

SKRIPSI

**ANALISIS PERBANDINGAN METODE SELEKSI GEN
DALAM IDENTIFIKASI KANDIDAT BIOMARKER
DIABETES MELITUS TIPE II**

Disusun dan diajukan oleh

MUKRIMAH RAMDAYANI

H131 14 322



**PROGRAM STUDI SISTEM INFORMASI
DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2021**

**ANALISIS PERBANDINGAN METODE SELEKSI GEN
DALAM IDENTIFIKASI KANDIDAT BIOMARKER
DIABETES MELITUS TIPE II**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains
pada Program Studi Sistem Informasi Departemen Matematika Fakultas
Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin**

MUKRIMAH RAMDAYANI

H131 14 322

**PROGRAM STUDI SISTEM INFORMASI
DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR**

2021

LEMBAR PERNYATAAN KEOTENTIKAN

Yang bertanda tangan dibawah ini :

Nama : Mukrimah Ramdayani

NIM : H13114322

Program Studi : Sistem Informasi

Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

ANALISIS PERBANDINGAN METODE SELEKSI GEN DALAM IDENTIFIKASI KANDIDAT BIOMARKER DIABETES MELITUS TIPE II

Adalah benar hasil karya saya sendiri bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini merupakan hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 24 Agustus 2021

Yang Menyatakan



Mukrimah Ramdayani

**ANALISIS PERBANDINGAN METODE SELEKSI GEN DALAM
IDENTIFIKASI KANDIDAT BIOMARKER
DIABETES MELITUS TIPE II**

Disetujui oleh :

Pembimbing Utama

Pembimbing Pertama


Dr. Eng. Armin Lawi, M.Eng.

NIP. 19720423 199512 1 001


Supri Bin Hj Amir, S.Si., M.Eng

NIP. 19880504 201903 1 012

Ketua Program Studi


Dr. Muhammad Hasbi, M.Sc

NIP. 19630720 198903 1 003



Pada Tanggal : 24 Agustus 2021

LEMBAR PENGESAHAN SKRIPSI

**ANALISIS PERBANDINGAN METODE SELEKSI GEN DALAM
IDENTIFIKASI KANDIDAT BIOMARKER
DIABETES MELITUS TIPE II**

Disusun dan diajukan oleh :

MUKRIMAH RAMDAYANI

H131 14 322

Telah dipertahankan dihadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana pada Program Studi Sistem Informasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

Pada tanggal 24 Agustus 2021

dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui

Pembimbing Utama

Pembimbing Pertama

Dr. Eng. Armin Lawi, M.Eng.

NIP. 19720423 199512 1 001

Supri Bin Hj Amir, S.Si., M.Eng

NIP. 19880504 201903 1 012

Ketua Program Studi

Dr. Muhamad Hasbi, M.Sc

NIP. 19630720 198903 1 003



KATA PENGANTAR



Alhamdulillah, segala puji hanyalah milik Allah, Tuhan yang Maha Rahman dan Maha Rahim, yang telah memberikan keberkahanNya berupa Ilmu kepada penulis sehingga dapat menyelesaikan penyusunan skripsi dengan judul “**Analisis Perbandingan Metode Seleksi Gen Dalam Identifikasi Kandidat Biomarker Diabetes Melitus Tipe II**” untuk memenuhi salah satu persyaratan dalam meraih gelar Sarjana Sains pada Program Studi Sistem Informasi, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin. Shalawat serta salam tak lupa penulis haturkan untuk Rasulullah *Shallallahu ‘Alaihi Wasallam*, insan pemberi Syafaat, insan yang sempurna akal dan budi pekerti sebagai Uswatun Hasanah.

Penulis menyadari sepenuhnya bahwa karya tulis ini dapat terwujud berkat bantuan dari berbagi pihak. Untuk itu pada kesempatan kali ini, penulis menghaturkan penghargaan dan terimakasih yang paling tulus kepada Ayahanda **Saabuddin** dan Ibunda **Rosmiati** yang telah membesarkan penulis dengan cinta, kasih sayang, serta penuh keikhlasan hati, yang telah mengiringi dan menyemangati setiap langkah penulis dengan doa dan restunya, demi kelanjutan belajar penulis.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada :

1. Ibu **Prof. Dr. Dwia Aries Tina Palubuhu, MA**, selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya, Bapak **Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya, Bapak **Prof. Dr. Nurdin, S.Si., M.Si**, selaku Ketua Departemen Matematika dan segenap **Dosen Pengajar dan Staf Departemen Matematika** yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen

Matematika serta semua pihak birokrasi atas ilmu dan kemudahan-kemudahan yang diberikan, baik dibidang akademik maupun dibidang kemahasiswaan.

2. Bapak **Dr. Eng. Armin Lawi, M.Eng.**, selaku pembimbing utama dan Bapak **Supri bin Hj. Amir, S.Si., M.Eng.**, selaku pembimbing pertama, yang telah memberikan banyak bantuan atas nasehat, dukungan, doa dan dengan setulus hati telah meluangkan waktunya ditengah berbagai kesibukan dan prioritasnya untuk membimbing penulis menyelesaikan tugas akhir ini.
3. Bapak **Dr. Hendra, S.Si., M.Kom.** dan Bapak **Edy Saputra, S.Si., M.Si.**, selaku penguji atas kesediaannya untuk memberikan kritik, saran dan dukungan yang membangun kepada penulis dalam perbaikan skripsi ini.
4. Sahabat-sahabat **Sekret Telkomas (Nita, Yuyu, Nuhi, Nanda, Ica, Murni, Firda)** juga **Sukma** dan **Iyam** yang telah banyak membantu dan memberi saran serta support dalam menyelesaikan skripsi ini.
5. Teman-teman **TRANSPOSE 2014 Annisaul Mukrmina, Munawir Djamaluddin, Arfyan Saputra, Muh. Sarwan, Irwan, Syahrul, Eko, Ridho** yang telah banyak membantu penulis dalam pengumpulan berkas dan informasi yang diberikan, juga untuk **Nurul Rahma, Meylina, Fitriana, Seltuti, April, Niar, Hikma, Lintang** serta teman-teman lainnya yang tidak bisa disebut satu per satu. Terima kasih untuk dukungan dan waktunya saat berjuang sama-sama di dunia perkuliahan dan di himpunan. Juga untuk ketua angkatan **Fandy Heribet** yang selalu memberi dukungan pada penulis untuk pengerjaan tugas akhir.
6. Teman-teman seperjuangan **Ilmu Komputer 2014 (Khalil, Yaumil, Jo, Luki, Ola, Ayu, Fuad, Dila, Danti, Syam, Titin, Yusri, dll)** yang selalu memberikan semangat serta saran untuk penulis
7. Seluruh teman-teman **KKN UNHAS GEL.96 Kec. Tinggimoncong, Kab. Gowa** terkhusus kepada teman posko **Kelurahan Garassi: Ica, Nisa, Niar, Nasir, Kak Azam dan Kak Sandy** serta warga kelurahan Garassi yang telah menjadi teman serta keluarga baru dan semoga kedepannya silaturahmi yang telah dibangun bersama tetap terjalin dengan baik.

8. Kepada semua pihak yang tidak dapat penulis sebutkan satu-persatu, semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah di sisi **Allah *Subhanahu Wata'ala***.

Semoga bantuan dan harapan dan doanya bukanlah suatu hal yang sia-sia dihadapan Allah *Subhanahu Wata'ala*. Akhir kata penulis mengharapkan kritikan dan saran yang bersifat membangun demi kesempurnaan tulisan ini. Sebab penulis menyadari kesalan dan kekurangan tidak pernah lepas dari setiap manusia sebagai makhluk Allah. Serta semoga skripsi ini dapat bermanfaat dan menambah ilmu pengetahuan bagi kita semua.

Semoga Allah *Subhanahu Wata'ala* meridhoi segala langkah kita semua. Akhirnya dengan segala kekurangan penulis mempersembahkan karya tulis ini.

Makassar, 24 Agustus 2021

Mukrimah Ramdayani

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Hasanuddin, saya yang bertanda tangan dibawah ini :

Nama : Mukrimah Ramdayani
NIM : H13114322
Program Studi : Sistem Informasi
Departemen : Matematika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Prediktor Royalti Noneklusif (*Non-exclusive Royalty-Free Right*)** atas tugas akhir saya yang berjudul :

**Analisis Perbandingan Metode Seleksi Gen Dalam Identifikasi Kandidat
Biomarker Diabetes Melitus Tipe II**

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal diatas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada 24 Agustus 2021

Yang menyatakan

(Mukrimah Ramdayani)

ABSTRAK

Penelitian mengenai penanganan yang tepat untuk penderita Diabetes Melitus (DM) tipe II telah banyak dilakukan. Salah satu penelitian mengenai penanganan yang tepat untuk penderita DM tipe II yaitu dengan mengidentifikasi kandidat biomarker yang berasal dari gen yang paling signifikan dengan memanfaatkan 25770 data ekspresi gen pada DM tipe II, selanjutnya hasil identifikasi gen informatif tersebut digunakan untuk meningkatkan hasil klasifikasi. Salah satu pendekatan yang dapat digunakan untuk mengidentifikasi biomarker adalah dengan menggunakan teknologi DNA *microarray*. Pengolahan dari data *microarray* dimulai dengan tahap preprocessing. Selanjutnya melalui tahap filtering menggunakan uji-t, *Singular Value Decomposition* (SVD), *Information Gain* (IG), Chi-Square, dan *Random Forest Importance*. Tahap terakhir dilakukan proses seleksi fitur menggunakan metode *Support Vector Machine* (SVM), Analisis Diskriminan Linier (LDA), *Random Forest-Recursive Feature Elimination* (RF-RFE) dan *Support Vector Machine-Recursive Feature Elimination* (SVM-RFE). Performansi dari semua metode seleksi fitur yang digunakan menunjukkan bahwa metode RF-RFE memperoleh nilai akurasi yang lebih baik dibandingkan dengan metode SVM, LDA dan SVM-RFE.

Kata kunci : Diabetes melitus Tipe II, *Microarray*, Biomarker, Support Vector Machine-Recursive Feature Elimination, Random Forest-Recursive Feature Elimination, SVM, Analisis Diskriminasi Linier

ABSTRACT

Many studies have been carried out on the appropriate treatment for type II Diabetes Mellitus (DM) patients. One of the studies regarding the appropriate treatment for patients with type II DM is to identify candidate biomarkers derived from the most significant genes by utilizing 25770 gene expression data in type II DM, then the results of the identification of informative genes are used to improve classification results. One approach that can be used to identify biomarkers is to use DNA microarray technology. Processing of microarray data begins with the preprocessing stage. Then through the filtering stage using the t-test, Singular Value Decomposition (SVD), Information Gain (IG), Chi-Square, and Random Forest Importance. The last stage is the feature selection process using the Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Random Forest-Recursive Feature Elimination (RF-RFE) and Support Vector Machine-Recursive Feature Elimination (SVM-RFE). The performance of all feature selection methods used shows that the RF-RFE method obtains better accuracy values than the SVM, LDA and SVM-RFE methods.

Keywords: *Diabetes mellitus Type II, Microarray, Biomarker, Support Vector Machine-Recursive Feature Elimination, Random Forest-Recursive Feature Elimination, SVM, Linear Discrimination Analysis*

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
LEMBAR PERNYATAAN KEOTENTIKAN.....	iii
LEMBAR PERSETUJUAN PEMBIMBING	iv
LEMBAR PENGESAHAN SKRIPSI.....	v
KATA PENGANTAR	vi
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	ix
ABSTRAK	x
<i>ABSTRACT</i>	xi
DAFTAR ISI.....	xii
DAFTAR GAMBAR	xv
DAFTAR TABEL.....	xvii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian	3
1.4. Manfaat Penelitian	4
1.5. Batasan Masalah.....	4
1.6. Organisasi Skripsi	4
BAB III TINJAUAN PUSTAKA	6
2.1. Landasan Teori.....	6
2.1.1. Diabetes Melitus Tipe 2	6
2.1.2. Biomarker.....	8
2.1.3. <i>Microarray</i>	9
2.1.4. Seleksi Fitur	11
2.1.5. Uji <i>Shapiro Wilk</i>	13
2.1.6. Transformasi Logaritma.....	13
2.1.7. Normalisasi <i>Quantile</i>	15

2.1.8. Uji-t Dua Sampel Bebas.....	16
2.1.9. Metode <i>Synthetic Minority Oversampling Technique</i>	17
2.1.10. <i>Chi-Square</i>	18
2.1.11. <i>Singular Value Decomposition</i>	20
2.1.12. <i>Information Gain</i>	21
2.1.13. Analisis Diskriminasi Linier	21
2.1.14. Hutan Acak (<i>Random Forest</i>)	25
2.1.15. <i>Super Vector Machine</i>	27
2.1.16. <i>Support Vector Machine-Recursive Feature Elimination</i>	32
2.1.17. <i>Cross Validation</i>	33
2.1.18. <i>Confussion Matrix</i>	35
2.1.19. <i>Receiver Operating Characteristic Curve</i>	37
2.2. Kerangka Konseptual	38
BAB IV METODE PENELITIAN	40
3.1. Sumber Data.....	40
3.2. Identifikasi Variabel.....	40
3.3. Metode Analisis	40
3.4. Diagram Alir Penelitian	41
BAB V HASIL DAN PEMBAHASAN	42
4.1. Pembahasan.....	42
4.1.1. Dekripsi Data	42
4.1.2. Pengolahan <i>Data Microarray</i>	45
4.1.3. <i>Preprocessing</i>	45
4.1.4. <i>Filtering</i>	46
4.1.5. Metode Klasifikasi	48
4.1.6. <i>Training dan Testing</i>	49
4.2. Hasil Perbandingan	59
4.3. Tabel Perbandingan.....	60
BAB VI KESIMPULAN DAN SARAN	61
5.1. Kesimpulan	61
5.2. Saran.....	61

DAFTAR PUSTAKA 63

DAFTAR GAMBAR

Gambar 2.1 Cara Kerja Teknologi <i>Microarray</i>	9
Gambar 2.2 <i>Affymetrix Genechip</i> (Cahyo, 2018)	11
Gambar 2.3 Visualisasi Cara Kerja LDA	22
Gambar 2.4 Gambaran Hasil Akhir Metode LDA	25
Gambar 2.5 Contoh <i>hyperplane</i> yang memisahkan dua kelas (Hall, 2012).....	28
Gambar 2.6 <i>Hyperplane</i> yang terbentuk diantara kelas -1 dan +1 (Hall, 2012)..	28
Gambar 2.7 Pemetaan data pada input space ke dimensi yang lebih tinggi menggunakan <i>hyperplane</i> (Munawarah dkk, 2016).....	31
Gambar 2.8 Kerangka Konseptual.....	39
Gambar 3.1 Diagram Alir Penelitian.....	41
Gambar 4.1 <i>Pie Chart</i> dari Pasien.....	44
Gambar 4.2 Kurva ROC Metode LDA untuk 150 Fitur Terbaik dengan Data Latih 70% dan Data Uji 30%.....	50
Gambar 4.3 Kurva ROC Metode LDA untuk 150 Fitur Gen dengan Data Latih 80% dan Data Uji 20%.....	51
Gambar 4.4 Kurva ROC Metode RF-RFE untuk 20 Fitur Terbaik dengan Data Latih 70% dan Data Uji 30%	51
Gambar 4.5 Kurva ROC Metode RF-RFE untuk 10 Fitur Terbaik dengan Data Latih 80% dan Data Uji 20%	52
Gambar 4.6 Kurva ROC Metode RF-RFE untuk 30 Fitur Terbaik dengan Data Latih 80% dan Data Uji 20%	52
Gambar 4.7 Kurva ROC Metode RF-RFE untuk 40 Fitur Terbaik dengan Data Latih 80% dan Data Uji 20%	53
Gambar 4.8 Kurva ROC Metode SVM-RFE untuk 200 Fitur Terbaik dengan Data Latih 70% dan Data Uji 30%	53
Gambar 4.9 Kurva ROC Metode SVM untuk 10 Fitur Terbaik dengan Data Latih 80% dan Data Uji 20%.....	54
Gambar 4.10 Kurva ROC Metode SVM untuk 40 Fitur Terbaik dengan Data Latih 80% dan Data Uji 20%.....	54

Gambar 4.11 Kurva ROC untuk 22 Fitur Gen dengan Data Latih 70% dan Data Uji 30%	57
Gambar 4.12 Kurva ROC untuk 50 Fitur Gen dengan Data Latih 70% dan Data Uji 30%	57
Gambar 4.13 Kurva ROC untuk 22 Fitur Gen dengan Data Latih 80% dan Data Uji 20%	58
Gambar 4.14 Kurva ROC untuk 22 Fitur Gen dengan Data Latih 80% dan Data Uji 20%	58

DAFTAR TABEL

Tabel 2.1 Bentuk Umum Tabel Kontingensi.....	19
Tabel 2.2 Skenario <i>10-Fold Cross Validation</i> (Kohavi, 1995)	34
Tabel 2.3 Format <i>Confusion Matrix</i>	35
Tabel 2.4 Tingkat Akurasi AUC	38
Tabel 4.1 Data Ekspresi Gen.....	43
Tabel 4.2 <i>Pheno Data</i> GSE18732	44
Tabel 4.3 Hasil Perbandingan Metode LDA, RF, dan SVM-RFE	50
Tabel 4.4 Hasil Perbandingan Akurasi dari Kernel yang Berbeda dengan Porsi Data 70% dan 30%	55
Tabel 4.5 Hasil Perbandingan Akurasi dari Kernel yang Berbeda dengan Porsi Data 80% dan 20%	56
Tabel 4.6 Tabel Perbandingan seluruh metode	60

BAB I

PENDAHULUAN

1.1. Latar Belakang

Diabetes Melitus (DM) tipe II merupakan penyakit kronik yang terjadi ketika pankreas tidak dapat lagi memproduksi insulin dalam jumlah yang cukup atau dapat juga disebabkan oleh berkurangnya kemampuan tubuh untuk merespon kerja insulin secara efektif (Soegondo dkk, 2009). Diabetes tipe ini tidak menunjukkan gejala yang dapat dilihat secara langsung sehingga sulit untuk mendeteksi penyakit tersebut. Jumlah penderita DM di dunia terus meningkat. Menurut International Diabetes Federation (IDF) pada tahun 2019, ada sekitar 463 juta orang di seluruh dunia menderita DM, diperkirakan prevalensinya akan terus meningkat mencapai 578 juta jiwa pada tahun 2030, dan 700 juta jiwa pada tahun 2045 dimana proporsi kejadian DM tipe 2 adalah 90% dari populasi dunia yang menderita DM (Saeedi dkk, 2019).

Peningkatan jumlah penderita DM terus meningkat setiap tahunnya terutama penderita DM tipe II namun hingga saat ini belum ada penanganan medis yang diketahui dapat menyembuhkan DM secara permanen. Tindakan pengobatan pada pasien hanya berfungsi untuk melakukan pencegahan komplikasi dengan cara menjaga kestabilan gula darah dengan pengobatan secara rutin seumur hidup karena DM merupakan penyakit seumur hidup yang tidak bisa disembuhkan secara permanen (Pratita, 2012). Oleh sebab itu, penelitian mengenai penanganan yang tepat untuk penderita DM tipe 2 terus dilakukan salah satunya adalah dengan mengidentifikasi biomarker dari penyakit DM tipe 2. Biomarker dapat membantu pendeteksian dini terhadap suatu penyakit sehingga dapat dilakukan penanganan yang tepat terhadap penyakit tersebut. Identifikasi biomarker dari penyakit DM tipe 2 terus berkembang salah satunya adalah dengan menggunakan teknologi DNA *microarray*.

Biomarker merupakan molekul penanda yang khas bagi sel, yang dapat digunakan untuk mendiagnosa suatu penyakit dan terapi target molekuler penyebab penyakit tertentu (Ni'mah, 2015). Identifikasi kandidat biomarker menggunakan

ekspresi gen dapat memberikan hasil yang sangat penting karena dapat digunakan untuk mendiagnosa lebih awal penyakit DM serta membantu pengembangan pengobatan personal (*personalized medicine*) dan juga membantu para peneliti untuk menemukan penanganan yang tepat terhadap penyakit DM tipe II, yang mengarah pada perbaikan terapi target molekuler yang telah ada serta pengembangan pendekatan terapi target molekuler yang baru.

Pengembangan biomarker terutama dalam kondisi atau kasus-kasus seperti diabetes setidaknya dapat membantu diagnosa lebih awal, akan tetapi dalam penelitian biomarker di laboratorium membutuhkan waktu yang cukup lama sehingga dibutuhkan suatu pendekatan yang dapat digunakan untuk menemukan biomarker suatu penyakit. Salah satu pendekatan yang dapat digunakan untuk mengidentifikasi biomarker adalah dengan menggunakan teknologi DNA *microarray* yang dapat mengukur tingkat ekspresi ribuan gen dari sebuah sampel (pasien) secara simultan dalam sebuah eksperimen (Stekel, 2003).

Perkembangan teknologi data *microarray* DNA dalam penelitian biomedis saat ini mencapai intensitas perkembangan yang tinggi. Data *microarray* DNA dapat digunakan untuk melakukan diagnosis suatu penyakit dengan melihat tingkat ekspresi dari sekian ribu gen dibawah kondisi dan lingkungan eksperimental tertentu. Dalam data *microarray* DNA terdapat matriks data ekspresi gen dengan setiap kolom di dalam matriksnya merepresentasikan tingkat ekspresi masing-masing gen dari sebuah eksperimen tunggal dan setiap baris merepresentasikan ekspresi gen di semua eksperimen. Gen dalam hal ini merupakan representasi fitur yang terdapat dalam data *microarray* DNA untuk selanjutnya dapat diolah dan diinterpretasikan secara utuh.

Teknologi DNA *microarray* digunakan untuk menentukan tingkat ekspresi ribuan gen yang dilakukan dalam sekali percobaan, dan secara simultan memantau proses biologi yang sedang berlangsung (Siang dkk, 2015). Kekurangan dari data *microarray* sendiri adalah jumlah dimensi yang besar dikarenakan jumlah sampel yang lebih sedikit dibanding jumlah gen yang mencapai ribuan. Besarnya dimensi

dapat berakibat pada tingkat performansi yang rendah. Untuk mengatasi permasalahan tersebut dapat dilakukan dengan mereduksi dimensi data. Salah satu cara untuk mereduksi dimensi data adalah seleksi fitur. Seleksi fitur atau juga disebut dengan seleksi gen data *microarray* memiliki peran penting untuk menentukan gen-gen informatif yang dapat membedakan kelas penyakit, sehingga gen-gen tersebut dapat digunakan sebagai penanda diagnostik untuk melakukan tindakan medis.

Terdapat metode statistik untuk melakukan klasifikasi telah digunakan untuk melakukan seleksi fitur, diantaranya menggunakan metode SVM, SVM-RFE, LDA dan RF-RFE. Hanya saja, keempat metode ini masih perlu dilakukan pengujian untuk menentukan metode mana yang paling efektif. Sehingga, dapat dilakukan deteksi dini pada DM tipe II dengan tingkat akurasi yang lebih baik. Dan diharapkan pula, penelitian ini bisa menjadi langkah awal untuk penelitian selanjutnya sehingga ditemukan penanganan medis yang tepat dan kesembuhan permanen bagi penderita DM tipe II.

1.2. Rumusan Masalah

Berdasarkan uraian dalam latar belakang, dapat dirumuskan masalah sebagai berikut:

1. Bagaimana cara mengolah data *microarray* untuk memprediksi kandidat biomarker pada penyakit DM tipe 2?
2. Bagaimana performansi metode seleksi fitur yang digunakan untuk menganalisis data *microarray* penyakit DM tipe 2?

1.3. Tujuan Penelitian

Tujuan dari penelitian ini sebagai berikut:

1. Mengolah data *microarray* untuk memprediksi kandidat biomarket pada penyakit DM tipe 2
2. Memperoleh performansi metode seleksi fitur yang digunakan untuk data *microarray* penyakit DM tipe 2.

1.4. Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Sistem pengklasifikasian penyakit DM tipe II yang efektif
2. Sebagai penelitian awal dan bahan acuan untuk penelitian selanjutnya

1.5. Batasan Masalah

Berdasarkan permasalahan yang ada, maka batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Berdasarkan penelitian pada skripsi (Nilamyani, 2018), (Auliah, 2018), dan (Sukmawati, 2018) Metode yang dibandingkan antara lain SVM dengan tiga kernel, SVM-RFE, RF-RFE dan LDA.
2. Jumlah data yang digunakan 25770 fitur dengan 118 sampel pasien
3. Pengolahan data menggunakan perangkat lunak R

1.6. Organisasi Skripsi

Organisasi skripsi berisi rincian tentang urutan penulisan dari setiap bab dan bagian bab dalam skripsi, mulai dari bab I hingga bab V.

BAB I : Pendahuluan

Bagian pendahuluan menjelaskan mengenai latar belakang melakukan penulisan, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan organisasi skripsi.

BAB II : Tinjauan Pustaka

Bagian ini berisi uraian tentang kajian daftar pustaka dan hipotesis penilitan. Kajian pustaka mempunyai peran yang sangat penting, kajian pustaka berfungsi sebagai landasan teoritik dalam menyusun pernyataan penelitian, tujuan, serta hipotesis. Bab ini terdiri dari landasan teori, penelitian terkait, dan kerangka konseptual.

BAB III : Metodologi Penelitian

Bagian ini membahas mengenai komponen dari metode penelitian yaitu tahapan penelitian, waktu dan tempat penelitian, perancangan sistem, sumber data, dan instrumen penelitian.

BAB IV : Hasil dan Pembahasan

Bagian ini membahas mengenai pencapaian hasil penelitian dan pembahasannya.

BAB V : Penutup

Bagian ini berisi simpulan dan saran membahas mengenai penafsiran dan pemaknaan penulis terhadap hasil analisis penelitian.