (K-NN)* in Identifying Tuberculosis Disease (Tb) By Utilizing Hog Feature Extraction. *International of Computer Science and Information*

Technology (AIOCSIT) Journal ISSN, 1(1), 33–38.

- Nabila, A., & Putra, E. P. (2022). *Perbedaan Supervised dan Unsupervised Pada Data Mining*. Sis.Binus. <https://sis.binus.ac.id/2022/02/10/yuk-cari-tahu-perbedaan-supervised-dan-unsupervised-pada-data-mining/>
- Nikmatun, I. A., & Waspada, I. (2019). Implementasi *Data Mining* Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma *K-Nearest Neighbor*. *Jurnal SIMETRIS*, 10(2), 421–432.
- Rahmani, M. A., Ratnawati, D. E., & Hanggara, B. T. (2021). *K-Nearest Neighbor untuk Memprediksi Pergantian Komputer di Bank X*. 5(7), 3198–3207. <http://j-ptiik.ub.ac.id>
- Ramadhani, F., Satria, A., & Sari, I. P. (2023). Implementasi Metode Fuzzy *K-Nearest Neighbor* dalam Klasifikasi Penyakit Demam Berdarah. *Hello World Jurnal Ilmu Komputer*, 2(2), 58–62. <https://doi.org/10.56211/helloworld.v2i2.253>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1). <https://doi.org/10.31294/ijcit.v5i1.7951>
- Sari, I. P., & Batubara, I. H. (2021, August). Perancangan Sistem Informasi Laporan Keuangan Pada Apotek Menggunakan Algoritma K-NN. In *Seminar Nasional Teknologi Edukasi Sosial dan Humaniora* (Vol. 1, No. 1, pp. 692-698)
- Salim, A. P., Laksitowening, K. A., & Asror, I. (2020). Time Series Prediction on College Graduation Using KNN Algorithm. 2020 8th International Conference on Information and Communication Technology (ICoICT), 1–4.
- Setiawati, L. (2015). Faktor-Faktor yang Mempengaruhi Prestasi Belajar Praktik Kejuruan Siswa SMK Program Studi Keahlian Teknik Komputer dan Informatika. *Jurnal pendidikan vokasi*, 5(1), 325–339. Available at: <https://journal.uny.ac.id/index.php/jpv/article/view/6487>
- Setiawan, H., & Ode, S. (2019). Implementasi Program Indonesia Pintar di SD Negeri Sukacai, Kabupaten Serang. *Journal of Government*, 4(2), 173–197.
- Sheth, V., Tripathi, U., & Sharma, A. (2022). A Comparative Analysis of Machine Learning Algorithms for Classification Purpose. *Procedia Computer Science*,

- 215, 422–431. <https://doi.org/10.1016/j.procs.2022.12.044>
- Sidik, Z. (2019). *Klasifikasi Kelancaran Kredit Furniture Menggunakan Algoritma K-Nearest Neighbor Berbasis Forward Selection*.
- Suyanto, 2017, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika.
- Syahrin, E., Santony, J., & Na'am, J. (2019). Pemodelan Penjualan Produk Herbal Menggunakan Metode Monte Carlo. *Jurnal KomtekInfo*, 5(3), 33–41. <https://doi.org/10.35134/komtekinfo.v5i3.29>
- Tangkelayuk, A. (2022). Klasifikasi Kualitas Air Menggunakan Metode K-NN, Naïve Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 1109–1119. <https://doi.org/10.35957/jatisi.v9i2.2048>
- Tarigan, P. M. S., Hardinata, J. T., Qurniawan, H., Safii, M., & Winanjaya, R. (2022). Implementasi *Data Mining* Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus:Toko Sinar Harahap). *Jurnal Janitra Informatika Dan Sistem Informasi*, 2(1), 9–19. <https://doi.org/10.25008/janitra.v2i1.142>
- Umaidah, Y., & Purwantaro. (2019). Penerapan Algoritma K-Nearest Neighbor (KNN) dengan Pencarian Optimal untuk Prediksi Siswa Berprestasi. *JISICOM (Journal of Information System, Informatics and Computing)*, 3(2), 1–8.
- Yovi, Ringgo Dwika, & Eka. (2022). Penerapan Metode Monte Carlo pada Simulasi Prediksi Jumlah Calon Mahasiswa Baru Universitas Muhammadiyah Bengkulu. *Jurnal PROCESSOR*, 17(2), 74–81. <https://doi.org/10.33998/processor.2022.17.2.1224>
- Vallat, R. (2018). Pingouin: statistics in Python. *Journal of Open Source Software*, 3(31), 1026. <https://doi.org/10.21105/joss.01026>

LAMPIRAN

Lampiran 1. SK – 1 Persetujuan Topik/Judul Penelitian



UMSU
Luhur | Cerdas | Terpercaya
Tidak menentang Allah dan agamanya, berakhlak mulia, berprestasi, bermartabat, dan berkeadilan.

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/11/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
Website: <http://www.umsu.ac.id> Email: fas@umsu.ac.id Instagram: [umsumedan](https://www.instagram.com/umsumedan) Facebook: [umsumedan](https://www.facebook.com/umsumedan) Twitter: [umsumedan](https://twitter.com/umsumedan) YouTube: [umsumedan](https://www.youtube.com/umsumedan)

PERSETUJUAN TOPIK/JUDUL PENELITIAN

Nomor Agenda :
Nama : AZZAHRAH
NPM : 2009020072
Tanggal Persetujuan : 30 Januari 2024
Topik Yang Disetujui Program Studi : Data Mining
Nama Dosen Pembimbing : Indah Purnama Sari, S.T., M.Kom
Judul Yang Disetujui Dosen Pembimbing : Perbandingan Sistem Prediksi Menggunakan Metode *Monte Carlo* Dengan Metode K-NN Pada Nilai Peserta Didik Uji Kompetensi Kejuruan (Studi kasus pada SMK Negeri 1 Stabat)

Medan, 05 Februari 2024

Disahkan oleh
Ketua Program Studi
Teknologi Informasi

(Fatma Sari Hutagalung, S.Kom., M.Kom)

Persetujuan
Dosen Pembimbing

(Indah Purnama Sari, S.T., M.Kom)



Lampiran 2. SK-2 Penetapan Dosen Pembimbing



UMSU

Unggul | Cerdas | Terpercaya

Universitas Muhammadiyah Sumatera Utara

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

<https://mkt.umhu.ac.id> fmkt@umhu.ac.id [umsumedan](https://www.facebook.com/umsumedan) [umsumedan](https://www.instagram.com/umsumedan) [umsumedan](https://www.youtube.com/umsumedan)

PENETAPAN DOSEN PEMBIMBING

PROPOSAL/SKRIPSI MAHASISWA

NOMOR : 154/IL.3-AU/UMSU-09/F/2024

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

Program Studi : Teknologi Informasi
Pada tanggal : 30 Januari 2024

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

Nama : AZZAHRAH
NPM : 2009020072
Semester : VII (Tujuh)
Program studi : Teknologi Informasi
Judul Proposal / Skripsi : Perbandingan Sistem Prediksi Menggunakan Metode Monte Carlo Dengan Metode K-NN Pada Nilai Peserta Didik Uji Kompetensi Kejuruan

Dosen Pembimbing : Indah Purnama Sari, S.T., M.Kom

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. Proyek Proposal / Skripsi dinyatakan " BATAL " bila tidak selesai sebelum Masa Kadaluaarsa tanggal : 30 Januari 2025
4. Revisi judul.....

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Ditetapkan di : Medan
Pada Tanggal : 18 Rajab 1445 H
30 Januari 2024, M

Dekan


Dr. Al-Khowarizmi, S.Kom., M.Kom
NIDN : 0127099201



Cc. File



Lampiran 3. SK-3 Formulir Berita Acara Pembimbing



UMSU

Unggul | Cerdas | Terpercaya
Diselesaikan sesuai isi agar diwujudkan nomor dan langgananya

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/II/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

<https://fikt.umsu.ac.id> fikt@umsu.ac.id [umsu](#) [umsu](#) [umsu](#) [umsu](#)

Berita Acara Pembimbingan Proposal

Nama Mahasiswa : AZZAHRAH
NPM : 2009020072
Program Studi : Teknologi Informasi
Konsentrasi : Data Mining
Nama Dosen Pembimbing : Indah Purnama Sari, S.T., M.Kom
Judul Penelitian : Perbandingan Sistem Prediksi Menggunakan Metode *Monte Carlo* Dengan Metode K-NN Pada Nilai Peserta Didik Uji Kompetensi Kejuruan (Studi kasus pada SMK Negeri 1 Stabat)

Tanggal Bimbingan	Hasil Evaluasi	Paraf Dosen
05 Februari 2024	Revisi Bab I & Bab II (Latar Belakang, Rumusan Masalah)	
21 Februari 2024	Acc Bab I & Bab II Revisi Bab III	
23 Februari 2024	Acc Sempro	
15 Juli 2024	Revisi Penulisan & Revisi Program	



UMSU

Unggul | Cerdas | Terpercaya

Bisa membuat surat ini agar dibutuhkan nomor dan langgananya

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/11/2019
 Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

<https://fiki.umsu.ac.id>

fiki@umsu.ac.id

[umsumedan](#)

[umsumedan](#)

[umsumedan](#)

[umsumedan](#)

31 / 07 - 2024	ACC Program & Revisi Daftar Pustaka	
08 / 08 - 2024	ACC Sidang	

Medan, 08 Agustus 2024

Disahkan oleh
Ketua Program Studi
Teknologi Informasi

(Fatma Sari Hutagalung, S.Kom., M.Kom.)

Persetujuan
Dosen Pembimbing

(Indah Purnama Sari, S.T., M.Kom.)

Lampiran 4. SK-4 Surat Permohonan Sempro



UMSU
Unggul | Cerdas | Terpercaya

Pila memuat surat mu agar di seluruh nomor dan tangganya

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

<https://fikt.umsu.ac.id>

fikti@umsu.ac.id

[umsu.medan](#)

[umsu.medan](#)

[umsu.medan](#)

[umsu.medan](#)

**PERMOHONAN
SEMINAR PROPOSAL SKRIPSI**

Kepada Yth.
Bapak Dekan FIKTI UMSU
Di
Medan

Medan, 26 Februari 2024

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dengan hormat, saya yang bertanda tangan di bawah ini mahasiswa Fakultas Ilmu Komputer dan Teknologi Informasi UMSU :

Nama Lengkap : AZZAHRAH
NPM : 2009020072
Program Studi : Teknologi Informasi

Mengajukan permohonan Mengikuti Seminar Proposal Skripsi yang ditetapkan dengan Surat Penetapan Judul Skripsi dan Pembimbing NomorII.3-AU/UMSU-09/F/2024 Tanggal 26 Februari 2024 dengan judul sebagai berikut :
Perbandingan Sistem Prediksi Menggunakan Metode *Monte Carlo* Dengan Metode K-NN Pada Nilai Peserta Didik Uji Kompetensi Kejuruan (Studi Kasus Pada SMK Negeri 1 Stabat)

Bersama permohonan ini saya lampirkan :

1. Surat Penetapan Judul Skripsi (SK-1),
2. Surat Penetapan Pembimbing (SK-2),
3. DEKAM yang telah disahkan,
4. Kartu Hasil Studi Semester 1 s/d terakhir ASLI,
5. Tanda Bukti Lunas Beban SPP tahap berjalan,
6. Tanda Bukti Lunas Biaya Seminar Proposal Skripsi,
7. Proposal Skripsi yang telah disahkan oleh Pembimbing (rangkap-3),
8. Semua berkas dimasukkan ke dalam MAP warna BIRU.

Demikian permohonan saya untuk pengurusan selanjutnya. Atas perhatian Bapak saya ucapkan terima kasih.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Menyetujui :
Pembimbing

(Indah Purnama Sari, S.T., M.Kom)

Pemohon

(Azzahrah)



Lampiran 5. SK-5 Surat Plagiasi

skripsi azzahrah.docx

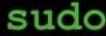
ORIGINALITY REPORT

25%	24%	11%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	etheses.uin-malang.ac.id Internet Source	2%
2	repository.umsu.ac.id Internet Source	2%
3	Submitted to Universitas Muhammadiyah Sumatera Utara Student Paper	2%
4	www.researchgate.net Internet Source	1%
5	journal.sinov.id Internet Source	1%
6	eprints.umpo.ac.id Internet Source	1%
7	repository.ub.ac.id Internet Source	1%
8	repository.teknokrat.ac.id Internet Source	1%
9	repository.upiypk.ac.id Internet Source	1%
10	Submitted to Badan PPSDM Kesehatan Kementerian Kesehatan	1%

Lampiran 7. Letter of Acceptance (LoA)



Jurnal Teknik Informatika

<https://jurnal.ilmubersama.com/index.php/sudo>

LETTER OF ACCEPTANCE (LoA)

Kepada Yth Bpk/Ibu/Sdr

Azzahrah, Indah Purnama Sari

Di

Tempat

Dengan ini kami sampaikan bahwa naskah dengan rincian berikut dinyatakan diterima untuk diterbitkan di sudo Jurnal Teknik Informatika, ISSN: 2829-7342 (Online) terbitan Volume 3 Nomor 4 Edisi Desember 2024.

Judul	Perbandingan Sistem Prediksi Menggunakan Metode Monte Carlo Dengan Metode K-NN Pada Nilai Peserta Didik Uji Kompetensi Kejuruan (Studi Kasus Pada SMK Negeri 1 Stabat)
Penulis Correspondent Email	Azzahrah, Indah Purnama Sari azzahrahzahrah11@gmail.com

Demikianlah surat keterangan ini kami buat untuk dapat digunakan seperlunya.







Medan, 08 Oktober 2024

Editor in Chief



Oris Krianto Sulaiman, S.T., M.Kom.

sudo Jurnal Teknik Informatika
Ilmu Bersama Center
Email: jurnalsudo@gmail.com



sudo Jurnal Teknik Informatika is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

Lampiran 8. Data Siswa Teknik Pemesinan SMK Negeri 1 Stabat

NO	NIS	NAMA SISWA	NILAI AKHIR	NILAI SIKAP	NILAI PENGETAHUAN	NILAI KETERAMPILAN
1	00020 17217	Abdul Hakim	85	Sangat Baik	Baik	Baik
2	00020 17218	Aldi Hardiansyah	92	Sangat Baik	Baik	Baik
3	00020 17219	Aldi Irwanda	81	Baik	Baik	Baik
4	00020 17221	Andreana Hasibuan	79	Baik	Baik	Baik
5	00020 17222	Andrian Pratama	88	Sangat Baik	Baik	Baik
6	00020 17224	Arman Haulian Ritonga	77	Baik	Baik	Baik
7	00020 17227	Dani Setiawan	80	Baik	Baik	Baik
8	00020 17228	David Pradana	93	Sangat Baik	Baik	Baik
9	00020 17229	Gita Afrizal	82	Baik	Baik	Baik
10	00020 17230	Gusti Pradana	83	Baik	Baik	Baik
11	00020 17231	Ibnu Khoir	78	Baik	Baik	Baik
12	00020 17232	Imam Hanafi	80	Baik	Baik	Baik
13	00020 17233	Irvan Syahputra	79	Baik	Baik	Baik
14	00020 17234	Irwansyah	90	Sangat Baik	Baik	Baik
15	00020 17235	Krisna Mukti	92	Sangat Baik	Baik	Baik
16	00020 17236	M. Ilham	82	Baik	Baik	Baik
17	00020 17237	Muhammmad Agtifa Sismu Rezky	84	Baik	Baik	Baik
18	00020 17238	Muhammmad Ali Arsyad	78	Baik	Baik	Baik
19	00020 17239	Muhammad Alwi Yuwanda	87	Sangat Baik	Baik	Baik
20	00020 17240	Muhammad Eko Pradana	90	Sangat Baik	Baik	Baik
21	00020 17241	Muhammad Fuad Fadhilah	88	Sangat Baik	Baik	Baik
22	00020 17242	Rio Prayuda	93	Sangat Baik	Baik	Baik
23	00020 17243	Rizky Ardiansyah	93	Sangat Baik	Baik	Baik
24	00020 17244	Sandipradika	89	Sangat Baik	Baik	Baik
25	00020 17245	Suhendri	80	Baik	Baik	Baik

NO	NIS	NAMA SISWA	NILAI AKHIR	NILAI SIKAP	NILAI PENGETAHUAN	NILAI KETERAMPILAN
26	00020 17246	Tri Budi Orasetyo	81	Baik	Baik	Baik
27	00020 17247	Wahyu Albani Rey	78	Baik	Baik	Baik
28	00020 17248	Wahyu Anggara	92	Sangat Baik	Baik	Baik
29	00020 17249	Wendi Alfani	88	Sangat Baik	Baik	Baik
30	00020 17250	Yohandi Maulana	94	Sangat Baik	Baik	Baik
31	00020 17251	Zakaria Alwie Rifai	91	Sangat Baik	Baik	Baik
32	00020 18397	Al Manda Nuzul Ramadhan	80	Baik	Baik	Baik
33	00020 18398	Algy Martan	82	Baik	Baik	Baik
34	00020 18399	Arif Kudadiri	78	Baik	Baik	Baik
35	00020 18400	Azwan Arianto	94	Sangat Baik	Baik	Baik
36	00020 18401	Bagus Satria	86	Sangat Baik	Baik	Baik
37	00020 18404	Dandi	86	Sangat Baik	Baik	Baik
38	00020 18405	Dimas Dayu Wenanda	84	Baik	Baik	Baik
39	00020 18406	Dimas Farizal Mukti	86	Sangat Baik	Baik	Baik
40	00020 18409	Harya Mitra	80	Baik	Baik	Baik
41	00020 18410	Hidayat	88	Sangat Baik	Baik	Baik
42	00020 18552	Irvan Efendi	88	Sangat Baik	Baik	Baik
43	00020 18412	Krisna Primadi Sitepu	88	Sangat Baik	Baik	Baik
44	00020 18413	M. Aditya Putra Rangga	86	Sangat Baik	Baik	Baik
45	00020 18414	M. Rizal	92	Sangat Baik	Baik	Baik
46	00020 18415	Mhd. Fahmi	90	Sangat Baik	Baik	Baik
47	00020 18416	Mhd. Faujul	92	Sangat Baik	Baik	Baik
48	00020 18418	Muhammad Ayub	92	Sangat Baik	Baik	Baik
49	00020 18419	Muhammad Nabil	80	Baik	Baik	Baik
50	00020 18422	Nanda Kesuma	80	Baik	Baik	Baik
51	00020 18423	Padil Akbar	88	Sangat Baik	Baik	Baik

NO	NIS	NAMA SISWA	NILAI AKHIR	NILAI SIKAP	NILAI PENGETAHUAN	NILAI KETERAMPILAN
52	00020 18424	Prabowo	94	Sangat Baik	Baik	Baik
53	00020 18425	Rangga Abde Firmansyah	78	Baik	Baik	Baik
54	00020 18426	Refal Amanda	90	Sangat Baik	Baik	Baik
55	00020 18427	Rendi Renaldi	90	Sangat Baik	Baik	Baik
56	00020 18428	Rico Rinanda	90	Sangat Baik	Baik	Baik
57	00020 18429	Rivaldo	94	Sangat Baik	Baik	Baik
58	00020 18430	Riyan Sutiyoso	90	Sangat Baik	Baik	Baik
59	00020 18431	Rizky Ramadan	94	Sangat Baik	Baik	Baik
60	00020 18432	Yuda Afandi	94	Sangat Baik	Baik	Baik
61	00020 19432	Abdila Azis	81	Baik	Baik	Baik
62	00020 19433	Abil Pratama	81	Baik	Baik	Baik
63	00020 19434	Achmad Al Chairi Lubis	87	Sangat Baik	Baik	Baik
64	00020 19435	Aditya Arif Yanto	85	Sangat Baik	Baik	Baik
65	00020 19436	Alvinda Sholikhin	81	Baik	Baik	Baik
66	00020 19437	Ardi Pratama	80	Baik	Baik	Baik
67	00020 19438	Arjuna Hutabarat	89	Sangat Baik	Baik	Baik
68	00020 19439	Danu Alfitrah	86	Sangat Baik	Baik	Baik
69	00020 19440	Dicky Pratama	75	Baik	Baik	Baik
70	00020 19441	Diki Candra	77	Baik	Baik	Baik
71	00020 19442	Dimas Aditia	81	Baik	Baik	Baik
72	00020 19443	Dimas Aldi Agustian	80	Baik	Baik	Baik
73	00020 19444	Dio Firmansyah	80	Baik	Baik	Baik
74	00020 19445	Dio Pratama	80	Baik	Baik	Baik
75	00020 19446	Fahri Aldi Pradika	91	Sangat Baik	Baik	Baik
76	00020 19447	Firman Dwi Kesuma	81	Baik	Baik	Baik
77	00020 19448	Irfan Fahreza	85	Sangat Baik	Baik	Baik

NO	NIS	NAMA SISWA	NILAI AKHIR	NILAI SIKAP	NILAI PENGETAHUAN	NILAI KETERAMPILAN
78	00020 19449	Irpan Afrija	77	Baik	Baik	Baik
79	00020 19450	Kornelius Triananda Sirait	85	Sangat Baik	Baik	Baik
80	00020 19451	M. Azriansyah	85	Sangat Baik	Baik	Baik
81	00020 19452	M. Ilham Juwanda	75	Baik	Baik	Baik
82	00020 19453	M.Rifal Tanjung	87	Sangat Baik	Baik	Baik
83	00020 19454	M.Taufik	80	Baik	Baik	Baik
84	00020 19455	Mhd. Dimas Saputra	80	Baik	Baik	Baik
85	00020 19456	Muhammad Akhyar Ramadan	87	Sangat Baik	Baik	Baik
86	00020 19457	Muhammad Ibnu Chaliq	80	Baik	Baik	Baik
87	00020 19458	Rifsky Ramadhan	87	Sangat Baik	Baik	Baik
88	00020 19459	Rizki Aji Kesuma	94	Sangat Baik	Baik	Baik
89	00020 19460	Roberto Permana Situmorang	84	Baik	Baik	Baik
90	00020 19461	Roy Arifin	75	Baik	Baik	Baik
91	00020 19462	Sigit Prasetyo	84	Baik	Baik	Baik
92	00020 19463	Suharianto	85	Sangat Baik	Baik	Baik
93	00020 19464	Suma Wijaya	75	Baik	Baik	Baik
94	00020 19465	Syahril Mahendra	80	Baik	Baik	Baik
95	00020 19466	Winarhadi Bimantoro	87	Sangat Baik	Baik	Baik
96	00020 19467	Yogi Pratama	86	Sangat Baik	Baik	Baik
97	00020 20433	Abdillah Firmansyah	90	Sangat Baik	Baik	Baik
98	00020 20434	Ade Rahmat Ramadhan Harahap	78	Baik	Baik	Baik
99	00020 20435	Agung Tanamal	90	Sangat Baik	Baik	Baik
100	00020 20436	Albert Rio Petra Simanjuntak	84	Baik	Baik	Baik
101	00020 20437	Aldo Syahputra	90	Sangat Baik	Baik	Baik
102	00020 20438	Andi Syahputra	88	Sangat Baik	Baik	Baik
103	00020 20439	Andriansah	77	Baik	Baik	Baik

NO	NIS	NAMA SISWA	NILAI AKHIR	NILAI SIKAP	NILAI PENGETAHUAN	NILAI KETERAMPILAN
104	00020 20440	Arman Prananta	77	Baik	Baik	Baik
105	00020 20441	Arya Fadlan Hafidztyar	80	Baik	Baik	Baik
106	00020 20442	Bambang Lesmana	78	Baik	Baik	Baik
107	00020 20443	Fadhil	90	Sangat Baik	Baik	Baik
108	00020 20444	Farhan Khairi Parinduri	88	Sangat Baik	Baik	Baik
109	00020 20445	Gledek Aljabir	77	Baik	Baik	Baik
110	00020 20446	Hafis Syahbani	82	Baik	Baik	Baik
111	00020 20447	Hasnan Nuzuli Al Furqon	85	Sangat Baik	Baik	Baik
112	00020 20449	Irfan Prayoga	90	Sangat Baik	Baik	Baik
113	00020 20450	Ivan Maulana	88	Sangat Baik	Baik	Baik
114	00020 20451	M. Furqon Azura Harahap	85	Sangat Baik	Baik	Baik
115	00020 20452	M. Jaya Praditto	90	Sangat Baik	Baik	Baik
116	00020 20453	M. Zulfakhrin Sembiring	78	Baik	Baik	Baik
117	00020 20454	M. Bagus Nugraha	80	Baik	Baik	Baik
118	00020 20456	Muhammad Irfan	77	Baik	Baik	Baik
119	00020 20457	Naqsa Sholadi	88	Sangat Baik	Baik	Baik
120	00020 20458	Nicholas Saptarossi	80	Baik	Baik	Baik
121	00020 20459	Radid Rizky Setiawan	82	Baik	Baik	Baik
122	00020 20461	Rico Syahputra	90	Sangat Baik	Baik	Baik
123	00020 20462	Rizky Andika Pratama S	90	Sangat Baik	Baik	Baik
124	00020 20463	Roky Rinaldo	78	Baik	Baik	Baik
125	00020 20464	Seftian Dwi Prayoga	88	Sangat Baik	Baik	Baik
126	00020 20465	Sumitra	90	Sangat Baik	Baik	Baik
127	00020 20466	Taufik Al Buqori	88	Sangat Baik	Baik	Baik
128	00020 20467	Tejo Agusti Utomo	90	Sangat Baik	Baik	Baik
129	00020 20468	Zikri Bagus Sumbada	77	Baik	Baik	Baik

NO	NIS	NAMA SISWA	NILAI AKHIR	NILAI SIKAP	NILAI PENGETAHUAN	NILAI KETERAMPILAN
130	00020 21146	Agus Ilhamdi	90	Sangat Baik	Baik	Baik
131	00020 21147	Ahmad Alfa Syahrin	82	Baik	Baik	Baik
132	00020 21148	Aidil Ilham	90	Sangat Baik	Baik	Baik
133	00020 21149	Aidil Syawal Aditama	88	Sangat Baik	Baik	Baik
134	00020 21150	Arjuna Wiratama	80	Baik	Baik	Baik
135	00020 21151	Bagas Arya Pratama	87	Sangat Baik	Baik	Baik
136	00020 21152	Bagus Tri Anggara	88	Sangat Baik	Baik	Baik
137	00020 21153	Calvin Dava Sudedi Putra	92	Sangat Baik	Baik	Baik
138	00020 21154	Fanny Fadila	91	Sangat Baik	Baik	Baik
139	00020 22793	Favian Zaidi	0	Kurang	Kurang	Kurang
140	00020 21155	Fery Setiawan	80	Baik	Baik	Baik
141	00020 21156	Gilang Saprizal	90	Sangat Baik	Baik	Baik
142	00020 21157	Ihsan Yuda Satrio	90	Sangat Baik	Baik	Baik
143	00020 21158	Ikmal Agustian Syahputra	88	Sangat Baik	Baik	Baik
144	00020 21159	Joko Prakoso	88	Sangat Baik	Baik	Baik
145	00020 21161	M Fikri Hutabarat	90	Sangat Baik	Baik	Baik
146	00020 21162	Mhd. Julham Syahputra	89	Sangat Baik	Baik	Baik
147	00020 21163	Muhamad Nur Afandi	88	Sangat Baik	Baik	Baik
148	00020 21164	Muhammad Alfa Rizi	88	Sangat Baik	Baik	Baik
149	00020 21165	Muhammad Fadli Iqbal Aroyan	92	Sangat Baik	Baik	Baik
150	00020 21166	Muhammad Fathin Rifai	92	Sangat Baik	Baik	Baik
151	00020 21167	Noer Muharif Ilham	92	Sangat Baik	Baik	Baik
152	00020 21168	Reno	93	Sangat Baik	Baik	Baik
153	00020 21169	Ridho Ramadhana	90	Sangat Baik	Baik	Baik
154	00020 21170	Rizky Afriandi	90	Sangat Baik	Baik	Baik
155	00020 21171	Rizky Ahmad Hamdhani	90	Sangat Baik	Baik	Baik

NO	NIS	NAMA SISWA	NILAI AKHIR	NILAI SIKAP	NILAI PENGETAHUAN	NILAI KETERAMPILAN
156	00020 21172	Rizky Suwanda	95	Sangat Baik	Baik	Baik
157	00020 21173	Shaff Adiyat Alfath	92	Sangat Baik	Baik	Baik
158	00020 21174	Sofiyanto Aditya Ginting	94	Sangat Baik	Baik	Baik
159	00020 21175	Tama Irawan	90	Sangat Baik	Baik	Baik
160	00020 21176	Taufik Hidayat Nst	90	Sangat Baik	Baik	Baik
161	00020 21177	Wanda Syahputra	90	Sangat Baik	Baik	Baik
162	00020 21178	Wendy Jay Prastyo	88	Sangat Baik	Baik	Baik
163	00020 21179	Yuda Firmansyah	87	Sangat Baik	Baik	Baik

Lampiran 9. Kode Program

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score,
f1_score

# Membaca dataset
file_path = 'ukkk.csv'
data = pd.read_csv(file_path)

# Menampilkan beberapa baris pertama dari dataset
print(data.head())

# Memprediksi nilai kolom 'target'
target_column = 'nilai_akhir' # nama kolom target di dataset
features = data.columns.difference([target_column])

# Memisahkan fitur dan target
X = data[features]
y = data[target_column]

# Fungsi Monte Carlo untuk prediksi
def monte_carlo_simulation(X, y, num_simulations=10):
    predictions = []
    for i in range(num_simulations):
        # Random sampling with replacement
        sample_indices = np.random.choice(len(X), size=len(X), replace=True)
        sample_X = X.iloc[sample_indices]
        sample_y = y.iloc[sample_indices]
        # Menghitung rata-rata dari sample_y sebagai prediksi (sederhana)
        prediction = sample_y.mean()
        predictions.append(prediction)
        print(f'Simulasi {i+1}: Prediksi = {prediction}')
    return predictions

# Melakukan simulasi Monte Carlo
predictions = monte_carlo_simulation(X, y)

# Menampilkan distribusi hasil prediksi
plt.hist(predictions, bins=30, edgecolor='k', alpha=0.7)
plt.xlabel('Prediksi')
plt.ylabel('Frekuensi')

```

```

plt.title('Distribusi Prediksi Monte Carlo')
plt.show()

# Menghitung statistik dari prediksi
pred_mean = np.mean(predictions)
pred_std = np.std(predictions)
print(f'Rata-rata Prediksi: {pred_mean}')
print(f'Standar Deviasi Prediksi: {pred_std}')

# Load dataset
df_ukk = pd.read_csv('test.csv')

# Data Cleaning missing value
df_ukk.isnull().sum

# Checkin if the data double
df_ukk.duplicated().sum

# Preprocessing data
le = LabelEncoder()
df_ukk['nama_siswa'] = le.fit_transform(df_ukk['nama_siswa'])

X = df_ukk.drop(['nilai_ket'], axis=1)
y = df_ukk['nilai_ket']

# Split data menjadi training dan testing set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25,
random_state=0)

# Buat model KNN
knn = KNeighborsClassifier(n_neighbors=2, metric='euclidean', p=2)

# Train model
knn.fit(X_train, y_train)

# Prediksi pada testing set
y_pred = knn.predict(X_test)

# Evaluasi model
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='macro')
recall = recall_score(y_test, y_pred, average='macro')
f1 = f1_score(y_test, y_pred, average='macro')
error_rate = 1 - accuracy

# Print evaluasi model
print("Akurasi: {:.2f}%".format(accuracy*100))
print("Presisi: {:.2f}%".format(precision*100))

```

```
print("Recall: {:.2f}%".format(recall*100))
print("F1 Score: {:.2f}%".format(f1*100))
print("Error Rate: {:.2f}%".format(error_rate*100))

#Exploratory Data Analysis (EDA)
sns.set_theme(style="darkgrid")
sns.countplot(x='nilai_ket', data=df_ukk, palette='rocket')
plt.ylabel("Jumlah")
plt.xlabel('0=Baik 1=Sangat Baik')
plt.title("Kompeten")
plt.show()

#Confusion Matrix
from sklearn.metrics import confusion_matrix, accuracy_score, precision_score,
recall_score, f1_score, classification_report
cm = confusion_matrix(y_test, y_pred)
print(cm)

#Visualisasi Confusion Matrix Bagian Testing
sns.heatmap(confusion_matrix(y_test, y_pred), annot= True, cmap='viridis',
fmt='.0f')
plt.xlabel('Predicted Values', fontdict={'size':20}, labelpad=10)
plt.ylabel('Actual Values', fontdict={'size':20}, labelpad=10)
plt.title('Confusion Matrix pada Testing')
plt.show()
```

Lampiran 10. Dokumentasi Penelitian

