

**Reduksi Dimensi Fitur Dengan Metode *Feature Selection*
Information Gain Pada Klasifikasi *Modified k-Nearest Neighbor***

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



Oleh :

Erindah Nuraprilliana Sari
NIM : 09021381621102

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2020**

LEMBAR PENGESAHAN TUGAS AKHIR

REDUKSI DIMENSI FITUR DENGAN METODE *FEATURE
SELECTION INFORMATION GAIN* PADA KLASIFIKASI
MODIFIED K-NEAREST NEIGHBOR

Oleh:

Erindah Nuraprilliana Sari
NIM: 09021381621102

Palembang, Desember 2020

Pembimbing I



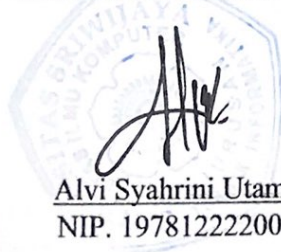
Rifkie Primartha, M.T.
NIP.197706012009121004

Pembimbing II



Ir. Muhammad Ihsan Jambak, M.Sc., M.M.
NIP. 196804052013081201

Mengetahui,
Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

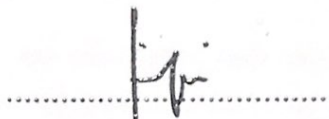
TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari jumat tanggal 4 Desember 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Erindah Nuraprilliana Sari
NIM : 09021381621102
Judul : Reduksi Dimensi dengan Metode *Feature Selection Information Gain* pada Klasifikasi *Modified K-Nearest Neighbor*

1. Pembimbing I

Rifkie Primartha, M.T
NIP. 19770612009121004



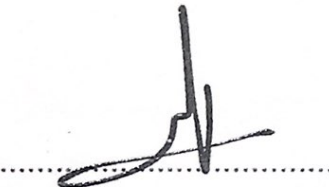
2. Pembimbing II

Ir. Muhammad Ihsan Jambak, M.Sc., M.M.
NIP. 196804052013081201



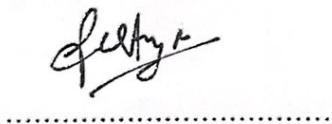
3. Penguji I

Dr. Abdiansah, S.Kom., M.CS.
NIP. 198410012009121005

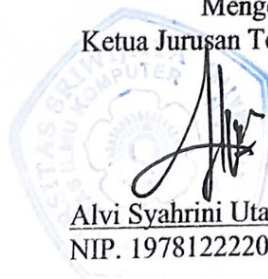


4. Penguji II

Desty Rodiah, MT
NIP. 1671016112890005



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Erindah Nuraprilliana Sari
NIM : 09021381621102
Program Studi : Teknik Informatika
Judul Skripsi : Reduksi Dimensi dengan Metode *Feature Selection Information Gain* pada Klasifikasi *Modified k-Nearest Neighbor*

Hasil Pengecekan Software *iThenticate/Turnitin* : 8%

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, Desember 2020



Erindah Nuraprilliana Sari
NIM. 09021381621102

MOTTO DAN PERSEMBAHAN

“Keep smiling, because life is a beautiful thing and there’s so much to smile about.”

(Marilyn Monroe)

Kupersembahkan karya tulis ini kepada :

- *Kedua orangtua*
- *Keluarga besar*
- *Teman - teman Seperjuangan*
- *Fakultas Ilmu Komputer*
- *Universitas Sriwijaya*

DIMENSIONAL REDUCTION WITH INFORMATION GAIN SELECTION
FEATURE METHOD USING THE MODIFIED K-NEAREST NEIGHBOR
CLASSIFICATION

By:
Erindah Nurapriyana Sari
09021381621102

ABSTRACT

The Classification algorithm Modified k-Nearest Neighbor (MkNN) is a development of the kNN algorithm where MkNN can solve the outlier problem in ordinary KNN. MkNN has several disadvantages such as requiring large computation and memory costs in its application and not good to dealing high-dimensional data. From this arises the question, whether dimension reduction by feature selection using information gain 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 dataset LSVT Voice Rehabilitation. MkNN classification using dimensional reduction with Information Gain results in an average accuracy of 83.46%, the average time of 6 seconds and the average memory of 120130765 bytes while MkNN classification without dimensional reduction results in an average accuracy of 80.78%, the average time is 12.5 seconds and the average memory is 121313689 bytes.

Keywords: Modified k-Nearest Neighbor, Information Gain, Classification, Feature Selection, Dimensional Reduction.

REDUKSI DIMENSI DENGAN METODE FEATURE SELECTION
INFORMATION GAIN PADA KLASIFIKASI MODIFIED K-NEAREST
NEIGHBOR

Oleh:

Erindah Nuraprilliana Sari

09021381621102

ABSTRAK

Algoritma klasifikasi *Modified k-Nearest Neighbor* (MkNN) merupakan pengembangan dari algoritma kNN dimana MkNN dapat mengatasi masalah *outlier* pada kNN biasa. MkNN memiliki beberapa kelemahan seperti memerlukan biaya komputasi dan memori yang besar dalam penerapannya serta kurang baik untuk menghadapi data dimensi tinggi. Dari sini muncul pertanyaan, apakah reduksi dimensi dengan seleksi fitur menggunakan *Information Gain* 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 dataset *LSVT Voice Rehabilitation*. Klasifikasi MkNN menggunakan reduksi dimensi dengan *Information Gain* menghasilkan rata-rata akurasi sebesar 83,46%, rata-rata waktu 6 detik dan rata-rata memori sebesar 120130765 *byte* sedangkan klasifikasi MkNN tanpa reduksi dimensi menghasilkan rata-rata akurasi sebesar 80,78%, rata-rata waktu 12,5 detik dan rata-rata memori sebesar 121313689 *byte*.

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

KATA PENGANTAR

Selama penelitian dan penyusunan laporan penelitian skripsi ini, penulis tidak luput dari kendala dan hambatan. Namun demikian kendala dan hambatan tersebut dapat penulis atasi berkat bantuan, bimbingan dan dukungan dari berbagai pihak. Oleh karena itu dalam kesempatan ini, penulis ingin menyampaikan rasa terimakasih sebesar-besarnya kepada:

1. Kedua orang tua saya dan keluarga yang telah memberikan dukungan baik moril maupun materil serta doa yang tiada henti-hentinya kepada penulis.
2. Bapak Rifkie Primartha, M.T selaku dosen pembimbing I dan Bapak Ir. Muhammad Ihsan Jambak, M.Sc. selaku pembimbing II yang telah membimbing, mengarahkan, dan memberikan motivasi kepada penulis dalam proses pengerjaan Tugas Akhir, hingga penulis dapat menyelesaikan Tugas Akhir dengan baik.
3. Ibu Mastura Diana Marieska, S.T., M.T. selaku dosen pembimbing akademik, yang telah membimbing, mengarahkan dan memberikan motivasi kepada penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
4. Bapak Abdiansah, S.Kom., M.Cs. selaku dosen penguji I, dan Ibu Desty Rodiah, M.T. selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
5. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Mbak Wiwin dan seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.

7. Adit, Melvin, dan Irfan yang telah banyak membantu, memberikan dukungan, informasi dan referensi selama proses pengerjaan Tugas Akhir hingga penulis dapat menyelesaikan Tugas Akhir dengan baik.
8. Rizky, Nadiah, Jeje, Vita serta seluruh teman jurusan Teknik Informatika yang tidak dapat disebutkan satu persatu yang telah berbagi keluh kesah, motivasi, semangat, dan hiburan selama masa perkuliahan.
9. Fatiah, Monica, Fero, Mkh, Mutek, Siti, dan Nujul teman-teman semasa SMA yang tetap mendukung dan menemani saya hingga saat 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.

Palembang, Desember 2020

Erindah Nuraprilliana Sari

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xvii
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	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 KAJIAN LITERATUR	II-1
2.1 Pendahuluan.....	II-1
2.2 Landasan Teori	II-1
2.2.1 Reduksi Dimensi	II-1
2.2.2 <i>Information Gain</i> (IG)	II-2
2.2.3 <i>Modified k-Nearest Neighbor</i> (MkNN).....	II-3
2.2.4 <i>k-Fold Cross Validation</i>	II-5
2.2.5 Evaluasi Hasil Pengujian Secara Statistik	II-6

2.3 Penelitian Lain yang Relevan.....	II-8
2.4 Kesimpulan.....	II-10
BAB III METODOLOGI PENELITIAN	III-1
3.1 Pendahuluan	III-1
3.2 Pengumpulan Data.....	III-1
3.2.1 Jenis dan Sumber Data	III-1
3.3 Tahapan Penelitian	III-2
3.3.1 Menetapkan Kerangka Kerja / Framework	III-2
3.3.2 Menetapkan Kriteria Pengujian.....	III-4
3.3.3 Menetapkan Format Data Pengujian	III-5
3.3.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian	III-6
3.3.5 Melakukan Pengujian Penelitian	III-7
3.3.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan.....	III-9
3.4 Metode Pengembangan Perangkat Lunak	III-9
3.4.1 Fase Insepsi	III-9
3.4.2 Fase Elaborasi.....	III-11
3.4.3 Fase Konstruksi	III-11
3.4.4 Fase Transisi.....	III-12
BAB IV PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1 Pendahuluan	IV-1
4.2 <i>Rational Unified Process (RUP)</i>	IV-1
4.2.1 Fase Insepsi	IV-1
4.2.2 Fase Elaborasi.....	IV-13
4.2.3 Fase Konstruksi	IV-27
4.2.4 Fase Transisi	IV-29
4.2.4.3 Pengujian	IV-30
4.3 Kesimpulan.....	IV-39
BAB V HASIL DAN ANALISIS PENELITIAN.....	V-1
5.1 Pendahuluan	V-1
5.2 Data Hasil Percobaan/Penelitian	V-1

5.2.1 Konfigurasi Percobaan	V-1
5.2.2 Pengujian Akurasi	V-2
5.2.3 Pengujian Waktu	V-15
5.2.4 Pengujian Memori	V-20
5.3 Analisa Hasil Pengujian.....	V-25
5.3.1 Analisis Hasil Pengujian Klasifikasi Modified k-Nearest Neighbor Tanpa Reduksi Dimensi dan Modified k-Nearest Neighbor Dengan Reduksi Dimensi Menggunakan <i>Information Gain</i>	V-25
5.3.2 Analisis Hasil Pengujian Klasifikasi Modified k-Nearest Neighbor Tanpa Reduksi Dimensi dan Modified k-Nearest Neighbor Dengan Reduksi Dimensi Menggunakan Algoritma Genetika.....	V-26
5.4 Kesimpulan.....	V-27
 BAB VI KESIMPULAN DAN SARAN	 VI-1
6.1 Kesimpulan.....	VI-1
6.2 Saran	VI-3
DAFTAR PUSTAKA	xvii
LAMPIRAN.....	L-1

DAFTAR TABEL

	Halaman
III-1. Rancangan Tabel Hasil Pengujian Prosentase Fitur Optimum	
<i>Information Gain</i> dan Parameter k Optimum.....	III-5
III-2. Rancangan Tabel Hasil Pengklasifikasian Data Sebelum dan	
Sesudah di Reduksi	III-6
III-3. Rancangan Tabel Hasil Analisis Pengklasifikasian	III-10
IV-1. Tabel Kebutuhan Fungsional	IV-3
IV-2. Tabel Kebutuhan Non-Fungsional	IV-3
IV-3. Definisi Aktor <i>Use Case</i>	IV-5
IV-4. Definisi <i>Use Case</i>	IV-5
IV-5. Skenario <i>Use Case</i> Melakukan Muat Data	IV-6
IV-6. Skenario <i>Use Case</i> Melakukan Proses Data	IV-8
IV-7. Skenario <i>Use Case</i> Melakukan Klasifikasi MkNN tanpa Reduksi	
Dimensi	IV-9
IV-8. Skenario <i>Use Case</i> Melakukan Klasifikasi MkNN dengan	
Reduksi Dimensi	IV-10
IV-9. Implementasi Kelas	IV-29
IV-10. Rencana Pengujian <i>Use Case</i> Melakukan Muat Data	IV-31
IV-11. Rencana Pengujian <i>Use Case</i> Melakukan Proses Data	IV-31
IV-12. Rencana Pengujian <i>Use Case</i> Klasifikasi MkNN tanpa Reduksi	
Dimensi	IV-32
IV-13. Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi MkNN	
dengan Reduksi Dimensi.....	IV-32
IV-14. Pengujian <i>Use Case</i> Melakukan Muat Data.....	IV-34
IV-15. Pengujian <i>Use Case</i> Melakukan Proses Data.....	IV-35
IV-16. Pengujian <i>Use Case</i> Klasifikasi MkNN tanpa Reduksi	
Dimensi	IV-36

IV-17. Pengujian <i>Use Case</i> Klasifikasi MkNN dengan Reduksi	
Dimensi	IV-37
V-1. Hasil Pengujian Prosentase Fitur <i>Information Gain</i> dan	
Parameter <i>k</i>	V-4
V-2. Hasil Uji Normalitas Parameter <i>k</i>	V-6
V-3. Hasil Uji Kruskal-Wallis Parameter <i>k</i>	V-7
V-4. Hasil Uji Normalitas Prosentase Fitur	V-8
V-5. Hasil Uji Homogenitas Prosentase Fitur	V-9
V-6. Hasil Uji Anova Prosentase Fitur	V-9
V-7. Hasil Rata-Rata <i>Ranking</i> Prosentase Fitur	V-10
V-8. Perbandingan Hasil Klasifikasi MkNN Tanpa Reduksi Dimensi	
dan Dengan Reduksi Dimensi	V-12
V-9. Hasil Uji Normalitas Akurasi	V-13
V-10. Hasil Uji Homogenitas Akurasi	V-14
V-11. Hasil Uji Anova Akurasi	V-14
V-12. Hasil Rata-Rata <i>Ranking</i> Akurasi	V-15
V-13. Hasil Uji Normalitas Waktu	V-16
V-14. Hasil Uji Kruskal-Wallis Waktu	V-16
V-15. Hasil Uji Man-Whitney Perbandingan Waktu pada Data Hasil	
Pengklasifikasian MkNN dan IG + MkNN	V-17
V-16. Hasil Uji Man-Whitney Perbandingan Waktu pada Data Hasil	
Pengklasifikasian MkNN dan GA + MkNN	V-18
V-17. Hasil Uji Man-Whitney Perbandingan Waktu pada Data Hasil	
Pengklasifikasian IG + MkNN dan GA + MkNN	V-19
V-18. Hasil Uji Normalitas Memori	V-20
V-19. Hasil Uji Kruskal-Wallis Memori	V-21
V-20. Hasil Uji Man-Whitney Perbandingan Memori pada Data Hasil ...	
Pengklasifikasian MkNN dan IG + MkNN	V-22
V-21. Hasil Uji Man-Whitney Perbandingan Memori pada Data Hasil ...	
Pengklasifikasian MkNN dan GA + MkNN	V-23

V-22. Hasil Uji Man-Whitney Perbandingan Memori pada Data Hasil ...
Pengklasifikasian IG + MkNN dan GA + MkNN..... V-24

DAFTAR GAMBAR

	Halaman
II-1. Ilustrasi <i>k-Fold Cross Validation</i>	II-6
III-1. Diagram Tahapan Penelitian	III-8
IV-1. Diagram <i>Use Case</i>	IV-4
IV-2. Diagram <i>Sequence</i> Melakukan Muat Data	IV-15
IV-3. Diagram <i>Sequence</i> Melakukan Proses Data	IV-15
IV-4. Diagram <i>Sequence</i> Melakukan Klasifikasi MkNN tanpa Reduksi Dimensi	IV-16
IV-5. Diagram <i>Sequence</i> Melakukan Klasifikasi MkNN dengan Reduksi . Dimensi	IV-17
IV-6. Rancangan Antarmuka Perangkat Lunak	IV-18
IV-7. Rancangan Fitur Muat Dataset	IV-18
IV-8. Rancangan Fitur Pilihan Proses Klasifikasi	IV-19
IV-9. Rancangan Fitur Input Parameter	IV-20
IV-10. Rancangan Fitur Tombol Mulai	IV-20
IV-11. Rancangan Fitur Hasil Klasifikasi.....	IV-21
IV-12. Rancangan Hasil Klasifikasi Setiap k-Fold.....	IV-21
IV-13. Diagram Aktivitas Melakukan Muat Data	IV-23
IV-14. Diagram Aktivitas Melakukan Proses Data	IV-24
IV-15. Diagram Aktivitas Melakukan Klasifikasi MkNN tanpa Reduksi..... Dimensi	IV-25
IV-16. Diagram Aktivitas Melakukan Klasifikasi MkNN dengan	
Reduksi Dimensi	IV-26
IV-17. Diagram Kelas.....	IV-28

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini memberikan penjelasan umum mengenai keseluruhan penelitian. Pendahuluan dimulai dengan penjelasan mengenai yang menjadi latar belakang dari penelitian ini. Antara lain membahas tentang latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan masalah yang menjadi dasar dalam penelitian ini.

1.2 Latar Belakang Masalah

Klasifikasi merupakan salah satu topik utama yang dibahas dalam *data mining* atau *machine learning*. Klasifikasi adalah pengelompokan data dengan cara mengenali pola tertentu dari data. Lalu pola tersebut digunakan untuk menentukan label yang belum diketahui dari data baru. Klasifikasi sebagai salah satu proses dalam *data mining* bertujuan untuk mendapatkan informasi yang berguna serta meningkatkan pengetahuan pengguna.

Modified K-Nearest Neighbor merupakan salah satu metode yang digunakan dalam melakukan klasifikasi dan metode ini adalah perkembangan dari metode *K-Nearest Neighbor*. Pada *Modified K-Nearest Neighbor*, setiap sampel harus divalidasi terlebih dahulu. Mekanisme kerja dari *Modified k-Nearest Neighbor* hampir sama dengan *k-Nearest Neighbor* yaitu menghitung jarak setiap

atribut yang ada. Proses komputasi klasifikasi membutuhkan waktu yang lama dan memori yang besar jika data memiliki dimensi yang tinggi. 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). Oleh karena itu sebelum melakukan proses klasifikasi akan dipilih terlebih dahulu atribut yang relevan dan bisa meningkatkan proses klasifikasi dari algoritma *Modified k-Nearest Neighbor*.

Untuk meningkatkan proses klasifikasi, perlu dilakukan proses seleksi fitur. Hatta, Arifin et al. (2013) mengungkapkan bahwa seleksi fitur adalah suatu cara yang dilakukan untuk memilih fitur-fitur yang relevan dan dapat mewakili suatu data. Penelitian Kurniawan (2018) mengenai pengaruh algoritma genetika terhadap hasil klasifikasi menggunakan algoritma MkNN, diperoleh hasil bahwa algoritma genetika sebagai seleksi fitur dapat mengurangi waktu komputasi dan memori yang diperlukan dalam melakukan klasifikasi. Kesimpulan yang didapat dari penelitian tersebut adalah metode seleksi fitur algoritma genetika dapat digunakan untuk mereduksi data berdimensi tinggi, maka pada penelitian ini akan melakukan reduksi dimensi dengan metode seleksi fitur yang lainnya. Ada banyak metode untuk seleksi fitur seperti metode *Information Gain*, metode *Particle Swarm Optimization*, metode *Chi Square*, metode *Ant Colony Optimization*, metode *Mutual Information* dan metode yang lainnya. Pada penelitian ini metode seleksi fitur yang digunakan adalah metode *Information Gain*.

Information Gain adalah metode seleksi fitur yang digunakan untuk menentukan seberapa penting atribut. *Information Gain* diperoleh dengan mengetahui nilai atribut, yang merupakan entropi distribusi sebelum perpecahan

dikurangi entropi distribusi setelahnya. *Information Gain* terbesar setara dengan entropi terkecil. *Information Gain* dapat mereduksi dimensi yang tinggi pada suatu data dengan mempertimbangkan atribut-atribut yang relevan untuk digunakan dalam proses klasifikasi.

Berdasarkan uraian tersebut, penelitian ini akan melihat pengaruh penerapan metode *Information Gain* dalam klasifikasi menggunakan algoritma *Modified k-Nearest Neighbor*.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang dikemukakan diatas, rumusan masalah dalam penelitian ini adalah bagaimana pengaruh reduksi dimensi terhadap klasifikasi menggunakan algoritma *Modified k-Nearest Neighbor*. Untuk menyelesaikan permasalahan yang disebutkan diatas, maka dalam penelitian ini dibagi menjadi 4 *research question* yaitu:

1. Bagaimana mekanisme kerja dari metode *Information Gain* sebagai metode reduksi dimensi?
2. Bagaimana cara kerja dari algoritma *Modified k-Nearest Neighbor* untuk melakukan klasifikasi?
3. Apa hubungan antara algoritma *Modified k-Nearest Neighbor* dan *Information Gain* dalam melakukan klasifikasi?
4. Apa perbedaan hasil reduksi oleh *Information Gain* dibandingkan dengan reduksi oleh Algoritma Genetika?

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Mengetahui mekanisme kerja dari *Information Gain* sebagai metode reduksi dimensi dan algoritma *Modified k-Nearest Neighbor* sebagai algoritma klasifikasi.
2. Mengetahui hubungan antara algoritma *Modified k-Nearest Neighbor* dan *Information Gain* dalam melakukan klasifikasi.
3. Mengetahui perbedaan hasil reduksi *Information Gain* dan Algoritma Genetika.
4. Menganalisis efektifitas metode reduksi dimensi menggunakan *Information Gain* untuk meningkatkan klasifikasi dibanding Algoritma Genetika.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah sebagai berikut :

1. Memahami hubungan antara algoritma *Modified k-Nearest Neighbor* dan *Information Gain* dalam melakukan klasifikasi.
2. Mampu menerapkan algoritma *Modified k-Nearest Neighbor* dan *Information Gain* dalam melakukan klasifikasi.
3. Menjadi rujukan bagi peneliti yang ini melanjutkan penelitian menggunakan reduksi dimensi dalam klasifikasi.

1.6 Batasan Masalah

Batasan masalah dalam penelitian ini adalah :

1. Metode reduksi dimensi yang digunakan adalah *Information Gain*.
2. Metode klasifikasi yang digunakan adalah *Modified k-Nearest Neighbor*.
3. Metode pengujian yang dilakukan dengan mengukur nilai akurasi, waktu komputasi dan memori yang dibutuhkan.

1.7 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah/ruang lingkup, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian, antara lain penelitian terkait, definisi-definisi Reduksi Dimensi, *Information Gain*, Algoritma *Modified k-Nearest Neighbor*, *K-Fold Cross Validation*, *Confusion Matrix* dan pada bab ini juga membahas penelitian-penelitian lain yang relevan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dibahas mengenai 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

Pada bab ini akan membahas tahapan-tahapan yang dilakukan dalam proses pengembangan perangkat lunak berdasarkan metode *Rational Unified Process* (RUP).

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini akan membahas hasil pengujian berdasarkan rencana pengujian yang telah ditentukan. Analisis hasil penelitian akan diberikan sebagai dasar dari kesimpulan yang diambil pada penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini berisikan kesimpulan dan saran dari penelitian ini. Kesimpulan dan saran diambil dari semua pembahasan yang dilakukan pada bab sebelumnya..

1.8 Kesimpulan

Penelitian ini akan melihat pengaruh reduksi dimensi terhadap klasifikasi menggunakan algoritma *Modified k-Nearest Neighbor* dan metode reduksi dimensi yang digunakan adalah *Information Gain*.

DAFTAR PUSTAKA

- Aini, S. H. A., Sari, Y. A., & Arwan, A. (2018). Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(9), 2546–2554.
- 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.
- Firmahsyah, & Gantini, T. (2016). Penerapan Metode Content-Based Filtering Pada Sistem Rekomendasi Kegiatan Ekstrakurikuler (Studi Kasus di Sekolah ABC). *Jurnal Teknik Informatika Dan Sistem Informasi*, 2(3). <https://doi.org/10.28932/jutisi.v2i3.548>.
- Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12): Springer Science & Business Media.
- Ghozali, Imam. (2009). *Aplikasi Analisis Multivariate dengan Program SPSS*. Semarang: Universitas Diponegoro.
- Hatta, H. R., Arifin, A. Z., & Yuniarti, A. (2013). *Information Gain Untuk Seleksi Fitur Pada Dokumen Teks Arab. VIII*, 1–4.
- 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.

- Idrees, F., et al. (2017). PIndroid: A novel Android malware detection system using ensemble learning methods. *Computers & Security*, 68, 36-46.
- Kurniawan, A., Jambak, M. I., & Yusliani, N. (n.d.). *The Effect of Dimensionality Reduction Using Genetic Algorithm on Classification Result of k-Nearest Neighbor and Modified k-Nearest Neighbor*. 3–8.
- 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.
- Naseriparsa, M., & Bigdoli, A. (2013). (Weka) Improving Performance of a Group of Classification Algorithms Using Resampling and Feature Selection. *World of Computer Science and Information Technology Journal*, 3(4), 70–76.
- 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.
- 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.
- 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.
- Shaltout, N. A., El-Hefnawi, M., Rafea, A., & Moustafa, A. (2014). Information gain as a feature selection method for the efficient classification of influenza based on viral hosts. *Lecture Notes in Engineering and Computer Science*, 1, 625–631.

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.