

**IMPLEMENTASI *K-NEAREST NEIGHBORS* (KNN) DAN FUNGSI
KEANGGOTAAN FUZZY UNTUK IMPUTASI *MISSING VALUES*
PADA KLASIFIKASI DATA**

(Skripsi)

Oleh

REVITA SETIANINGSIH

1917051058



**JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2023**

ABSTRAK

IMPLEMENTASI *K-NEAREST NEIGHBORS* (KNN) DAN FUNGSI KEANGGOTAAN FUZZY UNTUK IMPUTASI *MISSING VALUES* PADA KLASIFIKASI DATA

Oleh

REVITA SETIANINGSIH

Missing values merupakan suatu keadaan dimana terdapat nilai yang kosong atau hilang dalam sebuah dataset. Nilai yang hilang ini dapat terjadi karena kesalahan responden yang tidak menjawab semua pertanyaan dalam kuesioner, pengukuran dan eksperimen yang salah ketika input data secara manual, adanya beberapa data yang disensor, dan lainnya. Terdapat 3 cara untuk mengatasi *missing values* yaitu *listwise deletion*, *pairwise deletion*, dan imputasi. Pada penelitian ini dilakukan imputasi menggunakan KNN dan fungsi keanggotaan fuzzy untuk mengetahui apakah metode ini dapat menangani *missing values* dengan lebih baik dibanding dengan metode imputasi konvensional.

Fungsi keanggotaan fuzzy segitiga dan Gaussian digunakan dalam algoritma imputasi data untuk menghitung bobot nilai data tetangga terdekat sebelum digunakan selama proses imputasi. Adapun data penelitian pada penelitian ini menggunakan jenis data klasifikasi yang diperoleh dari *UCI Machine Learning Repository* yang berjumlah 5 dataset. Hasil yang diperoleh menunjukkan bahwa fungsi keanggotaan fuzzy memiliki nilai akurasi yang lebih tinggi dibandingkan dengan imputasi konvensional dimana fungsi keanggotaan fuzzy segitiga lebih unggul dari fungsi keanggotaan fuzzy Gaussian berdasarkan rata-rata akurasinya. Sehingga, fungsi keanggotaan fuzzy lebih baik untuk menangani *missing values* dalam meningkatkan akurasi.

Kata kunci: *Missing Values*, *K-Nearest Neighbor* (KNN), Fungsi Keanggotaan Fuzzy, Logika Fuzzy

ABSTRACT

IMPLEMENTATION OF K-NEAREST NEIGHBORS (KNN) AND FUZZY MEMBERSHIP FUNCTION FOR IMPUTATION OF MISSING VALUES IN DATA CLASSIFICATION

By

REVITA SETIANINGSIH

A missing value is a condition where there are any values that blank or missing in the dataset. This missing value can occur due to respondents' errors in not answering all questions in the questionnaire, incorrect measurements and experiments when entering data manually, the presence of some censored data, and others. There are 3 ways to overcome missing values, namely listwise deletion, pairwise deletion, and imputation. In this research, imputation was carried out using KNN and fuzzy membership functions to find out whether this method can handle missing values better than conventional imputation methods.

The triangular and Gaussian fuzzy membership functions are used in data imputation algorithms to calculate the weight for the data values of the nearest neighbor's before they are used during the imputation process. The research data in this study used the type of classification data obtained from the UCI Machine Learning Repository which amounted to 5 datasets. The results obtained show that the fuzzy membership function has a higher accuracy value compared to conventional imputation where the triangular fuzzy membership function is superior to the Gaussian fuzzy membership function based on its average accuracy. Thus, the fuzzy membership function is better for handling missing values in improving accuracy.

Keywords: Missing Values, K-Nearest Neighbor (KNN), Fuzzy Membership Function, Fuzzy Logic

**IMPLEMENTASI *K-NEAREST NEIGHBORS* (KNN) DAN FUNGSI
KEANGGOTAAN FUZZY UNTUK IMPUTASI *MISSING VALUES*
PADA KLASIFIKASI DATA**

Oleh

REVITA SETIANINGSIH

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA KOMPUTER

Pada

**Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG**

2023

Judul Skripsi : **IMPLEMENTASI *K-NEAREST NEIGHBORS* (KNN) DAN FUNGSI KEANGGOTAAN FUZZY UNTUK IMPUTASI *MISSING VALUES* PADA KLASIFIKASI DATA**

Nama Mahasiswa : **Revita Setianingsih**

Nomor Pokok Mahasiswa : **1917051058**

Program Studi : **S1 Ilmu Komputer**

Jurusan : **Ilmu Komputer**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**



1. Komisi Pembimbing


Dr. Ir. Kurnia Muludi, M.S.Sc.
NIP. 19640616 198902 1 001


Ridho Sholehurrohman, M.Mat.
NIP. 232111970128101

2. Ketua Jurusan Ilmu Komputer


Didik Kurniawan, S.Si., M.T.
NIP. 19800419 200501 1 004

MENGESAHKAN

1. Tim Penguji

Ketua : **Dr. Ir. Kurnia Muludi, M.S.Sc.**



Sekretaris : **Ridho Sholehurrohman, M.Mat.**



Penguji Bukan Pembimbing : **Dr. rer. nat. Akmal Junaidi, M.Sc**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam




Dr. Eng. Heri Satria, S.Si., M.Si.

NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi: **28 Juli 2023**

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Revita Setianingsih

NPM : 1917051058

Dengan ini menyatakan bahwa skripsi saya yang berjudul “Implementasi *K-Nearest Neighbors* (KNN) dan Fungsi Keanggotaan Fuzzy untuk Imputasi *Missing Values* pada Klasifikasi Data” merupakan karya saya sendiri, bukan karya orang lain. Semua tulisan yang tertulis dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Jika di kemudian hari terbukti bahwa karya tulis ilmiah saya terbukti hasil menjiplak karya orang lain, maka saya siap menerima sanksi berupa pencabutan gelar yang saya peroleh.

Randa Lampung, Juli 2023



Revita Setianingsih

NPM. 1917051058

RIWAYAT HIDUP

Penulis lahir di Sambikarto, pada 27 Agustus 2001, sebagai anak pertama dari dua bersaudara. Penulis menyelesaikan pendidikan formal di SDN 2 Sambikarto dan selesai pada tahun 2013. Kemudian pendidikan menengah pertama di SMPN 1 Bumi Agung yang diselesaikan pada tahun 2016, lalu melanjutkan ke pendidikan menengah atas di SMAN 1 Sekampung yang diselesaikan pada tahun 2019.

Pada tahun 2019 penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SBMPTN. Selama menjadi mahasiswa, penulis melakukan beberapa kegiatan antara lain.

1. Menjadi anggota muda bidang Internal Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2020/2021.
2. Menjadi anggota bidang Kaderisasi Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2020/2021.
3. Menjadi anggota bidang Kaderisasi Rois FMIPA Universitas Lampung Periode 2020/2021.
4. Menjadi Staff Ahli Kementerian Advokasi Publik BEM U KBM Universitas Lampung Periode 2020/2021.
5. Menjadi Bendahara Dinas Kajian dan Aksi Propaganda BEM FMIPA Universitas Lampung Periode 2021/2022.
6. Melaksanakan Kerja Praktik periode 2021/2022 di Dinas Perhubungan Provinsi Lampung.
7. Mengikuti program Studi Independen Bersertifikat Kampus Merdeka di PT MariBelajar Indonesia Cerdas pada tahun 2022.
8. Melaksanakan KKN periode II tahun 2022 di Desa Sri Gading, Kecamatan Labuhan Maringgai, Lampung Timur.

MOTTO

“The hardest choices in life aren't between what's right and what's wrong, but between what's right and what's best.”

- Jamie Ford -

“The harder you work, the luckier you get. But let them think it's all luck.”

- Thomas Jefferson –

PERSEMBAHAN

Alhamdulillahirobbilalamin

Puji dan syukur tercurahkan kepada Allah Subhanahu Wa Ta'alaah atas segala Rahmat dan Karunia-Nya sehingga saya dapat menyelesaikan skripsi ini. Shalawat serta salam selalu tercurahkan kepada Nabi Muhammad SAW.

Kupersembahkan karya ini kepada:

Kedua Orang Tuaku Tercinta

Yang senantiasa memberikan yang terbaik, dukungan dan doa yang selalu menyertaiku. Terima kasih banyak atas semangat, motivasi, dan apresiasi yang selalu mengiringi. Dan terima kasih banyak telah tanpa lelah mendidik dan membesarkanku dengan penuh cinta, kesabaran dan pengorbanan.

Seluruh Keluarga Besar Ilmu Komputer 2019

Yang selalu memberikan semangat dan dukungan.

Almamater Tercinta, Universitas Lampung dan Jurusan Ilmu Komputer

Tempat belajar mengemban seluruh ilmu untuk menjadi bekal hidup.

SANWACANA

Puji syukur kehadiran Allah SWT atas berkah, rahmat dan hidayat-Nya, serta petunjuk dan pedoman dari Rasulullah Nabi Muhammad SAW Penulis dapat menyelesaikan Skripsi yang berjudul “Implementasi *K-Nearest Neighbors* (KNN) dan Fungsi Keanggotaan Fuzzy untuk Imputasi *Missing Values* pada Klasifikasi Data” dengan baik.

Terima kasih Penulis ucapkan kepada semua pihak yang telah membantu dan berperan besar dalam penyusunan Skripsi ini, antara lain:

1. Kedua orang tua tercinta dan adik tersayang yang selalu memberi dukungan, do’a, semangat, dan kasih sayang yang tak terhingga. Semoga Allah SWT selalu memberikan kebahagiaan dan keberkahan dalam kehidupan kalian di dunia dan akhirat.
2. Bapak Dr. Ir. Kurnia Muludi, M.S.Sc. selaku Pembimbing Utama yang telah memberikan arahan, ide, serta saran kepada Penulis sehingga dapat menyelesaikan Skripsi ini dengan baik.
3. Bapak Ridho Sholehurrohman, M.Mat. selaku Pembimbing Kedua yang juga selalu dapat meluangkan waktunya untuk membimbing Penulis dalam memberikan ide, motivasi, serta saran untuk dapat menyelesaikan Skripsi ini.
4. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc. selaku Pembahas yang telah memberikan masukan yang bermanfaat dalam perbaikan Skripsi ini.

5. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc. selaku Pembimbing Akademik penulis yang selalu mendukung peningkatan akademik Penulis.
6. Bapak Didik Kurniawan, S.Si., M.T. selaku Ketua Jurusan Ilmu Komputer FMIPA Universitas Lampung.
7. Bapak Dr. Eng. Heri Satria, S.Si., M.Si. selaku Dekan FMIPA Universitas Lampung.
8. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc. selaku Sekretaris Jurusan Ilmu Komputer FMIPA Universitas Lampung.
9. Ibu Ade Noera Maela dan Bang Zainuddin yang telah membantu segala urusan administrasi Penulis di Jurusan Ilmu Komputer.
10. Bapak dan Ibu Dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung yang telah membantu penulis selama proses perkuliahan dan memberikan ilmu yang bermanfaat kepada penulis.
11. Pihak pemberi program beasiswa Bidikmisi yang telah memberikan bantuan finansial kepada penulis selama masa kuliah.
12. Achmad romadoni yang telah memberikan semangat dan membantu dalam pengerjaan skripsi ini.
13. Teman dekat di jurusan, yaitu Salsa dan Tasya. Terima kasih telah berteman baik dan banyak membantu Penulis dalam berbagai hal.
14. Fitri, Yulia, dan Sarah selaku teman yang selalu mendengar keluhan penulis, memberi dukungan dan canda tawa.
15. Rekan-rekan Ilmu Komputer 2019, kakak dan adik tingkat yang tidak dapat disebut satu persatu, rekan kelompok, rekan diskusi, rekan bercanda, dan rekan berjuang semasa duduk di bangku kuliah.

16. Teman-teman Himakom 2020, Rois 2020-2021, BEM U 2020, Pansus FMIPA 2020, dan Pimpinan BEM FMIPA 2021 yang tidak dapat disebutkan satu persatu. Terima kasih sudah mengajarkan banyak hal dalam berorganisasi.
17. Diri sendiri. Terima kasih karena telah tidak menyerah dalam banyak hal. Kamu hebat.

Penulis menyadari bahwa Skripsi ini masih jauh dari kata sempurna, semoga Skripsi ini membawa manfaat dan keberkahan bagi semua yang membacanya Aamiin Ya Rabbal Aalamiin.

Bandar Lampung, Juli 2023

Revita Setianingsih

NPM. 1917051058

DAFTAR ISI

	Halaman
DAFTAR ISI	xv
DAFTAR GAMBAR	xvii
DAFTAR TABEL	xix
DAFTAR KODE PROGRAM	xx
DAFTAR LAMPIRAN	xxi
I. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan.....	3
1.4 Manfaat.....	3
1.5 Batasan Masalah.....	4
II. TINJAUAN PUSTAKA	5
2.1 Logika Fuzzy	5
2.1.1. Himpunan Fuzzy.....	6
2.1.2. Fungsi Keanggotaan Fuzzy	6
2.2 Imputasi <i>K-Nearest Neighbors</i> (KNN).....	9
2.3 Klasifikasi <i>K-Nearest Neighbors</i> (KNN)	10
2.4 <i>Euclidean Distance</i>	11
2.5 <i>Min-Max Normalization</i>	12
2.6 <i>Confusion Matrix</i>	13

2.7 Penelitian Terdahulu.....	14
III. METODE PENELITIAN.....	18
3.1 Waktu dan Tempat Penelitian	18
3.2 Tahapan Penelitian	18
3.3 Alat dan Bahan	22
3.3.1 Alat Penelitian	22
3.3.2 Bahan Penelitian.....	23
IV. HASIL DAN PEMBAHASAN.....	27
4.1 Hasil dan Pembahasan.....	27
4.1.1 <i>Hepatitis dataset</i>	35
4.1.2 <i>Ozone Level Detection Dataset</i>	39
4.1.3 <i>Blood Transfusion Service Center Dataset</i>	44
4.1.4 <i>Parkinson Speech Dataset</i>	48
4.1.5 <i>Audit Risk Dataset</i>	53
4.2 Perbandingan Hasil Akurasi.....	58
V. KESIMPULAN DAN SARAN.....	62
5.1 Kesimpulan.....	62
5.2 Saran.....	63
DAFTAR PUSTAKA.....	64

DAFTAR GAMBAR

	Halaman
Gambar 1. Logika tegas (kiri) dan logika fuzzy (kanan).	5
Gambar 2. Grafik fungsi keanggotaan fuzzy segitiga.....	7
Gambar 3. Grafik fungsi keanggotaan trapesium.	8
Gambar 4. Grafik fungsi keanggotaan Gaussian.....	9
Gambar 5. Ilustrasi ketetanggaan KNN.	10
Gambar 6. Tahapan perhitungan nilai imputasi metode KNN dan fuzzy.....	20
Gambar 7. Tahapan perhitungan nilai imputasi metode konvensional.....	21
Gambar 8. Tahapan klasifikasi dan uji akurasi untuk setiap teknik imputasi.....	22
Gambar 9. Ilustrasi data untuk perhitungan jarak terdekat.	28
Gambar 10. Ilustrasi perolehan nilai terdekat.	31
Gambar 11. Ilustrasi imputasi dengan fungsi keanggotaan fuzzy segitiga.....	33
Gambar 12. Ilustrasi imputasi dengan fungsi keanggotaan fuzzy Gaussian.....	35
Gambar 13. Grafik persentase akurasi pada <i>hepatitis dataset</i>	37
Gambar 14. Grafik persentase akurasi pada <i>ozone level detection dataset</i>	42
Gambar 15. Grafik persentase akurasi pada <i>blood transfusion dataset 5%</i>	47
Gambar 16. Grafik persentase akurasi pada <i>blood transfusion dataset 10%</i>	47
Gambar 17. Grafik persentase akurasi pada <i>parkinson speech dataset 5%</i>	52
Gambar 18. Grafik persentase akurasi pada <i>parkinson speech dataset 10%</i>	52
Gambar 19. Grafik persentase akurasi pada <i>audit risk dataset 5%</i>	57
Gambar 20. Grafik persentase akurasi pada <i>audit risk dataset 10%</i>	57
Gambar 21. Grafik persentase rata-rata akurasi seluruh dataset.	59
Gambar 22. Perbandingan hasil akurasi dengan penelitian terdahulu.	60
Gambar 23. <i>Hepatitis dataset</i>	73
Gambar 24. <i>Ozone level detection dataset</i>	73
Gambar 25. <i>Parkinson speech dataset</i> dengan <i>missing 5%</i>	73

Gambar 26. <i>Parkinson speech dataset</i> dengan <i>missing</i> 10%.....	74
Gambar 27. <i>Blood transfusion dataset</i> dengan <i>missing</i> 5%.....	74
Gambar 28. <i>Blood transfusion dataset</i> dengan <i>missing</i> 10%.....	74
Gambar 29. <i>Audit risk dataset</i> dengan <i>missing</i> 5%.	75
Gambar 30. <i>Audit risk dataset</i> dengan <i>missing</i> 10%.....	75

DAFTAR TABEL

	Halaman
Tabel 1. <i>Confusion Matrix</i>	13
Tabel 2. Rincian persentase nilai hilang pada masing-masing dataset	23
Tabel 3. Rincian nilai hilang masing-masing fitur pada hepatitis dataset.....	36
Tabel 4. Hasil perhitungan akurasi pada hepatitis dataset	38
Tabel 5. Rincian nilai hilang masing-masing fitur pada <i>ozone dataset</i>	39
Tabel 6. Hasil perhitungan akurasi pada <i>ozone dataset</i> 10%.....	43
Tabel 7. Rincian nilai hilang masing-masing fitur pada <i>transfusion dataset</i>	44
Tabel 8. Hasil perhitungan akurasi pada <i>blood transfusion dataset</i> 5%	46
Tabel 9. Hasil perhitungan akurasi pada <i>blood transfusion dataset</i> 10%	46
Tabel 10. Rincian nilai hilang masing-masing fitur pada <i>parkinson dataset</i>	48
Tabel 11. Hasil perhitungan akurasi pada <i>parkinson speech dataset</i> 5%	51
Tabel 12. Hasil perhitungan akurasi pada <i>Parkinson speech dataset</i> 10%.....	51
Tabel 13. Rincian nilai hilang masing-masing fitur pada <i>audit dataset</i>	53
Tabel 14. Hasil perhitungan akurasi pada <i>audit risk dataset</i> 5%.....	56
Tabel 15. Hasil perhitungan akurasi pada <i>audit risk dataset</i> 10%.....	56
Tabel 16. Rata-rata akurasi metode konvensional dari seluruh dataset	58
Tabel 17. Rata-rata akurasi metode KNN dan fuzzy dari seluruh dataset	58

DAFTAR KODE PROGRAM

	Halaman
Kode Program 1. Penghapusan dataset lengkap dengan persentase 5%.....	67
Kode Program 2. Penghapusan dataset lengkap dengan persentase 10%.....	67
Kode Program 3. Mencari nilai terdekat berdasarkan jarak terdekat.....	68
Kode Program 4. Fungsi keanggotaan fuzzy segitiga.....	68
Kode Program 5. Fungsi keanggotaan fuzzy Gaussian.....	68
Kode Program 6. Menghitung bobot nilai dengan fungsi fuzzy segitiga.....	69
Kode Program 7. Menghitung bobot nilai dengan fungsi fuzzy Gaussian	70
Kode Program 8. Menjalankan seluruh fungsi yang telah didefinisikan	70
Kode Program 9. Imputasi menggunakan metode mean	71
Kode Program 10. Imputasi menggunakan metode median	71
Kode Program 11. <i>Listwise Deletion</i> (data asli)	71
Kode Program 12. Klasifikasi data hasil imputasi	72
Kode Program 13. Grafik persentase hasil akurasi	72

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Kode program yang digunakan untuk merekayasa dataset lengkap menjadi dataset dengan persentase missing 5% dan 10%.....	67
2. Kode program yang digunakan untuk melakukan imputasi dataset dengan <i>missing values</i>	68
3. Kode program yang digunakan untuk imputasi dengan metode konvensional dan klasifikasi data asli.....	71
4. Kode program yang digunakan untuk klasifikasi data hasil imputasi dan grafik persentase hasil akurasi.....	72
5. Dataset yang digunakan dalam penelitian.....	73

I. PENDAHULUAN

1.1 Latar Belakang

Ketika melakukan sebuah penelitian, data yang lengkap dan berkualitas sangat diharapkan dalam mengumpulkan data (Susanti et al., 2018). Pada bidang ilmu komputer berkisar tahun 1990-an, mulai terjadi peningkatan masalah kualitas data seiring dengan munculnya sistem *data warehouse*. Masalah ini kemudian dikembalikan ke kualitas data yang buruk ketika terjadi kegagalan proyek pada *database* (El-Bakry et al., 2021). Menurut (Pratama et al., 2016) definisi kualitas data yang sering digunakan oleh peneliti yaitu “*fitness for use*” artinya kualitas data dapat diringkas dalam bagaimana sistem *fit* dengan kenyataan atau bagaimana pengguna benar-benar memanfaatkan data dalam sistem.

Missing values atau dapat dikenal dengan nilai yang hilang adalah suatu keadaan dimana terdapat nilai yang kosong atau hilang dalam sebuah dataset. Nilai yang hilang ini dapat terjadi karena responden yang tidak menjawab semua pertanyaan dalam kuesioner, pengukuran dan eksperimen yang salah ketika input data secara manual, adanya beberapa data yang disensor, dan lainnya (Kaiser, 2014). Terdapat tiga jenis *missing values* yaitu *Missing Completely at Random* (MCAR), *Missing at Random* (MAR), dan *Not Missing at Random* (NMAR). MCAR adalah kondisi dimana nilai hilang sepenuhnya secara acak dan tidak berkaitan dengan nilai yang hilang atau tidak hilang. MAR adalah kondisi dimana nilai hilang secara acak juga namun nilai yang hilang berkaitan dengan variabel yang hilang dan variabel yang tidak hilang. Sedangkan NMAR adalah kondisi dimana terjadi nilai hilang secara tidak acak dan berkaitan dengan nilai yang hilang itu sendiri (Susanti et al., 2018).

Ada tiga cara dalam menangani nilai yang hilang (Davey & Savla, 2009). Pertama, *pairwise deletion* yaitu mengatasi nilai yang hilang dengan menghapus nilai yang mengandung *missing values* pada perhitungan yang spesifik. Dalam *pairwise deletion*, data yang dihapus hanya pada pasangan data yang dihitung dan nilai kosong pada fitur lain dapat dipertahankan, sehingga analisis yang dilakukan hanya pada data yang tersedia. Kedua, *listwise deletion* yaitu mengatasi nilai yang hilang dengan menghapus semua observasi yang terdapat setidaknya satu *missing values*. Namun penggunaan cara ini dianggap kurang efektif karena dapat menghilangkan informasi penting dalam data yang dihapus. Ketiga, imputasi yaitu mengatasi nilai yang hilang dengan mengisi nilai yang hilang tersebut menggunakan nilai yang memungkinkan berdasarkan informasi yang didapat dari nilai yang diketahui. Adapun beberapa metode imputasi konvensional yang dapat digunakan untuk mengisi nilai yang hilang menurut (Maula et al., 2022) antara lain rata-rata (*mean*), nilai yang sering muncul (*mode*), dan nilai tengah (*median*). Selain metode tersebut, *missing values* juga dapat dilakukan menggunakan metode-metode lain untuk meminimalkan dampak negatif dari *missing values*.

Dalam penelitian ini akan dilakukan imputasi *missing values* menggunakan metode *K-Nearest Neighbors* (KNN) dan fungsi keanggotaan fuzzy. Teknik yang dilakukan dalam memprediksi nilai yang hilang ini dicari berdasarkan observasi terdekat dari observasi lain yang memiliki kemiripan nilai dengan nilai yang hilang. Teknik yang dilakukan terlebih dahulu menentukan nilai k untuk tetangga terdekat. Kemudian menghitung nilai berdasarkan jarak terdekat menggunakan rumus *Euclidean distance*. Nilai tetangga terdekat tersebut dimasukkan ke dalam formula keanggotaan fuzzy dan dicari bobot nilainya. Imputasi *missing values* dilakukan dengan menghitung nilai terdekat dan bobot nilainya menggunakan fungsi rata-rata tertimbang. Data yang telah diimputasi selanjutnya diklasifikasi menggunakan metode klasifikasi *K-Nearest Neighbours* (KNN). Keakuratan teknik yang diusulkan akan dievaluasi dengan menghitung tingkat akurasinya. Nilai akurasi yang didapat digunakan untuk membandingkan metode terbaik antara teknik imputasi

menggunakan metode yang diusulkan dan teknik imputasi secara konvensional.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka didapat rumusan masalah yang perlu diselesaikan yaitu:

1. Bagaimana imputasi *missing values* menggunakan metode KNN dan fungsi keanggotaan fuzzy, serta metode konvensional.
2. Bagaimana perbandingan tingkat akurasi berdasarkan klasifikasi menggunakan KNN antara teknik imputasi konvensional dengan teknik imputasi menggunakan metode KNN dan fungsi keanggotaan fuzzy.

1.3 Tujuan

Tujuan dari penelitian adalah sebagai berikut.

1. Melakukan imputasi *missing values* menggunakan metode imputasi KNN dan fungsi keanggotaan fuzzy, serta metode konvensional.
2. Melakukan klasifikasi data hasil imputasi menggunakan metode KNN.
3. Menghitung dan membandingkan nilai akurasi antara teknik imputasi *missing values* konvensional dengan teknik imputasi menggunakan metode KNN dan fungsi keanggotaan fuzzy.

1.4 Manfaat

Manfaat yang didapat dari penelitian ini adalah:

1. Memperoleh hasil perbandingan tingkat akurasi antara metode imputasi *missing values* secara konvensional dengan metode imputasi *missing values* menggunakan metode KNN dan fungsi keanggotaan fuzzy pada klasifikasi 5 dataset.

2. Mengetahui metode terbaik antara metode imputasi konvensional dengan metode imputasi KNN dan fungsi keanggotaan fuzzy untuk permasalahan *missing values*.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini agar penelitian lebih terfokus dan tidak meluas adalah sebagai berikut.

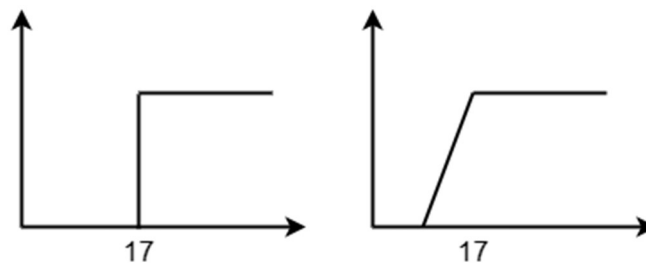
1. Dataset yang digunakan berjumlah 5 dataset dengan 2 dataset merupakan dataset dengan nilai yang hilang dan 3 dataset merupakan dataset lengkap dengan nilai yang dihilangkan.
2. Metode yang digunakan untuk melakukan imputasi *missing values* yaitu metode KNN dan fungsi keanggotaan fuzzy, serta metode konvensional.
3. Fungsi keanggotaan fuzzy yang digunakan terdiri dari fungsi keanggotaan fuzzy segitiga dan fungsi keanggotaan fuzzy Gaussian.
4. Metode imputasi *missing values* konvensional yang digunakan sebagai pembandingan adalah rata-rata (*mean*) dan nilai tengah (*median*).
5. Metode klasifikasi untuk menghitung tingkat akurasi yang digunakan adalah Metode KNN.
6. Dataset yang nilainya dihilangkan pada dataset lengkap dilakukan secara acak.

II. TINJAUAN PUSTAKA

2.1 Logika Fuzzy

Fuzzy dalam bahasa Inggris memiliki arti kabur atau tidak jelas. Yang artinya logika fuzzy merupakan logika yang kabur atau mengandung unsur ketidakpastian (Saelan, 2009). Pada pertengahan tahun 1960, Prof. Lotfi Zadeh dari Universitas California, mengembangkan sebuah ide bernama set fuzzy yang digunakan untuk penggolongan set. Logika fuzzy tidak seperti logika tegas (*Boolean*) yang memiliki 2 nilai yakni 0 atau 1, salah atau benar. Logika fuzzy memiliki banyak nilai yang terbagi ke dalam derajat keanggotaan dan derajat kebenaran. Derajat keanggotaan dan derajat kebenaran ini dapat menjadikan sebagian benar dan sebagian salah pada waktu yang sama (Wardhani & Haerani, 2011).

Seperti contoh dalam kehidupan sehari-hari, seseorang dikatakan dewasa ketika berusia 17 tahun ke atas. Jika menggunakan logika tegas, seseorang yang berusia 17 tahun kurang 1 hari akan didefinisikan sebagai seorang yang tidak dewasa. Namun dalam logika fuzzy akan dinyatakan bahwa seseorang tersebut hampir dewasa (Saelan, 2009). Gambar 1 menunjukkan perbedaan antara logika tegas dan logika fuzzy menurut (Saelan, 2009).



Gambar 1. Logika tegas (kiri) dan logika fuzzy (kanan).

2.1.1. Himpunan Fuzzy

Himpunan fuzzy merupakan teknik pengelompokan berdasarkan variabel bahasa (*linguistic variable*) yang dinyatakan sebagai fungsi keanggotaan dalam U semesta (Saelan, 2009). Himpunan fuzzy berdasar pada konsep memperluas fungsionalitas fungsi-fungsi tersebut sampai fungsinya mencakup bilangan real pada interval $[0, 1]$. Nilai keanggotaan himpunan fuzzy menyatakan bahwa suatu objek tidak hanya bernilai benar atau salah yang ditandai dengan 1 atau 0, tetapi masih ada nilai lain yang terletak antara benar dan salah tersebut (Nainggolan, 2016).

2.1.2. Fungsi Keanggotaan Fuzzy

Menurut logika tegas, fungsi keanggotaan dinyatakan sebagai keanggotaan pada suatu himpunan. Fungsi keanggotaan $X_A(x)$ bernilai 1 jika x merupakan anggota himpunan A, dan bernilai 0 jika x bukan anggota dari himpunan A (Saelan, 2009). Sehingga, fungsi keanggotaan tegas ini hanya beranggotakan 0 atau 1 yang dapat dinyatakan sebagai berikut.

$$X_A: x \rightarrow \{0, 1\}$$

Sedangkan menurut logika fuzzy, fungsi keanggotaan dinyatakan sebagai derajat keanggotaan dalam suatu himpunan yang nilainya terletak antara 0 dan 1. Derajat keanggotaan ini dinyatakan sebagai μ_A . Fungsi keanggotaan $\mu_A(x)$ bernilai 1 jika x merupakan anggota penuh himpunan A, dan bernilai 0 apabila x bukan anggota himpunan A. Sedangkan apabila derajat keanggotaan terletak antara (0, 1) misal $\mu_A(x) = \mu$ dinyatakan bahwa x sebagian anggota himpunan A dengan derajat keanggotaan sebesar μ (Saelan, 2009). Sehingga, fungsi keanggotaan fuzzy dapat dinyatakan dalam range nilai sebagai berikut.

$$\mu_A: x \rightarrow [0, 1]$$

Secara umum, fungsi keanggotaan pada himpunan fuzzy dapat ditentukan dengan fungsi keanggotaan segitiga (*triangle*), trapezium (*trapezoid*), dan fungsi keanggotaan Gaussian (*Gaussian*) (Purba, 2016).

a. Fungsi keanggotaan segitiga (*triangle*)

Fungsi keanggotaan segitiga digunakan untuk menggambarkan himpunan fuzzy dengan bentuk segitiga. Fungsi keanggotaan ini memiliki tiga parameter yakni titik awal (a), titik puncak (b), dan titik akhir (c). Bentuk fungsi keanggotaan segitiga ini mirip dengan kurva segitiga, dimana keanggotaan suatu nilai pada himpunan fuzzy naik secara linear dari a ke b dan turun secara linear dari b ke c. Adapun persamaan fungsi keanggotaan segitiga adalah

$$\mu(x; a, b, c) = \begin{cases} 0 & x \leq a \\ \frac{x - a}{b - a} & a < x \leq b \\ \frac{c - x}{c - b} & b < x < c \\ 0 & x \geq c \end{cases}$$

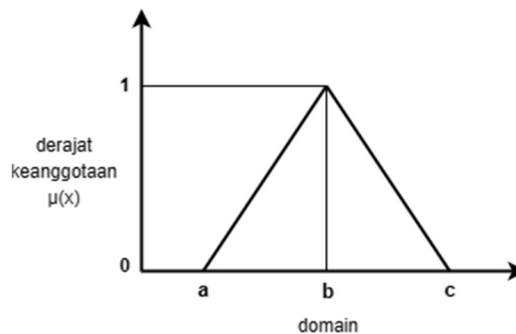
Keterangan:

a = nilai domain terkecil yang memiliki derajat keanggotaan 0

b = nilai domain yang mempunyai derajat keanggotaan 1

c = nilai domain terbesar yang mempunyai derajat keanggotaan 0

Gambar 2 menunjukkan representasi fungsi keanggotaan segitiga dalam bentuk grafik. Sumbu x berlaku sebagai nilai domain dan sumbu y berlaku sebagai derajat keanggotaan $\mu(x)$.



Gambar 2. Grafik fungsi keanggotaan fuzzy segitiga.

b. Fungsi keanggotaan trapesium (*trapezoid*)

Fungsi keanggotaan trapesium digunakan untuk menggambarkan himpunan fuzzy dengan bentuk trapesium. Fungsi keanggotaan ini memiliki empat parameter yakni titik awal (a), titik awal puncak (b), titik akhir puncak (c), dan titik akhir (d). Bentuk fungsi keanggotaan ini mirip dengan kurva trapesium dimana keanggotaan suatu nilai pada himpunan fuzzy naik secara linear dari a ke b, kemudian tetap di nilai maksimum 1 dari b ke c, dan turun secara linear dari c ke d. Adapun persamaan fungsi keanggotaan trapesium adalah

$$\mu(x; a, b, c, d) = \begin{cases} 0 & x \leq a \\ \frac{x - a}{b - a} & a < x \leq b \\ 1 & b < x \leq c \\ \frac{d - x}{d - c} & c < x \leq d \\ 0 & x > d \end{cases}$$

Keterangan:

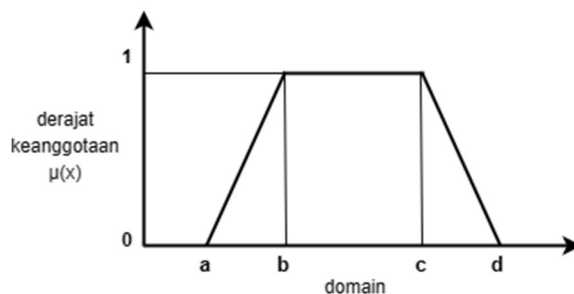
a = nilai domain terkecil yang memiliki derajat keanggotaan 0

b = nilai domain terkecil yang memiliki derajat keanggotaan 1

c = nilai domain terbesar yang memiliki derajat keanggotaan 1

d = nilai domain terbesar yang memiliki derajat keanggotaan 0

Gambar 3 menunjukkan representasi fungsi keanggotaan fuzzy trapesium dalam bentuk grafik. Sumbu x berlaku sebagai nilai domain dan sumbu y berlaku sebagai derajat keanggotaan $\mu(x)$.



Gambar 3. Grafik fungsi keanggotaan trapesium.

c. Fungsi keanggotaan Gaussian (*Gaussian*)

Fungsi keanggotaan Gaussian digunakan untuk menggambarkan himpunan fuzzy dengan bentuk kurva Gaussian atau *bell curve*. Fungsi keanggotaan ini memiliki dua parameter yakni *mean* atau pusat kurva (μ) dan standar deviasi atau lebar kurva (σ). Bentuk fungsi ini mirip dengan kurva lonceng dimana keanggotaan suatu nilai pada himpunan fuzzy naik dari 0 ke nilai maksimum di sekitar pusat kurva (μ) dan kemudian turun kembali ke 0. Adapun persamaan fungsi keanggotaan Gaussian adalah

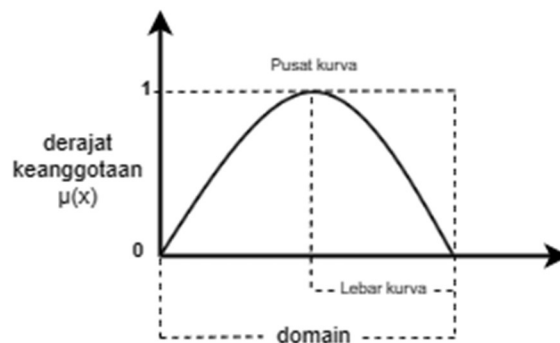
$$\mu(x; c, \sigma) = e^{-\frac{(x-c)^2}{2\sigma^2}}$$

Keterangan:

μ = nilai domain pada pusat kurva

σ = lebar kurva / standar deviasi

Gambar 4 menunjukkan representasi fungsi keanggotaan Gaussian dalam bentuk grafik. Sumbu x berlaku sebagai nilai domain pada pusat kurva dan sumbu y berlaku sebagai derajat keanggotaan $\mu(x)$.



Gambar 4. Grafik fungsi keanggotaan Gaussian.

2.2 Imputasi *K-Nearest Neighbors* (KNN)

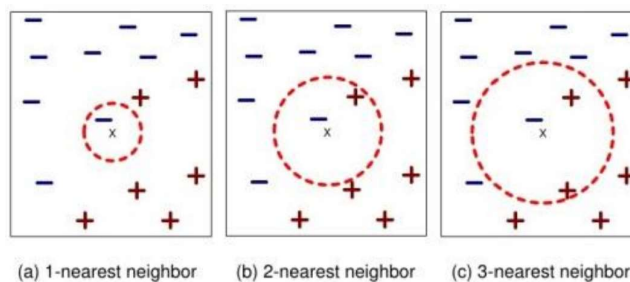
Algoritma KNN telah banyak digunakan untuk mengatasi permasalahan nilai yang hilang terutama pada dataset yang memiliki lebih dari satu fitur nilai yang

hilang (Mawarsari, 2016). Algoritma ini menggunakan observasi yang serupa dengan observasi yang memiliki nilai yang hilang dalam melakukan imputasi data. Observasi yang memiliki nilai yang hilang ini kemudian disebut dengan observasi target. Dari pendekatan yang akan dilakukan, ukuran jarak dihitung menggunakan rumus *euclidean distance* antara observasi target dengan masing-masing observasi lain. Algoritma KNN lebih detail dijelaskan pada tahapan penelitian beserta fungsi keanggotaan fuzzy.

2.3 Klasifikasi *K-Nearest Neighbors* (KNN)

K-Nearest Neighbors (KNN) merupakan suatu metode yang menggunakan algoritma *supervised learning*. Algoritma ini bertujuan untuk menemukan pola baru dalam data dengan mencari hubungan pola data yang telah ada dengan data yang baru (Susanti et al., 2018). KNN merupakan metode yang dapat digunakan untuk klasifikasi berdasarkan data pembelajaran dari sejumlah k tetangga terdekat. Pada klasifikasi data, KNN bekerja dengan menghitung jarak terdekat antara data baru dan data yang sudah diketahui kelasnya menggunakan jarak euclidean.

KNN menggunakan prinsip ketetanggaan (*neighbor*) untuk melakukan prediksi kelas baru (Siringoringo, 2018). Jumlah tetangga yang digunakan sebanyak k tetangga. Nilai k yang terbaik bergantung pada data yang ada. Nilai k yang bagus dapat dipilih dengan optimasi parameter seperti menggunakan *Cross-validation* (Isman et al., 2021). Gambar 5 menunjukkan ilustrasi prinsip ketetanggaan KNN (Siringoringo, 2018).



Gambar 5. Ilustrasi ketetanggaan KNN.

Pada metode KNN, hal pertama yang dilakukan yakni menentukan nilai k tetangga terdekat. Jumlah k paling baik merupakan angka ganjil, seperti $k = 1, 3, 5$, dan seterusnya. Cara kerja metode KNN yakni memproyeksi data training pada ruang berdimensi banyak. Ruang ini kemudian dibagi ke bagian-bagian yang merepresentasikan kriteria data. Setiap data *training* direpresentasi ke dalam titik-titik c pada ruang dimensi banyak tersebut. Selanjutnya data baru yang akan diklasifikasi diproyeksikan pada ruang dimensi banyak yang telah memuat titik-titik c data *training*. Proses klasifikasi ini dilakukan dengan mencari titik c terdekat dari c baru (Isman et al., 2021). Tahapan klasifikasi menggunakan metode KNN (Admojo & Fadhila, 2020) adalah sebagai berikut.

Algoritma klasifikasi <i>K-Nearest Neighbors</i> (KNN)
<ol style="list-style-type: none"> 1. Menentukan parameter k (jumlah tetangga terdekat). 2. Menghitung jarak antara data <i>training</i> dan data <i>testing</i> menggunakan perhitungan jarak Euclidean. 3. Mengurutkan jarak dari terkecil ke terbesar. 4. Menentukan jarak terdekat sampai urutan k. 5. Memasangkan kelas yang bersesuaian. 6. Mencari jumlah kelas dari tetangga terdekat dan tetapkan kelas sebagai kelas data yang akan dievaluasi.

2.4 *Euclidean Distance*

Euclidean distance adalah metode yang digunakan untuk mencari jarak antara dua titik variabel dan merepresentasikan tingkat kemiripan. Semakin kecil jaraknya maka semakin mirip kedua variabel tersebut (Riadi et al., 2020). Jika data baru memiliki jarak minimum dan memiliki kemiripan yang cukup tinggi maka *euclidean distance* dapat dikatakan baik (Fadlil & Saifudin, 2015).

Adapun rumus *Euclidean distance* (Yudhana et al., 2020) dapat ditunjukkan sebagai berikut.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_{training}^i - y_{testing}^i)^2}$$

Keterangan:

- $d(x, y)$ = Jarak
- $x_{training}^i$ = Data *training*
- $y_{testing}^i$ = Data *testing*
- i = Variabel data
- n = Dimensi data

2.5 *Min-Max Normalization*

Min-Max Normalization atau selanjutnya dapat disebut sebagai normalisasi minmax merupakan salah satu metode normalisasi dengan mentransformasi data asli secara linier guna menghasilkan nilai perbandingan yang seimbang antara data saat ini dan data setelah proses (Hanifa et al., 2017). Transformasi ini mengubah data dalam range 0 hingga 1, dengan 0 sebagai nilai terkecil dan 1 sebagai nilai terbesar. Adapun rumus yang digunakan menurut (Hanifa et al., 2017) adalah sebagai berikut.

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{range.max} - \text{range.min}) + \text{range.min}$$

Keterangan:

- v' = Data setelah normalisasi
- v = Data sebelum normalisasi
- \min_A = Nilai minimum pada kolom sebelum normalisasi
- \max_A = Nilai maksimum pada kolom sebelum normalisasi
- range.max = 1
- range.min = 0

2.6 Confusion Matrix

Confusion matrix merupakan salah satu alat populer yang digunakan untuk mengevaluasi performa klasifikasi (Siringoringo, 2018). *Confusion matrix* ini berisi informasi tentang hasil prediksi klasifikasi terhadap data aktual yang dilakukan oleh sistem klasifikasi. *Confusion matrix* digambarkan dengan 4 kombinasi berbeda yang merupakan representasi hasil proses klasifikasi yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Tabel 1 menunjukkan tabel *confusion matrix* yang dapat digunakan untuk menghitung performa sistem klasifikasi.

Tabel 1. *Confusion Matrix*

<i>Class</i>	<i>Predictive Positive</i>	<i>Predictive Negative</i>
<i>Actual Positive</i>	TP	FN
<i>Actual Negative</i>	FP	TN

Nilai *True Positive* (TP) merupakan jumlah kelas positif yang diklasifikasikan dengan tepat dan nilai *True Negative* (TN) merupakan jumlah kelas negatif yang diklasifikasikan dengan tepat. Sedangkan nilai *False Positive* (FP) merupakan jumlah kelas positif yang tidak diklasifikasikan dengan tepat dan nilai *False Negative* (FN) merupakan jumlah kelas negatif yang tidak diklasifikasikan dengan tepat. Berdasarkan *confusion matrix* ini dapat ditentukan kriteria performa seperti *Recall*, *Precision*, *Accuracy*, *specificity*, *F-Measure*, *G-Mean* dan lain sebagainya.

Dalam penelitian ini, performa klasifikasi yang akan digunakan yaitu Akurasi. Perhitungan tingkat akurasi merupakan pengujian yang dilakukan dengan tujuan untuk memperkirakan seberapa tepat hasil klasifikasi pada data (Maricar & Dian Pramana, 2019). Nilai akurasi didapatkan dari jumlah data yang dikategorikan tepat sesuai kategorinya (TP) ditambah dengan jumlah data negatif benar (TN) yang dibagi dengan jumlah semua data. Adapun cara

menghitung akurasi menurut (Ariansyah & Kusmira, 2021) dapat dilihat pada rumus perhitungan akurasi berikut.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

Keterangan:

TP = Data positif yang diklasifikasikan dengan benar

TN = Data negatif yang diklasifikasikan dengan benar

FP = Data positif yang diklasifikasikan dengan salah

FN = Data negatif yang diklasifikasikan dengan salah

2.7 Penelitian Terdahulu

Berikut beberapa penelitian yang telah dilakukan oleh peneliti terdahulu, diantaranya:

2.7.1. *Fuzzy Based Techniques for Handling Missing Values*

Dalam penelitian ini, dilakukan imputasi *missing values* pada enam set data *time series* menggunakan teknik fungsi keanggotaan fuzzy Gaussian dan fungsi keanggotaan *triangular Fuzzy* untuk mengidentifikasi manakah imputasi terbaik untuk menangani *missing values* dimana fungsi keanggotaan fuzzy digunakan untuk mencari bobot nilai data berdasarkan tetangga terdekat. Penelitian ini memiliki tujuan untuk memastikan kelengkapan dimensi data *time series* yang mengalami *missing values*. Dua teknik yang diusulkan tersebut nantinya akan dibandingkan dengan teknik tradisional yakni menggunakan nilai rata-rata dan rata-rata tertimbang. Keakuratan teknik ini kemudian dievaluasi menggunakan tiga metode klasifikasi tradisional yaitu *Neural Network*, *Naïve Bayes*, dan *Decision Tree*. Hasil dari evaluasi menunjukkan bahwa kedua teknik yang diusulkan

memiliki tingkat akurasi yang lebih tinggi dibanding dengan teknik imputasi tradisional karena perilakunya dalam menangani *missing values* sebagaimana mestinya. Hasil evaluasi ini disajikan dalam bentuk diagram per masing-masing dataset. Dalam penelitian ini pula dijelaskan bahwa fungsi keanggotaan *triangular fuzzy* memiliki hasil imputasi lebih baik dari fungsi keanggotaan fuzzy Gaussian (El-Bakry et al., 2021).

2.7.2. Perbandingan Performa antara Imputasi Metode Konvensional dan Imputasi dengan Algoritma *Mutual Nearest Neighbor*

Dalam penelitian ini, dilakukan imputasi nilai yang hilang menggunakan algoritma *Mutual Nearest Neighbor* (MNN). Algoritma ini merupakan algoritma klasifikasi hasil pengembangan dari algoritma *K-Nearest Neighbor* (KNN). Algoritma MNN bekerja dengan menggunakan tetangga mutual terdekat dari suatu *instance* untuk mendapatkan label kelas dari *instance* tersebut. Adapun dataset yang digunakan terdiri dari dataset *wisconsin prognostic breast cancer (wpbc)*, *large soybean database (soybean)*, *1984 United States Congressional voting records database (voting)*, *credit approval (credit)*, *cleveland heart disease database (cleveland)* dan *hepatitis domain (hepatitis)* yang mana keenam data tersebut diperoleh dari *UCI Machine Learning Repository*. Tujuan dari penelitian ini adalah membandingkan akurasi antara algoritma MNN dengan metode imputasi konvensional. Metode imputasi konvensional yang digunakan adalah imputasi dengan nilai *mean* untuk tipe data atribut kontinu dan nilai *modus* untuk tipe data atribut kategorikal.

Imputasi dengan algoritma MNN diawali dengan memasukkan nilai *k*. Selanjutnya apabila ditemukan nilai hilang pada *instance x*, maka dilakukan pencarian *k* tetangga mutual dari *instance x* dengan menghitung jarak antara *instance x* dengan *instance* lain. Dari *k*

tetangga terdekat yang diperoleh, misal $y(x)$, maka dicari k tetangga terdekat dari masing-masing *instance* dalam $y(x)$. Jika x adalah salah satu dari k tetangga terdekat *instance* tersebut, *instance* dimasukkan ke dalam himpunan tetangga mutual x , misal $M(x)$. Kemudian menghitung rata-rata nilai data atribut yang memiliki *missing value* untuk data kontinu dan *modus* untuk data kategorikal dari *instance-instance* dalam $M(x)$. Jika *instance* x tidak punya tetangga mutual maka imputasikan hasil dari metode konvensional ke *missing value*. Untuk mendapatkan performa metode imputasi, dataset hasil imputasi diklasifikasikan menggunakan algoritma KNN, MNN, RIPPER, pohon keputusan C4.5, *Naïve Bayes Classifier* (NBC), dan *Support Vector Machine* (SVM). Uji coba dilakukan dengan *k-fold cross validation* dengan nilai $k=10$. Hasil yang diperoleh dalam penelitian ini yaitu imputasi dengan algoritma MNN mempunyai performa lebih baik daripada metode konvensional. Pada hasil uji coba, rata-rata metode imputasi MNN mempunyai nilai akurasi yang lebih besar daripada metode imputasi konvensional. Hal ini terjadi pada uji coba terhadap dataset *wdbc*, *voting*, *credit*, *Cleveland*, dan *hepatitis*. Sedangkan, pada uji coba terhadap dataset *soybean* memiliki nilai rata-rata akurasi cenderung berimbang (Alfarisi et al., 2013).

2.7.3. *K Nearest Neighbor* dalam Imputasi *Missing Data*

Dalam penelitian ini, dilakukan imputasi *missing data* menggunakan metode *K Nearest Neighbor* (KNN) dan NN untuk mengatasi permasalahan *missing data*. Tujuan dari penelitian ini yaitu membandingkan hasil imputasi mana yang lebih baik dalam mengatasi *missing data*. Data yang digunakan adalah data iris *versicolour* yang diambil dari *UCI Machine Learning Repository*. Data ini terdiri dari 50 observasi dengan 4 fitur yaitu v_1 , v_2 , v_3 , dan v_4 . Penelitian ini terlebih dahulu melakukan simulasi *missing data* dengan mekanisme MAR dan pola *missing data univariate*. Scenario *missing data* terjadi di fitur v_1

dimana persentase *missing data* yang digunakan yakni 10%, 20%, dan 30%. Setiap persentase pada *missing data* ini dilakukan 10 kali replikasi. Jumlah k yang digunakan yaitu $k=1$, $k=5$, $k=10$, $k=15$, dan $k=20$. Setelah simulasi *missing data* dilakukan, kemudian menghitung jarak antara data yang memiliki *missing data* dengan data lengkap. Kemudian mengurutkan nilai jarak dari terbesar ke terkecil dan lakukan imputasi menggunakan rumus *weight mean estimation* berdasarkan jumlah k yang digunakan. Setelah proses imputasi dilakukan, hasil imputasi *missing data* dievaluasi dengan menghitung nilai MSE dan MAPE untuk setiap persentase *missing*. Hasil dari penelitian ini disimpulkan bahwa imputasi *missing data* menggunakan KNN lebih baik daripada menggunakan NN karena nilai rata-rata MSE dan MAPE dari KNN lebih kecil dibandingkan NN. Besaran persentase *missing data* disini mempengaruhi jumlah k terbaik untuk imputasi. Semakin besar persentasenya maka semakin besar pula nilai rata-rata MSE dan MAPE untuk setiap k . Berdasarkan nilai rata-rata MSE dan MAPE ini pula dari 10 replikasi, KNN terbaik terjadi pada saat $k=10$ pada *missing data* 10% dan 20%, sedangkan untuk *missing data* 30% terjadi pada saat $k=15$ (Susanti et al., 2018).

III. METODE PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada Januari 2023 sampai Juni 2023 bertempat di Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Tahapan Penelitian

Tahapan pertama yang dilakukan yaitu pengumpulan data. Data dikumpulkan dengan metode pengumpulan data sekunder dari portal *UCI Machine Learning Repository*. Kemudian dataset yang lengkap akan dihilangkan beberapa nilainya secara acak dengan persentase 5% dan 10%. Selanjutnya identifikasi berapa banyak nilai yang hilang untuk setiap fitur. Metode yang diusulkan untuk mengatasi permasalahan nilai yang hilang ini dimulai dengan menentukan nilai k titik data tetangga terdekat untuk setiap nilai yang hilang pada observasi tertentu. Nilai k yang digunakan pada penelitian ini yaitu 3, 5, 7, dan 9. Setelah itu menghitung jarak untuk nilai yang hilang berdasarkan nilai observasi lain. Jarak yang telah diperoleh nantinya akan diurutkan dan dicari jarak minimum. Nilai data pada jarak minimum tersebut digunakan pada fungsi keanggotaan fuzzy untuk mencari bobotnya. Fungsi keanggotaan fuzzy yang digunakan pada penelitian ini yaitu fungsi keanggotaan fuzzy segitiga dan fungsi keanggotaan fuzzy Gaussian. Algoritma 1 menunjukkan detail algoritma menggunakan metode KNN dan fungsi keanggotaan fuzzy segitiga.

Algoritma 1: Imputasi KNN dan fungsi keanggotaan fuzzy segitiga

1. Menentukan nilai k, yaitu jumlah observasi terdekat.

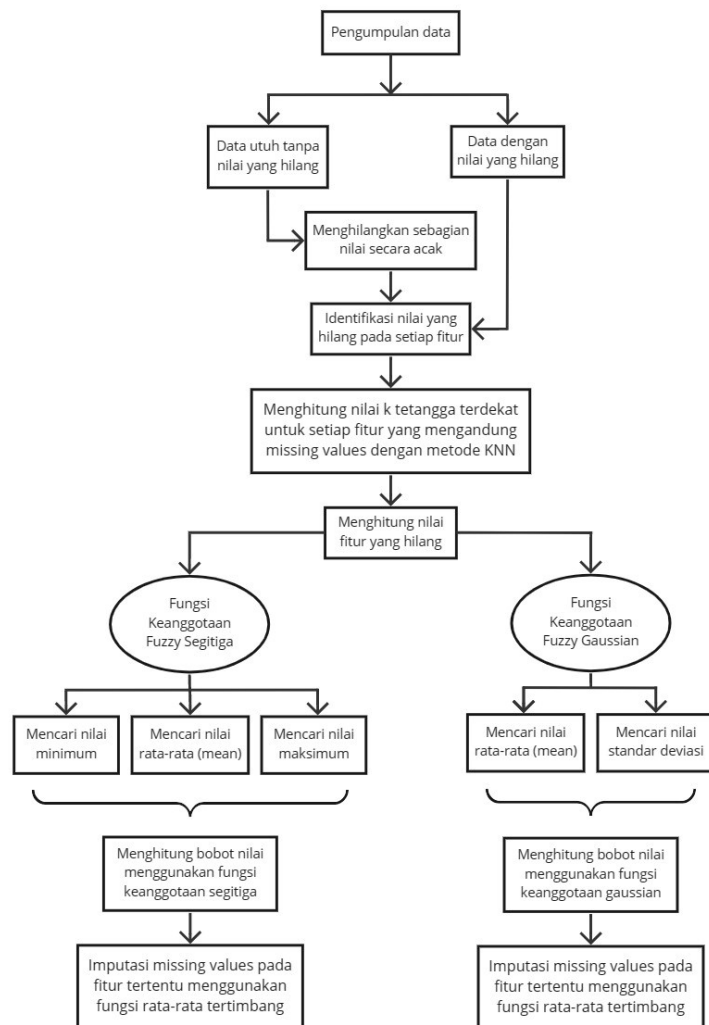
2. Menghitung jarak terdekat antara observasi target dengan observasi yang tidak memiliki nilai yang hilang.
 3. Mencari nilai observasi terdekat dengan jarak minimum.
 4. Menghitung nilai minimum (nilai observasi tetangga terdekat).
 5. Menghitung nilai maksimum (nilai observasi tetangga terdekat).
 6. Menghitung nilai rata-rata (nilai observasi tetangga terdekat).
 7. Menghitung bobot nilai menggunakan fungsi keanggotaan fuzzy segitiga.
 8. Menghitung nilai rata-rata tertimbang dari bobot dan nilai tetangga terdekat.
- $$\frac{\sum_{a \in A} \text{Nilai Tetangga Terdekat } (a) \text{ Bobot nilai fuzzy segitiga } (a)}{\sum_{a \in A} \text{Bobot nilai fuzzy segitiga } (a)}$$
9. Selesai.

Fungsi keanggotaan segitiga memiliki banyak keunggulan seperti mudah untuk diimplementasikan, lebih nyaman, respon cepat, dan perhitungan cepat (Sadollah, 2018). Fungsi keanggotaan ini bekerja dengan mencari nilai minimum, maksimum, dan nilai rata-rata. Sedangkan fungsi keanggotaan Gaussian bekerja dengan mencari nilai rata-rata dan standar deviasi. Imputasi nilai yang hilang ini dihitung menggunakan rata-rata tertimbang antara nilai tetangga terdekat dan bobotnya. Nilai yang telah didapat untuk masing-masing kolom nilai yang hilang dimasukkan berdasarkan fitur yang ada. Algoritma 2 menunjukkan detail algoritma menggunakan metode KNN dan fungsi keanggotaan Gaussian.

- | Algoritma 2: Imputasi KNN dan fungsi keanggotaan fuzzy Gaussian |
|--|
| <ol style="list-style-type: none"> 1. Menentukan nilai k, yaitu jumlah observasi terdekat. 2. Menghitung jarak terdekat antara observasi target dengan observasi yang tidak memiliki nilai yang hilang. 3. Mencari nilai observasi terdekat dengan jarak minimum. 4. Menghitung nilai rata-rata (nilai observasi tetangga terdekat). 5. Menghitung standar deviasi (nilai observasi tetangga terdekat). |

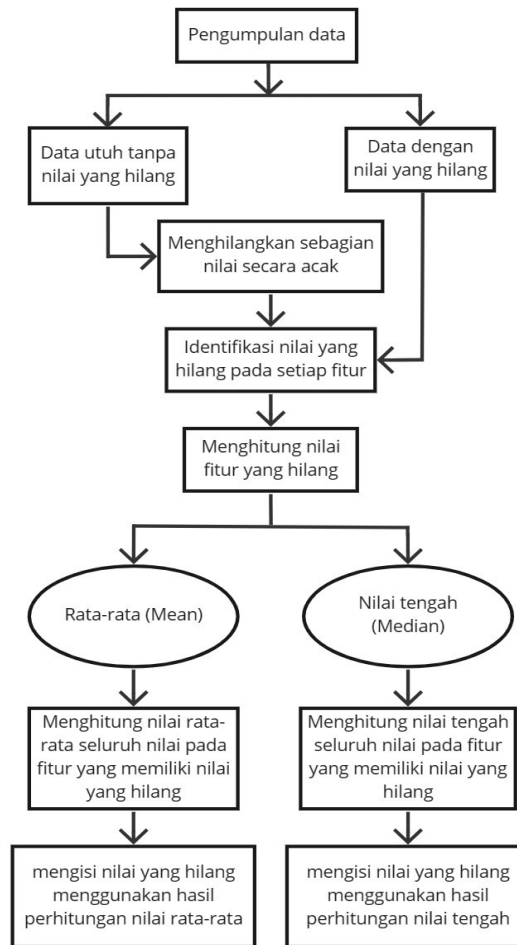
6. Menghitung bobot nilai menggunakan fungsi keanggotaan fuzzy Gaussian.
 7. Menghitung nilai rata-rata tertimbang dari bobot dan nilai tetangga terdekat.
- $$\frac{\sum_{a \in A} \text{Nilai Tetangga Terdekat } (a) \text{ Bobot nilai fuzzy gauss } (a)}{\sum_{a \in A} \text{Bobot nilai fuzzy gauss } (a)}$$
8. Selesai.

Selanjutnya imputasi menggunakan teknik konvensional dilakukan dengan menghitung nilai rata-rata dan nilai tengah untuk setiap fitur yang memiliki nilai hilang. Nilai yang didapat kemudian pula dimasukkan ke masing-masing fitur.



Gambar 6. Tahapan perhitungan nilai imputasi metode KNN dan fuzzy.

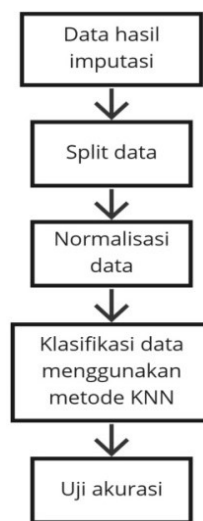
Gambar 6 menunjukkan tahapan perhitungan nilai yang hilang menggunakan metode KNN dan fungsi keanggotaan fuzzy. Metode KNN bekerja dengan mencari jarak terdekat berdasarkan k yang ditentukan untuk memprediksi nilai yang hilang. Sedangkan fungsi keanggotaan fuzzy bekerja untuk mencari bobot nilai berdasarkan tetangga terdekat. Imputasi dilakukan dengan menghitung nilai yang hilang dengan fungsi rata-rata tertimbang.



Gambar 7. Tahapan perhitungan nilai imputasi metode konvensional.

Gambar 7 menunjukkan tahapan perhitungan nilai yang hilang menggunakan metode konvensional. Metode konvensional bekerja dengan mencari nilai rata-rata dan nilai tengah pada masing-masing fitur. Imputasi nilai yang hilang dilakukan dengan mengganti nilai yang hilang pada masing-masing fitur menggunakan nilai rata-rata dan nilai tengah yang telah diperoleh.

Setelah nilai data *missing values* diperoleh, tahapan selanjutnya yaitu klasifikasi menggunakan metode KNN. Pada tahapan ini dilakukan pembagian data menjadi data *training* dan data *testing* dengan persentase 75% untuk data *training* dan 25% untuk data *testing*. Data kemudian di normalisasi menggunakan teknik minmax dan diklasifikasi menggunakan metode KNN. Tahap terakhir yakni uji akurasi untuk memperoleh hasil perbandingan antara teknik yang diusulkan dan teknik tradisional. Gambar 8 menunjukkan tahapan klasifikasi data dan uji akurasi terhadap masing-masing teknik imputasi yang telah diperoleh *missing values nya*.



Gambar 8. Tahapan klasifikasi dan uji akurasi untuk setiap teknik imputasi.

3.3 Alat dan Bahan

Alat dan bahan yang diperlukan dalam penelitian ini sebagai berikut.

3.3.1 Alat Penelitian

a. Perangkat Keras

Perangkat keras yang digunakan adalah laptop, dengan spesifikasi sebagai berikut.

- Processor : Intel® Core™ i5-4210M CPU @ 2.60GHz
- RAM : 8 GB

- System Type: *64-bit operating system, x64-based processor*

b. Perangkat Lunak

Perangkat lunak yang digunakan sebagai berikut.

- Sistem Operasi *Windows 10 Pro 64 bit*.
- *Jupyter Notebook (Anaconda)* sebagai alat untuk melakukan imputasi dan klasifikasi.
- *Microsoft Excel 2019* sebagai alat untuk menyimpan dataset.

3.3.2 Bahan Penelitian

Bahan yang digunakan dalam penelitian ini terfokus pada jenis data klasifikasi yang diperoleh dari *UCI Machine Learning Repository*. Total keseluruhan data yang digunakan yaitu 5.253 jumlah observasi dengan 5 dataset. Tabel 2 menunjukkan rincian dataset yang digunakan beserta persentase nilai hilangnya.

Tabel 2. Rincian persentase nilai hilang pada masing-masing dataset

No.	Nama Dataset	Jumlah Fitur	Jumlah Observasi	Persentase Missing
1	<i>Hepatitis Dataset</i>	20	155	5.39%
2	<i>Ozone Level Detection Dataset</i>	73	2.534	8.07%
3	<i>Blood Transfusion Service Center Dataset</i>	5	748	5% dan 10%
4	<i>Parkinson Speech Dataset with Multiple Types of Sound Recordings</i>	26	1.040	5% dan 10%
5	<i>Audit Risk Dataset</i>	26	776	5% dan 10%

a. *Hepatitis Dataset*

Hepatitis dataset merupakan dataset hepatitis yang mengandung kumpulan data klasifikasi penyakit hepatitis yang memiliki *missing values*. Dataset ini memiliki 20 fitur dan 155 jumlah observasi. Dataset didonasikan pada 01-11-1988 oleh Gail Gong dari *Carnegie-Mellon University*, Yugoslavia (Hepatitis, 1988). Adapun 20 fitur yang terdapat dalam dataset ini yaitu *class, age, sex, steroid, antivirals, fatigue, malaise, anorexia, liver big, liver firm, spleen palpable, spiders, ascites, varices, bilirubin, alk phosphate, sgot, albumin, protime, dan histology*. Persentase nilai yang hilang dalam dataset ini yaitu 5.39%.

b. *Ozone Level Detection Dataset*

Ozone level detection dataset merupakan dataset deteksi tingkat ozon dengan set puncak 8 jam. Dataset ini mengandung *missing values* dengan persentase *missing* sebesar 8.07%. Data yang ada dikumpulkan dari tahun 1998 hingga 2004 di daerah Houston, Galveston dan Brazoria. Dataset didonasikan oleh Kun Zhang, Wei Fan, dan XiaoJing Yuan (Zhang et al., 2008). Dataset memiliki 73 fitur dengan 2.534 jumlah observasi. Adapun 73 fitur tersebut antara lain WSR0-WSR23, T0-T23, WSR_PK, WSR_AV, T_PK, T_AV, T85, RH85, U85, V85, HT85, T70, RH70, U70, V70, HT70, T50, RH50, U50, V50, HT50, KI, TT, SLP, SLP_, *Precp*, dan *Class*. Fitur kelas menunjukkan keadaan perhari apakah merupakan hari ozon atau hari yang normal. Semua fitur yang dimulai dengan huruf T menunjukkan suhu diukur pada waktu yang berbeda sepanjang hari, WS menunjukkan kecepatan angin pada berbagai waktu, RH menunjukkan kelembaban relatif, U menunjukkan angin arah timur-barat pada 850 hpa, V menunjukkan angin arah utara-selatan pada 850 hpa, HT menunjukkan tinggi geopotensial pada 850 hpa, KI menunjukkan K indeks, TT menunjukkan T-Total, SLP menunjukkan tekanan permukaan laut, dan *Precp* menunjukkan *precipitation* atau cuaca.

c. *Blood Transfusion Service Center Dataset*

Pusat layanan transfusi darah merupakan pusat layanan untuk transfusi darah yang berada di Kota Hsin-Chu, Taiwan. Dataset ini merupakan kumpulan data yang diambil dari pusat layanan transfusi darah Hsin-Chu guna mendemonstrasikan model pemasaran RFMTC (versi modifikasi dari RFM). Darah yang disumbangkan dikumpulkan setiap tiga bulan ke universitas di kota Hsin-Chu. Untuk membangun model FRMTC, mereka memilih 748 donor secara acak dari database donor. Keseluruhan data donor yang diambil ini masing-masing termasuk R (*Recency* – bulan sejak donasi terakhir), F (Frekuensi – jumlah total donasi), M (*Monetary* – total darah yang disumbangkan dalam cc), T (*Time* – bulan sejak donasi pertama), dan variabel biner. Variabel biner ini mewakili apakah pendonor menyumbangkan darah pada Maret 2007 (1 untuk menyumbangkan darah; 0 untuk tidak menyumbangkan darah). Kemudian dataset ini tergolong ke dalam dataset lengkap tanpa *missing values* dengan 5 fitur dan 748 jumlah observasi. Dataset ini didonasikan pada 02-10-2008 oleh Prof. I-Cheng Yeh (Yeh, 2008).

d. *Parkinson Speech Dataset with Multiple Types of Sound Recordings*

Parkinson Speech Dataset merupakan dataset yang diambil dari rekaman suara milik 20 pasien dengan penyakit *Parkinson's Disease* (PD) dan 20 individu sehat. Beberapa jenis rekaman suara (26 sampel) yang diambil termasuk vokal berkelanjutan, angka, kata-kata dan kalimat pendek. 20 pasien PD ini diantaranya 6 wanita dan 14 pria sedangkan 20 individu sehat diantaranya 10 wanita dan 10 dengan sebelumnya telah mengajukan banding di Departemen Neurologi di Fakultas Kedokteran Cerrahpasa, Universitas Istanbul. Seluruh sampel yang diambil berbasis frekuensi linier dan waktu diekstraksi dari setiap sampel suara. Skor UPDRS (*Unified Parkinson's Disease Rating Scale*) dari setiap pasien yang ditentukan oleh dokter ahli juga tersedia pada dataset ini. Dataset ini didonasikan pada 06-11-2014 dengan jumlah observasi yaitu 1.040 observasi (Kursun et al., 2014).

e. Audit Risk Dataset

Audit risk dataset merupakan dataset non-rahasia dari suatu perusahaan Auditor India yang dikumpulkan secara lengkap pada tahun 2015 hingga 2016 guna membantu auditor dalam membangun model klasifikasi yang dapat memprediksi perusahaan penipuan berdasarkan faktor resiko saat ini dan historis. Dataset ini memiliki 26 fitur dan 776 jumlah observasi. Dataset didonasikan pada 14-07-2018 oleh Nishtha Hooda, CSED, TIET, dan Patiala (Hooda et al., 2018). Adapun 26 fitur yang ada mencakup faktor resiko yang diperiksa dari berbagai bidang seperti catatan masa lalu kantor audit, audit-paras, laporan kondisi lingkungan, ringkasan reputasi perusahaan, laporan masalah yang sedang berlangsung, catatan nilai laba, catatan nilai kerugian, laporan tindak lanjut, dan lainnya. Setelah wawancara mendalam dengan auditor, faktor-faktor risiko penting dievaluasi dan probabilitas keberadaannya dihitung dari catatan masa kini dan masa lalu.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini memiliki hasil penelitian yang dapat disimpulkan sebagai berikut.

1. Imputasi *missing values* menggunakan metode KNN dan fungsi keanggotaan fuzzy berhasil dilakukan menggunakan Bahasa pemrograman python dengan *library Pandas, NumPy, sklearn, SciPy, scikit-learn*, dan *library matplotlib*.
2. Klasifikasi data hasil imputasi berhasil dilakukan menggunakan metode klasifikasi KNN dimana k yang digunakan yaitu 3, 5, 7, dan 9.
3. Hasil penelitian pada keseluruhan dataset diperoleh bahwa imputasi dengan fungsi keanggotaan segitiga mencapai rata-rata akurasi antara 91.72% sampai 92.07% dengan maksimum rata-rata akurasi berada pada $k=7$ dan imputasi dengan fungsi keanggotaan Gaussian antara 89.90% sampai 90.28% dengan maksimum rata-rata berada pada $k=9$. Sedangkan imputasi konvensional mencapai rata-rata akurasi untuk imputasi mean sebesar 82.24%, imputasi median sebesar 82.23%, dan rata-rata akurasi data asli sebesar 76.02%.
4. Berdasarkan perbandingan dataset yang sama dengan penelitian terdahulu, metode imputasi KNN dan fungsi keanggotaan fuzzy memberikan hasil akurasi yang lebih tinggi pada klasifikasi KNN dibandingkan dengan klasifikasi *Decision Tree, Neural Network*, dan *Naïve Bayes* yang sebelumnya telah dilakukan pada penelitian terdahulu. Hal ini menunjukkan bahwa metode imputasi ini dapat digunakan untuk menangani permasalahan *missing values* karena dapat menyesuaikan dan

memberikan hasil akurasi yang baik pada metode klasifikasi yang berbeda.

5. Secara keseluruhan, imputasi dengan fungsi keanggotaan fuzzy memiliki nilai akurasi lebih tinggi dibanding dengan imputasi konvensional. Dimana fungsi keanggotaan fuzzy segitiga memiliki nilai rata-rata akurasi yang lebih tinggi dibandingkan dengan fungsi keanggotaan fuzzy Gaussian. Sehingga, fungsi keanggotaan fuzzy dapat lebih baik untuk menangani *missing values* dibanding imputasi konvensional dalam meningkatkan akurasi.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, diperoleh beberapa saran untuk menjadi bahan pertimbangan penelitian masa mendatang diantaranya:

1. Data kelas serta nilai k untuk imputasi dan klasifikasi dapat diperluas.
2. Menggunakan metode konvensional lain seperti modus.
3. Pembagian opsi data latih yang lebih banyak dan menggunakan metode klasifikasi yang berbeda.
4. Imputasi dilakukan dengan menggunakan perhitungan *distance* yang lain.

DAFTAR PUSTAKA

DAFTAR PUSTAKA

- Admojo, T., & Fadhila. (2020). Klasifikasi Aroma Alkohol Menggunakan Metode KNN. *Indonesian Journal of Data an Science*, 1(2), 34–38.
- Alfarisi, A. R., Tjandrasa, H., & Arieshanti, I. (2013). Perbandingan Performa antara Imputasi Metode Konvensional dan Imputasi dengan Algoritma Mutual Nearest Neighbor. *JURNAL TEKNIK POMITS*, 2(1), A73–A76.
- Ariansyah, & Kusmira, M. (2021). Analisis Sentimen Pengaruh Pembelajaran Daring Terhadap Motivasi Belajar Di Masa Pandemi Menggunakan Naive Bayes Dan Svm. *Faktor Exacta*, 14(3), 100.
<https://doi.org/10.30998/faktorexacta.v14i3.10325>
- Davey, A., & Savla, J. “Tina.” (2009). A Structural Equation Modeling Approach. In *Statistical Power Analysis with Missing Data* (1st Editio, p. 384). <https://doi.org/https://doi.org/10.4324/9780203866955>
- El-Bakry, M., El-Kilany, A., Mazen, S., & Ali, F. (2021). Fuzzy based Techniques for Handling Missing Values. *International Journal of Advanced Computer Science and Applications*, 12(3), 50–55.
<https://doi.org/10.14569/IJACSA.2021.0120306>
- Fadlil, A., & Saifudin. (2015). Sistem Identifikasi Citra Kayu Berdasarkan Tekstur Menggunakan Gray Level Co-occurrence Matrix (GLCM) dengan Klasifikasi Jarak Euclidean. *Sinergi*, 19, 181–186.
- Hanifa, T., Tesha, & Al-Faraby, S. (2017). Analisis Churn Prediction pada Data Pelanggan PT. Telekomunikasi dengan Logistic Regression dan Underbagging. *E-Proceeding of Engineering*, 4(2), 3210–3225.
- Hepatitis. (1988). *UCI Machine Learning Repository*.
<https://doi.org/https://doi.org/10.24432/C5Q59J>
- Hooda, N., Bawa, S., & Rana, P. S. (2018). Fraudulent Firm Classification: A Case Study of an External Audit. *Applied Artificial Intelligence*, 32(1), 48–64. <https://doi.org/10.1080/08839514.2018.1451032>
- Isman, Andani Ahmad, & Abdul Latief. (2021). Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(3), 557–564.
<https://doi.org/10.29207/resti.v5i3.3006>

- Kaiser, J. (2014). Dealing with Missing Values in Data. *Journal of Systems Integration*, 42–51. <https://doi.org/10.20470/jsi.v5i1.178>
- Kursun, O., Sakar, B., Isenkul, M., Sakar, C., Sertbas, A., & Gurgun, F. (2014). *Parkinson Speech Dataset with Multiple Types of Sound Recordings*. UCI Machine Learning Repository. <https://doi.org/https://doi.org/10.24432/C5NC8M>
- Maricar, M. A., & Dian Pramana. (2019). Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali. *Jurnal Sistem Dan Informatika (JSI)*, 14(1), 16–22. <https://doi.org/10.30864/jsi.v14i1.233>
- Maula, R. A., Gunawan, A. I., Bayu Dewantara, B. S., Al Rasyid, M. U. H., Setiawardhana, S., Saputra, F. A., & Ispianto, J. (2022). Handling Missing Value dengan Pendekatan Regresi pada Dataset Akuakultur Berukuran Kecil. *Jurnal Rekayasa Elektrika*, 18(3), 175–184. <https://doi.org/10.17529/jre.v18i3.25903>
- Mawarsari, U. (2016). Imputasi Missing Data Dengan K-Nearest Neighbor Dan algoritma Genetika. *AdMathEdu : Jurnal Ilmiah Pendidikan Matematika, Ilmu Matematika Dan Matematika Terapan*, 6(1). <https://doi.org/10.12928/admathedu.v6i1.4764>
- Nainggolan, J. M. (2016). *Logika Fuzzy (Fuzzy Logic): Teori dan Penerapan Pada Sistem Daya (Kajian Pengaruh Induksi Medan Magnet)*. 23. <http://member.unila.ac.id/~ftelektro/lab/ltp/dokumen/Fuzzy Logic Paper.doc>
- Pratama, I., Permanasari, A. E., Ardiyanto, I., & Indrayani, R. (2016). A review of missing values handling methods on time-series data. *2016 International Conference on Information Technology Systems and Innovation (ICITSI)*, 1–6. <https://doi.org/10.1109/ICITSI.2016.7858189>
- Purba, D. (2016). *ANALISA PENINGKATAN SOFTSKILL DAN HARDSKILL MAHASISWA MELALUI PEMBELAJARAN SCL DAN RC DENGAN MACHINE LEARNING*. 01(2), 19–24.
- Riadi, I., Fadlil, A., & Annisa, P. (2020). Identifikasi Tulisan Tangan Huruf Katakana Jepang Dengan Metode Euclidean. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 4(1), 29. <https://doi.org/10.30645/j-sakti.v4i1.184>
- Sadollah, A. (2018). Bab pengantar: fungsi keanggotaan mana yang sesuai dalam sistem fuzzy? In *Logika Fuzzy berdasarkan metode optimasi dan sistem kontrol serta aplikasinya*. IntechOpen.

- Saelan, A. (2009). Logika Fuzzy. *Makalah If2091 Struktur Diskrit Tahun 2009*, 1(13508029), 1–5.
- Siringoringo, R. (2018). KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN k-NEAREST NEIGHBOR. *Jurnal ISD*, 3(1), 44–49.
- Susanti, Martha, S., & Sulistianingsih INTISARI, E. (2018). K Nearest Neighbor Dalam Imputasi Missing Data. *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, 07(1), 9–14. <http://archive.ics.uci.edu/ml/datas/Iris>.
- Wardhani, L. K., & Haerani, E. (2011). Analisis Pengaruh Pemilihan Fuzzy Membership Function Terhadap Output Sebuah Sistem Fuzzy Logic. *Sntiki Iii*, 326–333.
- Yeh, I.-C. (2008). *Blood Transfusion Service Center*. UCI Machine Learning Repository. <https://doi.org/https://doi.org/10.24432/C5GS39>
- Yudhana, A., Sunardi, S., & Hartanta, A. J. S. (2020). Algoritma K-Nn Dengan Euclidean Distance Untuk Prediksi Hasil Penggajian Kayu Sengon. *Transmisi*, 22(4), 123–129. <https://doi.org/10.14710/transmisi.22.4.123-129>
- Zhang, K., Fan, W., & Yuan, X. (2008). *Ozone Level Detection*. UCI Machine Learning Repository. <https://doi.org/https://doi.org/10.24432/C5NG6W>