

**PERBANDINGAN METODE PENGUKURAN JARAK PADA
ALGORITMA *K-NEAREST NEIGHBOR* UNTUK
KLASIFIKASI**

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



Oleh:

Taca Rosa
NIM : 09021281520100

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2019**

LEMBAR PENGESAHAN TUGAS AKHIR

PERBANDINGAN METODE PENGUKURAN JARAK PADA ALGORITMA *K-NEAREST NEIGHBOR* UNTUK KLASIFIKASI

Oleh :

TACA ROSA
NIM : 09021281520100

Palembang, 25 Juli 2019

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP. 197706012009121004

Pembimbing,

Rifkie Primartha, M.T.
NIP. 197706012009121004


TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari **Jumat** tanggal **19 Juli 2019** telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Taca Rosa
N I M : 09021281520100
Judul : Perbandingan Metode Pengukuran Jarak pada Algoritma *K-Nearest Neighbor* untuk Klasifikasi

1. Pembimbing

Rifkie Primartha, M.T.
NIP. 197706012009121004



.....

2. Penguji I

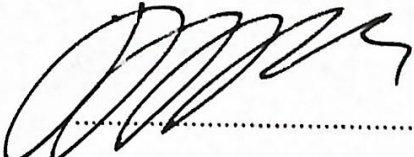
Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



.....

3. Penguji II

Danny Matthew Saputra, M.Sc
NIP. 198505102015041002



.....

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha S.T. M.T.
NIP 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Taca Rosa
NIM : 09021281520100
Program Studi : Teknik Informatika
Judul Skripsi : Perbandingan Metode Pengukuran Jarak pada Algoritma
K-Nearest Neighbor untuk Klasifikasi
Hasil Pengecekan Software *iThenticate/Turnitin* : 19%

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, 25 Juli 2019



Taca Rosa
NIM. 09021281520100

Motto :

- *Be better than you were yesterday*
- *Let them talk anything about you, then silence them with your successful.*

Kupersembahkan karya tulis ini kepada :

- *Orang tuaku tersayang*
- *Keluarga besarku*
- *Sahabat dan teman seperjuanganku*
- *Fakultas Ilmu Komputer*
- *Universitas Sriwijaya*

PERBANDINGAN METODE PENGUKURAN JARAK PADA ALGORITMA *K-NEAREST NEIGHBOR* UNTUK KLASIFIKASI

Oleh :
Taca Rosa
09021281520100

ABSTRACT

K-Nearest Neighbor is a non-parametric classification algorithm that does not use training data and does not use initial assumptions or models in the calculation process. The quality of the classification results of the k-Nearest Neighbor algorithm is very dependent on the distance between object and value of k specified, so the selection of method for distance measurement determines the results of classification. In this study, several methods of measuring Euclidean distance, Manhattan distance, Tchebychev distance and Cosine distance were examined to see distance measurement methods that can be used optimally on the k-Nearest Neighbor algorithm. The selection of k values also determines the results of the classification of the k-Nearest Neighbor algorithm, for which the determination of the value of k also needs to be considered. This study uses a cervical cancer dataset and provides the highest accuracy results in the Cosine distance measurement method that is equal to 92.559% at the value of k=9, while for the Manhattan distance measurement method is 91.666% with a value of k=13, the Tchebychev distance measurement method is 91.666% with the value of k=13 and the lowest accuracy value obtained by the Euclidean method that is equal to 91.071% at the value of k=13. Based on the value of accuracy, the most compatible distance measurement method used is Cosine distance even with the best k value obtained is k=9 though it has the highest computation time which is 0.898 s compared to the Euclidean method of 0.555 s, the Manhattan method is 0.5882 s and Tchebychev has the lowest computation time which is 0.49 s.

Key Word: *Distance measurement, K-Nearest Neighbor, Euclidean Distance, Manhattan Distance, Tchebychev Distance, Cosine Distance*

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP. 197706012009121004

Pembimbing,

Rifkie Primartha, M.T.
NIP. 197706012009121004

PERBANDINGAN METODE PENGUKURAN JARAK PADA ALGORITMA *K-NEAREST NEIGHBOR* UNTUK KLASIFIKASI

Oleh :
Taca Rosa
09021281520100

ABSTRAK

K-Nearest Neighbor merupakan algoritma klasifikasi non-parametik yang tidak menggunakan *training data* dan tidak menggunakan asumsi atau model awal pada proses perhitungan. Kualitas hasil klasifikasi algoritma *k-Nearest Neighbor* sangat bergantung pada jarak kedekatan antar objek dan nilai dari *k* yang ditetapkan, maka pemilihan metode untuk perhitungan jarak sangat menentukan hasil pengelompokan. Pada penelitian ini, dilakukan perbandingan beberapa metode pengukuran jarak *Euclidean distance*, *Manhattan distance*, *Tchebychev distance* dan *Cosine distance* untuk melihat metode pengukuran jarak yang dapat bekerja secara optimal pada algoritma *k-Nearest Neighbor*. Pemilihan nilai *k* juga menentukan hasil klasifikasi algoritma *k-Nearest Neighbor*, untuk itu penetapan nilai *k* juga perlu diperhatikan. Penelitian ini menggunakan dataset kanker serviks dan memberikan hasil akurasi tertinggi pada metode pengukuran jarak *Cosine distance* yaitu sebesar 92,559% pada nilai *k*=9, sedangkan untuk metode pengukuran jarak *Manhattan* sebesar 91,666% dengan nilai *k*=13, pada metode pengukuran jarak *Tchebychev* sebesar 91,666% dengan nilai *k*=13 dan nilai akurasi terendah diperoleh oleh metode *Euclidean* yaitu sebesar 91,071% pada nilai *k*=13. Berdasarkan nilai akurasi tersebut diperoleh metode pengukuran jarak yang tepat digunakan adalah *Cosine distance* dengan nilai *k* terbaik yaitu *k*=9 meskipun memiliki waktu komputasi tertinggi yaitu 0,898 s dibandingkan metode *Euclidean* sebesar 0,555 s, metode *Manhattan* sebesar 0,5882 s dan *Tchebychev* memiliki waktu komputasi yang terendah yaitu 0,49 s.

Key Word: Pengukuran Jarak,, *K-Nearest Neighbor*, *Euclidean Distance*, *Manhattan Distance* ,*Tchebychev Distance*,*Cosine Distance*

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP. 197706012009121004

Pembimbing,

Rifkie Primartha, M.T.
NIP. 197706012009121004

KATA PENGANTAR



Puji syukur kepada Allah SWT 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. Allah Subhanahu wa Ta'ala yang telah memberikan hamba keimanan, kesehatan, kecerdasan, kemudahan, dan kelancaran sehingga hamba dapat menyelesaikan tugas-tugas hamba sebagai seorang mahasiswa.
2. Nabi Muhammad Shallallahu „Alaihi wa Sallam yang telah memberikan jalan yang terang dan memberikan ilmu yang bermanfaat kepada umatnya atas seizin Allah.
3. Bapak Jaidan Jauhari, S.Pd., M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Rifkie Primartha, M.T selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya dan sekaligus pembimbing yang telah mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir serta Ibu Hardini Novianti, M.T. selaku Sekretaris Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Yunita, S.SI, M.CS. selaku Pembimbing Akademik pada Jurusan Teknik Informatika Universitas Sriwijaya.
6. Ibu Alvi Syahrini Utami, M.Kom selaku dosen penguji I dan Bapak Danny Matthew Saputra, M.Sc selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
7. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
8. Orang tua saya, Jaya Utami dan Laily Nasroh yang telah memberikan dorongan dan dukungan baik secara moril maupun materil, serta kepada kakak perempuan saya yang saya cintai dan sayangi dengan segenap jiwa dan raga, Putri Fahadis.
9. Mbak Winda dan Kak Ricy serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
10. Mega Rizky, Arfah Anggraina, Vira Melinda, Faiz Muhammad, Ahmad Halim Prabowo, Arief Rachmatullah, Gian Agnar Moslem, Muhammad Ikrom serta seluruh teman-teman jurusan Teknik Informatika baik angkatan 2015, 2014, dan 2013 yang telah berbagi keluh kesah, motivasi, dan canda tawa di masa-masa perkuliahan ini.
11. BEM KM Fasilkom Unsri.

12. Mikyal Marshalita, Mardiah Aqidah Islamiah, dan Zainul Arifin, teman sejawat yang telah menemani penulis dalam penulisan Tugas Akhir.

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.

Palembang, 25 Juli 2019

Taca Rosa

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN 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.....	xv
BAB I.....	I-1
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 Sistematika Penulisan	I-5
1.8 Kesimpulan	I-7
BAB II.....	II-1
2.1 Pendahuluan.....	II-1
2.2 Penelitian Terkait.....	II-1
2.3 <i>K-Nearest Neighbor</i>	II-3
2.4 <i>Euclidean Distance</i>	II-5
2.4 <i>Manhattan Distance</i>	II-6
2.4 <i>Tchebychev Distance</i>	II-7
2.4 <i>Cosine Distance</i>	II-7
2.4 <i>Confussion Matrix</i>	II-7
2.4 Kesimpulan.....	II-9
BAB III	III-1
3.1 Pendahuluan.....	III-1
3.2 Unit Penelitian	III-1
3.3 Data.....	III-1
3.3.1 Jenis dan Sumber Data	III-1
3.3.2 Metode Pengumpulan Data	III-2

3.4 Tahapan Penelitian.....	I-2
3.4.1 Menetapkan Kerangka Kerja / Framework	III-2
3.4.2 Menetapkan Kriteria Pengujian	III-4
3.4.3 Menetapkan Format Data Pengujian	III-5
3.4.4 Menetapkan Alat Yang Digunakan Dalam Pelaksanaan Penelitian	III-8
3.4.5 Melakukan Pengujian Penelitian	III-8
3.4.6 Melakukan Analisis Hasil Pengujian dan Kesimpulan	III-9
3.5 Metode Pengembangan Perangkat Lunak	III-9
3.5.1 Fase Insepsi	III-10
3.5.2 Fase Elaborasi.....	III-11
3.5.3 Fase Konstruksi	III-11
3.5.4 Fase Transisi	III-12
3.5 Manajemen Perangkat Lunak	III-12
BAB IV	IV-1
4.1 Pendahuluan.....	IV-1
4.2 Fase Insepsi.....	IV-1
4.2.1 Permodelan Bisnis	IV-2
4.2.2 Kebutuhan Sistem.....	IV-3
4.2.2.1 Fitur Praproses	IV-3
4.2.2.2 Fitur Pengukuran Jarak	IV-3
4.2.2.3 Fitur Klasifikasi.....	IV-4
4.2.3 Analisis dan Desain	IV-5
4.2.3.1 Analisis Kebutuhan Perangkat Lunak	IV -5
4.2.3.2 Analisis Data	IV -6
4.2.3.3 Analisis Praproses Data	IV -7
4234 Analisis Perhitungan Metode <i>Euclidean Distance</i>	IV -9
4235 Analisis Perhitungan Metode <i>Manhattan Distance</i> ..	IV -10
4236 Analisis Perhitungan Metode <i>Tchebychev Distance</i> .	IV -10
4237 Analisis Perhitungan Metode <i>Cosine Distance</i>	IV -11
4.2.3.8 Analisis <i>K-Nearest Neighbor</i>	IV -12
4.2.3.9 Desain Perangkat Lunak	IV -13
4.3 Fase Elaborasi	IV -27
4.3.1 Pemodelan Bisinis	IV-27
4.3.1.1 Perancangan Data.....	IV-28
4.3.1.2 Perancangan Antar Muka.....	IV-28
4.3.2 Kebutuhan Sistem.....	IV-29
4.3.2 Diagram <i>Sequence</i>	IV-29
4.4 Fase Konstruksi	IV-32
4.4.1 Kebutuhan Sistem.....	IV-32

4.4.2	Diagram Kelas.....	V-32
4.4.3	Implementasi	IV-33
4.4.3.1	Implementasi Kelas.....	IV-33
4.4.3.2	Implementasi Antarmuka.....	IV-35
4.5	Fase Transisi	IV-35
4.5.1	Pemodelan Bisnis	IV-36
4.5.2	Kebutuhan Sistem.....	IV-36
4.5.3	Rencana Pengujian	IV-36
4.5.3.1	Rencana Pengujian Use Case Melakukan Praproses Data	IV-36
4.5.3.2	Rencana Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Euclidean</i>	IV-37
4.5.3.3	Rencana Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Manhattan</i>	IV-37
4.5.3.4	Rencana Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Tchebychev</i>	IV-38
4.5.3.5	Rencana Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Cosine</i>	IV-39
4.5.4	Implementasi	IV-39
4541	Pengujian Use Case Melakukan Praproses Data.....	IV-40
4542	Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Euclidean</i>	IV-41
4543	Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Manhattan</i>	IV-42
4544	Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Tchebychev</i>	IV-44
4545	Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Cosine</i>	IV-45
4.6	Kesimpulan.....	IV-47
BAB V	V-1
5.1	Pendahuluan.....	V-1
5.2	Data Hasil Percobaan/ Penelitian.....	V-2
5.2.1	Konfigurasi Percobaan	V-1
5.2.2	Data Hasil Klasifikasi.....	V-1
5.2.3	Nilai Akurasi, <i>Sensitivity</i> , <i>Specificity</i> , dan waktu Komputasi ...	V-4
5.3	Analisis Hasil Penelitian.....	V-10
5.4	Kesimpulan.....	V-13

BAB VI.....	I-1
6.1 Pendahuluan.....	VI-1
6.2 Kesimpulan.....	VI-1
6.3 Saran	VI-2
DAFTAR PUSTAKA	xx

DAFTAR TABEL

Tabel III-1 Rancangan Tabel <i>Confussion Matrix</i> Hasil klasifikasi.....	I-5
Tabel III-2 Rancangan Tabel Hasil Analisa Akurasi Algoritma <i>K-Nearest Neighbor</i> Terhadap Metode Pengukuran Jarak.....	III-6
Tabel III-3 Rancangan Tabel Hasil Analisa <i>Sensitivity</i> Algoritma <i>K-Nearest Neighbor</i> Terhadap Metode Pengukuran Jarak.....	III-6
Tabel III-4 Rancangan Tabel Hasil Analisa <i>Specificity</i> Algoritma <i>K-Nearest Neighbor</i> Terhadap Metode Pengukuran Jarak.....	III-7
Tabel III-5 Rancangan Tabel Hasil Analisa Waktu Komputasi Algoritma <i>K-Nearest Neighbor</i> Terhadap Metode Pengukuran Jarak	III-7
Tabel III-6 Rancangan Tabel Hasil Analisa Klasifikasi	III-9
Tabel III-7 Tabel Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure</i> (WBS)	III-13
Tabel IV-1. Tabel Kebutuhan Fungsional.....	IV-4
Tabel IV-2. Tabel Kebutuhan Non Fungsional.....	IV-4
Tabel IV-3 . Contoh Data Uji Coba	IV-7
Tabel IV-4 . Hasil Penghilangan Data <i>Outlier</i>	IV-8
Tabel IV-5 . Hasil Penghilangan <i>Missing Value</i>	IV-8
Tabel IV-6 . Sampel data <i>testing</i>	IV-9
Tabel IV-7 . Sampel data <i>training</i>	IV-9
Tabel IV-8. Hasil perhitungan <i>Euclidean distance</i>	IV-10
Tabel IV-9. Hasil perhitungan <i>Manhattan distance</i>	IV-10

Tabel IV-10. Hasil perhitungan <i>Chebyshev distance</i>	V-11
Tabel IV-11. Hasil perhitungan <i>Coseine distance</i>	IV-11
Tabel IV-12. Pengelompokkan Data Hasil Pengurutan Berdasarkan Jarak Kedekatan.....	IV-12
Tabel IV-13. Hasil Klasifikasi Menggunakan Algoritma <i>K-Nearest Neighbor</i>	IV-12
Tabel IV – 14. Definisi Aktor <i>Use Case</i>	IV-14
Tabel IV – 15. Definisi <i>Use Case</i>	IV-15
Tabel IV – 16. Skenario Melakukan Praproses Data	IV-16
Tabel IV-17. Skenario <i>Use Case</i> Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Euclidean</i>	IV-17
Tabel IV-18. Skenario <i>Use Case</i> Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Manhattan</i>	IV-19
Tabel IV-19. Skenario <i>Use Case</i> Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Tchebychev</i>	IV-20
Tabel IV-20. Skenario <i>Use Case</i> Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Cosine</i>	IV-22
Tabel IV-21 Tabel Implementasi Kelas	IV-34
Tabel IV-22. Rencana Pengujian <i>Use Case</i> Melakukan Praproses Data	IV-36
Tabel IV-23. Rencana Pengujian <i>Use Case</i> Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Euclidean</i>	IV-37
Tabel IV-24. Rencana Pengujian <i>Use Case</i> Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Manhattan</i>	IV-38
Tabel IV- 25. Rencana Pengujian <i>Use Case</i> Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Tchebychev</i>	IV-38

Tabel IV- 26. Rencana Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Cosine</i>	V-39
Tabel IV- 27. Pengujian Use Case Melakukan Praproses Data	IV-40
Tabel IV-28. Rencana Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Euclidean</i>	IV-41
Tabel IV- 29 Rencana Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Manhattan</i>	IV-42
Tabel IV -30. Rencana Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Tchebychev</i>	IV-44
Tabel IV- 31. Rencana Pengujian Use Case Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Cosine</i>	IV-45
Tabel V-1 Tabel <i>Confussion Matrix</i> Hasil klasifikasi.....	V-3
Tabel V-2 Tabel Hasil Analisa Akurasi	V-5
Tabel V-3 Tabel Hasil Analisa <i>Specificity</i>	V-6
Tabel V-4 Tabel Hasil Analisa <i>Sensitivity</i>	V-8
Tabel V-5 Tabel Hasil Analisa Waktu Komputasi	V-9
Tabel V-6 Tabel Hasil Analisa Klasifikasi	V-22

DAFTAR GAMBAR

Gambar II-1. Metode <i>k-nearest neighbor</i>	I-4
Gambar II-2 Perbedaan <i>Manhattan distance</i> dan <i>Euclidean distance</i>	II-6
Gambar II-3. Contoh tabel <i>confusion matrix</i>	II-8
Gambar III-1. Tahapan Pengujian Penelitian.....	III-8
Gambar IV-1. Diagram <i>Use Case</i>	IV-14
Gambar IV-2. Diagram Aktivitas <i>Use Case</i> Melakukan Praproses Data	IV-23
Gambar IV-3. Diagram Aktivitas <i>Use Case</i> Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Euclidean</i>	IV-24
Gambar IV-4. Diagram Aktivitas <i>Use Case</i> Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Manhattan</i>	IV-25
Gambar IV- 5. Diagram Aktivitas <i>Use Case</i> Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Tchebychev</i>	IV-26
Gambar IV-6. Diagram Aktivitas <i>Use Case</i> Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan <i>Cosine</i>	IV-27
Gambar IV-7. Rancangan Antarmuka Perangkat Lunak	IV-28
Gambar IV-8. Diagram <i>Sequence</i> Melakukan Praproses Data	IV-30
Gambar IV-9. Diagram <i>Sequence</i> Melakukan Proses Klasifikasi dengan <i>K-Nearest Neighbor</i> dan 4 Metode Pengukuran Jarak <i>Euclidean</i> , <i>Manhattan</i> , <i>Tchebvchev</i> , dan <i>Cosine</i>	IV-31
Gambar IV-9. Diagram Kelas	IV-33
Gambar IV-15. Antarmuka Perangkat Lunak	IV-35
Gambar V-1. Perbandingan Nilai Akurasi	V-18

Gambar V-2. Perbandingan Nilai <i>Specificity</i>	V-19
Gambar V-3. Perbandingan Nilai <i>Sensitivity</i>	V-20
Gambar V-4. Perbandingan Waktu Komputasi	V-21

DAFTAR LAMPIRAN

LAMPIRAN 1 Kode Program	L-1
-------------------------------	-----

BAB I

PENDAHULUAN

1.1. Pendahuluan

Pada bab ini membahas latar belakang masalah, rumusan masalah, tujuan, manfaat penelitian, batasan masalah, dan sistematika penulisan. Kemudian akan dijelaskan secara umum mengenai keseluruhan penelitian.

Pendahuluan dimulai dengan penjelasan mengenai perkembangan data, pengolahan data dibidang data mining, 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

Banyaknya data yang dihasilkan setiap harinya dan semakin pesatnya perkembangan teknologi media penyimpanan digital mengakibatkan terjadi ledakan data. Kondisi ini membuat teknik analisis data secara tradisional sudah tidak dapat diterapkan dalam pengolahan data tersebut. Semakin pesatnya perkembangan teknologi, proses pencarian informasi dari sejumlah data yang banyak dilakukan dengan *data mining*. Data mining merupakan metode yang melakukan proses ekstraksi dari sejumlah data sehingga didapatkan pola data yang nantinya akan menjadi suatu informasi atau pengetahuan baru (Kurniawan & Ivandari, 2018). Dalam proses *data mining*, terdapat data berjenis *supervised* dan

unsupervised dimana data *supervised* ditandai dengan adanya label sedangkan data *unsupervised* tidak memiliki label data.

Teknik klasifikasi yang merupakan salah satu fungsi utama data mining, dapat digunakan untuk proses pengelompokkan data dari data yang telah ada dengan menggunakan data berlabel atau data *supervised*. Salah satu algoritma yang banyak digunakan untuk klasifikasi adalah *k-Nearest Neighbor*. *K-Nearest Neighbor* merupakan algoritma non-parametik yang tidak menggunakan training data dan tidak menggunakan asumsi atau model awal pada proses perhitungan, melainkan menggunakan hipotesis secara langsung berdasarkan data training yang disediakan (Primartha, 2018; Walters-Williams & Li, 2010) . Algoritma *k-Nearest Neighbor* menggunakan pendekatan *supervised learning* dimana data yang digunakan merupakan data berlabel. Selain itu, algoritma ini sederhana dan mudah diimplementasikan. Meskipun sederhana, algoritma ini telah diuji di beberapa kasus dan menghasilkan performa yang cukup tinggi.

Pada penelitian Binabar and Ivandari (2018) menggunakan algoritma *k-Nearest Neighbor* dalam mencari parameter k yang paling optimal terhadap dataset *breast cancer wisconsin* yang diakses pada *UCI machine learning repository*, menunjukkan bahwa algoritma *k-Nearest Neighbor* dengan nilai $k = 13$ memiliki tingkat akurasi yaitu 97,28% dengan nilai error sebesar 1,5% dan nilai mikro sebesar 97,28%. Ini menunjukkan bahwa penggunaan algoritma *k-nearest neighbour* untuk data medis *breast cancer*, dapat menghasilkan performa yang cukup tinggi meskipun masih dapat ditingkatkan lagi.

Kualitas hasil pengelompokkan algoritma *k-Nearest Neighbor* sangat bergantung pada jarak kedekatan antar objek dan nilai dari k yang ditetapkan (Medjahed, Saadi, & Benyettou, 2013; Weinberger & Saul, 2009). Medjahed et al. (2013) melakukan penelitian pada algoritma *k-nearest neighbour* dengan membandingkan beberapa fungsi pengukuran jarak. Pada penelitian tersebut, dilakukan perbandingan metode pengukuran jarak antara *euclidean distance*, *manhattan distance*, *tchebychev distance*, *cosine distance* dan *correlation distance*. Penelitian ini memberikan hasil akurasi terbaik pada 2 metode yaitu *euclidean distance* dan *manhattan distance*, dimana pengukuran dengan metode tersebut berhasil memberikan tingkat akurasi sebesar 98,70% pada $k=1$. Dikarenakan metode *k-Nearest Neighbor* sangat bergantung pada hasil perhitungan jarak antar objek, maka pemilihan metode untuk perhitungan jarak sangat menentukan hasil pengelompokkan.

Berdasarkan hal tersebut, maka penelitian ini akan menekankan pada pemilihan metode pengukuran jarak antara *euclidean*, *manhattan*, *tchebychev* dan *cosine* yang dapat meningkatkan kinerja algoritma *k-Nearest Neighbor* untuk klasifikasi.

1.3. Rumusan Masalah

Rumusan masalah yang dibahas dalam penulisan adalah metode pengukuran jarak apa yang dapat bekerja secara optimal terhadap kinerja algoritma *k-Nearest Neighbor* untuk klasifikasi. Untuk menjawab rumusan

masalah tersebut, maka pada penelitian ini dibagi menjadi beberapa *research question* sebagai berikut:

1. Bagaimana cara mengetahui metode pengukuran jarak yang dapat bekerja optimal terhadap kinerja algoritma *k-Nearest Neighbor* pada data mining klasifikasi?
2. Bagaimana cara menetapkan nilai *k* yang sesuai untuk optimalisasi perhitungan algoritma *k-Nearest Neighbor* ?

1.4. Tujuan Penelitian

Tujuan dilakukannya penelitian adalah sebagai berikut:

1. Mengetahui metode pengukuran jarak apa yang dapat bekerja optimal terhadap kinerja algoritma *k-Nearest Neighbor* pada data mining klasifikasi.
2. Mengetahui nilai *k* yang sesuai untuk optimalisasi perhitungan algoritma - *k-Nearest Neighbor*.

1.5. Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut :

1. Mengetahui metode pengukuran jarak yang tepat terhadap kinerja algoritma *k-Nearest Neighbor*.
2. Memahami cara menentukan nilai *k* yang sesuai pada perhitungan algoritma *k-Nearest Neighbor*.

3. Mampu menerapkan algoritma *k-Nearest Neighbor* pada data mining klasifikasi.

1.6. Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Penelitian menggunakan 4 metode pengukuran jarak *Euclidean*, *Manhattan*, *Tchebychev* dan *Cosine*.
2. Data yang dijadikan objek penelitian merupakan dataset kanker serviks yang diperoleh pada situs *UCI Machine Learning Repository*.
3. Evaluasi hasil klasifikasi berdasarkan perhitungan *confussion matrix* dan waktu komputasi. Kemudian dari *confussion matrix*, akan ditetapkan *accuracy*, *sensitivity* dan *specificity*.

1.7. Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut :

BAB I. PENDAHULUAN

Pada bab pendahuluan akan diuraikan mengenai latar belakang, rumusan masalah, manfaat dan tujuan penelitian, batasan masalah atau ruang lingkup penelitian dan juga sistematika penulisan.

BAB II. KAJIAN LITERATUR

Dalam bab ini akan diuraikan *literature review* yang dilakukan terhadap jurnal, buku, dan artikel, berkaitan dengan metode *k-*

Nearest Neighbor, metode-metode pengukuran jarak *Euclidean*, *Manhattan*, *Tchebychev* dan *Cosine* serta pengukuran performa pada algoritma klasifikasi dan penelitian yang berkaitan dengan metode yang digunakan.

BAB III. METODOLOGI PENELITIAN

Dalam bab ini akan diuraikan secara jelas tentang data, tahapan-tahapan penelitian, metode pengembangan perangkat lunak, serta manajemen proyek penelitian yang berupa penjadwalan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Dalam bab ini akan dibahas pengembangan alat penelitian metode pengukuran jarak *Euclidean*, *Manhattan*, *Tchebychev*, dan *Cosine* pada algoritma *K-Nearest Neighbor* yang berupa perancangan dan lingkungan implementasi, kebutuhan, pemodelan bisnis, analisis dan desain, implementasi serta pengujian.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini, hasil pengujian dipaparkan berdasarkan langkah-langkah yang telah direncanakan. Proses analisis dilakukan dengan menarik kesimpulan dari hasil penelitian yang dilakukan.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini dipaparkan kesimpulan dari penelitian yang telah dilakukan dan berisi saran-saran yang diharapkan berguna untuk pengembangan selanjutnya.

1.8. Kesimpulan

Berdasarkan uraian pada bab I ini, dijelaskan bahwa masalah yang dihadapi algoritma *k-Nearest Neighbor* adalah tingkat performansi algoritma yang masih dapat dikembangkan. Terdapat sebuah pendekatan yang dianggap mampu mengatasi hal tersebut yaitu dengan pemilihan metode pengukuran jarak yang tepat. Sehingga, penelitian ini menekankan pada perbandingan pemilihan metode pengukuran jarak antara *Euclidean*, *Manhattan*, *Tchebychev* dan *Cosine* terhadap peningkatan kinerja algoritma *k-Nearest Neighbor* pada data mining klasifikasi. Kemudian dijelaskan rumusan masalah, tujuan, manfaat dan batasan masalah yang nantinya akan dijadikan panduan dalam penulisan proposal penelitian selanjutnya.

DAFTAR PUSTAKA

- Binabar, S. W., & Ivandari, I. (2018). Optimasi Parameter K pada Algoritma KNN untuk Deteksi Penyakit Kanker Payudara. *IC-Tech*, 13(1).
- Kurniawan, M. F., & Ivandari, I. (2018). KOMPARASI ALGORITMA DATA MINING UNTUK KLASIFIKASI PENYAKIT KANKER PAYUDARA. *IC-Tech*, 12(1).
- Lukito, Y., & Chrismanto, A. R. (2015). Perbandingan Metode-Metode Klasifikasi Untuk Indoor Positioning System. *Jurnal Teknik Informatika dan Sistem Informasi*, 1(2).
- Medjahed, S. A., Saadi, T. A., & Benyettou, A. (2013). Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules. *International Journal of Computer Applications*, 62(1).
- Pressman, R. S. (2005). *Software engineering: a practitioner's approach*: Palgrave Macmillan.
- Primartha, R. (2018). *Belajar Machine Learning Teori dan Praktik*.
- Wajeed, M. A., & Adilakshmi, T. (2011). *Different similarity measures for text classification using KNN*. Paper presented at the Computer and Communication Technology (ICCCT), 2011 2nd International Conference on.
- Walters-Williams, J., & Li, Y. (2010). Comparative study of distance functions for nearest neighbors *Advanced Techniques in Computing Sciences and Software Engineering* (pp. 79-84): Springer.

Weinberger, K. Q., & Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research, 10*(Feb), 207-244.