

Perbandingan Metode Perhitungan Jarak dalam Algoritma *K-Nearest Neighbour* untuk Klasifikasi Data Penyakit pada Manusia

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



Oleh :

Hezar Muhammad
NIM : 09021381722143

**Jurusan Teknik Informatika Bilingual
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2022**

Perbandingan Metode Perhitungan Jarak dalam Algoritma *K-Nearest Neighbour* untuk Klasifikasi Data Penyakit pada Manusia

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



Oleh :

Hezar Muhammad
NIM : 09021381722143

**Jurusan Teknik Informatika Bilingual
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2022**

LEMBAR PENGESAHAN SKRIPSI

Perbandingan Metode Perhitungan Jarak dalam Algoritma
K-Nearest Neighbour untuk Klasifikasi Data Penyakit pada Manusia

Oleh :

Hezar Muhammad
NIM : 09021381722143

Palembang, 20 Juni 2022

Pembimbing I



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

Pembimbing II,



Danny Matthew Saputra, M.Sc.
NIP. 198505102015041002

Mengetahui,
Ketua Jurusan Teknik Informatika



Aly Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Jum'at tanggal 3 Juni 2022 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Hezar Muhammad
NIM : 09021381722143
Judul : Perbandingan Metode Perhitungan Jarak dalam Algoritma *K-Nearest Neighbour* untuk Klasifikasi Data Penyakit pada Manusia

dan dinyatakan **LULUS**.

1. Ketua

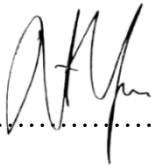
Dr. M. Fachrurrozi, M.T.
NIP. 198005222008121002



.....

2. Penguji I

Novi Yusliani, M.T.
NIP. 198211082012122001



.....

3. Penguji II

Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001



.....

4. Pembimbing I

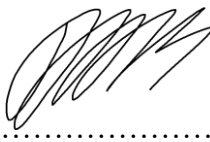
Dian Palupi Rini, M.Kom., Ph.D.
NIP. 197802232006042002



.....

5. Pembimbing II

Danny Matthew Saputra, M.Sc.
NIP. 198505102015041002



.....

Mengetahui,
Ketua Jurusan Teknik Informatika



Alyi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan di bawah ini :

Nama : Hezar Muhammad

NIM : 09021381722143

Program Studi : Teknik Informatika Bilingual

Judul Skripsi : Perbandingan Metode Perhitungan Jarak dalam Algoritma *K-Nearest Neighbour* untuk Klasifikasi Data Penyakit pada Manusia

Hasil Pengecekan Software *iThenticate/Turnitin* : 6 %

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek 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, 20 Juni 2022



Hezar Muhammad

NIM. 09021381722143

MOTTO DAN PERSEMBAHAN

Motto:

“Yesterday is History, Tomorrow is Mistery, and Today is a Gift“
-Master Oogway

Kupersembahkan karya tulis ini kepada:

- Allah SWT
- Nabi Muhammad SAW
- Ibu dan Ayah
- Seluruh Keluarga Besar
- Dosen Pembimbing dan Dosen Penguji
- Para Dosen Universitas Sriwijaya
- Guru-guru
- Sahabat
- Almamater

ABSTRACT

One of the algorithms in *Machine Learning* principles that can be used to classify large amounts of data by producing clear and effective results is the *K-Nearest Neighbour* Algorithm. This research aims to prove which of the distance calculation formulas in the *K-Nearest Neighbour* Algorithm has an accuracy value with good performance in classifying human disease datasets. The proof is done by comparing the existing distance calculation formulas, namely *Euclidean distance*, *Manhattan distance*, *Minkowsky distance* and *Chebyshev distance*. The dataset uses data on diseases that are often experienced by humans such as Heart Attack, Breast Cancer, Diabetes and Stroke. The results show that *Euclidean distance* is the distance calculation formula that gets the most accuracy values with good performance compared to other distance calculation formulas. Even the highest accuracy results obtained can reach 90% with performance values namely *Precision* of 0.921053, *Recall* of 0.833333 and *F-Measure* of 0.875 with a K value of 3 for the classification of breast cancer datasets. So it is concluded that Euclidean distance is a distance calculation formula that is able to produce high accuracy values with good performance for classifying human disease datasets.

Keywords: *Machine Learning, K-Nearest Neighbour Algorithm, Euclidean Distance, Manhattan Distance, Minkowsky Distance, Chebyshev Distance, Heart Attack, Breast Cancer, Diabetes, Stroke.*

ABSTRAK

Salah satu algoritma dalam prinsip *Machine Learning* yang dapat digunakan untuk mengklasifikasikan data dalam jumlah besar dengan menghasilkan hasil yang jelas serta efektif adalah Algoritma *K-Nearest Neighbour*. Penelitian ini bertujuan untuk membuktikan di antara rumus perhitungan jarak pada Algoritma *K-Nearest Neighbour* manakah yang memiliki nilai akurasi dengan kinerja yang baik dalam mengklasifikasikan dataset penyakit pada manusia. Pembuktian dilakukan dengan membandingkan rumus perhitungan jarak yang ada yaitu *Euclidean distance*, *Manhattan distance*, *Minkowsky distance* dan *Chebyshev distance*. Dataset menggunakan data penyakit yang sering dialami oleh manusia seperti Serangan Jantung, Kanker Payudara, Diabetes dan Stroke. Hasil penelitian menunjukkan *Euclidean distance* merupakan rumus perhitungan jarak yang paling banyak memperoleh nilai akurasi dengan kinerja yang baik dibandingkan rumus perhitungan jarak lainnya. Bahkan hasil akurasi tertinggi yang diperoleh bisa mencapai 90% dengan nilai kinerja yaitu *Precision* sebesar 0.921053, *Recall* sebesar 0.833333 dan *F-Measure* sebesar 0.875 pada nilai K bernilai 3 untuk klasifikasi dataset penyakit Kanker Payudara. Sehingga disimpulkan bahwa *Euclidean distance* merupakan rumus perhitungan jarak yang mampu menghasilkan nilai akurasi tinggi dengan kinerja yang baik untuk pengklasifikasian dataset penyakit pada manusia.

Kata Kunci: *Machine Learning*, Algoritma *K-Nearest Neighbour*, *Euclidean distance*, *Manhattan distance*, *Minkowsky distance*, *Chebyshev distance*, Serangan Jantung, Kanker Payudara, Diabetes, Stroke.

KATA PENGANTAR

Alhamdulillahirobbil'alamin, puji dan syukur kita panjatkan kehadirat Allah SWT. Karena atas segala limpahan rahmat dan karunia-Nya penulis dapat menyelesaikan Laporan Tugas Akhir ini. Serta tak lupa shalawat dan salam kita curahkan kepada junjungan kita Nabi Muhammad SWT beserta keluarganya, yang telah membawa kita ke zaman yang terang benderang ini. Seperti yang kita tahu Tugas Akhir ini dilaksanakan sebagai salah satu syarat dalam menyelesaikan pendidikan program Strata-1 Program Studi Teknik Informatika Bilingual pada Fakultas Ilmu Komputer, Universitas Sriwijaya.

Pada kesempatan kali ini izinkan penulis mengucapkan banyak terima kasih kepada seluruh pihak yang telah berkontribusi serta membantu penulis selama penyusunan Tugas Akhir hingga sampai pembuatan pada laporan Tugas Akhir yang berjudul "PERBANDINGAN METODE PERHITUNGAN JARAK DALAM ALGORITMA K-NEAREST NEIGHBOUR UNTUK KLASIFIKASI DATA PENYAKIT PADA MANUSIA". Adapun ucapan terima kasih tersebut saya ucapkan kepada:

1. Tuhan yang Maha Esa lagi Maha Kuasa, Allah SWT. Karena atas rahmat serta kuasa-Nya lah penulis dapat menyelesaikan Laporan Kerja Praktik ini sebaik mungkin.
2. Baginda Nabi Muhammad SAW yang telah membawa kita dari zaman jahiliyah ke zaman yang terang benderang ini.
3. Yang terhormat Ayah M.Hosarudin dan Ibu Husnaini yang telah memberikan dukungan serta doa yang tiada henti sehingga penulis dapat menyelesaikan laporan kerja praktik ini.
4. Yang terhormat Bapak Jaidan Jauhari, S.Pd., M.T. Selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Yang terhormat Ibu Alvi Syahrini Utami, M.Kom. Selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Yang terhormat Ibu Dian Palupi Rini, M.Kom., Ph.D. Selaku Pembimbing 1 Tugas Akhir dan Bapak Danny Matthew Saputra, M.SC., Selaku

pembimbing 2 Tugas Akhir sekaligus pembimbing akademik di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

7. Segenap staff pengajar Di Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Mbak Wiwin Juliani, S.Si, selaku staff administrasi Jurusan Teknik Informatika Bilingual.
9. Serta tak lupa sahabat saya beserta teman-teman seperjuangan di Jurusan Teknik Informatika Bilingual tentunya.

Penulis menyadari bawasannya masih banyak kesalahan serta kekurangan yang terdapat pada laporan Tugas Akhir ini, baik itu dalam segi bahasa maupun segi teknis penyajian penulisan. Oleh karena itu, Penulis mengharapkan kritik serta saran dari segenap pembaca semua yang bersifat membangun agar penulis dapat memperbaiki kesalahan tersebut dikemudian harinya. Namun dibalik itu semua, penulis juga berharap bawasannya laporan Tugas Akhir ini dapat di gunakan semaksimal mungkin serta bermanfaat bagi kita semua.

Palembang, 20 Juni 2022



Hezar Muhammad

DAFTAR ISI

Halaman

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN SIDANG SKRIPSI	iii
PERNYATAAN BEBAS PLAGIAT.....	iv
MOTTO DAN PERSEMBAHAN	v
<i>ABSTRACT</i>	vi
ABSTRAK.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xv
DAFTAR GAMBAR	xxiii
DAFTAR LAMPIRAN.....	xxviii
BAB I PENDAHULUAN	
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-3
1.4 Tujuan Penelitian	I-4
1.5 Manfaat Penelitian	I-4
1.6 Batasan Masalah.....	I-5
1.7 Sistematis Penulisan.....	I-5
BAB II KAJIAN LITERATUR	
2.1 Pendahuluan	II-1
2.2 Landasan Teori.....	II-1
2.2.1 Penyakit Serangan Jantung	II-1
2.2.2 Penyakit Kanker Payudara.....	II-2
2.2.3 Penyakit Diabetes	II-3
2.2.4 Penyakit Stroke.....	II-4
2.2.5 <i>Machine Learning</i>	II-5
2.2.6 <i>Klasifikasi</i>	II-6
2.2.7 Algoritma <i>K-Nearest Neighbour</i>	II-10

2.2.8	Nilai K	II-11
2.2.9	Rumus Perhitungan Jarak <i>Euclidean Distance</i>	II-11
2.2.10	Rumus Perhitungan Jarak <i>Manhattan Distance</i>	II-12
2.2.11	Rumus Perhitungan Jarak <i>Minkowsky Distance</i>	II-13
2.2.12	Rumus Perhitungan Jarak <i>Chebyshev Distance</i>	II-14
2.2.13	<i>Confusion Matrix</i>	II-14
2.2.14	<i>Rational United Process(RUP)</i>	II-16
2.3	Penelitian Lain yang Relevan.....	II-17
2.4	Kesimpulan	II-21
BAB III METODOLOGI PENELITIAN		
3.1	Pendahuluan	III-1
3.2	Unit Penelitian.....	III-1
3.3	Pengumpulan Data	III-1
3.3.1	Jenis Data.....	III-1
3.3.2	Sumber Data	III-1
3.4	Tahapan Penelitian	III-2
3.4.1	Kerangka Kerja.....	III-2
3.4.2	Kriteria Pengujian.....	III-4
3.4.3	Format Data Pengujian	III-4
3.4.4	Alat yang Digunakan Dalam Penelitian	III-4
3.4.5	Pengujian Penelitian	III-5
3.4.6	Analisis Hasil Penelitian dan Kesimpulan Penelitian.....	III-5
3.5	Metode Pengembangan Perangkat Lunak.....	III-6
3.5.1	<i>Rational United Process(RUP)</i>	III-1
3.5.1.1	<i>Fase Inception</i>	III-1
3.5.1.2	<i>Fase Elaboration</i>	III-2
3.5.1.3	<i>Fase Construction</i>	III-3
3.5.1.4	<i>Fase Transition</i>	III-3
3.6	Manajemen Proyek Penelitian	III-8
BAB IV PENGEMBANGAN PERANGKAT LUNAK		
4.1	Pendahuluan	IV-1
4.2	<i>Rational United Process(RUP)</i>	IV-1
4.2.1	Tahap <i>Incaption</i>	IV-1
4.2.1.1	<i>Business Modeling</i>	IV-1
4.2.1.2	<i>Requirement</i>	IV-2
4.2.1.3	Analisis dan Desain	IV-3
4.2.1.3.1	Analisis Kebutuhan Perangkat Lunak	IV-3
4.2.1.3.2	Analisis Data.....	IV-9
4.2.1.3.3	Analisis Hasil dan Evaluasi	IV-9
4.2.1.3.4	Desain <i>Software</i>	IV-10
4.2.2	Tahap <i>Elaboration</i>	IV-12

4.2.2.1	<i>Business Modeling</i>	IV-12
4.2.2.1.1	Perancangan Data	IV-13
4.2.2.1.2	Perancangan <i>Interface</i>	IV-13
4.2.2.2	Kebutuhan Sistem	IV-13
4.2.2.3	Diagram Aktifitas	IV-14
4.2.2.4	<i>Sequence Diagram</i>	IV-19
4.2.3	Tahap <i>Construction</i>	IV-23
4.2.3.1	Kebutuhan Sistem	IV-24
4.2.3.2	Diagram Kelas	IV-24
4.2.3.3	Implementasi	IV-25
4.2.4	Tahap <i>Transition</i>	IV-27
4.2.4.1	<i>Business Modeling</i>	IV-27
4.2.4.2	Kebutuhan Sistem	IV-27
4.2.4.3	Rencana Pengujian	IV-27
4.2.4.3.1	Rencana Pengujian Program dengan Memuat File Data <i>Training</i>	IV-28
4.2.4.3.2	Rencana Pengujian Sistem dalam Penerapan Algoritma <i>K-Nearest Neighbour</i>	IV-28
4.2.4.4	Penerapan Rencana Pengujian	IV-29
4.2.4.4.1	Pengujian Program dengan Memuat File Data <i>Training</i>	IV-29
4.2.4.4.2	Pengujian Sistem dalam Penerapan Algoritma <i>K-Nearest Neighbour</i>	IV-30
4.3	Kesimpulan	IV-31
BAB V ANALISIS DAN HASIL		
5.1	Pendahuluan	V-1
5.2	Data Hasil Percobaan Penelitian	V-1
5.2.1	Konfigurasi Percobaan	V-1
5.2.2	Data Hasil Konfigurasi	V-2
5.2.2.1	Hasil Pengujian pada Algoritma <i>K-Nearest Neighbour</i> Menggunakan Rumus <i>Euclidean Distance</i>	V-2
5.2.2.1.1	Hasil Pengujian Rumus <i>Euclidean Distance</i> Menggunakan Dataset Penyakit serangan Jantung	V-3
5.2.2.1.2	Hasil Pengujian Rumus <i>Euclidean Distance</i> Menggunakan Dataset Penyakit Kanker Payudara	V-5
5.2.2.1.3	Hasil Pengujian Rumus <i>Euclidean Distance</i> Menggunakan Dataset Penyakit Diabetes ...	V-8
5.2.2.1.4	Hasil Pengujian Rumus <i>Euclidean Distance</i> Menggunakan Dataset Penyakit Stroke	V-10

5.2.2.2	Hasil Pengujian pada Algoritma <i>K-Nearest Neighbour</i> Menggunakan Rumus <i>Manhattan Distance</i>	V-13
5.2.2.2.1	Hasil Pengujian Rumus <i>Manhattan Distance</i> Menggunakan Dataset Penyakit serangan Jantung.....	V-14
5.2.2.2.2	Hasil Pengujian Rumus <i>Manhattan Distance</i> Menggunakan Dataset Penyakit Kanker Payudara.....	V-16
5.2.2.2.3	Hasil Pengujian Rumus <i>Manhattan Distance</i> Menggunakan Dataset Penyakit Diabetes .	V-19
5.2.2.2.4	Hasil Pengujian Rumus <i>Manhattan Distance</i> Menggunakan Dataset Penyakit Stroke	V-21
5.2.2.3	Hasil Pengujian pada Algoritma <i>K-Nearest Neighbour</i> Menggunakan Rumus <i>Minkosky Distance</i>	V-24
5.2.2.3.1	Hasil Pengujian Rumus <i>Minkosky Distance</i> Menggunakan Dataset Penyakit serangan Jantung.....	V-24
5.2.2.3.2	Hasil Pengujian Rumus <i>Minkosky Distance</i> Menggunakan Dataset Penyakit Kanker Payudara.....	V-27
5.2.2.3.3	Hasil Pengujian Rumus <i>Minkosky Distance</i> Menggunakan Dataset Penyakit Diabetes .	V-29
5.2.2.3.4	Hasil Pengujian Rumus <i>Minkosky Distance</i> Menggunakan Dataset Penyakit Stroke	V-32
5.2.2.4	Hasil Pengujian pada Algoritma <i>K-Nearest Neighbour</i> Menggunakan Rumus <i>Chebyshev Distance</i>	V-34
5.2.2.4.1	Hasil Pengujian Rumus <i>Chebyshev Distance</i> Menggunakan Dataset Penyakit serangan Jantung.....	V-35
5.2.2.4.2	Hasil Pengujian Rumus <i>Chebyshev Distance</i> Menggunakan Dataset Penyakit Kanker Payudara.....	V-37
5.2.2.4.3	Hasil Pengujian Rumus <i>Chebyshev Distance</i> Menggunakan Dataset Penyakit Diabetes .	V-40
5.2.2.4.4	Hasil Pengujian Rumus <i>Chebyshev Distance</i> Menggunakan Dataset Penyakit Stroke	V-42
5.3	Analisis Hasil Penelitian	V-45
5.3.1	Analisis Hasil Pengujian pada Algoritma <i>K-Nearest Neighbour</i> untuk Dataset Penyakit Serangan Jantung	V-45

5.3.2	Analisis Hasil Pengujian pada Algoritma <i>K-Nearest Neighbour</i> untuk Dataset Penyakit Kanker Payudara	V-54
5.3.3	Analisis Hasil Pengujian pada Algoritma <i>K-Nearest Neighbour</i> untuk Dataset Penyakit Diabetes	V-63
5.3.4	Analisis Hasil Pengujian pada Algoritma <i>K-Nearest Neighbour</i> untuk Dataset Penyakit Stroke	V-72
5.4	Kesimpulan	V-81
BAB VI KESIMPULAN DAN SARAN		
6.1	Pendahuluan	VI-1
6.2	Kesimpulan	VI-1
6.3	Saran.....	VI-3
DAFTAR PUSTAKA		xxix
LAMPIRAN		

DAFTAR TABEL

Halaman

Tabel II-1. Tabel Tabel <i>Confussion Matrix</i>	III-15
Tabel III-1. Tabel Jumlah Data pada Dataset Penyakit.....	III-2
Tabel IV-1. Tabel Kebutuhan Fungsional.....	IV-2
Tabel IV-2. Tabel Kebutuhan Non-Fungsional	IV-2
Tabel IV-3. Tabel Data Training.....	IV-3
Tabel IV-4. Tabel Hasil Perhitungan <i>Euclidean Distance</i>	IV-4
Tabel IV-5. Tabel Peringkat jarak Terdekat dengan Nilai K=3.....	IV-4
Tabel IV-6. Tabel Data Training.....	IV-5
Tabel IV-7. Tabel Hasil Perhitungan <i>Manhattan Distance</i>	IV-5
Tabel IV-8. Tabel Peringkat jarak Terdekat dengan Nilai K=3.....	IV-6
Tabel IV-9. Tabel Data Training.....	IV-6
Tabel IV-10. Tabel Hasil Perhitungan <i>Minkowsky Distance</i>	IV-7
Tabel IV-11. Tabel Peringkat jarak Terdekat dengan Nilai K=3.....	IV-7
Tabel IV-12. Tabel Data Training.....	IV-8
Tabel IV-13. Tabel Hasil Perhitungan <i>Chebyshev Distance</i>	IV-8
Tabel IV-14. Tabel Peringkat jarak Terdekat dengan Nilai K=3.....	IV-9
Tabel IV-15. Tabel Definisi <i>Actor</i>	IV-11
Tabel IV-16. Tabel Kegunaan pada Setiap Bagian <i>Use Case</i>	IV-11
Tabel IV-17. Tabel Implementasi Kelas Program	IV-26
Tabel IV-18. Tabel Rencana Pengujian Program Memuat File Data Training	IV-28
Tabel IV-19. Tabel Rencana Pengujian Sistem dalam Penerapan Algoritma K-Nearest Neighbour	IV-28
Tabel IV-20. Tabel Pengujian Program Memuat File Data Training	IV-29
Tabel IV-21. Tabel Pengujian Sistem dalam Penerapan Algoritma K-Nearest Neighbour	IV-30
Tabel V-1. Tabel Data Benar Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Euclidean Distance</i> dengan 100 Data <i>Testing</i>	V-3
Tabel V-2. Tabel Data Salah Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Euclidean Distance</i> dengan 100 Data <i>Testing</i>	V-3

Tabel V-3. Tabel Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing.....	V-3
Tabel V-4. Tabel Data Benar Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Euclidean Distance</i> dengan Data Testing 20% dari Dataset	V-4
Tabel V-5. Tabel Data Salah Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Euclidean Distance</i> dengan Data Testing 20% dari Dataset	V-4
Tabel V-6. Tabel Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Euclidean Distance</i> dengan Data Testing 20% dari Dataset	V-5
Tabel V-7. Tabel Data Benar Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing	V-6
Tabel V-8. Tabel Data Salah Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing	V-6
Tabel V-9. Tabel Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing.....	V-6
Tabel V-10. Tabel Data Benar Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Euclidean Distance</i> dengan Data Testing 20% dari Dataset	V-7
Tabel V-11. Tabel Data Salah Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Euclidean Distance</i> dengan Data Testing 20% dari Dataset	V-7
Tabel V-12. Tabel Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Euclidean Distance</i> dengan Data Testing 20% dari Dataset	V-7
Tabel V-13. Tabel Data Benar Terklasifikasi Penyakit Diabetes Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing	V-8
Tabel V-14. Tabel Data Salah Terklasifikasi Penyakit Diabetes Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing	V-8
Tabel V-15. Tabel Hasil Pengujian Penyakit Diabetes Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing	V-9
Tabel V-16. Tabel Data Benar Terklasifikasi Penyakit Diabetes Menggunakan <i>Euclidean Distance</i> dengan Data Testing 20% dari Dataset	V-9
Tabel V-17. Tabel Data Salah Terklasifikasi Penyakit Diabetes Menggunakan <i>Euclidean Distance</i> dengan Data Testing 20% dari Dataset	V-10
Tabel V-18. Tabel Hasil Pengujian Penyakit Diabetes Menggunakan <i>Euclidean Distance</i> dengan Data Testing 20% dari Dataset	V-10
Tabel V-19. Tabel Data Benar Terklasifikasi Penyakit Stroke Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing	V-11
Tabel V-20. Tabel Data Salah Terklasifikasi Penyakit Stroke Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing	V-11

Tabel V-21. Tabel Hasil Pengujian Penyakit Stroke Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing	V-11
Tabel V-22. Tabel Data Benar Terklasifikasi Penyakit Stroke Menggunakan <i>Euclidean Distance</i> dengan Data Testing 20% dari Dataset	V-12
Tabel V-23. Tabel Data Salah Terklasifikasi Penyakit Stroke Menggunakan <i>Euclidean Distance</i> dengan Data Testing 20% dari Dataset	V-12
Tabel V-24. Tabel Hasil Pengujian Penyakit Stroke Menggunakan <i>Euclidean Distance</i> dengan Data Testing 20% dari Dataset	V-12
Tabel V-25. Tabel Data Benar Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Manhattan Distance</i> dengan 100 Data Testing.....	V-14
Tabel V-26. Tabel Data Salah Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Manhattan Distance</i> dengan 100 Data Testing.....	V-14
Tabel V-27. Tabel Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Manhattan Distance</i> dengan 100 Data Testing	V-14
Tabel V-28. Tabel Data Benar Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Manhattan Distance</i> dengan Data Testing 20% dari Dataset	V-15
Tabel V-29. Tabel Data Salah Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Manhattan Distance</i> dengan Data Testing 20% dari Dataset	V-15
Tabel V-30. Tabel Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Manhattan Distance</i> dengan Data Testing 20% dari Dataset.....	V-15
Tabel V-31. Tabel Data Benar Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Manhattan Distance</i> dengan 100 Data Testing.....	V-16
Tabel V-32. Tabel Data Salah Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Manhattan Distance</i> dengan 100 Data Testing.....	V-17
Tabel V-33. Tabel Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Manhattan Distance</i> dengan 100 Data Testing	V-17
Tabel V-34. Tabel Data Benar Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Manhattan Distance</i> dengan Data Testing 20% dari Dataset	V-18
Tabel V-35. Tabel Data Salah Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Manhattan Distance</i> dengan Data Testing 20% dari Dataset	V-18
Tabel V-36. Tabel Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Manhattan Distance</i> dengan Data Testing 20% dari Dataset.....	V-18
Tabel V-37. Tabel Data Benar Terklasifikasi Penyakit Diabetes Menggunakan <i>Manhattan Distance</i> dengan 100 Data Testing.....	V-19
Tabel V-38. Tabel Data Salah Terklasifikasi Penyakit Diabetes Menggunakan <i>Manhattan Distance</i> dengan 100 Data Testing.....	V-19

Tabel V-39. Tabel Hasil Pengujian Penyakit Diabetes Menggunakan <i>Manhattan Distance</i> dengan 100 Data Testing	V-19
Tabel V-40. Tabel Data Benar Terklasifikasi Penyakit Diabetes Menggunakan <i>Manhattan Distance</i> dengan Data Testing 20% dari Dataset.....	V-20
Tabel V-41. Tabel Data Salah Terklasifikasi Penyakit Diabetes Menggunakan <i>Manhattan Distance</i> dengan Data Testing 20% dari Dataset.....	V-20
Tabel V-42. Tabel Hasil Pengujian Penyakit Diabetes Menggunakan <i>Manhattan Distance</i> dengan Data Testing 20% dari Dataset	V-20
Tabel V-43. Tabel Data Benar Terklasifikasi Penyakit Stroke Menggunakan <i>Manhattan Distance</i> dengan 100 Data <i>Testing</i>	V-21
Tabel V-44. Tabel Data Salah Terklasifikasi Penyakit Stroke Menggunakan <i>Manhattan Distance</i> dengan 100 Data <i>Testing</i>	V-22
Tabel V-45. Tabel Hasil Pengujian Penyakit Stroke Menggunakan <i>Manhattan Distance</i> dengan 100 Data Testing	V-22
Tabel V-46. Tabel Data Benar Terklasifikasi Penyakit Stroke Menggunakan <i>Manhattan Distance</i> dengan Data Testing 20% dari Dataset.....	V-23
Tabel V-47. Tabel Data Salah Terklasifikasi Penyakit Stroke Menggunakan <i>Manhattan Distance</i> dengan Data Testing 20% dari Dataset.....	V-23
Tabel V-48. Tabel Hasil Pengujian Penyakit Stroke Menggunakan <i>Manhattan Distance</i> dengan Data Testing 20% dari Dataset	V-23
Tabel V-49. Tabel Data Benar Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Minkowsky Distance</i> dengan 100 Data <i>Testing</i>	V-24
Tabel V-50. Tabel Data Salah Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Minkowsky Distance</i> dengan 100 Data <i>Testing</i>	V-24
Tabel V-51. Tabel Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Minkowsky Distance</i> dengan 100 Data Testing	V-25
Tabel V-52. Tabel Data Benar Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Minkowsky Distance</i> dengan Data Testing 20% dari Dataset	V-25
Tabel V-53. Tabel Data Salah Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Minkowsky Distance</i> dengan Data Testing 20% dari Dataset	V-26
Tabel V-54. Tabel Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Minkowsky Distance</i> dengan Data Testing 20% dari Dataset	V-26
Tabel V-55. Tabel Data Benar Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Minkowsky Distance</i> dengan 100 Data <i>Testing</i>	V-27
Tabel V-56. Tabel Data Salah Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Minkowsky Distance</i> dengan 100 Data <i>Testing</i>	V-27
Tabel V-57. Tabel Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Minkowsky Distance</i> dengan 100 Data Testing	V-27

Tabel V-58. Tabel Data Benar Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Minkowsky Distance</i> dengan Data Testing 20% dari Dataset	V-28
Tabel V-59. Tabel Data Salah Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Minkowsky Distance</i> dengan Data Testing 20% dari Dataset	V-28
Tabel V-60. Tabel Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Manhattan Minkowsky</i> dengan Data Testing 20% dari Dataset	V-28
Tabel V-61. Tabel Data Benar Terklasifikasi Penyakit Diabetes Menggunakan <i>Minskowsky Distance</i> dengan 100 Data Testing.....	V-29
Tabel V-62. Tabel Data Salah Terklasifikasi Penyakit Diabetes Menggunakan <i>Minskowsky Distance</i> dengan 100 Data Testing.....	V-30
Tabel V-63. Tabel Hasil Pengujian Penyakit Diabetes Menggunakan <i>Minkowsky Distance</i> dengan 100 Data Testing.....	V-30
Tabel V-64. Tabel Data Benar Terklasifikasi Penyakit Diabetes Menggunakan <i>Minskowsky Distance</i> dengan Data Testing 20% dari Dataset.....	V-31
Tabel V-65. Tabel Data Salah Terklasifikasi Penyakit Diabetes Menggunakan <i>Minskowsky Distance</i> dengan Data Testing 20% dari Dataset.....	V-31
Tabel V-66. Tabel Hasil Pengujian Penyakit Diabetes Menggunakan <i>Minkowsky Distance</i> dengan Data Testing 20% dari Dataset	V-31
Tabel V-67. Tabel Data Benar Terklasifikasi Penyakit Stroke Menggunakan <i>Minkowsky Distance</i> dengan 100 Data Testing	V-32
Tabel V-68. Tabel Data Salah Terklasifikasi Penyakit Stroke Menggunakan <i>Minkowsky Distance</i> dengan 100 Data Testing	V-32
Tabel V-69. Tabel Hasil Pengujian Penyakit Stroke Menggunakan <i>Minkowsky Distance</i> dengan 100 Data Testing	V-32
Tabel V-70. Tabel Data Benar Terklasifikasi Penyakit Stroke Menggunakan <i>Minkowsky Distance</i> dengan Data Testing 20% dari Dataset	V-33
Tabel V-71. Tabel Data Salah Terklasifikasi Penyakit Stroke Menggunakan <i>Minkowsky Distance</i> dengan Data Testing 20% dari Dataset	V-33
Tabel V-72. Tabel Hasil Pengujian Penyakit Stroke Menggunakan <i>Minkowsky Distance</i> dengan Data Testing 20% dari Dataset	V-33
Tabel V-73. Tabel Data Benar Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing	V-35
Tabel V-74. Tabel Data Salah Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing	V-35
Tabel V-75. Tabel Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Euclidean Distance</i> dengan 100 Data Testing.....	V-35

Tabel V-76	Tabel Data Benar Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Chebyshev Distance</i> dengan Data Testing 20% dari Dataset	V-36
Tabel V-77.	Tabel Data Salah Terklasifikasi Penyakit Serangan Jantung Menggunakan <i>Chebyshev Distance</i> dengan Data Testing 20% dari Dataset	V-36
Tabel V-78.	Tabel Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Chebyshev Distance</i> dengan Data Testing 20% dari Dataset	V-36
Tabel V-79	Tabel Data Benar Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-37
Tabel V-80	Tabel Data Salah Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-38
Tabel V-81.	Tabel Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-38
Tabel V-82.	Tabel Data Benar Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Chebyshev Distance</i> dengan Data Testing 20% dari Dataset	V-39
Tabel V-83.	Tabel Data Salah Terklasifikasi Penyakit Kanker Payudara Menggunakan <i>Chebyshev Distance</i> dengan Data Testing 20% dari Dataset	V-39
Tabel V-84.	Tabel Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Chebyshev Distance</i> dengan Data Testing 20% dari Dataset	V-39
Tabel V-85.	Tabel Data Benar Terklasifikasi Penyakit Diabetes Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-40
Tabel V-86.	Tabel Data Salah Terklasifikasi Penyakit Diabetes Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-40
Tabel V-87.	Tabel Hasil Pengujian Penyakit Diabetes Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-40
Tabel V-88.	Tabel Data Benar Terklasifikasi Penyakit Diabetes Menggunakan <i>Chebyshev Distance</i> dengan Data Testing 20% dari Dataset	V-41
Tabel V-89.	Tabel Data Salah Terklasifikasi Penyakit Diabetes Menggunakan <i>Chebyshev Distance</i> dengan Data Testing 20% dari Dataset	V-41
Tabel V-90.	Tabel Hasil Pengujian Penyakit Diabetes Menggunakan <i>Minkowsky Distance</i> dengan Data Testing 20% dari Dataset	V-41
Tabel V-91.	Tabel Data Benar Terklasifikasi Penyakit Stroke Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-42
Tabel V-92.	Tabel Data Salah Terklasifikasi Penyakit Stroke Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-43
Tabel V-93.	Tabel Hasil Pengujian Penyakit Stroke Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-43

Tabel V-94. Tabel Data Benar Terklasifikasi Penyakit Stroke Menggunakan <i>Chebyshev Distance</i> dengan Data Testing 20% dari Dataset	V-44
Tabel V-95. Tabel Data Salah Terklasifikasi Penyakit Stroke Menggunakan <i>Chebyshev Distance</i> dengan Data Testing 20% dari Dataset	V-44
Tabel V-96. Tabel Hasil Pengujian Penyakit Stroke Menggunakan <i>Chebyshev Distance</i> dengan Data Testing 20% dari Dataset	V-44
Tabel V-97. Tabel Perbandingan Hasil Akurasi untuk Penyakit Serangan Jantung dengan 100 Data Testing	V-45
Tabel V-98. Tabel Perbandingan Hasil Akurasi untuk Penyakit Serangan Jantung dengan Data Testing 20% dari Dataset.....	V-46
Tabel V-99. Tabel Perbandingan Hasil <i>Precision</i> untuk Penyakit Serangan Jantung dengan 100 Data Testing.....	V-48
Tabel V-100. Tabel Perbandingan Hasil <i>Precision</i> untuk Penyakit Serangan Jantung dengan Data Testing 20% dari Dataset	V-49
Tabel V-101. Tabel Perbandingan Hasil <i>Recall</i> untuk Penyakit Serangan Jantung dengan 100 Data Testing	V-50
Tabel V-102. Tabel Perbandingan Hasil <i>Recall</i> untuk Penyakit Serangan Jantung dengan Data Testing 20% dari Dataset.....	V-51
Tabel V-103. Tabel Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Serangan Jantung dengan 100 Data Testing.....	V-52
Tabel V-104. Tabel Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Serangan Jantung Data Testing 20% dari Dataset.....	V-53
Tabel V-105. Tabel Perbandingan Hasil Akurasi untuk Penyakit Kanker Payudara dengan 100 Data Testing	V-55
Tabel V-106. Tabel Perbandingan Hasil Akurasi untuk Penyakit Kanker Payudara dengan Data Testing 20% dari Dataset.....	V-56
Tabel V-107. Tabel Perbandingan Hasil <i>Precision</i> untuk Penyakit Kanker Payudara dengan 100 Data Testing	V-57
Tabel V-108. Tabel Perbandingan Hasil <i>Precision</i> untuk Penyakit Kanker Payudara dengan Data Testing 20% dari Dataset.....	V-58
Tabel V-109. Tabel Perbandingan Hasil <i>Recall</i> untuk Penyakit Kanker Payudara dengan 100 Data Testing	V-59
Tabel V-110. Tabel Perbandingan Hasil <i>Recall</i> untuk Penyakit Kanker Payudara dengan Data Testing 20% dari Dataset.....	V-59
Tabel V-111. Tabel Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Kanker Payudara dengan 100 Data Testing	V-60
Tabel V-112. Tabel Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Kanker Payudara dengan Data Testing 20% dari Dataset.....	V-61

Tabel V-113. Tabel Perbandingan Hasil Akurasi untuk Penyakit Diabetes dengan 100 Data Testing.....	V-63
Tabel V-114. Tabel Perbandingan Hasil Akurasi untuk Penyakit Diabetes dengan Data Testing 20% dari Dataset	V-65
Tabel V-115. Tabel Perbandingan Hasil <i>Precision</i> untuk Penyakit Diabetes dengan 100 Data Testing	V-66
Tabel V-116. Tabel Perbandingan Hasil <i>Precision</i> untuk Penyakit Diabetes dengan Data Testing 20% dari Dataset.....	V-67
Tabel V-117. Tabel Perbandingan Hasil <i>Recall</i> untuk Penyakit Diabetes dengan 100 Data Testing.....	V-68
Tabel V-118. Tabel Perbandingan Hasil <i>Recall</i> untuk Penyakit Diabetes dengan Data Testing 20% dari Dataset	V-69
Tabel V-119. Tabel Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Diabetes dengan 100 Data Testing	V-70
Tabel V-120. Tabel Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Diabetes Data Testing 20% dari Dataset.....	V-71
Tabel V-121. Tabel Perbandingan Hasil Akurasi untuk Penyakit Stroke dengan 100 Data Testing.....	V-73
Tabel V-122. Tabel Perbandingan Hasil Akurasi untuk Penyakit Stroke dengan Data Testing 20% dari Dataset	V-74
Tabel V-123. Tabel Perbandingan Hasil <i>Precision</i> untuk Penyakit Stroke dengan 100 Data Testing	V-75
Tabel V-124. Tabel Perbandingan Hasil <i>Precision</i> untuk Penyakit Stroke dengan Data Testing 20% dari Dataset	V-76
Tabel V-125. Tabel Perbandingan Hasil <i>Recall</i> untuk Penyakit Stroke dengan 100 Data Testing.....	V-77
Tabel V-126. Tabel Perbandingan Hasil <i>Recall</i> untuk Penyakit Stroke dengan Data Testing 20% dari Dataset	V-77
Tabel V-127. Tabel Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Stroke dengan 100 Data Testing.....	V-78
Tabel V-128. Tabel Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Stroke Data Testing 20% dari Dataset.....	V-80

DAFTAR GAMBAR

Halaman

Gambar II-1. Proses memetakan klasifikasi atribut x kedalam lebel Kelas y.....	II-6
Gambar II-2. Contoh Proses <i>Learning</i> dalam klasifikasi data	II-9
Gambar II-3. Contoh Proses <i>Classification</i> dalam klasifikasi data.....	II-9
Gambar II-4. Contoh Proses <i>Rational United Process(RUP)</i>	II-9
Gambar III-1. Kerangka Kerja Penelitian	III-2
Gambar III-2. Penjadwalan Penelitian dengan Menggunakan Gantt Chard	III-8
Gambar IV-1. <i>Use Case Program</i>	IV-10
Gambar IV-2. Rancangan <i>Interface Program</i>	IV-13
Gambar IV-3. Proses Memuat Dataset <i>Training</i>	IV-14
Gambar IV-4. Proses Klasifikasi dengan Menggunakan Algoritma <i>K-Nearest Neighbour</i> dan Rumus <i>Euclidean Distance</i>	IV-15
Gambar IV-5. Proses Klasifikasi dengan Menggunakan Algoritma <i>K-Nearest Neighbour</i> dan Rumus <i>Manhattan Distance</i>	IV-16
Gambar IV-6. Proses Klasifikasi dengan Menggunakan Algoritma <i>K-Nearest Neighbour</i> dan Rumus <i>Minkowsky Distance</i>	IV-17
Gambar IV-7. Proses Klasifikasi dengan Menggunakan Algoritma <i>K-Nearest Neighbour</i> dan Rumus <i>Chebyshev Distance</i>	IV-18
Gambar IV-8. <i>Sequence Diagram</i> Memuat Dataset <i>Training</i>	IV-19
Gambar IV-9. <i>Sequence Diagram</i> Proses Klasifikasi dengan Menggunakan Algoritma <i>K-Nearest Neighbour</i> dan Rumus <i>Euclidean Distance</i>	IV-20
Gambar IV-10. <i>Sequence Diagram</i> Proses Klasifikasi dengan Menggunakan Algoritma <i>K-Nearest Neighbour</i> dan Rumus <i>Manhattan Distance</i>	IV-21
Gambar IV-11. <i>Sequence Diagram</i> Proses Klasifikasi dengan Menggunakan Algoritma <i>K-Nearest Neighbour</i> dan Rumus <i>Minkowsky Distance</i>	IV-22
Gambar IV-12. <i>Sequence Diagram</i> Proses Klasifikasi dengan Menggunakan Algoritma <i>K-Nearest Neighbour</i> dan Rumus <i>Chebyshev Distance</i>	IV-23
Gambar IV-13. Diagram Kelas	IV-25
Gambar IV-14. Gambar Tampilan Akhir.....	IV-26

Gambar V-1. Grafik Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Euclidean Distance</i> dengan 100 Data <i>Testing</i>	V-4
Gambar V-2. Grafik Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Euclidean Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-5
Gambar V-3. Grafik Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Euclidean Distance</i> dengan 100 Data <i>Testing</i>	V-6
Gambar V-4. Grafik Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Euclidean Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-8
Gambar V-5. Grafik Hasil Pengujian Penyakit Diabetes Menggunakan <i>Euclidean Distance</i> dengan 100 Data <i>Testing</i>	V-9
Gambar V-6. Grafik Hasil Pengujian Penyakit Diabetes Menggunakan <i>Euclidean Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-10
Gambar V-7. Grafik Hasil Pengujian Penyakit Stroke Menggunakan <i>Euclidean Distance</i> dengan 100 Data <i>Testing</i>	V-12
Gambar V-8. Grafik Hasil Pengujian Penyakit Stroke Menggunakan <i>Euclidean Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-13
Gambar V-9. Grafik Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Manhattan Distance</i> dengan 100 Data <i>Testing</i>	V-15
Gambar V-10. Grafik Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Manhattan Distance</i> dengan Data <i>Testing</i> 20% dari dataset.....	V-16
Gambar V-11. Grafik Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Manhattan Distance</i> dengan 100 Data <i>Testing</i>	V-17
Gambar V-12. Grafik Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Manhattan Distance</i> dengan Data <i>Testing</i> 20% dari dataset.....	V-18
Gambar V-13. Grafik Hasil Pengujian Penyakit Diabetes Menggunakan <i>Manhattan Distance</i> dengan 100 Data <i>Testing</i>	V-20
Gambar V-14. Grafik Hasil Pengujian Penyakit Diabetes Menggunakan <i>Manhattan Distance</i> dengan Data <i>Testing</i> 20% dari dataset.....	V-21
Gambar V-15. Grafik Hasil Pengujian Penyakit Stroke Menggunakan <i>Manhattan Distance</i> dengan 100 Data <i>Testing</i>	V-22
Gambar V-16. Grafik Hasil Pengujian Penyakit Stroke Menggunakan <i>Manhattan Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-23
Gambar V-17. Grafik Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Minkowsky Distance</i> dengan 100 Data <i>Testing</i>	V-25
Gambar V-18. Grafik Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Minkowsky Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-26
Gambar V-19. Grafik Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Minkowsky Distance</i> dengan 100 Data <i>Testing</i>	V-28

Gambar V-20. Grafik Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Minkowsky Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-29
Gambar V-21. Grafik Hasil Pengujian Penyakit Diabetes Menggunakan <i>Minkowsky Distance</i> dengan 100 Data <i>Testing</i>	V-30
Gambar V-22. Grafik Hasil Pengujian Penyakit Diabetes Menggunakan <i>Minkowsky Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-31
Gambar V-23. Grafik Hasil Pengujian Penyakit Stroke Menggunakan <i>Minkowsky Distance</i> dengan 100 Data <i>Testing</i>	V-33
Gambar V-24. Grafik Hasil Pengujian Penyakit Stroke Menggunakan <i>Minkowsky Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-34
Gambar V-25. Grafik Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-36
Gambar V-26. Grafik Hasil Pengujian Penyakit Serangan Jantung Menggunakan <i>Chebyshev Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-37
Gambar V-27. Grafik Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-38
Gambar V-28. Grafik Hasil Pengujian Penyakit Kanker Payudara Menggunakan <i>Chebyshev Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-39
Gambar V-29. Grafik Hasil Pengujian Penyakit Diabetes Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-41
Gambar V-30. Grafik Hasil Pengujian Penyakit Diabetes Menggunakan <i>Chebyshev Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-42
Gambar V-31. Grafik Hasil Pengujian Penyakit Stroke Menggunakan <i>Chebyshev Distance</i> dengan 100 Data <i>Testing</i>	V-43
Gambar V-32. Grafik Hasil Pengujian Penyakit Stroke Menggunakan <i>Chebyshev Distance</i> dengan Data <i>Testing</i> 20% dari dataset	V-44
Gambar V-33. Grafik Perbandingan Hasil Akurasi untuk Penyakit Serangan Jantung dengan 100 Data <i>Testing</i>	V-46
Gambar V-34. Grafik Perbandingan Hasil Akurasi untuk Penyakit Serangan Jantung dengan Data <i>Testing</i> 20% dari dataset	V-47
Gambar V-35. Grafik Perbandingan Hasil <i>Precision</i> untuk Penyakit Serangan Jantung dengan 100 Data <i>Testing</i>	V-48
Gambar V-36. Grafik Perbandingan Hasil <i>Precision</i> untuk Penyakit Serangan Jantung dengan Data <i>Testing</i> 20% dari dataset	V-49
Gambar V-37. Grafik Perbandingan Hasil <i>Recall</i> untuk Penyakit Serangan Jantung dengan 100 Data <i>Testing</i>	V-50
Gambar V-38. Grafik Perbandingan Hasil <i>Recall</i> untuk Penyakit Serangan Jantung dengan Data <i>Testing</i> 20% dari dataset	V-51

Gambar V-39. Grafik Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Serangan Jantung dengan 100 Data <i>Testing</i>	V-52
Gambar V-40. Grafik Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Serangan Jantung dengan Data <i>Testing</i> 20% dari dataset	V-53
Gambar V-41. Grafik Perbandingan Hasil Akurasi untuk Penyakit Kanker Payudara dengan 100 Data <i>Testing</i>	V-55
Gambar V-42. Grafik Perbandingan Hasil Akurasi untuk Penyakit Kanker Payudara dengan Data <i>Testing</i> 20% dari dataset.....	V-56
Gambar V-43. Grafik Perbandingan Hasil <i>Precision</i> untuk Penyakit Kanker Payudara dengan 100 Data <i>Testing</i>	V-57
Gambar V-44. Grafik Perbandingan Hasil <i>Precision</i> untuk Penyakit Kanker Payudara dengan Data <i>Testing</i> 20% dari dataset.....	V-58
Gambar V-45. Grafik Perbandingan Hasil <i>Recall</i> untuk Penyakit Kanker Payudara dengan 100 Data <i>Testing</i>	V-59
Gambar V-46. Grafik Perbandingan Hasil <i>Recall</i> untuk Penyakit Kanker Payudara dengan Data <i>Testing</i> 20% dari dataset.....	V-60
Gambar V-47. Grafik Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Kanker Payudara dengan 100 Data <i>Testing</i>	V-61
Gambar V-48. Grafik Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Kanker Payudara dengan Data <i>Testing</i> 20% dari dataset.....	V-62
Gambar V-49. Grafik Perbandingan Hasil Akurasi untuk Penyakit Diabetes dengan 100 Data <i>Testing</i>	V-64
Gambar V-50. Grafik Perbandingan Hasil Akurasi untuk Penyakit Diabetes dengan Data <i>Testing</i> 20% dari dataset.....	V-65
Gambar V-51. Grafik Perbandingan Hasil <i>Precision</i> untuk Penyakit Diabetes dengan 100 Data <i>Testing</i>	V-66
Gambar V-52. Grafik Perbandingan Hasil <i>Precision</i> untuk Penyakit Diabetes dengan Data <i>Testing</i> 20% dari dataset.....	V-67
Gambar V-53. Grafik Perbandingan Hasil <i>Recall</i> untuk Penyakit Diabetes 100 Data <i>Testing</i>	V-68
Gambar V-54. Grafik Perbandingan Hasil <i>Recall</i> untuk Penyakit Diabetes dengan Data <i>Testing</i> 20% dari dataset	V-69
Gambar V-55. Grafik Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Diabetes dengan 100 Data <i>Testing</i>	V-70
Gambar V-56. Grafik Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Diabetes dengan Data <i>Testing</i> 20% dari dataset.....	V-71
Gambar V-57. Grafik Perbandingan Hasil Akurasi untuk Penyakit Stroke dengan 100 Data <i>Testing</i>	V-73

Gambar V-58. Grafik Perbandingan Hasil Akurasi untuk Penyakit Stroke dengan Data <i>Testing</i> 20% dari dataset	V-74
Gambar V-59. Grafik Perbandingan Hasil <i>Precision</i> untuk Penyakit Stroke dengan 100 Data <i>Testing</i>	V-75
Gambar V-60. Grafik Perbandingan Hasil <i>Precision</i> untuk Penyakit Stroke dengan Data <i>Testing</i> 20% dari dataset.....	V-76
Gambar V-61. Grafik Perbandingan Hasil <i>Recall</i> untuk Penyakit Stroke dengan 100 Data <i>Testing</i>	V-77
Gambar V-62. Grafik Perbandingan Hasil <i>Recall</i> untuk Penyakit Stroke dengan Data <i>Testing</i> 20% dari dataset	V-78
Gambar V-63. Grafik Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Stroke dengan 100 Data <i>Testing</i>	V-79
Gambar V-64. Grafik Perbandingan Hasil <i>F-Measure</i> untuk Penyakit Stroke dengan Data <i>Testing</i> 20% dari dataset.....	V-80

DAFTAR LAMPIRAN

Lampiran 1. *Source Code Program*

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini berisikan tentang pokok-pokok pikiran yang melandasi pembuatan skripsi seperti latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian serta batasan masalah.

1.2 Latar Belakang

Di zaman sekarang ini perkembangan teknologi sangatlah pesat terutama pada sistem komputer yang menggunakan prinsip *Machine learning*. Ada banyak bidang dalam kehidupan yang mulai menerapkan prinsip *Machine learning* didalamnya mulai dari bidang teknologi informasi sampai ke bidang kesehatan. Pada dasarnya prinsip *Machine Learning* merupakan pembelajaran sistem dalam mengambil keputusan menggunakan banyak sekali data sehingga sangat berkaitan erat dengan yang disebut klasifikasi data.

Klasifikasi data terbilang sangatlah penting terutama untuk dataset yang berhubungan langsung dengan kelangsungan hidup manusia sebagai contoh yaitu dataset penyakit yang sering dialami oleh manusia. Dengan adanya klasifikasi data pada sistem nantinya, diharapkan dapat memudahkan paramedis dalam mengidentifikasi penyakit yang sedang dialami oleh pasien sehingga pasien tersebut mendapatkan penanganan serta pengobatan yang lebih tepat terhadap penyakitnya. Oleh Karena itu, dibutuhkan sebuah algoritma yang dapat mendukung pada penerapan klasifikasi data tersebut.

Algoritma klasifikasi pada *Machine Learning* yang sangat populer untuk digunakan dalam mengklasifikasikan dataset yaitu Algoritma *K-Nearest Neighbour*. Hal ini disebabkan Algoritma *K-Nearest Neighbour* memiliki kelebihan yaitu dapat menghasilkan data yang kuat atau jelas dan efektif jika digunakan pada data dengan jumlah yang cukup besar (Rio, 2015). Selain itu juga Algoritma *K-Nearest Neighbour* memiliki konsistensi yang kuat. Ketika Jumlah data mendekati tak terhingga, maka algoritma ini dapat menjamin *Error Rate* yang tidak lebih dari dua kali *Bayes Error Rate*.

Pada Algoritma *K-Nearest Neighbour* terdapat empat jenis rumus perhitungan jarak yang paling sering digunakan yaitu *Euclidean distance*, *Manhattan distance*, *Minkowsky distance* dan *Chebyshev distance* (Wahyono, 2020). Dengan berpedoman pada hasil penelitian pada jurnal sebelumnya yang secara khusus membahas tentang keempat rumusan perhitungan jarak yang ada pada Algoritma *K-Nearest Neighbour* tersebut. Menunjukkan bahwa jarak *Euclidean* dan *Minkowski* pada Algoritma *K-Nearest Neighbour* dapat menghasilkan akurasi yang lebih baik dibandingkan *Chebyshev* maupun *Manhattan* dalam mengklasifikasi data tekstual. Hasil terbaik tersebut diperoleh ketika nilai *K* bernilai 3 (Wahyono, 2020).

Selain itu juga adapun sebuah penelitian yang secara khusus membandingkan antara dua rumusan perhitungan jarak yaitu *Euclidean distance* dan *Manhattan distance* mengatakan jika hasil akurasi tertinggi dari kedua rumusan perhitungan jarak tersebut sebesar 84% dengan nilai *K* bernilai 3. Hasil penelitian ini menunjukkan bahwa Metode Pendekatan *Manhattan Distance* memiliki nilai selisih

2,04% lebih tinggi dibandingkan dengan *Euclidean Distance*. Persentase tersebut menunjukkan bahwa *Manhattan Distance* lebih akurat dibandingkan dengan *Euclidean Distance* sehingga *Manhattan Distance* bekerja dengan baik dalam memberikan rekomendasi untuk klasifikasi transportasi bus (Rozzi, 2020).

Berdasarkan penelitian-penelitian yang telah dilakukan sebelumnya maka penelitian kali ini bertujuan untuk membuktikan dari beberapa jenis rumus perhitungan jarak yang ada pada Algoritma *K-Nearest Neighbour* manakah di antaranya yang memiliki akurasi dengan nilai kinerja yang baik jika digunakan dalam mengklasifikasikan dataset tentang berbagai jenis penyakit.

1.3 Rumusan Masalah

Berdasarkan permasalahan dari latar belakang di atas, maka adapun permasalahan yang dapat diambil adalah:

1. Apakah perbedaan rumusan perhitungan jarak akan mempengaruhi hasil dari Algoritma *K-Nearest Neighbour*?
2. Berapakah nilai K yang dapat memberikan nilai kinerja terbaik untuk masalah klasifikasi data penyakit pada manusia dengan menggunakan perhitungan jarak yang berbeda-beda pada Algoritma *K-Nearest Neighbour*?
3. Bagaimana hasil perbandingan dari metode perhitungan jarak yang berbeda dalam Algoritma *K-Nearest Neighbour* pada masalah klasifikasi data penyakit pada manusia?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Untuk mengetahui apakah dengan menggunakan berbagai jenis perhitungan jarak yang berbeda-beda pada Algoritma *K-Nearest Neighbour* dapat mempengaruhi hasil klasifikasi terhadap dataset berbagai jenis penyakit pada manusia.
2. Bertujuan untuk menentukan pada nilai K beberapa yang mampu memberikan hasil yang memiliki kinerja terbaik pada klasifikasi dataset berbagai penyakit pada manusia.
3. Untuk mengetahui dari hasil perbandingan antar setiap rumusan perhitungan jarak manakah yang paling banyak memiliki nilai kinerja terbaik untuk digunakan pada masalah klasifikasi penyakit pada manusia.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah:

1. Dengan adanya sistem ini akan memberikan pemahaman tentang bagaimana cara Algoritma *K-Nearest Neighbour* dapat bekerja dalam menyelesaikan suatu masalah yang ada dengan menggunakan rumusan perhitungan jarak yang berbeda-beda.
2. Juga dengan adanya penelitian ini diharapkan dapat mengetahui dari keempat rumus penghitungan jarak yang ada pada Algoritma *K-Nearest Neighbour*, manakah di antaranya yang memiliki nilai kinerja terbaik dalam mengklasifikasikan data sehingga cocok digunakan untuk memecahkan semua masalah yang ada nantinya.

1.6 Batasan Masalah

Batasan-batasan masalah dalam tugas akhir ini dibuat agar dalam pengerjaannya lebih terarah dan tertuju pada tujuan yang tertera diatas, adapun batasan-batasan tersebut di antaranya:

1. Penelitian ini menggunakan empat jenis rumus perhitungan jarak pada Algoritma *K-Nearest Neighbour* yaitu *Euclidean distance*, *Manhattan distance*, *Minkowsky distance* dan *Chebyshev distance*.
2. Serta menggunakan empat macam dataset jenis penyakit yang berbeda-beda yaitu Penyakit Serangan Jantung, Kanker Payudara, Diabetes, dan Stroke.

1.7 Sistematika Penulisan

BAB I. PENDAHULUAN

Bab ini berisikan pokok-pokok pikiran yang menjadi landasan dalam penelitiannya seperti latar belakang, rumusan masalah, tujuan dan manfaat penelitian serta batasan masalah.

BAB II. KAJIAN LITERATUR

Bab ini berisikan teori-teori dasar yang digunakan dalam proses penelitian pada skripsi. Sebagai gambarannya bab ini akan menerangkan definisi dari berbagai teori seperti apa itu Algoritma *K-Nearest Neighbour*, *Euclidean distance*, *Manhattan distance*, *Minkowsky distance*, *Chebyshev distance* dan lain-lain.

BAB III. METODOLOGI PENELITIAN

Bab ini membahas tentang tahapan-tahapan yang ada pada penelitian. Setiap tahapan penelitian yang ada akan dijelaskan

dengan secara rinci dalam sebuah kerangka kerja. Hingga pada bagian akhir bab ini berisikan tentang perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan membahas tentang bagaimana cara pengembangan dari sistem yang akan kami buat tersebut mulai dari mulai dari rancangan, fungsional, non-fungsional, ataupun use-case yang akan digunakan dalam sistem tersebut.

BAB V. HASIL DAN ANALISIS

Bab ini berisikan tentang hasil yang dicapai selama melakukan penelitian. Lalu dianalisis untuk menarik kesimpulan dari hasil yang didapat selama melakukan penelitian tersebut.

BAB VI. KESIMPULAN DAN SARAN

Pada Bab ini berisikan kesimpulan yang didapat setelah melakukan analisis terhadap hasil yang pada bab sebelumnya. Selain berisikan kesimpulan bab ini juga berisikan saran agar sistem yang kami rancang serta dibuat ini dapat dikembangkan menjadi lebih baik lagi.

DAFTAR PUSTAKA

- Amir, Amiza dkk. 2017. *The Performance Analysis of K-Nearest Neighbors (K-NN) Algorithm for Motor Imagery Classification Based on EEG Signal*.
- Baharuddin, Mus Mulyadi., Hasanuddin, Tasrif., Azis, Huzain. 2019. *Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca*.
- Dinata, Rozzi Kesuma., Hasdyna, Novia. 2020. *Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus*.
- Gie, Wen., Jollyta, Deny. 2020. *Perbandingan Euclidean dan Manhattan untuk Optimasi Cluster Menggunakan Davies Bouldin Index: Status Covid-19 Wilayah Riau*.
- Guo, Gongde., Wang, Hui., Bi, Yaxin., Bell, David. 2004. *KNN Model-Based Approach in Classification*.
- Leidiyana, Henny. 2013. *Penerapan Algoritma K-Nearest Neighbor untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor*.
- Nishom, Muhammad. 2019. *Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square*.
- Prasath, Surya dkk. 2019. *Distance and similarity measures effect on the performance of k-nearest neighbor classifier - a review*.
- Primartha, Rifkie. 2018. *Belajar Machine Learning Teori dan Praktik*. Bandung: Informatika.
- Sani, Ramadhan Rakhmat., Zeniarja, Junta., Luthfiarta, Ardytha. 2016. *Penerapan Algoritma K-Nearest Neighbor pada Information Retrieval dalam Penentuan Topik Referensi Tugas Akhir*.
- Sinwar, Deepak., Kaushik, Rahul. 2014. *Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering*. International Journal for Research in Applied Science and Engineering Technology (IJRASET), vol. 2, no. 5, pp. 270-274.
- Wahyono., Trisna, I Nyoman Prayana., Sariwening, Sarah Lintang., Fajar, Muhammad., Wijayanto, Danur. 2020. *Perbandingan Perhitungan Jarak pada K-Nearest Neighbour dalam Klasifikasi Data Tekstual*.
- Wiyono, Slamet., Abidin, Taufiq. 2018. *Implementation of K-Nearest Neighbor (KNN) Algorithm to Predict Student's Performance*.