

**PREDIKSI STADIUM KANKER PAYUDARA BERBASIS  
DATA MIRNA DAN GEN DENGAN SELEKSI  
FITUR DAN VOTING CLASSIFIER**



**UMN**

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**SKRIPSI**

**RENFRED LEEMAN  
00000056836**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG  
2025**

**PREDIKSI STADIUM KANKER PAYUDARA BERBASIS  
DATA MIRNA DAN GEN DENGAN SELEKSI  
FITUR DAN VOTING CLASSIFIER**



**SKRIPSI**

Diajukan sebagai salah satu syarat untuk memperoleh  
Gelar Sarjana Komputer (S.Kom.)

**RENFRED LEEMAN**

**00000056836**

**UMN**  
**UNIVERSITAS**  
**MULTIMEDIA**  
**NUSANTARA**  
**PROGRAM STUDI INFORMATIKA**  
**FAKULTAS TEKNIK DAN INFORMATIKA**  
**UNIVERSITAS MULTIMEDIA NUSANTARA**  
**TANGERANG**  
**2025**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Renfred Leeman  
Nomor Induk Mahasiswa : 00000056836  
Program Studi : Informatika

Skripsi dengan judul:

**Prediksi Stadium Kanker Payudara Berbasis Data miRNA dan Gen dengan Seleksi Fitur dan Voting Classifier**

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 01 Juli 2025



(Renfred Leeman)

UMM  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## HALAMAN PENGESAHAN

Skripsi dengan judul

### **PREDIKSI STADIUM KANKER PAYUDARA BERBASIS DATA MIRNA DAN GEN DENGAN SELEKSI FITUR DAN VOTING CLASSIFIER**

oleh

Nama : Renfred Leeman  
NIM : 00000056836  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Jumat, 18 Juli 2025

Pukul 10.00 s/d 12.00 dan dinyatakan

**LULUS**

Dengan susunan penguji sebagai berikut

Ketua Sidang

Penguji

  
(Dr. Maria Irminda Prasetyowati, S.Kom.,  
M.T.)

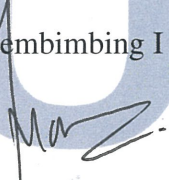
  
(Moeljono Widjaja, B.Sc., M.Sc.,  
Ph.D)

NIDN: 0725057201

NIDN: 0311106903

Pembimbing I

Pembimbing II

  
(Marlinda Vasty Overbeek, S.Kom,  
M.Kom)

  
(David Agustriawan, S.Kom.,  
M.Sc.,Ph.D)

NIDN: 0818038501

NIDN: 0525088601

Ketua Program Studi Informatika,

  
(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)

NIDN: 0315109103

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK  
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Renfred Leeman  
NIM : 00000056836  
Program Studi : Informatika  
Jenjang : S1  
Judul Karya Ilmiah : Prediksi Stadium Kanker Payudara  
Berbasis Data miRNA dan Gen dengan  
Seleksi Fitur dan Voting Classifier

Menyatakan dengan sesungguhnya bahwa saya bersedia:

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) \*\*.

Tangerang, 01 Juli 2025

Yang menyatakan



Renfred Leeman

\*\*Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

**HALAMAN PERSEMBAHAN / MOTTO**

”Überm Sternenzelt richtet Gott, wie wir gerichtet.”

Friedrich Schiller



## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga skripsi yang berjudul "Prediksi Stadium Kanker Payudara Berbasis Data miRNA dan Gen dengan Seleksi Fitur dan Voting Classifier" ini dapat diselesaikan dengan baik. Penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Marlinda Vasty Overbeek, S.Kom, M.Kom, sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Bapak David Agustriawan, S.Kom., M.Sc., Ph.D, sebagai Pembimbing kedua yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
6. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Harapannya, penelitian ini dapat memberikan kontribusi dalam pengembangan metode deteksi dini kanker payudara serta menjadi acuan bagi penelitian lanjutan di bidang bioinformatika dan ilmu kedokteran.

Tangerang, 01 Juli 2025



Renfred Leeman



# **PREDIKSI STADIUM KANKER PAYUDARA BERBASIS DATA MIRNA DAN GEN DENGAN SELEKSI FITUR DAN VOTING CLASSIFIER**

Renfred Leeman

## **ABSTRAK**

Kanker payudara merupakan salah satu penyebab utama kematian pada perempuan di seluruh dunia. Deteksi dini terhadap stadium kanker payudara sangat penting untuk meningkatkan peluang keberhasilan pengobatan. Penelitian ini bertujuan untuk membangun model klasifikasi stadium kanker payudara (*early stage* dan *advanced stage*) dengan memanfaatkan data ekspresi gen dan miRNA dari basis data *The Cancer Genome Atlas* (TCGA). Proses penelitian melibatkan tahapan praproses data, seleksi gen dan miRNA berdasarkan dan tanpa analisis *Differentially Expressed Genes* (DEG) menggunakan metode *limma*, dilanjutkan dengan seleksi fitur menggunakan kombinasi *Logistic Regression* (L1) dan *Recursive Feature Elimination* (RFE). Model klasifikasi dilatih menggunakan algoritma *Support Vector Machine* (SVM), *Random Forest* (RF), *Logistic Regression* (LR), dan *Voting Classifier*. Evaluasi dilakukan secara realistis menggunakan pembagian data *train* dan *test*. Hasil penelitian menunjukkan bahwa data ekspresi gen dengan model *Voting Classifier* mampu mencapai akurasi hingga 92,6% pada data pengujian. Selain itu, penelitian ini berhasil mengidentifikasi kombinasi biomarker potensial sebanyak 41 fitur, yang menunjukkan kontribusi kolektif yang lebih kuat dibandingkan penggunaan fitur tunggal. Temuan ini menegaskan bahwa seleksi fitur terstruktur dan pendekatan *multivariabel* dapat meningkatkan akurasi klasifikasi stadium kanker payudara serta mendukung identifikasi biomarker molekuler secara lebih efisien dan aplikatif.

**Kata kunci:** Biomarker, Differentially Expressed Genes, Gen, Kanker payudara, miRNA

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A



**BREAST CANCER STAGE PREDICTION BASED ON MIRNA AND GENE  
EXPRESSION DATA USING FEATURE SELECTION AND VOTING  
CLASSIFIER**

Renfred Leeman

**ABSTRACT**

*Breast cancer is one of the leading causes of mortality among women worldwide. Early detection of breast cancer stages is crucial to improve treatment success rates. This study aims to develop a classification model to distinguish breast cancer stages (early stage and advanced stage) using gene expression and miRNA data from The Cancer Genome Atlas (TCGA) database. The research process involved data preprocessing, selection of genes and miRNAs with and without Differentially Expressed Genes (DEG) analysis using the limma method, followed by feature selection combining Logistic Regression (L1) and Recursive Feature Elimination (RFE). Classification models were trained using Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and Voting Classifier algorithms. Evaluation was conducted realistically using train-test data splits. The results demonstrated that gene expression data classified with the Voting Classifier achieved an accuracy of up to 92.6% on the test set. Furthermore, this study identified a potential biomarker combination consisting of 41 features, which collectively provided stronger predictive power compared to single-feature approaches. These findings highlight that structured feature selection and multivariable approaches can improve the accuracy of breast cancer stage classification while supporting a more efficient and practical identification of molecular biomarkers.*

**Keywords:** *Biomarkers, Breast cancer, Differentially Expressed Genes, Gene, miRNA*

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

## DAFTAR ISI

|                                                                        |      |
|------------------------------------------------------------------------|------|
| HALAMAN JUDUL . . . . .                                                | i    |
| PERNYATAAN TIDAK MELAKUKAN PLAGIAT . . . . .                           | ii   |
| HALAMAN PENGESAHAN . . . . .                                           | iii  |
| HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH . . . . .                   | iv   |
| HALAMAN PERSEMBAHAN/MOTO . . . . .                                     | v    |
| KATA PENGANTAR . . . . .                                               | vi   |
| ABSTRAK . . . . .                                                      | vii  |
| ABSTRACT . . . . .                                                     | viii |
| DAFTAR ISI . . . . .                                                   | ix   |
| DAFTAR TABEL . . . . .                                                 | xi   |
| DAFTAR GAMBAR . . . . .                                                | xii  |
| DAFTAR RUMUS . . . . .                                                 | xiv  |
| DAFTAR LAMPIRAN . . . . .                                              | xv   |
| BAB 1 PENDAHULUAN . . . . .                                            | 1    |
| 1.1 Latar Belakang Masalah . . . . .                                   | 1    |
| 1.2 Rumusan Masalah . . . . .                                          | 4    |
| 1.3 Batasan Permasalahan . . . . .                                     | 4    |
| 1.4 Tujuan Penelitian . . . . .                                        | 5    |
| 1.5 Manfaat Penelitian . . . . .                                       | 6    |
| 1.6 Sistematika Penulisan . . . . .                                    | 6    |
| BAB 2 LANDASAN TEORI . . . . .                                         | 8    |
| 2.1 Tinjauan Teori . . . . .                                           | 8    |
| 2.1.1 Breast Cancer . . . . .                                          | 8    |
| 2.1.2 Breast Cancer Staging . . . . .                                  | 8    |
| 2.2 Dataset . . . . .                                                  | 10   |
| 2.2.1 Gene Expression RNAseq - STAR - FPKM-UQ . . . . .                | 10   |
| 2.2.2 Stem Loop Expression - miRNA Expression Quantification . . . . . | 11   |
| 2.3 Differential Expression Gene (LIMMA) . . . . .                     | 12   |
| 2.4 Seleksi Fitur . . . . .                                            | 13   |
| 2.4.1 Regularisasi L1 (Lasso) dengan Logistic Regression . . . . .     | 13   |
| 2.4.2 Recursive Feature Elimination (RFE) . . . . .                    | 15   |
| 2.5 Algoritma Klasifikasi . . . . .                                    | 17   |
| 2.5.1 Support Vector Machine (SVM) . . . . .                           | 18   |
| 2.5.2 Random Forest (RF) . . . . .                                     | 25   |
| 2.5.3 Voting Classifier . . . . .                                      | 28   |
| 2.6 Evaluasi . . . . .                                                 | 30   |
| 2.6.1 Confusion Matrix . . . . .                                       | 30   |
| 2.6.2 Accuracy . . . . .                                               | 30   |
| 2.6.3 Precision . . . . .                                              | 31   |
| 2.6.4 Recall . . . . .                                                 | 31   |
| 2.6.5 F1-Score . . . . .                                               | 32   |
| BAB 3 METODOLOGI PENELITIAN . . . . .                                  | 33   |
| 3.1 Gambaran Umum Penelitian . . . . .                                 | 33   |
| 3.2 Spesifikasi Perangkat . . . . .                                    | 34   |
| 3.3 Studi Literatur . . . . .                                          | 35   |
| 3.4 Pengumpulan Data . . . . .                                         | 35   |
| 3.5 Preprocessing Data . . . . .                                       | 36   |
| 3.6 Feature Selection dan Differentially Expressed Genes . . . . .     | 38   |

|                |                                                                                 |    |
|----------------|---------------------------------------------------------------------------------|----|
| 3.7            | Pembangunan Model . . . . .                                                     | 39 |
| 3.8            | Evaluasi . . . . .                                                              | 39 |
| 3.9            | Analisis Biomarker . . . . .                                                    | 39 |
| BAB 4          | HASIL DAN DISKUSI . . . . .                                                     | 41 |
| 4.1            | Preprocessing . . . . .                                                         | 41 |
| 4.1.1          | Penggabungan Dataset . . . . .                                                  | 41 |
| 4.1.2          | Menangani Missing Value . . . . .                                               | 42 |
| 4.1.3          | Klasifikasi dan Filter Stage . . . . .                                          | 43 |
| 4.1.4          | Filter Race . . . . .                                                           | 44 |
| 4.2            | Tahapan Eksekusi DEG dan Feature Selection . . . . .                            | 45 |
| 4.2.1          | Seleksi Fitur Tanpa DEG . . . . .                                               | 45 |
| 4.2.2          | Differentially Expressed Genes (DEG) dengan Feature<br>Selection (FS) . . . . . | 48 |
| 4.3            | Eksperimen Pembangunan Model . . . . .                                          | 51 |
| 4.3.1          | Logistic Regression . . . . .                                                   | 52 |
| 4.3.2          | Support Vector Machine (SVM) . . . . .                                          | 53 |
| 4.3.3          | Random Forest (RF) . . . . .                                                    | 55 |
| 4.3.4          | Voting Classifier . . . . .                                                     | 56 |
| 4.4            | Hasil dan Evaluasi Skenario . . . . .                                           | 58 |
| 4.4.1          | Skenario DEG dan Feature Selection . . . . .                                    | 58 |
| 4.4.2          | Skenario Feature Selection Only . . . . .                                       | 60 |
| 4.4.3          | Diskusi Penelitian . . . . .                                                    | 68 |
| BAB 5          | SIMPULAN DAN SARAN . . . . .                                                    | 80 |
| 5.1            | Simpulan . . . . .                                                              | 80 |
| 5.2            | Saran . . . . .                                                                 | 81 |
| DAFTAR PUSTAKA | . . . . .                                                                       | 82 |



## DAFTAR TABEL

|            |                                                                                                                                           |    |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tabel 2.1  | Klasifikasi Stadium Kanker Payudara berdasarkan Edisi ke-8 dari Sistem Staging <i>American Joint Committee on Cancer</i> (AJCC) . . . . . | 9  |
| Tabel 2.2  | Confusion Matrix . . . . .                                                                                                                | 30 |
| Tabel 3.1  | Daftar Skenario Penelitian <i>Feature Selection</i> dan <i>Differentially Expressed Genes</i> . . . . .                                   | 39 |
| Tabel 4.1  | Distribusi Stadium Kanker berdasarkan Data miRNA . . . . .                                                                                | 44 |
| Tabel 4.2  | Distribusi Stadium Kanker berdasarkan Data FPKM-UQ . . . . .                                                                              | 44 |
| Tabel 4.3  | Top 5 Akurasi Terbaik dari Semua Model dengan <i>Feature Selection</i> dan <i>Differential Expression Gene</i> . . . . .                  | 59 |
| Tabel 4.4  | Top 5 Akurasi Terbaik <i>Logistic Regression</i> dengan <i>Feature Selection Only</i> (Data Test) . . . . .                               | 61 |
| Tabel 4.5  | Top 5 Akurasi Terbaik SVM dengan <i>Feature Selection Only</i> (Data Test) . . . . .                                                      | 63 |
| Tabel 4.6  | Top 5 Akurasi Terbaik <i>Random Forest</i> dengan <i>Feature Selection Only</i> (Data Test) . . . . .                                     | 65 |
| Tabel 4.7  | Top 5 Akurasi Terbaik <i>Voting Classifier</i> dengan <i>Feature Selection Only</i> (Data Test) . . . . .                                 | 67 |
| Tabel 4.8  | Uji Coba Metode <i>Feature Selection</i> Alternatif pada Model Klasifikasi <i>Voting Classifier</i> . . . . .                             | 69 |
| Tabel 4.9  | Daftar <i>ENSEMBL</i> Gene ID, <i>GenBank/Scaffold</i> IDm, dan Referensi <i>Overlap</i> . . . . .                                        | 70 |
| Tabel 4.10 | Daftar <i>GenBank/Scaffold</i> ID dan Keterangan . . . . .                                                                                | 72 |



## DAFTAR GAMBAR

|             |                                                                                                                                                            |    |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Gambar 2.1  | Hyperplane SVM yang memaksimalkan margin pemisah antar kelas dalam ruang 2 dimensi . . . . .                                                               | 20 |
| Gambar 2.2  | Ilustrasi <i>kernel trick</i> yang memetakan data ke ruang fitur berdimensi lebih tinggi untuk memungkinkan pemisahan non-linear . . . . .                 | 23 |
| Gambar 2.3  | Mekanisme <i>Hard Voting</i> dan <i>Soft Voting</i> . . . . .                                                                                              | 30 |
| Gambar 3.1  | <i>Research Pipeline</i> untuk Klasifikasi Tahapan Kanker Payudara Berdasarkan Data Ekspresi Gen dan miRNA. . .                                            | 34 |
| Gambar 3.2  | <i>Gene Expression RNAseq STAR - FPKM-UQ Dataset</i> . . .                                                                                                 | 36 |
| Gambar 3.3  | <i>Stem Loop Expression - miRNA Expression Quantification Dataset</i> . . . . .                                                                            | 36 |
| Gambar 3.4  | <i>Phenotype Dataset</i> . . . . .                                                                                                                         | 36 |
| Gambar 4.1  | <i>Feature Importance 10 Feature</i> . . . . .                                                                                                             | 47 |
| Gambar 4.2  | <i>Feature Importance 41 Feature</i> . . . . .                                                                                                             | 48 |
| Gambar 4.3  | Plot <i>Logistic Regression - Accuracy</i> terhadap Jumlah Fitur Terseleksi ( <i>Data Test</i> ) . . . . .                                                 | 61 |
| Gambar 4.4  | Plot <i>Logistic Regression - F1-Score (Macro Avg)</i> terhadap Jumlah Fitur Terseleksi ( <i>Data Test</i> ) . . . . .                                     | 62 |
| Gambar 4.5  | Plot SVM - <i>Accuracy</i> terhadap Jumlah Fitur Terseleksi ( <i>Data Test</i> ) . . . . .                                                                 | 63 |
| Gambar 4.6  | Plot SVM - <i>F1-Score (Macro Avg)</i> terhadap Jumlah Fitur Terseleksi ( <i>Data Test</i> ) . . . . .                                                     | 64 |
| Gambar 4.7  | Plot <i>Random Forest - Accuracy</i> terhadap Jumlah Fitur Terseleksi ( <i>Data Test</i> ) . . . . .                                                       | 65 |
| Gambar 4.8  | Plot <i>Random Forest - F1-Score (Macro Avg)</i> terhadap Jumlah Fitur Terseleksi ( <i>Data Test</i> ) . . . . .                                           | 66 |
| Gambar 4.9  | Plot <i>Voting Classifier - Accuracy</i> terhadap Jumlah Fitur Terseleksi ( <i>Data Test</i> ) . . . . .                                                   | 67 |
| Gambar 4.10 | Plot <i>Voting Classifier - F1-Score (Macro Avg)</i> terhadap Jumlah Fitur Terseleksi ( <i>Data Test</i> ) . . . . .                                       | 68 |
| Gambar 4.11 | Kurva ROC untuk 31 Fitur Baru yang Belum Teridentifikasi dalam Penelitian Sebelumnya (Tidak Menunjukkan <i>Overlap</i> dengan Basis Data Publik) . . . . . | 77 |
| Gambar 4.12 | Kontribusi relatif 31 fitur baru dalam mendukung performa klasifikasi bersama 10 fitur validasi. . . . .                                                   | 79 |

## DAFTAR KODE

|           |                                                                                                                                         |    |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------|----|
| Kode 4.1  | <i>Load data</i> ekspresi <i>miRNA</i> , gen ( <i>FPKM-UQ</i> ), dan <i>phenotype</i> . . .                                             | 41 |
| Kode 4.2  | Mengambil kolom <i>race</i> dan <i>stage</i> dari dataset <i>phenotype</i> . . . . .                                                    | 42 |
| Kode 4.3  | Transpose dataset <i>miRNA</i> dan Gen ( <i>FPKM-UQ</i> ) . . . . .                                                                     | 42 |
| Kode 4.4  | Mengabungkan data dari <i>phenotype</i> ke dataset <i>miRNA</i> dan Gen ( <i>FPKM-UQ</i> ) . . . . .                                    | 42 |
| Kode 4.5  | Penghapusan data NaN atau '-' pada kolom <i>ajcc_pathologic_stage</i> . <i>diagnoses</i> dan <i>race</i> . <i>demographic</i> . . . . . | 43 |
| Kode 4.6  | Feature selection L1 Logistic Regression dan Recursive Feature Elimination (RFE) . . . . .                                              | 46 |
| Kode 4.7  | Membentuk desain model berdasarkan stadium kanker . . . . .                                                                             | 48 |
| Kode 4.8  | Proses identifikasi gen dengan ekspresi berbeda . . . . .                                                                               | 49 |
| Kode 4.9  | Menyimpan hasil DEG ke file . . . . .                                                                                                   | 49 |
| Kode 4.10 | Filterisasi fitur berdasarkan Differentially Expressed Genes (DEG) dengan $\text{adj.P.Val} < 0.5$ . . . . .                            | 50 |
| Kode 4.11 | Mengambil 10 Upregulated dan 10 Downregulated . . . . .                                                                                 | 50 |
| Kode 4.12 | Menggabungkan fitur overlap antar Differentially Expressed Genes (DEG) dengan Feature selection . . . . .                               | 51 |
| Kode 4.13 | Model Logistic Regression . . . . .                                                                                                     | 53 |
| Kode 4.14 | Model Support Vector Machine (SVM) . . . . .                                                                                            | 54 |
| Kode 4.15 | Model Random Forest (RF) . . . . .                                                                                                      | 55 |
| Kode 4.16 | Model Voting Classifier . . . . .                                                                                                       | 57 |





## DAFTAR RUMUS

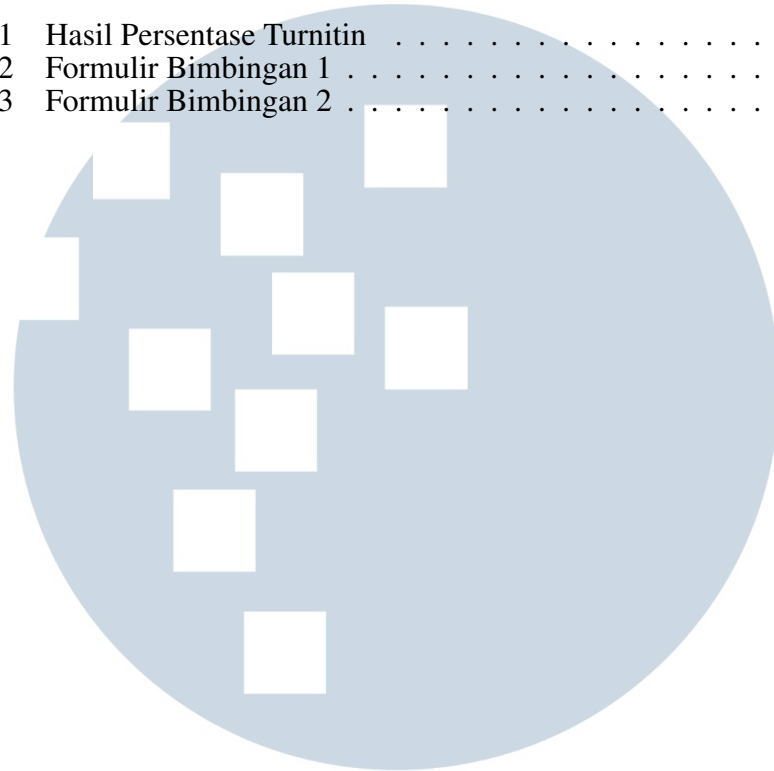
|            |                                                                                  |    |
|------------|----------------------------------------------------------------------------------|----|
| Rumus 2.1  | <i>FPKM-UQ</i>                                                                   | 10 |
| Rumus 2.2  | <i>Linear</i>                                                                    | 14 |
| Rumus 2.3  | <i>Logistic Regression</i>                                                       | 14 |
| Rumus 2.4  | <i>L1 Regularization (Lasso)</i>                                                 | 15 |
| Rumus 2.5  | <i>Support Vector Machine (SVM)</i>                                              | 16 |
| Rumus 2.6  | <i>Recursive Feature Elimination (RFE)</i>                                       | 17 |
| Rumus 2.7  | <i>Support Vector Machine (SVM) – Decision Boundary Condition</i>                | 18 |
| Rumus 2.8  | <i>Support Vector Machine (SVM) – Linear Separability Condition</i>              | 19 |
| Rumus 2.9  | <i>Support Vector Machine (SVM) – Hard Margin Constraint</i>                     | 19 |
| Rumus 2.10 | <i>Support Vector Machine (SVM) – Soft Margin Optimization Objective</i>         | 19 |
| Rumus 2.12 | <i>Support Vector Machine (SVM) – Margin of the Hyperplane</i>                   | 20 |
| Rumus 2.13 | <i>Support Vector Machine (SVM) – Fungsi Objektif untuk Memaksimalkan Margin</i> | 20 |
| Rumus 2.14 | <i>Support Vector Machine (SVM) – Dual Problem Optimization</i>                  | 21 |
| Rumus 2.17 | <i>Support Vector Machine (SVM) – Kernel Trick</i>                               | 22 |
| Rumus 2.18 | <i>Support Vector Machine (SVM) – Mercer's Condition</i>                         | 22 |
| Rumus 2.19 | <i>Support Vector Machine (SVM) – Kernel Linear</i>                              | 23 |
| Rumus 2.20 | <i>Support Vector Machine (SVM) – Kernel Polinomial</i>                          | 24 |
| Rumus 2.21 | <i>Support Vector Machine (SVM) – Radial Basis Function (RBF)</i>                | 24 |
| Rumus 2.22 | <i>Support Vector Machine (SVM) – Fungsi Keputusan dengan Kernel</i>             | 24 |
| Rumus 2.23 | <i>Voting Majority (Majority Voting)</i>                                         | 25 |
| Rumus 2.24 | <i>Gini Impurity</i>                                                             | 26 |
| Rumus 2.25 | <i>Entropy</i>                                                                   | 27 |
| Rumus 2.26 | <i>Information Gain (IG)</i>                                                     | 27 |
| Rumus 2.27 | <i>Voting Classifier</i>                                                         | 28 |
| Rumus 2.28 | <i>Soft Voting</i>                                                               | 29 |
| Rumus 2.29 | <i>Accuracy</i>                                                                  | 31 |
| Rumus 2.30 | <i>Precision</i>                                                                 | 31 |
| Rumus 2.31 | <i>Recall</i>                                                                    | 31 |
| Rumus 2.32 | <i>F-1 Score</i>                                                                 | 32 |

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



## DAFTAR LAMPIRAN

|            |                                     |     |
|------------|-------------------------------------|-----|
| Lampiran 1 | Hasil Persentase Turnitin . . . . . | 89  |
| Lampiran 2 | Formulir Bimbingan 1 . . . . .      | 100 |
| Lampiran 3 | Formulir Bimbingan 2 . . . . .      | 104 |



UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA