

Perbandingan Metode Pengukuran Jarak pada Algoritma Klasifikasi *k-Nearest Neighbor* dengan Dataset *Titanic*

*Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika*



Oleh :

Rizqi Septian Dwinanda
NIM : 09021281520109

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2021

LEMBAR PENGESAHAN TUGAS AKHIR

*Perbandingan Metode Pengukuran Jarak pada Algoritma
Klasifikasi k-Nearest Neighbor dengan Dataset Titanic*

Oleh :

Rizqi Septian Dwinanda
NIM : 09021281520109

Indralaya, Januari 2020

Pembimbing I,

Pembimbing II,



Dian Palupi Rini, M.Kom., Ph.D.
NIP. 197802232006042002



Kanda Januar Miraswan, M.T
NIP. 199001092019031012

Mengetahui,
Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Kamis tanggal 31 Desember 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Rizqi Septian Dwinanda

NIM : 09021281520109

Judul : Perbandingan Metode Pengukuran Jarak pada Algoritma Klasifikasi *k-Nearest Neighbor* dengan Dataset *Titanic*.

1. Pembimbing I.

Dian Palupi Rini, M.Kom., Ph.D
NIP. 197802232006042002



2. Pembimbing II.

Kanda Januar Miraswan, M.T
NIP. 199001092019031012



3. Penguji I.

M. Fachrurrozi, M.T.
NIP 1980052220081221002




4. Penguji II.

M Ourhanul Rizqie, S.Kom. M.T
NIP 1671060312870008



Mengetahui,
Ketua Jurusan Teknik Informatika




Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Rizqi Septian Dwinanda
NIM : 09021281520109
Program Studi : Teknik Informatika
Judul Skripsi : Perbandingan Metode Pengukuran Jarak pada Algoritma
Klasifikasi k-Nearest Neighbor dengan Dataset Titanic.

Hasil Pengecekan Software *iThenticate/Turnitin* : 10 %

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, Januari 2020



Rizqi Septian Dwinanda
NIM. 09021281520109

Motto :

- *Do it. If it's doesn't work, just leave*

Kupersembahkan karya tulis ini kepada :

- *Orang tuaku tercinta*
- *Saudaraku*
- *Sahabat dan teman seperjuanganku*
- *Jurusan Teknik Informatika*
- *Fakultas Ilmu Komputer*
- *Universitas Sriwijaya*

COMPARSION OF DISTANCE MEASUREMENT METHODS ON K-NEARSET NEIGHBOR CLASSIFICATION ALGORITHM WITH DATASET TITANIC

By :

Rizqi Septian Dwinanda

NIM : 09021281520109

ABSTRACT

K-Nearest Neighbor algorithm is a classification algorithm that can be used to classify a data with good result. one of them is to classify the titanic dataset. The quality of the classification result of the k - Nearest Neighbor is very dependent on the distance between object and value of k specified, so the selection of method for distance measurement determines the result of classification.in this research a comparison of several methods of measuring distances, including Manhattan distance, Euclidean distance and Chebyshev distance were examined to see distance measurement methods that can be used optimally on the k - Nearest Neighbor algorithm with the predefined titanic dataset. This study produces a classification value with the highest accuracy in the Chebyshev distance method with an average accuracy of 58.89%. Meanwhile, for the measurement of the distance, the Manhattan distance with an average value of 54.60% and the Euclidean distance with an average value of 52.95%.

Keyword : Classification, Distance measurement, Manhattan Distance, Euclidean Distance, Chebyshev Distance.

Perbandingan Metode Pengukuran Jarak pada Algoritma Klasifikasi k-Nearest Neighbor dengan Dataset Titanic

Oleh :

Rizqi Septian Dwinanda

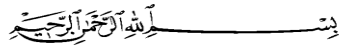
NIM : 09021281520109

ABSTRAK

Algoritma k – Nearest Neighbor merupakan salah satu algoritma klasifikasi yang dapat digunakan untuk mengklasifikasi suatu data dengan hasil yang cukup baik, salah satunya untuk mengklasifikasi dataset *titanic*. Kualitas hasil keluaran klasifikasi algoritma k – Nearest Neighbor sangat bergantung pada jarak kedekatan antar objek dan nilai dari k yang ditetapkan, maka pemilihan metode untuk perhitungan jarak akan menentukan hasil dari pengklasifikasian. Pada penelitian ini, dilakukan perbandingan beberapa metode pengukuran jarak antara lain Manhattan distance, Euclidean distance, dan Chebyshev distance untuk melihat metode pengukuran jarak yang dapat bekerja dengan optimal pada algoritma k-Nearest Neighbor dengan dataset *titanic* yang sudah ditetapkan sebelumnya. Penelitian ini menghasilkan nilai klasifikasi dengan akurasi tertinggi pada metode Chebyshev distance yaitu dengan rata-rata akurasi sebesar 58,89% sedangkan untuk pengukuran jarak Manhattan distance dengan rata-rata sebesar 54,60% dan Euclidean distance dengan nilai rata-rata 52,95%.

Kata kunci : Klasifikasi, Pengukuran Jarak, Manhattan Distance, Euclidean Distance, Chebyshev Distance.

KATA PENGANTAR



Puji syukur kepada Semesta atas berkat dan rahmat-Nya yang telah diberikan kepada penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan Pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Orang tuaku, yang selalu mendokan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Ibu Alvi Syahrini Utami, M.Kom selaku Ketua Jurusan Teknik Informatika dan Ibu Hardini Novianti, M.T selaku Sekretaris Jurusan Teknik Informatika.
3. Ibu Dian Palupi Rini, M.Kom., Ph.D selaku dosen pembimbing I dan Bapak Kanda Januar Miraswan, M.T selaku pembimbing II, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
4. Bapak Kanda Januar Miraswan, M.T selaku dosen pembimbing akademik, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Bapak M. Fachrurrozi, M.T. selaku dosen penguji I, dan Bapak M Qurhanul Rizqie, S.Kom, M.T selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Kak Ricy serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Teman – teman Ghibah Pagi serta seluruh teman-teman jurusan Teknik Informatika yang telah saling berbagi selama masa perkuliahan ini.
9. BPH HMIF yang telah memberikan ruang bagi Penulis untuk berprestasi dan berkarya.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, Januari 2021

Penulis

DAFTAR ISI

Halaman

HALAMAN JUDUL	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xix

BAB I PENDAHULUAN

1.1	Pendahuluan.....	I-1
1.2	Latar Belakang Masalah.....	I-1
1.3	Rumusan Masalah.....	I-4
1.4	Tujuan Penelitian.....	I-4
1.5	Manfaat Penelitian.....	I-5
1.6	Batasan Masalah.....	I-5
1.7	Sistematika Penulisan.....	I-5
1.8	Kesimpulan.....	I-6

BAB II TINJAUAN PUSTAKA

2.1	Pendahuluan.....	II-1
2.2	Landasan Teori.....	II-1
2.2.1	Chebyshev Distance.....	II-1
2.2.2	Euclidean Distance.....	II-1
2.2.3	Manhattan Distance.....	II-2

2.2.4	k-Nearest Neighbor.....	II-2
2.2.5	Pengujian Penelitian.....	II-3
2.2.5	Rational Unified Process (RUP).....	II-4
2.3	Penelitian Lain yang Relevan	II-6
2.3.1	Implementasi Metode k-Nearest Neighbor (KNN) dan Simple Additive Weighting (SAW) dalam Pengambilan keputusan Seleksi Penerimaan Anggota Paskibraka.....	II-6
2.3.2	Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square.....	II-6
2.3.3	Metode Manhattan, Euclidean dan Chebyshev pada Algoritma K-Means untuk Pengelompokan Status Desa.....	II-7
2.4	Kesimpulan.....	II-7

BAB III METODOLOGI PENELITIAN

3.1	Pendahuluan.....	III-1
3.2	Pengumpulan Data.....	III-1
3.2.1	Jenis dan Sumber Data.....	III-1
3.2.2	Metode Pengumpulan Data.....	III-1
3.3.1	Menetapkan Kerangka Kerja / Framework.....	III-2
3.3.2	Menetapkan Kriteria Pengujian.....	III-4
3.3.3	Menetapkan Format data Pengujian.....	III-5
3.4.4	Menentukan Alat yang Digunakan Dalam Pelaksanaan Penelitian.....	III-10
3.4.5	Melakukan Pengujian Penelitian.....	III-10

3.3.6	Melakukan Analisa Hasil dan Membuat Kesimpulan Penelitian.....	III-11
3.4	Metode Pengembangan Perangkat Lunak.....	III-11
3.4.1	RUP.....	III-12
3.5	Manajemen Proyek Penelitian.....	III-14

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1	Pendahuluan.....	IV-1
4.2	Rational Unified Process (RUP).....	IV-1
4.2.1	Fase Insepsi.....	IV-1
4.2.1.1	Pemodelan Bisnis.....	IV-1
4.2.1.2	Kebutuhan Sistem.....	IV-2
4.2.1.3	Analisis dan Desain.....	IV-3
4.2.1.3.1	Analisis Kebutuhan Perangkat Lunak.....	IV-4
4.2.1.3.2	Analisis Data.....	IV-4
4.2.1.3.3	Analisis Metode Pengukuran Jarak (distance).....	IV-6
4.2.1.3.4	Desain Perangkat Lunak.....	IV-9
4.2.2	Fase Elaborasi.....	IV-18
4.2.2.1	Pemodelan Bisnis.....	IV-18
4.2.2.1.1	Perancangan Data.....	IV-18
4.2.2.1.2	Perancangan Antar Muka.....	IV-18
4.2.2.2	Kebutuhan Sistem.....	IV-19
4.2.2.3	Diagram Aktivitas.....	IV-20
4.2.2.4	Diagram Sequence.....	IV-22
4.2.3	Fase Konstruksi.....	IV-27
4.2.3.1	Kebutuhan Sistem.....	IV-27
4.3.2.2	Diagram Kelas.....	IV-28

4.2.3.3	Diagram Kelas Analisis.....	IV-28
4.2.3.4	Implementasi.....	IV-29
4.2.4	Fase Transisi.....	IV-31

BAB V HASIL DAN ANALISIS PENELITIAN

5.1	Pendahuluan.....	V-1
5.2	Data Hasil Percobaan Penelitian.....	V-1
5.2.2.1	Hasil Pengujian Klasifikasi Menggunakan Metode Manhattan k – Nearest Neighbor.....	V-2
5.2.2.2	Hasil Pengujian Klasifikasi Menggunakan Metode Euclidean k – Nearest Neighbor.....	V-6
5.2.2.3	Hasil Pengujian Klasifikasi Menggunakan Metode Chebyshev k – Nearest Neighbor.....	V-9
5.3.1	Analisis Hasil Pengujian Klasifikasi dataset titanic dengan Metode pengukuran jarak Manhattan dan algoritma k – Nearest Neighbor.....	V-13
5.3.2	Analisis Hasil Pengujian Klasifikasi dataset titanic dengan Metode pengukuran jarak Euclidean dan algoritma k – Nearest Neighbor.....	V-14
5.3.2	Analisis Hasil Pengujian Klasifikasi dataset titanic dengan Metode pengukuran jarak Chebyshev dan algoritma k – Nearest Neighbor.....	V-14
5.3.3	Analisis Hasil Pengujian Klasifikasi Dataset Titanic dengan Metode Pengukuran jarak Manhattan, Euclidean dan Chebyshev dan algoritma k – Nearest Neighbor.....	V-15
5.4	Kesimpulan.....	V-15

BAB VI	KESIMPULAN DAN SARAN	
6.1	Pendahuluan.....	VI-1
6.2	Kesimpulan.....	VI-1
6.3	Saran.....	VI-2
	DAFTAR PUSTAKA.....	xiv

DAFTAR TABEL

	Halaman
Tabel III-1. Tabel Hasil Pengujian Nilai Akurasi Algoritma <i>k-NN</i> metode <i>Euclidean Distance</i>	III-5
Tabel III-2. Tabel Hasil Pengujian Nilai Presisi Algoritma <i>k-NN</i> metode <i>Euclidean Distance</i>	III-6
Tabel III-3. Tabel Hasil Pengujian Nilai Recall Algoritma <i>k-NN</i> metode <i>Euclidean Distance</i>	III-6
Tabel III-4. Tabel Hasil Pengujian Nilai Akurasi Algoritma <i>k-NN</i> metode <i>Manhattan Distance</i>	III-7
Tabel III-5. Tabel Hasil Pengujian Nilai Presisi Algoritma <i>k-NN</i> metode <i>Manhattan Distance</i>	III-7
Tabel III-6. Tabel Hasil Pengujian Nilai Recall Algoritma <i>k-NN</i> metode <i>Manhattan Distance</i>	III-8
Tabel III-7. Tabel Hasil Pengujian Nilai Akurasi Algoritma <i>k-NN</i> metode <i>Chebyshev Distance</i>	III-8
Tabel III-8. Tabel Hasil Pengujian Nilai Presisi Algoritma <i>k-NN</i> metode <i>Chebyshev Distance</i>	III-9
Tabel III-9. Tabel Hasil Pengujian Nilai Recall Algoritma <i>k-NN</i> metode <i>Chebyshev Distance</i>	III-9
Tabel III-10. Rancangan Tabel Analisa Hasil Klasifikasi Algoritma <i>k-NN</i> dengan metode <i>Euclidean</i> , <i>Manhattan</i> dan <i>Chebyshev</i>	III-11
Tabel III-11. Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure</i> (WBS).....	III-15
Tabel IV-1. Tabel Kebutuhan Fungsional Sistem.....	IV-2
Tabel IV-2. Tabel Kebutuhan Non-Fungsional Sistem.....	IV-3
Tabel IV-3. Tabel Fitur – Fitur pada Dataset.....	IV-4
Tabel IV-4. Tabel Gambaran Dataset yang Digunakan.....	IV-5

Tabel IV- 5. Tabel Contoh Data Latih.....	IV-6
Tabel IV-6. Tabel Contoh Data Uji.....	IV-6
Tabel IV-7. Tabel Hasil Perhitungan Jarak Euclidean Distance.....	IV-7
Tabel IV-8. Tabel Hasil Pengurutan Data Perhitungan Jarak.....	IV-7
Tabel IV-9. Tabel Hasil Perhitungan Jarak Manhattan Distance.....	IV-8
Tabel IV-10. Tabel Hasil Pengurutan Data Perhitungan Jarak.....	IV-8
Tabel IV-11. Tabel Hasil Perhitungan Jarak Chebyshev Distance.....	IV-9
Tabel IV-12. Tabel Hasil Pengurutan Data Perhitungan Jarak.....	IV-9
Tabel IV-13. Tabel Definisi Aktor <i>Use Case</i>	IV-10
Tabel IV-14. Tabel Definisi <i>Use Case</i>	IV-11
Tabel IV-15. Tabel Skenario <i>Use Case</i> memuat dataset <i>titanic</i>	IV-12
Tabel IV-16. Skenario <i>Use Case</i> Klasifikasi dengan metode Manhattan-k-NN.....	IV-14
Tabel IV-17. Skenario <i>Use Case</i> Klasifikasi dengan metode Euclidean-k-NN.....	IV-15
Tabel IV-18. Skenario <i>Use Case</i> Klasifikasi dengan metode Chebyshev-k-NN.....	IV-17
Tabel IV-19. Implementasi Kelas.....	IV-30
Tabel IV-20. Rencana Pengujian <i>Use Case</i> Memuat dataset <i>titanic</i>	IV-33
Tabel IV-21. Rencana Pengujian <i>Use Case</i> Klasifikasi dengan Metode Manhattan-k-NN.....	IV-33
Tabel IV-22. Rencana Pengujian <i>Use Case</i> Klasifikasi dengan Metode Euclidean-k-NN.....	IV-34
Tabel IV-23. Rencana Pengujian <i>Use Case</i> Klasifikasi dengan Metode Chebyshev-k-NN.....	IV-35
Tabel IV-24. Pengujian <i>Use Case</i> Memuat dataset <i>titanic</i>	IV-36
Tabel IV-25. Pengujian <i>Use Case</i> Klasifikasi dengan Metode Manhattan- k-NN	IV-37
Tabel IV-26. Pengujian <i>Use Case</i> Klasifikasi dengan Metode Euclidean- k-NN	IV-38

Tabel IV-27. Pengujian <i>Use Case</i> Klasifikasi dengan Metode Chebyshev- k-NN	IV-40
Tabel V-1. Hasil Pengujian Nilai Akurasi Klasifikasi Menggunakan Metode Manhattan <i>k – Nearest Neighbour</i>	V-2
Tabel V-2, Hasil Pengujian nilai presisi Klasifikasi Menggunakan Metode Manhattan <i>k – Nearest Neighbour</i>	V-3
Tabel V-3. Hasil Pengujian nilai recall Klasifikasi Menggunakan Metode Manhattan <i>k – Nearest Neighbour</i>	V-4
Tabel V-4. Hasil Pengujian Nilai Akurasi Klasifikasi Menggunakan Metode Euclidean <i>k – Nearest Neighbour</i>	V-6
Tabel V-5. Hasil Pengujian Nilai Presisi Klasifikasi Menggunakan Metode Euclidean <i>k – Nearest Neighbour</i>	V-7
Tabel V-6. Hasil Pengujian Nilai Recall Klasifikasi Menggunakan Metode Euclidean <i>k – Nearest Neighbour</i>	V-8
Tabel V-7. Hasil Pengujian Nilai Akurasi Klasifikasi Menggunakan Metode Chebyshev <i>k – Nearest Neighbour</i>	V-10
Tabel V-8. Hasil Pengujian Nilai Presisi Klasifikasi Menggunakan Metode Chebyshev <i>k – Nearest Neighbour</i>	V-11
Tabel V-9. Hasil Pengujian Nilai Recall Klasifikasi Menggunakan Metode Chebyshev <i>k – Nearest Neighbour</i>	V-12
Tabel V-10. Perbandingan Hasil Pengujian Dataset <i>Titanic</i> dengan Metode Pengukuran jarak <i>Manhattan, Euclidean dan Chebyshev</i> dan algoritma <i>k – Nearest Neighbour</i>	V-15

DAFTAR GAMBAR

	Halaman
Gambar II-1. Arsitektur <i>Rational Unified Process (RUP)</i>	II-5
Gambar III-1. Diagram Tahap Penelitian.....	III-2
Gambar III-2. Diagram Tahapan Pengujian Penelitian.....	III-10
Gambar III-3. <i>Gantt Chart</i> Penjadwalan Penelitian Tahap Menentukan Ruang Lingkup dan Unit Penelitian.....	III-20
Gambar III-4. <i>Gantt Chart</i> Penjadwalan Penelitian Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian.....	III-21
Gambar III-5. <i>Gantt Chart</i> Penjadwalan Penelitian Tahap Menentukan Kriteria Pengujian.....	III-21
Gambar III-6. <i>Gantt Chart</i> Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan Untuk Pelaksanaan Penelitian pada Fase Insepsi.....	III-22
Gambar III-7. <i>Gantt Chart</i> Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan Untuk Pelaksanaan Penelitian pada Fase Elaborasi.....	III-22
Gambar III-8. <i>Gantt Chart</i> Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan Untuk Pelaksanaan Penelitian pada Fase Konstruksi.....	III-23
Gambar III-9. <i>Gantt Chart</i> Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan Untuk Pelaksanaan Penelitian pada Fase Transisi.....	III-23
Gambar III-10. <i>Gantt Chart</i> Penjadwalan Penelitian Tahap Melakukan Pengujian Penelitian.....	III-24
Gambar III-11. <i>Gantt Chart</i> Penjadwalan Penelitian Tahap Melakukan Analisa Hasil Pengujian dan Pembuatan Kesimpulan.....	III-24
Gambar IV-1. Diagram <i>Use Case</i>	IV-10
Gambar IV-2. Rancangan Antar Muka Perangkat Lunak.....	IV-19
Gambar IV-3. Diagram Aktivitas Memuat dataset <i>titanic</i>	IV-20
Gambar IV-4. Diagram Aktivitas Klasifikasi dengan metode Manhattan-k-NN.....	IV-21

Gambar IV-5. Diagram Aktivitas Klasifikasi dengan metode Euclidean-k-NN.....	IV-21
Gambar IV-6. Diagram Aktivitas Klasifikasi dengan metode Chebyshev-k-NN.....	V-22
Gambar IV-7. Diagram <i>Sequence</i> Memuat Dataset <i>titanic</i>	V-23
Gambar IV-8. Diagram <i>Sequence</i> Klasifikasi dengan metode Manhattan-K-NN.....	V-24
Gambar IV-9. Diagram <i>Sequence</i> Klasifikasi dengan metode Euclidean-K-NN.....	V-25
Gambar IV-10. Diagram <i>Sequence</i> Klasifikasi dengan metode Chebyshev-K-NN.....	V-26
Gambar IV-11. Diagram Kelas.....	V-28
Gambar IV-12. Diagram Kelas Analisis Memuat Dataset <i>titanic</i>	V-28
Gambar IV-13. Diagram Kelas Klasifikasi dengan metode Manhattan-k-NN.....	V-29
Gambar IV-14. Diagram Kelas Klasifikasi dengan metode Euclidean-k-NN.....	V-29
Gambar IV-15. Diagram Kelas Klasifikasi dengan metode Chebyshev-k-NN.....	V-29
Gambar IV-16. Antar Muka Perangkat Lunak.....	V-31

DAFTAR LAMPIRAN

	Halaman
Kode Program.....	L1-1
Diagram Kelas.....	L2-1

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini memberikan penjelasan umum tentang latar belakang masalah, rumusan masalah, tujuan, manfaat penelitian, batasan masalah, dan sistematika penulisan.

Pendahuluan dimulai dengan memberikan penjelasan mengenai perkembangan data, teknik klasifikasi, penelitian terkait dengan algoritma *k-Nearest Neighbor* serta metode-metode pengukuran jarak yang nantinya akan dibandingkan hasil perhitungan pada algoritma *k-Nearest Neighbor*.

1.2 Latar Belakang Masalah

Data mining merupakan suatu bidang ilmu yang menggunakan teknik dengan memanfaatkan data lama untuk mendapatkan suatu informasi pengetahuan yang baru (Binabar & Ivandari, 2017). *Data mining* lebih unggul untuk teknik pengolahan data dikarenakan *data mining* akan memecahkan masalah dengan menganalisis data yang telah ada dalam basis data (Rohman, 2015). Dalam proses *data mining*, terdapat data supervised dan unsupervised dimana data supervised ditandai dengan adanya label, sementara data unsupervised tidak memiliki label data.

Teknik klasifikasi yang merupakan salah satu fungsi utama *data mining*, dapat digunakan untuk proses pengelompokan data dari data yang telah ada dengan menggunakan data berlabel atau data *supervised*. Teknik klasifikasi sangat cocok untuk diterapkan pada dataset yang memiliki variabel target, dan algoritma ini juga memiliki akurasi yang tinggi. Sementara itu *Clustering* adalah salah satu alat bantu pada data mining yang bertujuan mengelompokkan objek-objek ke dalam *cluster - cluster*. *Cluster* adalah sekumpulan objek - objek data yang similar satu sama lain dalam cluster yang sama dan disimilar terhadap objek-objek yang berbeda cluster (Defiyanti, 2017)

Algoritma *k-Nearest Neighbor* melakukan klasifikasi terhadap objek berdasarkan data latih yang menggunakan jarak terdekat atau kemiripan terhadap objek tersebut (Wahyono, Trisna, Sariwening, Fajar, & Wijayanto, 2020). Algoritma *k-Nearest Neighbor* menggunakan pendekatan *supervised learning* dimana data yang digunakan merupakan data berlabel. Algoritma *k-Nearest Neighbor* memiliki keunggulan yaitu paling umum, paling sederhana dan non parametrik yang tidak mendasarkan pada distribusi data tertentu (Ekinci, Acun, & Omurca, 2018), sehingga algoritma ini cocok untuk digunakan pada dataset *titanic* pada penelitian ini.

Salah satu komponen esensial dari algoritma *k-Nearest Neighbor* adalah perhitungan jarak, dimana perhitungan jarak suatu data akan digunakan untuk mencocokkan antara data hasil pelatihan dengan data baru sebagai pembanding (Wahyono et al., 2020). Perhitungan distance digunakan untuk melakukan pembobotan pada suatu data agar data tersebut bisa ditempatkan berdasarkan

kelas yang mempunyai bobot paling mendekati. Perhitungan distance mempunyai arti penting, yaitu jika hasil pembobotan jarak tersebut mempunyai selisih tertentu terhadap nilai data latih, maka hasil dari perhitungan distance tersebut bisa berubah. Pada perhitungan nilai jarak di algoritma *k-Nearest Neighbor* Euclidean distance umumnya paling banyak digunakan dalam penelitian, meskipun algoritma *k-Nearest Neighbor* salah satu kelemahan dari algoritma *k-Nearest Neighbor* adalah nilai *k* yang bias. Nilai *k* menyatakan berapa banyak jumlah neighbor atau data yang terdekat dengan suatu objek. Jumlah neighbor yang berbeda tentu akan mempengaruhi hasil klasifikasi terhadap satu objek (Angreni, Adisasmita, & Ramli, 2018). Adapun cara untuk mendapatkan nilai *k* yang optimal adalah dengan menyesuaikan metode pengukuran jarak antar neighbor terdekat. Pada penelitian (Latifah, 2015), dilakukan perhitungan jarak menggunakan algoritma Manhattan distance untuk menentukan pemenang lelang, dan di dapatkan hasil akurasi sebesar 86%, sementara pada penelitian (Wahyono et al., 2020) dilakukan perbandingan pengukuran jarak antara Euclidean distance, Manhattan distance, minkowski distance dan chebycev distance untuk mengklasifikasi data tekstual dan di dapatkan hasil masing-masing sebesar 79%, 59%, 79%, dan 76%. Hal ini membuktikan adanya pengaruh tertentu perhitungan nilai jarak data terhadap suatu algoritma. Untuk itulah penelitian ini dilakukan untuk mencari metode pengukuran jarak antar distance yaitu Manhattan distance, Euclidean distance dan minkowski distance terhadap dataset yang ditentukan untuk mengetahui metode mana yang paling optimal untuk jenis dataset *titanic*.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang dikemukakan diatas, rumusan masalah yang dibahas dalam penelitian ini adalah mengenai metode pengukuran jarak apa yang dapat bekerja paling optimal untuk algoritma *k-Nearest Neighbor* jika digunakan untuk klasifikasi dataset *titanic*.

Untuk menyelesaikan permasalahan yang disebutkan diatas, maka dalam penelitian ini dibagi dalam beberapa *research question*, yaitu sebagai berikut.

1. Apakah pengaruh dari perbandingan metode pengukuran jarak yang dilakukan terhadap algoritma klasifikasi *k-Nearest Neighbor*?
2. Bagaimana perbandingan hasil klasifikasi algoritma *k-Nearest Neighbor* jika menggunakan metode pengukuran jarak yang berbeda-beda?

1.4 Tujuan Penelitian

Tujuan dilakukanya penelitian Tugas Akhir ini adalah sebagai berikut.

1. Mengetahui pengaruh perbandingan metode pengukuran jarak yang dilakukan untuk mengklasifikasi dataset *titanic* dengan menggunakan algoritma *k-Nearest Neighbor*.
2. Mengetahui perbedaan perbandingan hasil klasifikasi algoritma *k-Nearest Neighbor* terhadap data yang ditentukan jika dilakukan pengukuran jarak dengan metode yang berbeda-beda.

1.5 Manfaat Penelitian

Manfaat yang didapat dari penelitian Tugas Akhir ini adalah sebagai berikut.

1. Memahami cara kerja algoritma *k-Nearest Neighbor* untuk mengklasifikasi dataset *titanic* dengan metode pengukuran jarak yang berbeda-beda.
2. Mengetahui metode pengukuran jarak yang paling optimal untuk digunakan dalam teknik klasifikasi algoritma *k-Nearest Neighbor* pada dataset *titanic*.

1.6 Batasan Masalah

Batasan masalah pada penelitian Tugas Akhir ini adalah sebagai berikut.

1. Penelitian menggunakan 3 metode pengukuran jarak yaitu *Euclidean*, *Chebyshev*, dan *Manhattan*.
2. Data yang dijadikan objek penelitian diambil dari dataset *open source* yang diunduh dari situs *kaggle*.

1.7 Sistematika Penulisan

Sistematika penulisan penelitian Tugas Akhir ini adalah sebagai berikut.

BAB I. PENDAHULUAN

Pada bab pendahuluan akan dijelaskan mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penulisan yang berdasarkan jurnal, buku, maupun artikel yang berkaitan dengan penelitian.

BAB III. METODOLOGI PENELITIAN

Dalam bab ini akan dijelaskan yang dilakukan dalam penelitian ini. Tiap tahap akan dijelaskan secara rinci sehingga membentuk sebuah kerangka kerja termasuk perancangan manajemen proyek pada pelaksanaan penelitian yang akan ditulis pada bagian akhir bab ini.

1.8 Kesimpulan

Penelitian mengenai perbandingan metode pencarian jarak pada algoritma *k-Nearest Neighbor* untuk mengklasifikasi dataset *titanic* dilakukan bertujuan untuk menjawab rumusan masalah dan menggali potensi terhadap algoritma *k-Nearest Neighbor* untuk dikembangkan agar dapat ditemukan gabungan teknik pencarian terhadap algoritma yang tepat yang dapat meningkatkan hasil performa pengklasifikasian algoritma *k-Nearest Neighbor*.

DAFTAR PUSTAKA

- Angreni, I. A. A., Adisasmita, S. A., & Ramli, M. I. (2018). Terhadap Tingkat Akurasi Identifikasi Kerusakan Jalan, *7(2)*, 63–70.
- Binabar, S. W., & Ivandari. (2017). Optimasi Parameter K pada Algoritma KNN untuk Deteksi Penyakit Kanker Payudara. *IC-Tech, XII(2)*.
- Defiyanti, S. (2017). Integrasi Metode Clustering dan Klasifikasi untuk Data Numerik. *Citee*, 256–261.
- Ekinci, E., Acun, N., & Omurca, S. İ. (2018). A Comparative Study on Machine Learning Techniques Using Titanic Dataset. *7th International Conference on Advanced Technologies- ICAT'18*, (April), 411–416. Retrieved from <https://www.researchgate.net/publication/324909545>
- Latifah, K. (2015). Kombinasi Algoritma K-NN dan Manhattan Distance untuk Menentukan Pemenang Lelang. *Jurnal Informatika Upgris (JIU)*, 49–58. Retrieved from <https://docplayer.info/34038947-Kombinasi-algoritma-k-nn-dan-manhattan-distance-untuk-menentukan-pemenang-lelang.html>
- Mustofa, Z., & Suasana, I. S. (2018). Algoritma Clustering K-Medoids pada E-Government Bidang Information and Communication. *Jurnal Teknologi Informasi Dan Komunikasi*, *9(1)*, 1–10.
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, *4(1)*,

20–24. <https://doi.org/10.30591/jpit.v4i1.1253>

Religia, Y. (2016). Metode Manhattan , Euclidean Dan Chebyshev Pada Algoritma K-Means Untuk Pengelompokan Status Desa.

Rohman, A. (2015). Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa. *Neo Teknika*, 1(1).
<https://doi.org/10.37760/neoteknika.v1i1.350>

T, A. J., Yanosma, D., & Anggriani, K. (2017). Implementasi Metode K-Nearest Neighbor (Knn) Dan Simple Additive Weighting (Saw) Dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibraka. *Pseudocode*, 3(2), 98–112. <https://doi.org/10.33369/pseudocode.3.2.98-112>

Tharwat, A., Mahdi, H., Elhoseny, M., & Hassanien, A. E. (2018). Recognizing human activity in mobile crowdsensing environment using optimized k-NN algorithm. *Expert Systems with Applications*, 107, 32–44.
<https://doi.org/10.1016/j.eswa.2018.04.017>

Thunem, A. P., Fredriksen, R., & Thunem, H. P. (2006). *Management of Requirements in NPP Modernisation Projects - Project Report 2005. Management.*

Wahyono, W., Trisna, I. N. P., Sariwening, S. L., Fajar, M., & Wijayanto, D. (2020). Comparison of distance measurement on k-nearest Neighbor in textual data classification. *Jurnal Teknologi Dan Sistem Komputer*, 8(1), 54–58. <https://doi.org/10.14710/jtsiskom.8.1.2020.54-58>