

**PENGARUH REDUKSI DIMENSI MENGGUNAKAN
ALGORITMA GENETIKA TERHADAP HASIL
PENGKLASIFIKASIAN ALGORITMA *K-NEAREST
NEIGHBOR* DAN *MODIFIED K-NEAREST NEIGHBOR***

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



Oleh:

Agusti Kurniawan
09021281520117

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2019**

LEMBAR PENGESAHAN TUGAS AKHIR

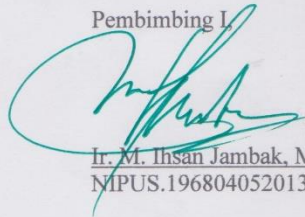
PENGARUH REDUKSI DIMENSI MENGGUNAKAN ALGORITMA
GENETIKA TERHADAP HASIL PENGKLASIFIKASIAN ALGORITMA K-
NEAREST NEIGHBOR DAN MODIFIED K-NEAREST NEIGHBOR

Oleh :

AGUSTI KURNIAWAN
NIM : 09021281520117

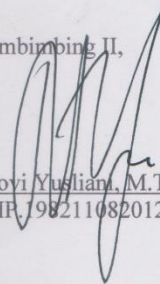
Indralaya, September 2019

Pembimbing I,



Ir. M. Ihsan Jambak, M.Sc
NIPUS.196804052013081201

Pembimbing II,



Novi Yuliant, M.T
NIP.198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika,



Rifkie Primartha, M.T.
NIP.197706012009121004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Jumat tanggal 23 Agustus 2019 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Agusti Kurniawan
N I M : 09021281520117
Judul : Pengaruh Reduksi Dimensi Menggunakan Algoritma Genetika Terhadap Hasil Pengklasifikasian Algoritma *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor*

1. Pembimbing I

Ir. M. Ihsan Jambak, M.Sc
NIPUS. 196804052013081201

2. Pembimbing II

Novi Yusliani, M.T
NIP. 198211082012122001

3. Penguji I

Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

4. Penguji II

M. Ali Buchari, M.T
NIPUS. 1671043003880004

Mengetahui,

Ketua Jurusan Teknik Informatika



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

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Agusti Kurniawan
NIM : 09021281520117
Program Studi : Teknik Informatika
Judul Skripsi : Pengaruh Reduksi Dimensi Menggunakan Algoritma Genetika Terhadap Hasil Pengklasifikasian Algoritma *k-Nearest Neighbor* dan *Modified k-Nearest Neighbor*
Hasil Pengecekan Software *iThenticate/Turnitin* : 7 %

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, 20 September 2019



Agusti Kurniawan
NIM. 09021281520117

Motto :

“ Jika mereka bisa, kenapa saya tidak.”

“ Ilmu adalah harta yang tak akan pernah habis.”

“ Masalah akan terasa ringan dengan bersabar dan berlapang dada.”

“ Learn from yesterday, Live for today, And hope for tomorrow.”

Kupersembahkan karya tulis ini kepada :

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

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. Orang tuaku, Husni Syahri dan Sulastiah serta seluruh keluarga besarku yang selalu mendokan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Rifkie Primartha, M.T. selaku Ketua Jurusan Teknik Informatika
4. Bapak Ir. M. Ihsan Jambak, M.Sc selaku dosen pembimbing I yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Ibu Novi Yusliani, M.T selaku dosen pembimbing II, yang telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
6. Bapak Alfarissi, M.Comp.Sc. selaku dosen pembimbing akademik.
7. Ibu Alvi Syahrini Utami, M.Kom selaku dosen penguji I, dan Bapak M. Ali Buchari, M.T selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
8. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
9. Mbak Winda dan Kak Ricy serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
10. Seluruh teman-teman jurusan Teknik Informatika angkatan 2015 yang telah berbagi keluh kesah, motivasi, dan canda tawa di masa-masa perkuliahan ini.

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, September 2019

Agusti Kurniawan

DAFTAR ISI

LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xvii
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian	I-5
1.5 Manfaat Penelitian	I-6
1.6 Batasan Masalah.....	I-6
1.7 Sistematika Penulisan.....	I-7
1.8 Kesimpulan	I-8
BAB II KAJIAN TEORITITS	II-1
2.1 Pendahuluan	II-1
2.2 Reduksi Dimensi	II-1
2.3 Algoritma Genetika.....	II-2
2.3.1 Algoritma Genetika untuk Seleksi Fitur	II-5
2.4 <i>k-Nearest Neighbor</i> (kNN).....	II-10
2.5 <i>Modified k-Nearest Neighbor</i> (MkNN).....	II-11
2.6 <i>Euclidean Distance</i>	II-14
2.7 <i>k-Fold Cross Validation</i>	II-14
2.8 Evaluasi Menggunakan <i>Receiver Operating Characteristic</i> (ROC).....	II-15
2.9 <i>Rational Unified Process</i>	II-18
2.8.1 Fase-fase <i>Rational Unified Process</i> (RUP)	II-19
2.9 Penelitian Lain yang Relevan.....	II-20

2.10 Kesimpulan	II-23
BAB III METODOLOGI PENELITIAN	III-1
3.1 Pendahuluan.....	III-1
3.2 Data.....	III-1
3.2.1 Jenis dan Sumber Data.....	III-1
3.2.2 Metode Pengumpulan Data.....	III-1
3.3 Tahapan Penelitian.....	III-2
3.3.1 Menetapkan Kerangka Kerja / Framework.....	III-2
3.3.1.1 Praproses	III-2
3.3.1.2 Pengklasifikasian dengan Algoritma kNN dan MkNN pada Data yang Belum di Reduksi	III-3
3.3.1.3 Reduksi Dimensi Menggunakan Algoritma Genetika.....	III-4
3.3.1.4 Pengklasifikasian Menggunakan Algoritma <i>k-Nearest Neighbor</i>	III-5
3.3.1.5 Pengklasifikasian Menggunakan Algoritma <i>Modified k-Nearest Neighbor</i>	III-6
3.3.1.6 Evaluasi Klasifikasi.....	III-6
3.3.2 Menetapkan Kriteria Pengujian	III-7
3.3.3 Menetapkan Format Data Pengujian	III-7
3.3.3.1 Hasil Pengujian Pengklasifikasian Data Sebelum di Reduksi.....	III-8
3.3.3.2 Hasil Pengujian Parameter Algoritma Genetika	III-8
3.3.3.3 Hasil Pengujian Pengklasifikasian Data Sesudah di Reduksi	III-10
3.3.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian.....	III-10
3.3.5 Melakukan Pengujian Penelitian	III-11
3.3.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan	III-13
3.4 Metode Pengembangan Perangkat Lunak	III-15
3.4.1 Fase Insepsi.....	III-15
3.4.2 Fase Elaborasi.....	III-15
3.4.3 Fase Konstruksi	III-16
3.4.4 Fase Transisi	III-17
3.5 Manajemen Proyek Penelitian.....	III-17
3.6 Kesimpulan	III-29
BAB IV PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1 Pendahuluan	IV-1

4.2 Fase Insepsi	IV-1
4.2.1 Pemodelan Bisnis.....	IV-2
4.2.2 Kebutuhan Sistem.....	IV-5
4.2.3 Analisis dan Desain	IV-15
4.2.3.1 Analisis Praproses Data.....	IV-15
4.2.3.2 Analisis Reduksi Dimensi Menggunakan Algoritma Genetika	IV-16
4.2.3.3 Analisis Klasifikasi.....	IV-17
4.2.3.4 Analisis Kualitas Klasifikasi	IV-18
4.3 Fase Elaborasi	IV-18
4.3.1 Pemodelan Bisnis.....	IV-18
4.3.2 Kebutuhan Sistem.....	IV-24
4.3.3 Analisis dan Desain	IV-28
4.4 Fase Konstruksi	IV-34
4.4.1 Pemodelan Bisnis.....	IV-34
4.4.2 Kebutuhan Sistem.....	IV-36
4.5 Fase Transisi.....	IV-36
4.5.1 Pemodelan Bisnis.....	IV-37
4.5.2 Kebutuhan Sistem.....	IV-37
4.5.3 Pengujian	IV-37
4.5.3.1 Rencana Pengujian	IV-38
4.5.3.2 Implementasi Pengujian	IV-41
4.6 Kesimpulan	IV-48

BAB V HASIL DAN ANALISIS PENELITIAN..... V-1

5.1 Pendahuluan	V-1
5.2 Konfigurasi Percobaan	V-1
5.3 Hasil Pengujian	V-3
5.3.1 Hasil Pengujian Klasifikasi kNN dan MkNN Tanpa Reduksi Dimensi ..	V-3
5.3.2 Hasil Pengujian Parameter Algoritma Genetika.....	V-10
5.3.2.1 Hasil Pengujian Jumlah Generasi.....	V-10
5.3.2.2 Hasil Pengujian Jumlah <i>popsiz</i> e	V-13
5.3.2.3 Hasil Pengujian Parameter <i>cr</i> (<i>crossover rate</i>)	V-15
5.3.2.4 Hasil Pengujian Parameter <i>mr</i> (<i>mutation rate</i>).....	V-17
5.3.3 Hasil Pengujian Klasifikasi kNN dan MkNN dengan Reduksi Dimensi	V-21

5.3.4 Hasil Pengklasifikasian Sebelum dan Sesudah Reduksi Dimensi.....	V-28
5.4 Olah Data	V-35
5.4.1 Olah Data Hasil Perbandingan Nilai AUC	V-35
5.4.2 Olah Data Hasil Perbandingan Waktu Komputasi	V-41
5.4.3 Olah Data Hasil Perbandingan Jumlah Memori	V-46
5.4 Analisis Hasil Pengujian	V-52
5.4.1 Analisis Hasil Pengujian Klasifikasi <i>k-Nearest Neighbor</i> dan <i>Modified k-Nearest Neighbor</i> pada Keseluruhan Data	V-52
5.4.2 Analisis Hasil Pengujian Klasifikasi <i>k-Nearest Neighbor</i> pada Data Tanpa Reduksi dengan Data Hasil Reduksi oleh Algoritma Genetika.....	V-54
5.4.3 Analisis Hasil Pengujian Klasifikasi <i>Modified k-Nearest Neighbor</i> pada Data Tanpa Reduksi dengan Data Hasil Reduksi oleh Algoritma Genetika.....	V-55
5.4.4 Analisis Hasil Pengujian Klasifikasi <i>k-Nearest Neighbor</i> dan <i>Modified k-Nearest Neighbor</i> pada Data Hasil Reduksi Algoritma Genetika	V-56
5.5 Kesimpulan	V-57
BAB VI KESIMPULAN DAN SARAN.....	VI-1
6.1 Pendahuluan	VI-1
6.2 Kesimpulan	VI-1
6.3 Saran.....	VI-3
DAFTAR PUSTAKA	xix
CODE	xxii

DAFTAR TABEL

Tabel II-1. Terminologi Genetika Manusia dan Algoritma Genetika.....	II-6
Tabel II-2. Confusion Matrix	II-16
Tabel III-1. Rancangan Tabel Hasil Pengklasifikasian kNN dan MkNN Tanpa Reduksi Dimensi	III-8
Tabel III-2. Rancangan Tabel Hasil Pengujian Jumlah Iterasi	III-9
Tabel III-3. Rancangan Tabel Hasil Pengujian Jumlah popsize.....	III-9
Tabel III-4. Rancangan Tabel Hasil Pengujian Jumlah cr (crossover rate).....	III-9
Tabel III-5. Rancangan Tabel Hasil Pengujian Jumlah mr (mutation rate).....	III-9
Tabel III-6. Rancangan Tabel Hasil Pengklasifikasian Dengan Data yang di Reduksi.....	III-10
Tabel III-7. Rancangan Tabel Hasil Analisa Pengklasifikasian	III-14
Tabel III-8. Tabel Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure</i> (WBS).....	III-18
Tabel IV-1. Definisi Aktor Use Case	IV-4
Tabel IV-2. Definisi Use Case	IV-4
Tabel IV-3. Tabel Kebutuhan Fungsional.....	IV-6
Tabel IV-4. Tabel Kebutuhan Non-Fungsional.....	IV-6
Tabel IV-5. Skenario Melakukan praproses pada data masukan	IV-7
Tabel IV-6. Skenario Use Case Mereduksi dimensi data dengan Algoritma Genetika	IV-8
Tabel IV-7. Tabel Skenario Mengklasifikasi data menggunakan kNN	IV-11
Tabel IV-8. Tabel Skenario Mengklasifikasi data menggunakan MkNN.....	IV-13
Tabel IV-9. Tabel Skenario Evaluasi Hasil Klasifikasi	IV-15
Tabel IV-10. Implementasi Kelas	IV-36
Tabel IV-11. Rencana Pengujian Use Case Melakukan Praproses pada Data Masukan	IV-38
Tabel IV-12. <i>Rencana Pengujian Use Case</i> Melakukan Klasifikasi dengan kNN	IV-38
Tabel IV-13. <i>Rencana Pengujian Use Case</i> Melakukan Klasifikasi dengan MkNN	IV-39
Tabel IV-14. <i>Rencana Pengujian Use Case</i> Melakukan Klasifikasi dengan kNN pada Data yang Telah di Reduksi Dimensi oleh Algoritma Genetika	IV-39
Tabel IV-15. <i>Rencana Pengujian Use Case</i> Melakukan Klasifikasi dengan kNN pada Data yang Telah di Reduksi Dimensi oleh Algoritma Genetika	IV-40
Tabel IV-16. Pengujian Fitur Praproses.....	IV-42
Tabel IV-17. Pengujian Fitur Klasifikasi dengan kNN.....	IV-43
Tabel IV-18. Pengujian Fitur Klasifikasi dengan MkNN	IV-44
Tabel IV-19. Pengujian Fitur Klasifikasi dengan kNN pada Data yang Telah di Reduksi Dimensi oleh Algoritma Genetika	IV-45
Tabel IV-20. Pengujian Fitur Klasifikasi dengan MkNN pada Data yang Telah di Reduksi Dimensi oleh Algoritma Genetika	IV-46

Tabel V-1. Hasil Pengujian Klasifikasi dengan kNN dan MkNN Tanpa Reduksi Dimensi Untuk Parameter k=3 dan k=5 pada Datasets Wisconsin prognosis breast cancer (WPBC)	V-4
Tabel V-2. Hasil Pengujian Klasifikasi dengan kNN dan MkNN Tanpa Reduksi Dimensi Untuk Parameter k=7 dan k=9 pada Datasets Wisconsin prognosis breast cancer (WPBC)	V-5
Tabel V-3. Hasil Pengujian Klasifikasi dengan kNN dan MkNN Tanpa Reduksi Dimensi Untuk Parameter k=3 dan k=5 pada Dataset LSVT Voice Rehabilitation	V-6
Tabel V-4. Hasil Pengujian Klasifikasi dengan kNN dan MkNN Tanpa Reduksi Dimensi Untuk Parameter k=7 dan k=9 pada Dataset LSVT Voice Rehabilitation	V-7
Tabel V-5. Hasil Pengujian Klasifikasi dengan kNN dan MkNN Tanpa Reduksi Dimensi Untuk Parameter k=3 dan k=5 pada Dataset HCC Survival	V-8
Tabel V-6. Hasil Pengujian Klasifikasi dengan kNN dan MkNN Tanpa Reduksi Dimensi Untuk Parameter k=7 dan k=9 pada Dataset HCC Survival	V-9
Tabel V-7. Hasil Pengujian Jumlah Generasi Dataset WPBC	V-11
Tabel V-8. Hasil Pengujian Jumlah Generasi Dataset LSVT Voice Rehabilitation	V-11
Tabel V-9. Hasil Pengujian Jumlah Generasi Dataset HCC Survival.....	V-12
Tabel V-10. Hasil Pengujian Jumlah popsize Dataset WPBC	V-13
Tabel V-11. Hasil Pengujian Jumlah popsize Dataset LSVT Voice Rehabilitation	V-14
Tabel V-12. Hasil Pengujian Jumlah popsize Dataset HCC Survival.....	V-14
Tabel V-13. Hasil Pengujian Parameter cr Dataset WPBC.....	V-15
Tabel V-14. Hasil Pengujian Parameter cr Dataset LSVT Voice Rehabilitation	V-16
Tabel V-15. Hasil Pengujian Parameter cr Dataset HCC Survival	V-17
Tabel V-16. Hasil Pengujian Parameter mr Dataset WPBC	V-18
Tabel V-17. Hasil Pengujian Parameter mr Dataset LSVT Voice Rehabilitation	V-19
Tabel V-18. Hasil Pengujian Parameter mr Dataset HCC Survival.....	V-20
Tabel V-19. Hasil Pengujian Klasifikasi dengan kNN dan MkNN Dengan Reduksi Dimensi Untuk Parameter k=3 dan k=5 pada Datasets WPBC	V-22
Tabel V-20. Hasil Pengujian Klasifikasi dengan kNN dan MkNN Dengan Reduksi Dimensi Untuk Parameter k=7 dan k=9 pada Datasets WPBC	V-23
Tabel V-21. Hasil Pengujian Klasifikasi dengan kNN dan MkNN Dengan Reduksi Dimensi Untuk Parameter k=3 dan k=5 pada Dataset LSVT Voice Rehabilitation	V-24
Tabel V-22. Hasil Pengujian Klasifikasi dengan kNN dan MkNN Dengan Reduksi Dimensi Untuk Parameter k=7 dan k=9 pada Dataset LSVT Voice Rehabilitation	V-25
Tabel V-23. Hasil Pengujian Klasifikasi dengan kNN dan MkNN Dengan Reduksi Dimensi Untuk Parameter k=3 dan k=5 pada Dataset HCC Survival	V-26
Tabel V-24. Hasil Pengujian Klasifikasi dengan kNN dan MkNN Dengan Reduksi Dimensi Untuk Parameter k=7 dan k=9 pada Dataset HCC Survival	V-27
Tabel V-25. Tabel Hasil Perbandingan Klasifikasi pada Dataset WPBC	V-29

Tabel V-26. Tabel Hasil Perbandingan Klasifikasi pada Dataset LSVT Voice Rehabilitation	V-31
Tabel V-27. Tabel Hasil Perbandingan Klasifikasi pada Dataset HCC Survival	V-33
Tabel V-28. Deskripsi Nilai AUC	V-35
Tabel V-29. Hasil Uji Normalitas Nilai AUC	V-36
Tabel V-30. Hasil Uji Kruskal-Wallis Perbandingan Nilai AUC	V-36
Tabel V-31. Hasil Uji Mann-Whitney Perbandingan Nilai AUC pada Data Hasil Pengklasifikasian kNN dan MkNN	V-37
Tabel V-32. Hasil Uji Mann-Whitney Perbandingan Nilai AUC pada Data Hasil Pengklasifikasian kNN dan GA + kNN	V-38
Tabel V-33. Hasil Uji Mann-Whitney Perbandingan Nilai AUC pada Data Hasil Pengklasifikasian kNN dan GA + MkNN	V-38
Tabel V-34. Hasil Uji Mann-Whitney Perbandingan Nilai AUC pada Data Hasil Pengklasifikasian MkNN dan GA + kNN	V-39
Tabel V-35. Hasil Uji Mann-Whitney Perbandingan Nilai AUC pada Data Hasil Pengklasifikasian MkNN dan GA + MkNN	V-39
Tabel V-36. Hasil Uji Mann-Whitney Perbandingan Nilai AUC pada Data Hasil Pengklasifikasian GA + kNN dan GA + MkNN	V-40
Tabel V-37. Deskripsi Waktu Komputasi	V-41
Tabel V-38. Hasil Uji Normalitas Waktu Komputasi	V-42
Tabel V-39. Hasil Uji Kruskal-Wallis Perbandingan Waktu Komputasi	V-42
Tabel V-40. Hasil Uji Mann-Whitney Perbandingan Waktu Komputasi pada Data Hasil Pengklasifikasian kNN dan MkNN	V-43
Tabel V-41. Hasil Uji Mann-Whitney Perbandingan Waktu Komputasi pada Data Hasil Pengklasifikasian kNN dan GA + kNN	V-44
Tabel V-42. Hasil Uji Mann-Whitney Perbandingan Waktu Komputasi pada Data Hasil Pengklasifikasian kNN dan GA + MkNN	V-44
Tabel V-43. Hasil Uji Mann-Whitney Perbandingan Waktu Komputasi pada Data Hasil Pengklasifikasian MNN dan GA + kNN	V-45
Tabel V-44. Hasil Uji Mann-Whitney Perbandingan Waktu Komputasi pada Data Hasil Pengklasifikasian MNN dan GA + MkNN	V-45
Tabel V-45. Hasil Uji Mann-Whitney Perbandingan Waktu Komputasi pada Data Hasil Pengklasifikasian GA + kNN dan GA + MkNN	V-46
Tabel V-46. Deskripsi Jumlah Memori	V-46
Tabel V-47. Hasil Uji Normalitas Jumlah Memori	V-47
Tabel V-48. Hasil Uji Kruskal-Wallis Perbandingan Jumlah Memori	V-48
Tabel V-49. Hasil Uji Mann-Whitney Perbandingan Jumlah Memori pada Data Hasil Pengklasifikasian kNN dan MkNN	V-49
Tabel V-50. Hasil Uji Mann-Whitney Perbandingan Jumlah Memori pada Data Hasil Pengklasifikasian kNN dan GA + kNN	V-49
Tabel V-51. Hasil Uji Mann-Whitney Perbandingan Jumlah Memori pada Data Hasil Pengklasifikasian kNN dan GA + MkNN	V-50
Tabel V-52. Hasil Uji Mann-Whitney Perbandingan Jumlah Memori pada Data Hasil Pengklasifikasian MkNN dan GA + kNN	V-50
Tabel V-53. Hasil Uji Mann-Whitney Perbandingan Jumlah Memori pada Data Hasil Pengklasifikasian MkNN dan GA + MkNN	V-51

Tabel V-54. Hasil Uji Mann-Whitney Perbandingan Jumlah Memori pada Data Hasil Pengklasifikasian GA + kNN dan GA + MkNN V-51
Error! Bookmark not defined.

DAFTAR GAMBAR

Gambar II-1. Langkah Dasar Algoritma Genetika.....	II-3
Gambar II-2. Ruang Pencarian Algoritma Genetika.....	II-4
Gambar II-3. Alur Kerja Algoritma Genetika untuk Seleksi Fitur	II-9
Gambar II-4. Alur Kerja Proses Klasifikasi Oleh MkNN.....	II-12
Gambar II-5. Ilustrasi k-Fold Cross Validation	II-15
Gambar II-6. Arsitektur RUP (Kruchten, 2004)	II-19
Gambar III-1. Tahapan Pengujian Penelitian	III-12
Gambar III-2. Gantt Chart Penjadwalan Penelitian Tahap Menentukan Ruang Lingkup dan Unit Penelitian	III-22
Gambar III-3. <i>Gantt Chart</i> Penjadwalan Penelitian Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian	III-23
Gambar III-4. Gantt Chart Penjadwalan Penelitian Tahap Menentukan Kriteria Pengujian.....	III-23
Gambar III-5. Gantt Chart Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan Untuk Pelaksanaan Penelitian pada Fase Insepsi	III-24
Gambar III-6. Gantt Chart Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan Untuk Pelaksanaan Penelitian pada Fase Elaborasi	III-25
Gambar III-7. Gantt Chart Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan Untuk Pelaksanaan Penelitian pada Fase Konstruksi	III-26
Gambar III-8. Gantt Chart Penjadwalan Penelitian Tahap Menentukan Alat yang Digunakan Untuk Pelaksanaan Penelitian pada Fase Transisi.....	III-27
Gambar III-9. <i>Gantt Chart</i> Penjadwalan Penelitian Tahap Melakukan Pengujian Penelitian.....	III-28
Gambar III-10. <i>Gantt Chart</i> Penjadwalan Penelitian Tahap Melakukan Analisa Hasil Pengujian dan Pembuatan Kesimpulan	III-28
Gambar IV-1. Diagram Use Case.....	IV-3
Gambar IV-2. Diagram Sequence Melakukan Praproses.....	IV-19
Gambar IV-3. Diagram Sequence Melakukan Klasifikasi dengan kNN.....	IV-20
Gambar IV-4. Diagram Sequence Melakukan Klasifikasi dengan MkNN	IV-21
Gambar IV-5. Diagram Sequence Melakukan Reduksi Dimensi dengan Algoritma Genetika dan Klasifikasi dengan kNN	IV-22
Gambar IV-6. Diagram Sequence Melakukan Reduksi Dimensi dengan Algoritma Genetika dan Klasifikasi dengan MkNN	IV-23
Gambar IV-7. Rancangan Antar Muka Perangkat Lunak	IV-24
Gambar IV-8. Rancangan Fitur Muat Dataset.....	IV-24
Gambar IV-9. Fitur Pilihan Proses Klasifikasi.....	IV-25
Gambar IV-10. Rancangan Fitur Input Parameter	IV-27
Gambar IV-11. Rancangan Fitur Tombol Mulai.....	IV-27
Gambar IV-12. Rancangan Fitur Hasil Klasifikasi	IV-28
Gambar IV-13. Rancangan Hasil Klasifikasi Setaip k-Fold	IV-28
Gambar IV-14. Diagram Aktivitas Melakukan Praproses Data.....	IV-29
Gambar IV-15. Diagram Aktivitas Melakukan Klasifikasi dengan kNN	IV-30
Gambar IV-16. Diagram Aktivitas Melakukan Klasifikasi dengan MkNN...IV-31	

Gambar IV-17. Diagram Aktivitas Melakukan Reduksi Dimensi Algoritma Genetika dan Klasifikasi dengan kNN.....	IV-32
Gambar IV-18. Diagram Aktivitas Melakukan Reduksi Dimensi Algoritma Genetika dan Klasifikasi dengan kNN.....	IV-33
Gambar IV-19. Diagram Kelas	IV-35

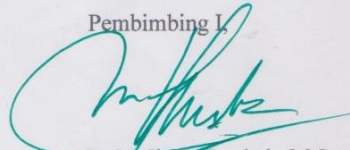
ABSTRACT

The Classification algorithm k-Nearest Neighbor (kNN) has many weaknesses, one of these weaknesses can be overcome by the development of kNN algorithm, namely Modified k-Nearest Neighbor (MkNN) where MkNN can overcome the problem of outliers in traditional kNN. On the other hand kNN still leaves some weaknesses such as requiring large computing and memory costs in its application, it is better to use all the features or parts of it while it is known that the performance of this kNN is apparently not good in dealing with the curse of dimensionality. From this arises the question, whether dimension reduction by feature selection using genetic algorithms has an effect that can overcome these weaknesses. Moreover, it is known that feature selection has a direct effect with reduced processing time for data mining algorithms, improves performance in classification and also results that are easier to understand. To determine the effect of this dimension reduction, the method will be tested on 3 datasets namely the LSVT Voice Rehabilitation dataset, Wisconsin breast cancer prognosis, and HCC Survival. From the results of the study showed that dimensional reduction with genetic algorithms has a good effect on the classification of kNN and MkNN algorithms for each test data, which is indicated by an increased level of accuracy with the AUC, and decreasing computing time and memory used.

Keywords: k-Nearest Neighbor, Modified k-Nearest Neighbor, Genetic Algorithm, Classification, Dimensional Reduction, Feature Selection.

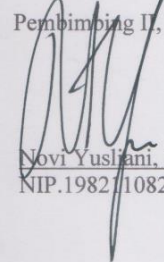
Indralaya, September 2019

Pembimbing I,



Ir. M. Ihsan Jambak, M.Sc
NIPUS.196804052013081201

Pembimbing II,



Novi Yuliani, M.T
NIP.198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika,



Rifkie Primartha, M.T.
NIP.197706012009121004

ABSTRAK

Algoritma klasifikasi *k-Nearest Neighbor* (kNN) mempunyai banyak kelemahan, salah satu kelemahan ini dapat diatasi oleh algoritma pengembangan dari kNN yaitu *Modified k-Nearest Neighbor* (MkNN) dimana MkNN dapat mengatasi masalah *outlier* pada kNN biasa. Namun, disisi lain kNN masih menyisakan beberapa kelemahan seperti memerlukan biaya komputasi dan memori yang besar dalam penerapannya, lebih baik menggunakan seluruh fitur atau sebagiannya saja sedangkan diketahui bahwa performa dari kNN ini ternyata kurang baik dalam menghadapi kutukan dimensi. Dari sini muncul pertanyaan, apakah reduksi dimensi dengan seleksi fitur menggunakan algoritma genetika mempunyai pengaruh yang dapat mengatasi kelemahan-kelemahan tersebut. Apalagi diketahui bahwa seleksi fitur berefek langsung dengan penurunan waktu pemrosesan algoritma *data mining*, meningkatkan performa dalam melakukan klasifikasi dan juga hasil yang lebih mudah dipahami. Untuk mengetahui pengaruh dari reduksi dimensi ini, metode akan diujikan pada 3 dataset yaitu dataset *LSVT Voice Rehabilitation*, *Wisconsin prognosis breast cancer*, dan *HCC Survival*. Dari hasil penelitian menunjukkan bahwa reduksi dimensi dengan algoritma genetika mempunyai pengaruh yang baik terhadap algoritma klasifikasi kNN maupun MkNN bagi setiap data uji, yang ditunjukkan melalui peningkatan tingkat akurasi dengan AUC, dan menurunnya waktu komputasi dan memori yang digunakan.

Kata Kunci: *k-Nearest Neighbor*, *Modified k-Nearest Neighbor*, Algoritma Genetika, Klasifikasi, Reduksi Dimensi, Seleksi Fitur.

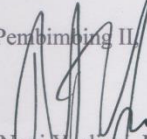
Indralaya, September 2019

Pembimbing I,



Ir. M. Ihsan Jambak, M.Sc
NIPUS.196804052013081201

Pembimbing II,



Novi Yuslimah, M.T
NIP.198211082012122001

Mengetahui,
Ketua Jurusan Teknik Informatika,



Rifkie Primartha, M.T.
NIP.197706012009121004

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian serta batasan masalah. Bab ini akan memberikan penjelasan umum mengenai keseluruhan penelitian. Pendahuluan dimulai dengan penjelasan mengenai klasifikasi dan di teruskan ke masalah algoritma klasifikasi *Modified k-Nearest Neighbor* (MkNN) yang merupakan pengembangan dari algoritma *k-Nearest Neighbor* (kNN). Selanjutnya di jelaskan penyelesaian masalah tersebut dengan menggunakan algoritma genetika.

1.2 Latar Belakang

Klasifikasi merupakan salah satu peran utama dari data mining. Dalam pemodelannya, klasifikasi membutuhkan data training untuk mengenali pola tertentu dari data dengan label atau kelas yang sudah ditentukan sebelumnya. Kemudian pola tersebut dipakai untuk menentukan label yang belum diketahui dari data baru. Terdapat banyak teknik klasifikasi data mining salah satu yang paling dasar dan sangat sederhana adalah kNN (Santosa & Umam, 2018). kNN bekerja dengan mengklasifikasikan suatu label berdasarkan dengan label yang paling sering muncul dalam k data latih terdekat.

Menurut Duda, Hart, dan Stork (2012), algoritma kNN memiliki beberapa kekurangan seperti: a) biaya komputasi yang besar; b) Dalam penerapannya

memerlukan memori yang besar; c) Akurasi yang rendah dalam dataset *multidimensional*; d) Apakah lebih baik menggunakan seluruh fitur, atau sebagiannya saja.

Salah satu penelitian yang mencoba mengatasi salah satu kelemahan ini adalah penelitian yang dilakukan oleh Parvin, Alizadeh, dan Minaei-Bidgoli (2008). Mereka mengajukan algoritma baru yang dapat mengatasi masalah rendahnya tingkat akurasi kNN dalam dataset *multidimensional* yaitu algoritma *Modified k-Nearest Neighbor* (MkNN). Di mana dalam MkNN ditambahkan perhitungan *validity* untuk mempertimbangkan validitas antar data latih dan juga *weighted voting* untuk menghitung bobot dalam penentuan label data uji. Hasil dari penelitian ini pun terbukti lebih baik jika dibandingkan dengan algoritma kNN biasa. Akan tetapi karena hanya bisa mengatasi masalah *outlier* algoritma MkNN ini masih membawa kelemahan algoritma kNN seperti dalam penerapannya memerlukan memori yang besar, komputasinya yang kompleks sehingga memakan waktu yang lama dan juga dalam penentuan fitur, apakah kNN ini lebih baik menggunakan keseluruhan fitur atau hanya sebagiannya saja. Apalagi diketahui bahwa performa dari algoritma kNN kurang baik dalam menghadapi *curse of dimensionality* (Pestov, 2013).

Dalam kNN dan MkNN konsep *nearest neighbor* akan kacau ketika dimensi dari ruang fitur terlalu tinggi. Pada data berdimensi tinggi titik data akan jauh dari satu sama lain sehingga membuat tetangga terdekat maupun yang terjauh hanya mempunyai sedikit perbedaan satu sama lain. Hal ini mengakibatkan dengan semakin tingginya jumlah dimensi maka pilihan dari *nearest neighbor* akan terlihat

acak (Beyer et al., 1999; Houle et al., 2010). Selain itu ketika dimensi pada data terlalu tinggi, keseluruhan obyek dari data akan ikut di evaluasi sehingga membuat waktu komputasi dan memori yang digunakan menjadi lebih besar (Sivasankar, Nair, & Judy, 2015). Sehingga bisa dikatakan *curse of dimensionality* dapat membuat proses pengolahan data menjadi kurang efektif dan efisien. Karena itu untuk mengatasi hal ini diperlukan teknik untuk mereduksi dimensi (Rufaidha, Nhita, & Murdiansyah, 2016). Ada dua teknik reduksi dimensi yang dapat digunakan untuk mengatasi masalah ini yaitu ekstraksi fitur dan seleksi fitur.

Pada penelitian ini reduksi dimensi menggunakan seleksi fitur akan digunakan. Seleksi fitur sendiri merupakan teknik yang dapat mengurangi jumlah fitur, menghapus data yang tidak relevan, data ganda, atau derau dan data yang tidak terlalu informatif. Sehingga akan berefek langsung dengan penurunan waktu pemrosesan algoritma data mining, meningkatkan performa dalam melakukan klasifikasi dan juga hasil yang lebih mudah di pahami (Gui et al., 2017; Huan & Lei, 2005). Meskipun begitu pemetaan yang akurat terhadap fitur-fitur dari ruang dimensi rendah diperlukan sehingga tidak ada informasi yang hilang dengan membuang fitur-fitur yang penting dan utama (Aziz et al., 2013). Algoritma seleksi fitur terbagi menjadi tiga tipe yaitu metode *wrapper*, *filter*, dan *embedded*. Dimana pada penelitian ini akan digunakan metode seleksi fitur berbasis *wrapper* dengan menggunakan algoritma genetika.

Penelitian yang menggabungkan algoritma genetika untuk mereduksi dimensi dan algoritma kNN untuk proses klasifikasi pernah dilakukan sebelumnya oleh Deekshatulu dan Chandra (2013). Pada penelitian ini peneliti mereduksi

dimensi dari data “*Heart Disease A.P*” dan beberapa dataset lainnya menggunakan algoritma genetika. Dari hasil penelitian menunjukkan bahwa penggabungan kedua metode ini dapat meningkatkan kemampuan pengklasifikasian dari algoritma kNN. Pada data “*Heart Disease A.P*” terjadi peningkatan tingkat akurasi dari yang sebelumnya 95% menjadi 100% pada saat di reduksi dimensinya. Sedangkan rata-rata tingkat akurasi untuk seluruh dataaset yang di uji terjadi peningkatan dari yang sebelumnya sebesar 92.14% menjadi 95.73%. Jadi bisa disimpulkan bahwa reduksi dimensi dengan menggabungkan algoritma genetika dengan kNN dapat meningkatkan tingkat akurasi dari algoritma klasifikasi kNN. Sedangkan waktu komputasi maupun memori yang digunakan pada penelitian ini tidak ikut di analisis sehingga tidak di ketahui pengaruh reduksi dimensi terhadap waktu komputasi maupun memori yang diperlukan oleh kNN dalam melakukan klasifikasi.

Oleh karena itu berdasarkan masalah yang sudah di ungkapkan sebelumnya, maka pada penelitian ini akan menguji pengaruh dari reduksi dimensi yang akan dilakukan dengan menggunakan algoritma genetika terhadap proses pengklasifikasian oleh algoritma kNN dan MkNN pada data berdimensi tinggi.

1.3 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah apa pengaruh reduksi dimensi dengan menggunakan algoritma genetika terhadap proses pengklasifikasian algoritma *k-Nearest Neighbor* (kNN) dan *Modified k-Nearest Neighbor* (MkNN) dalam data berdimensi tinggi.

Untuk menyelesaikan permasalahan yang ada pada rumusan masalah maka dalam penelitian ini memiliki beberapa *research question* sebagai berikut:

1. Bagaimana mekanisme algoritma genetika agar dapat digunakan dalam mereduksi dimensi?
2. Bagaimana pengaruh dari reduksi dimensi dengan menggunakan algoritma genetika terhadap proses pengklasifikasian algoritma *k-Nearest Neighbor* (kNN)?
3. Bagaimana pengaruh dari reduksi dimensi dengan menggunakan algoritma genetika terhadap proses pengklasifikasian algoritma *Modified k-Nearest Neighbor* (MkNN)?

1.4 Tujuan Penelitian

Tujuan dilakukannya penelitian adalah sebagai berikut:

1. Mengetahui cara pereduksian dimensi dengan menggunakan algoritma genetika;
2. Mengetahui pengaruh dari reduksi dimensi dengan menggunakan algoritma genetika terhadap proses pengklasifikasian algoritma *k-Nearest Neighbor* (kNN);
3. Mengetahui pengaruh dari reduksi dimensi dengan menggunakan algoritma genetika terhadap proses pengklasifikasian algoritma *Modified k-Nearest Neighbor* (MkNN);
4. Mengetahui apakah teknik reduksi dimensi dengan menggunakan algoritma genetika bisa di bilang baik untuk mendapatkan fitur-fitur yang relevan bagi proses klasifikasi;

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Memahami mekanisme algoritma genetika sebagai metode reduksi dimensi dan mekanisme algoritma *k-Nearest Neighbor* (kNN) dan *Modified k-Nearest Neighbor* (MkNN) sebagai metode klasifikasi jika diterapkan dalam data berdimensi tinggi;
2. Memahami pengaruh dari metode reduksi dimensi dengan menggunakan algoritma genetika terhadap proses pengklasifikasian data oleh algoritma *k-Nearest Neighbor* (kNN) dan *Modified k-Nearest Neighbor* (MkNN).
3. Mampu menerapkan metode klasifikasi *k-Nearest Neighbor* (kNN) dan *Modified k-Nearest Neighbor* (MkNN) dengan teknik reduksi dimensi algoritma genetika pada sistem pengklasifikasian data.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Metode perhitungan jarak yang akan digunakan adalah *Euclidean Distance*.
2. Pengujian metode penelitian akan diimplementasikan pada data berdimensi tinggi.
3. Evaluasi performa dari model klasifikasi akan dilakukan dengan mengukur nilai AUC yang di dapat dari kurva ROC, waktu komputasi dan memori yang dibutuhkan. Kemudian hasil klasifikasi akan dibandingkan antara metode klasifikasi kNN dengan dan tanpa reduksi dimensi dan MkNN dengan dan tanpa reduksi dimensi.

1.7 Sistematika Penulisan

BAB I. PENDAHULUAN

Bab ini akan membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, sistematika penulisan dan kesimpulan.

BAB II. KAJIAN LITERATUR

Bab ini akan menjelaskan dasar-dasar teori yang digunakan dalam penelitian. Pembahasan dalam bab ini meliputi hasil penelitian terdahulu yang relevan dengan penelitian ini, reduksi dimensi, penjelasan mengenai metode reduksi dimensi dan metode klasifikasi serta penjelasan lainnya yang berkaitan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini akan membahas tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini akan membahas perancangan perangkat lunak yang akan dibangun pada penelitian ini.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini akan menampilkan hasil pengujian berdasarkan langkah-langkah yang telah direncanakan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam pengembangan perangkat lunak ini selanjutnya.

1.8 Kesimpulan

Algoritma kNN mempunyai banyak kelemahan dimana salah satu kelemahan tersebut dapat di atasi oleh algoritma hasil pengembangan dari kNN yaitu MkNN. Algoritma MkNN ini terbukti dapat mengatasi masalah *outlier* pada data sehingga dapat meningkatkan akurasi hasil klasifikasi. Namun MkNN ini masih membawa kelemahan dari algoritma kNN seperti komputasi yang semakin kompleks, memakan banyak memori serta masalah penentuan jumlah fitur. Apalagi di ketahui bahwa performansi algoritma kNN kurang baik dalam mengatasi masalah *curse of dimensionality* sehingga di perlukan metode lain untuk mengurangi dimensi data. Salah satu algoritma reduksi dimensi yang dapat digunakan adalah algoritma genetika dimana diharapkan dari hasil pereduksian dimensi dapat mengurangi waktu komputasi dan juga dapat meningkatkan kualitas klasifikasi dari kedua algoritma klasifikasi kNN dan MkNN.

DAFTAR PUSTAKA

- Aziz, A. S. A., et al. (2013). *Genetic algorithm with different feature selection techniques for anomaly detectors generation*. Paper presented at the Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on.
- Beyer, K., et al. (1999). *When is "nearest neighbor" meaningful?* Paper presented at the International conference on database theory.
- Deekshatulu, B., & Chandra, P. (2013). Classification of heart disease using k-nearest neighbor and genetic algorithm. *Procedia Technology*, 10, 85-94.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*: John Wiley & Sons.
- Gui, J., et al. (2017). Feature selection based on structured sparsity: A comprehensive study. *IEEE transactions on neural networks and learning systems*, 28(7), 1490-1507.
- Houle, M. E., et al. (2010). *Can shared-neighbor distances defeat the curse of dimensionality?* Paper presented at the International Conference on Scientific and Statistical Database Management.
- Huan, L., & Lei, Y. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4), 491-502. doi:10.1109/TKDE.2005.66
- Parvin, H., Alizadeh, H., & Minaei-Bidgoli, B. (2008). *MKNN: Modified k-nearest neighbor*. Paper presented at the Proceedings of the World Congress on Engineering and Computer Science.
- Pestov, V. (2013). Is the k-NN classifier in high dimensions affected by the curse of dimensionality? *Computers & Mathematics with Applications*, 65(10), 1427-1437.
- Rufaidha, H., Nhita, F., & Murdiansyah, D. T. (2016). Prediksi Penyakit Menggunakan Algoritma K-nearest Neighbour Dan Algoritma Genetika Untuk Data Berdimensi Tinggi. *eProceedings of Engineering*, 3(2).
- Santosa, B., & Umam, A. (2018). *Data Mining dan Big Data Analytics: Teori dan Implementasi Menggunakan Python & Apache Spark* (Isa Ed.): Penebar Media Pustaka.
- Aalaei, S., et al. (2016). Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets. *Iranian journal of basic medical sciences*, 19(5), 476.
- Astuti, F. D., Ratnawati, D. E., & Widodo, A. W. (2017). Deteksi Penyakit Kucing dengan Menggunakan Modified K-Nearest Neighbor Teroptimasi (Studi Kasus: Puskesmas Klinik Hewan dan Satwa Sehat Kota Kediri). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- Babatunde, O. H., et al. (2014). A genetic algorithm-based feature selection.
- Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12): Springer Science & Business Media.

- Gunavathi, C., & Premalatha, K. (2014). Performance analysis of genetic algorithm with kNN and SVM for feature selection in tumor classification. *Int J Comput Electr Autom Control Inf Eng*, 8(8), 1490-1497.
- Han, J., Kamber, M., & Pei, J. (2012). Data mining concepts and techniques third edition. *Waltham: Elsevier*.
- Hossin, M., & Sulaiman, M. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2), 1.
- Hunter, A. (2000). Feature selection using probabilistic neural networks. *Neural Computing & Applications*, 9(2), 124-132.
- Idrees, F., et al. (2017). PIndroid: A novel Android malware detection system using ensemble learning methods. *Computers & Security*, 68, 36-46.
- Kruchten, P. (2004). *The rational unified process: an introduction*: Addison-Wesley Professional.
- Leidiana, H. (2013). Penerapan algoritma k-nearest neighbor untuk penentuan resiko kredit kepemilikan kendaraan bermotor. *PIKSEL (Penelitian Ilmu Komputer Sistem Embedded dan Logic)*, 1(1), 65-76.
- Prasetyo, V. R., Hartanto, B., & Mulyono, A. A. (2019). Penentuan Pembimbing Tugas Akhir Mahasiswa Jurusan Teknik Informatika Universitas Surabaya Dengan Metode Dice Coefficient. *Teknika*, 8(1), 44-51.
- Putri, N. M. E. P., & Ermatita. (2017). *Review : Data Mining menggunakan Algoritma Genetika*. Paper presented at the Prosiding Annual Research Seminar 2017.
- Raymer, M. L., et al. (2000). Dimensionality reduction using genetic algorithms. *IEEE transactions on evolutionary computation*, 4(2), 164-171.
- Sarmilah, M., Adiwijaya, A., & Atiqi, A. (2018). Analisis Seleksi Fitur Genetic Algorithm Dan Ekstraksi Fitur Wavelet Pada Klasifikasi Microarray Data Menggunakan Naïve Bayes. *eProceedings of Engineering*, 5(1).
- Shin, K.-s., Kim, K.-j., & Han, I. (1998). *Financial data mining using genetic algorithms technique: Application to KOSPI 200*. Paper presented at the 한국전문가시스템학회'98 추계학술대회.
- Simanjuntak, T. H., Mahmudy, W. F., & Sutrisno, S. (2017). Implementasi Modified K-Nearest Neighbor Dengan Otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(2), 75-79.
- Singh, D. A. A. G., et al. (2016). Dimensionality reduction using genetic algorithm for improving accuracy in medical diagnosis. *International Journal of Intelligent Systems and Applications*, 8(1), 67.
- Sivasankar, S., Nair, S., & Judy, M. (2015). Feature reduction in clinical data classification using augmented genetic algorithm. *International Journal of Electrical and Computer Engineering (IJECE)*, 5(6), 1516-1524.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427-437.
- Wafiyah, F., Hidayat, N., & Perdana, R. S. (2017). Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit

Demam. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.

Yu, L., et al. (2007). Application and Comparison of Classification Techniques in Controlling Credit Risk. *Recent Advances in Data Mining of Enterprise Data: Algorithm and Application*, 6, 111-145.

Indrayanti, I., Sugianti, D., & Al Karomi, A. (2017). Optimasi Parameter K pada Algoritma K-nearest Neighbour untuk Klasifikasi Penyakit Diabetes Mellitus. *Prosiding SNATIF*, 823-829.

Penev, I., Karova, M., & Todorova, M. (2016). On the optimum choice of the K Parameter in Hand-Written Digit Recognition by kNN in comparison to SVM. *International journal of neural networks and advanced applications*, 3.

Rivki, M., & Bachtiar, A. M. (2017). Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia. *Jurnal Sistem Informasi*, 13(1), 31-37.

Trevor, H., Robert, T., & JH, F. (2009). *The elements of statistical learning: data mining, inference, and prediction*: New York, NY: Springer.