

**IMPLEMENTASI METODE TF-IDF DAN ALGORITMA
NAIVE BAYES RANDOM FOREST DALAM APLIKASI
DIABETIC BERBASIS ANDROID**

SKRIPSI



Oleh :

I WAYAN ALSTON ARGODI

19081010163

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2023**

**IMPLEMENTASI METODE TF-IDF DAN ALGORITMA
NAIVE BAYES RANDOM FOREST DALAM APLIKASI
DIABETIC BERBASIS ANDROID**

SKRIPSI

Diajukan Untuk Memenuhi Sebagai Persyaratan Dalam Menempuh Gelar Sarjana
Komputer Program Studi Informatika



Oleh :

I WAYAN ALSTON ARGODI

19081010163

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR**

2023

LEMBAR PENGESAHAN SKRIPSI

Judul : IMPLEMENTASI METODE TF-IDF DAN ALGORITMA
NAIVE BAYES RANDOM FOREST DALAM APLIKASI
DIABETIC BERBASIS ANDROID

Oleh : I WAYAN ALSTON ARGODI

NPM : 19081010163

Telah Diseminarkan Dalam Ujian Skripsi Pada:

Hari Selasa, Tanggal 11 Juli 2023

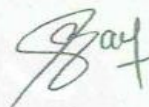
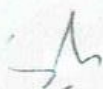
Mengetahui

Dosen Pembimbing

Dosen Penguji

1.

1.



Eva Yulia Puspaningrum, S.Kom., M.Kom

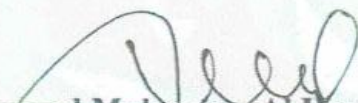
Made Hanindia Prami Swari, S.Kom,
M.Cs

NIP : 19890705 2021212 002

NIP : 19890205 2018032 001

2.

2.



Muhammad Muhafron Al Haromainy,
S.Kom., M.Kom.

Andreas Nugroho Sihananto, S.Kom,
M.Kom.

NIP : 19950601 202203 1 006

NPT : 211199 00 412271

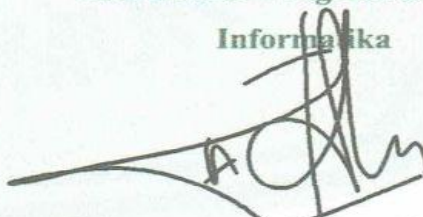
Menyetujui

Dekan Fakultas

Koordinator Program Studi

Ilmu Komputer

Informatika



Dr. Ir. Novirina Hendrasarie, MT.

Fetty Tri Anggraeny, S.Kom, M.Kom

NIP : 19681126 199403 2 001

NIP : 19820211 2021212 005

SURAT PERNYATAAN ANTI PLAGIAT

Saya mahasiswa program studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan dibawah ini:

Nama : I Wayan Alston Argodi

NPM : 19081010163

Menyatakan bahwa judul Skripsi / Tugas Akhir / Penelitian yang saya ajukan dan kerjakan, yang berjudul:

“IMPLEMENTASI METODE TF-IDF DAN ALGORITMA NAIVE BAYES RANDOM FOREST DALAM APLIKASI DIABETIC BERBASIS ANDROID”

Bukan merupakan plagiat dari Skripsi / Tugas Akhir / Penelitian orang lain juga bukan merupakan produk dan atau software yang beli dari pihak lain. Saya juga menyatakan bahwa Skripsi / Tugas Akhir ini adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun di institusi Pendidikan lain,

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka saya siap menerima sebagai konsekuensinya.

Surabaya, 17 Juli 2023

Penulis,



I Wayan Alston Argodi

NPM. 190181010163

IMPLEMENTASI METODE TF-IDF DAN ALGORITMA NAIVE BAYES RANDOM FOREST DALAM APLIKASI DIABETIC BERBASIS ANDROID

Nama Mahasiswa : I Wayan Alston Argodi
NPM : 19081010163
Program Studi : Informatika
Dosen Pembimbing : Eva Yulia Puspaningrum, S.Kom., M.Kom.
Muhammad Muharrom Al Haromainy, S.Kom.,
M.Kom

ABSTRAK

Diabetes merupakan penyakit serius yang terjadi ketika pankreas tidak menghasilkan cukup Insulin untuk mengatur gula darah, teknologi dapat berperan penting dalam penanganan memabntu mendiagnosa penyakit dan membantu manajemen perawatan penyakit diabetes seperti aplikasi berbasis selular. Metode *Term Frequency Inverse Document Frequency* digunakan dengan menghitung setiap kehadiran kata pada kumpulan dokumen. Algoritma Naive Bayes menggunakan probabilitas dalam menyelesaikan sebuah kasus klasifikasi perhitungan yang efisien dan cepat. Algoritma Random Forest dapat menyelesaikan masalah klasifikasi dengan membangun pohon keputusan mengambil suara mayoritas untuk klasifikasi. Berdasarkan penelitian ini diketahui Algoritma Naive Bayes dalam aplikasi diabetic berbasis android ini menghasilkan akurasi sebesar 66% dengan menempuh waktu komputasi selama 39 detik dengan konsumsi memori sebesar 80 sampai dengan 351 MB, algortima Random Forest memiliki keunggulan terhadap akurasi yang dihasilkan sebesar 88% dalam percobaan dengan menggunakan 691 data latih dan 71 data uji, waktu komputasi yang ditempuh selama 14 menit 29 detik dengan penggunaan memori rentang 136 sampai dengan 292 MB penggunaan CPU 23 sampai dengan 95%.

Kata Kunci: Diabetes, *Term Frequency Inverse Document Frequency*, Algoritma Random Forest, Algoritma Naive Bayes, Aplikasi Android.

KATA PENGANTAR

Puji syukur kita panjatkan kepada Tuhan Yang Maha Esa dalam kesempatan yang berbahagia ini, saya ingin mempersembahkan skripsi ini yang berjudul Implementasi Metode TF-IDF dan algoritma Naive Bayes dan algoritma Random Forest dalam aplikasi Diabetic berbasis Android. Skripsi ini merupakan hasil dari perjalanan penelitian dan eksplorasi yang telah saya lakukan selama beberapa waktu.

Banyak dukungan serta bantuan yang didapatkan selama melakukan penelitian hingga menyelesaikan laporan ini. Dengan rasa hormat serta banyak terima kasih penulis ucapkan kepada seluruh pihak yang telah membantu proses penyelesaian penelitian dari awal hingga akhir.

Penulis menyadari masih terdapat banyaknya kekurangan pada penelitian ini dikarenakan keterbatasan pengetahuan dan pemikiran. Oleh karena itu, penulis mengharapkan dan menerima kritik dan saran yang membangun dari semua pihak dalam penyempurnaan laporan ini. Semoga skripsi ini barokah, bermanfaat, dan berguna serta mampu memenuhi harapan berbagai pihak

Surabaya, 27 Juni 2023

Penulis,

I Wayan Alston Argodi

UCAPAN TERIMA KASIH

Ucapan terima kasih yang sebesar-besarnya diucapkan kepada seluruh pihak terkait yang telah membantu atas selesainya laporan skripsi ini. Secara khusus penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada semua pihak yang terlibat yang telah membantu. Pada kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Sang Hyang Widhi Wasa yang telah memberikan petunjuk dan kesehatan serta atas izin-Nya penulis mampu mengerjakan skripsi ini hingga selesai.
2. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Dr. Novirina Hendrasarie, S.T., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom, selaku koordinator program studi jurusan Informatika Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
5. Ibu Eva Yulia Puspaningrum, S.Kom., M.Kom, selaku Pembimbing yang telah bersedia membimbing dan mengarahkan penulis selama menyusun skripsi dan memberikan banyak ilmu serta solusi pada setiap permasalahan atas kesulitan dalam penulisan skripsi ini
6. Bapak Muhammad Muharrom Al Haromainy, S.Kom., M.Kom. selaku Dosen Pembimbing II yang telah bersedia meluangkan waktu, tenaga dan pikiran untuk memberikan penulis masukan pada proses bimbingan selama penyusunan skripsi ini
7. Seluruh Bapak dan Ibu dosen program studi Informatika UPN “Veteran” Jawa Timur yang telah mendidik dan memberikan ilmunya sehingga Penulis memiliki bekal untuk dapat melakukan penelitian ini.
8. Ibu Dian, selaku rekam medis Rumah Sakit Islam Surabaya yang telah membantu memberikan informasi seputar Diabetes.
9. Ibu Choirun Nisa, S.Kom, selaku Komite Koordinasi Pendidikan Rumah Sakit Islam Surabaya yang telah membantu dalam akses informasi terkait penyakit Diabetes.

10. Kedua Orang Tua, Adik, dan Kerabat yang telah memberikan doa, kasih sayang, serta semangat pada saat melaksanakan penelitian dan juga dalam pengerjaan laporan.
11. Teman-teman Angkatan 2019 dan anggota Flouws khususnya yang senantiasa menemani dan memberi motivasi dan bantuan ketika penulis sedang kesusahan dalam mengerjakan penelitian ini.
12. Teman-teman Bangkit Academy dan aviation channel yang senantiasa menemani dan memberi motivasi dan bantuan ketika penulis sedang kesusahan dalam mengerjakan penelitian ini.

DAFTAR ISI

| | |
|---|-------|
| LEMBAR PENGESAHAN..... | i |
| SURAT PERNYATAAN ANTI PLAGIAT..... | ii |
| ABSTRAK..... | iii |
| UCAPAN TERIMA KASIH..... | iv |
| DAFTAR ISI..... | v |
| DAFTAR TABEL..... | vii |
| DAFTAR GAMBAR..... | xviii |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Rumusan masalah..... | 3 |
| 1.3 Batasan Masalah..... | 4 |
| 1.4 Tujuan Penelitian..... | 4 |
| 1.5 Manfaat Penelitian..... | 5 |
| BAB II STUDI LITERATUR..... | 6 |
| 2.1 Penelitian sebelumnya..... | 6 |
| 2.2 Diabetes..... | 9 |
| 2.3 Machine Learning..... | 10 |
| 2.3.1 Supervised Learning..... | 12 |
| 2.4 Text Mining..... | 13 |
| 2.5 Metode Term Frequency Inverse Document Frequency..... | 14 |
| 2.5.1 Term Weighting..... | 14 |
| 2.5.2 Term Frequency..... | 15 |
| 2.5.3 Inverse Document Frequency..... | 15 |
| 2.5.4 Term Frequency Inverse Document Frequency..... | 16 |
| 2.5.5 Cosine similarty..... | 16 |
| 2.6 Algoritma Naive Bayes..... | 17 |
| 2.7 Algoritma Random Forest..... | 19 |

| | |
|---|-----------|
| 2.8 Android..... | 26 |
| 2.8.1 Android Studio..... | 27 |
| 2.9 Kotlin..... | 28 |
| 2.10 Jetpack..... | 28 |
| 2.10.1 Room database..... | 30 |
| 2.11 Pengujian..... | 30 |
| 2.11.1 Accuracy..... | 31 |
| 2.11.2 Precision..... | 31 |
| 2.11.3 Recall..... | 32 |
| 2.11.4 Specify..... | 32 |
| 2.11.5 F1 Score..... | 32 |
| 2.11.6 Pengujian metode dan algoritma..... | 33 |
| 2.11.7 Pengujian aplikasi..... | 34 |
| BAB III METODOLOGI PENELITIAN..... | 37 |
| 3.1 Tahapan penelitian..... | 37 |
| 3.2 Studi Literatur..... | 37 |
| 3.3 Akusisi dataset..... | 38 |
| 3.3.1 Preprocessing dataset..... | 41 |
| 3.4 Implementasi..... | 42 |
| 3.4.1 Preprocessing Kalimat..... | 44 |
| 3.4.2 Fungsi Term Frequency Inverse Document Frequency..... | 46 |
| 3.4.2.1 Term Frequency Inverse Document Frequency Similarity..... | 48 |
| 3.4.2.2 Cosine Similarity..... | 53 |
| 3.4.3 Fungsi algoritma Naive Bayes..... | 55 |
| 3.4.4 Fungsi algoritma Random Forest..... | 70 |
| 3.4.4.1 Akar 1..... | 72 |
| 3.4.4.2 Akar 2..... | 89 |
| 3.4.4.3 Akar 3..... | 94 |
| 3.4.5 Implementasi aplikasi Android..... | 101 |

| | |
|--|------------|
| 3.4.5.1 Implementasi layout aplikasi..... | 103 |
| 3.4.5.2 Implementasi modul utama..... | 104 |
| 3.5 Pengujian..... | 105 |
| 3.5.1 Pengujian metode dan algoritma..... | 105 |
| 3.5.2 Pengujian aplikasi..... | 107 |
| BAB IV HASIL DAN PEMBAHASAN..... | 110 |
| 4.1 Pre-processing..... | 111 |
| 4.1.1 Pre-processing diabetes dataset..... | 111 |
| 4.1.1.1 Explore Dataset..... | 111 |
| 4.1.1.2 Dropping duplicate..... | 113 |
| 4.1.1.3 Dropping Missing..... | 114 |
| 4.1.1.4 Outliers Detection..... | 115 |
| 4.1.1.5 Menyimpan sebagai CSV..... | 116 |
| 4.1.2 Preprocessing dataset kalimat..... | 119 |
| 4.1.2.1 Tokenise..... | 120 |
| 4.1.2.2 Capital removal..... | 121 |
| 4.1.2.2 Punctuation Removal..... | 122 |
| 4.2 Implementasi Term Frequency Inverse Document Frequency..... | 124 |
| 4.2.1 Query database..... | 124 |
| 4.2.2 Term Frequency Inverse Document Frequency..... | 126 |
| 4.3 Implementasi algoritma Naive Bayes..... | 131 |
| 4.4 Implementasi algoritma Random Forest..... | 134 |
| 4.5 Pengujian..... | 143 |
| 4.5.1 Pengujian metode Term Frequency Inverse Document Frequency dan algoritma Random Forest..... | 143 |
| 4.5.1.1 Skenario Pengujian Random Forest 90:10..... | 143 |
| 4.5.1.2 Skenario Pengujian Random Forest 80:20..... | 147 |
| 4.5.1.3 Skenario Pengujian Random Forest 70:30..... | 151 |
| 4.5.1.4 Skenario Pengujian Random Forest 60:40..... | 155 |

| | |
|--|-----|
| 4.5.1.5 Skenario Pengujian Random Forest 50:50..... | 159 |
| 4.5.1.6 Skenario Pengujian Random Forest 40:60..... | 162 |
| 4.5.2 Pengujian metode Term Frequency Inverse Document Frequency dan algoritma Naive Bayes..... | 165 |
| 4.5.2.1 Skenario Pengujian Naive Bayes 90:10..... | 166 |
| 4.5.2.2 Skenario Pengujian Naive Bayes 80:20..... | 169 |
| 4.5.2.3 Skenario Pengujian Naive Bayes 70:30..... | 171 |
| 4.5.2.4 Skenario Pengujian Naive Bayes 60:40..... | 173 |
| 4.5.2.5 Skenario Pengujian Naive Bayes 50:50..... | 175 |
| 4.5.2.6 Skenario Pengujian Naive Bayes 40:60..... | 177 |
| 4.5.3 Hasil Pengujian dengan data uji..... | 179 |
| 4.5.4 Pengujian skenario Chat..... | 182 |
| 4.5.4.1 Pengujian kalimat pengingat..... | 182 |
| 4.5.4.2 Skenario satu kalimat pengingat..... | 183 |
| 4.5.4.3 Skenario dua kalimat pengingat..... | 184 |
| 4.5.4.4 Skenario tiga kalimat pengingat..... | 185 |
| 4.5.4.5 Skenario empat kalimat pengingat..... | 187 |
| 4.5.4.6 Hasil skenario masukan kalimat pengingat..... | 188 |
| 4.5.4.7 Pengujian skenario pertanyaan Informasi..... | 189 |
| 4.5.4.8 Skenario satu pengujian pertanyaan informasi..... | 190 |
| 4.5.4.9 Skenario dua pengujian pertanyaan informasi..... | 191 |
| 4.5.4.10 Skenario tiga pengujian pertanyaan informasi..... | 192 |
| 4.5.4.11 Skenario empat pengujian kalimat pertanyaan informasi... | 193 |
| 4.5.4.12 Skenario lima pengujian kalimat pertanyaan informasi..... | 194 |
| 4.5.4.13 Skenario enam pengujian kalimat pertanyaan informasi.... | 195 |
| 4.5.4.14 Hasil skenario masukan kalimat Informasi..... | 196 |
| 4.5.4.14 Skenario Klasifikasi diabetes..... | 197 |
| 4.5.4.15 Skenario pertama klasifikasi diabetes..... | 198 |
| 4.5.4.16 Skenario kedua klasifikasi diabetes..... | 199 |

| | |
|---|------------|
| 4.5.4.17 Skenario ketiga klasifikasi diabetes..... | 200 |
| 4.5.4.18 Skenario empat klasifikasi diabetes..... | 202 |
| 4.5.4.19 Skenario lima klasifikasi diabetes..... | 203 |
| 4.5.4.20 Skenario enam klasifikasi diabetes..... | 204 |
| 4.5.4.21 Skenario tujuh klasifikasi diabetes..... | 206 |
| 4.5.4.22 Skenario delapan klasifikasi diabetes..... | 208 |
| 4.5.4.23 Hasil skenario kalimat klasifikasi diabetes..... | 210 |
| 4.5.5 Hasil Pengujian skenario Chat..... | 212 |
| BAB V PENUTUP..... | 215 |
| 5.1 Kesimpulan..... | 215 |
| 5.2 Saran..... | 216 |
| DAFTAR PUSTAKA..... | 217 |
| LAMPIRAN..... | 220 |
| BIODATA PENULIS..... | 224 |

DAFTAR TABEL

| | |
|--|----|
| Tabel 2.1. Penderita diabetes kawasan asia..... | 9 |
| Tabel 2.2 Skenario pengujian aplikasi..... | 35 |
| Tabel 3.1 Pima Diabetes dataset..... | 38 |
| Tabel 3.2 Atribut Pima Diabetes dataset..... | 39 |
| Tabel 3.3 Dataset kalimat..... | 39 |
| Tabel 3.4 Tokenisasi kalimat..... | 45 |
| Tabel 3.5 Capital removal kalimat..... | 45 |
| Tabel 3.6 Punctuation Removal kalimat..... | 46 |
| Tabel 3.7 Contoh kalimat input dengan kalimat lama..... | 47 |
| Tabel 3.8 Pembobotan Term Frequency..... | 49 |
| Tabel 3.9 Pembobotan Inverse Document Frequency..... | 50 |
| Tabel 3.10 Pembobotan term apa..... | 52 |
| Tabel 3.11 Pembobotan term itu..... | 52 |
| Tabel 3.12 Pembobotan term sakit..... | 52 |
| Tabel 3.13 Perhitungan cosine..... | 54 |
| Tabel 3.14 Data contoh perhitungan Naive Bayes..... | 56 |
| Tabel 3.15 Kondisi kalimat masukan baru..... | 57 |
| Tabel 3.16 Kondisi masukan baru..... | 57 |
| Tabel 3.17 Perhitungan fitur Pregnancies dengan outcome..... | 58 |
| Tabel 3.18 Perhitungan fitur Glucose dengan Outcome..... | 59 |
| Tabel 3.19 Perhitungan fitur Blood Pressure dan outcome..... | 61 |
| Tabel 3.20 Perhitungan fitur Skin thickness dan outcome..... | 62 |
| Tabel 3.21 Perhitungan fitur Insulin dan outcome..... | 63 |
| Tabel 3.22 perhitungan fitur BMI dan outcome..... | 64 |
| Tabel 3.23 Perhitungan fitur pedigree dan outcome..... | 65 |
| Tabel 3.24 Perhitungan fitur age dan outcome..... | 67 |
| Tabel 3.25 Hasil klasifikasi Naive Bayes..... | 69 |

| | |
|---|----|
| Tabel 3.26 Dataset bootstrap untuk algoritma Random Forest..... | 71 |
| Tabel 3.27 kontingensi pregnancies..... | 72 |
| Tabel 3.28 Kontingensi glucose..... | 73 |
| Tabel 3.29 Kontingensi blood pressure..... | 74 |
| Tabel 3.30 Kontingensi skin thickness..... | 74 |
| Tabel 3.31 Kontingensi insulin..... | 75 |
| Tabel 3.32 Kontingensi BMI..... | 75 |
| Tabel 3.33 kontingensi pedigree..... | 76 |
| Tabel 3.34 Kontingensi age..... | 76 |
| Tabel 3.35 Gini indeks pregnancies..... | 77 |
| Tabel 3.36 Gini indeks glucose..... | 78 |
| Tabel 3.37 Gini indeks blood pressures..... | 80 |
| Tabel 3.38 Gini indeks skin thickness..... | 81 |
| Tabel 3.39 Gini indeks insulin..... | 82 |
| Tabel 3.40 Gini indeks BMI..... | 83 |
| Tabel 3.41 Gini indeks pedigree..... | 85 |
| Tabel 3.42 Gini indeks age..... | 86 |
| Tabel 3.43 Tabel dataset bootstrap akar 2..... | 89 |
| Tabel 3.44 Kontingensi pregnancies akar 2..... | 90 |
| Tabel 3.45 Kontingensi glukosa akar 2..... | 90 |
| Tabel 3.46 Kontingensi blood preasure akar 2..... | 91 |
| Tabel 3.47 Kontingensi skin thickness akar 2..... | 92 |
| Tabel 3.48 Kontingensi BMI akar 2..... | 92 |
| Tabel 3.49 Kontingensi pedigree akar 2..... | 93 |
| Tabel 3.50 Kontingensi age akar 2..... | 93 |
| Tabel 3.51 Dataset bootstrap akar 3..... | 94 |
| Tabel 3.52 Kontingensi glukosa..... | 95 |
| Tabel 3.53 Kontingensi tekanan darah..... | 96 |
| Tabel 3.54 Kontingensi skin thickness..... | 96 |

| | |
|--|-----|
| Tabel 3.55 Kontingensi insulin..... | 97 |
| Tabel 3.56 Kontingensi BMI..... | 97 |
| Tabel 3.57 Kontingensi pedigree..... | 98 |
| Tabel 3.58 Kontingensi age..... | 98 |
| Tabel 3.59 skenario menggunakan data pengujian..... | 105 |
| Tabel 3.60 Skenario pengujian aplikasi..... | 109 |
| Tabel 4.1 Kode untuk mengambil dataset..... | 111 |
| Tabel 4.2 Kode untuk menampilkan bentuk data..... | 112 |
| Tabel 4.3 Kode untuk menampilkan informasi data..... | 112 |
| Tabel 4.4 Kode untuk menampilkan bentuk data..... | 113 |
| Tabel 4.5 Kode untuk mencari nilai kosong..... | 114 |
| Tabel 4.6 Kode untuk mencari nilai outliers..... | 115 |
| Tabel 4.7 Kode untuk menyimpan data sebagai csv..... | 116 |
| Tabel 4.8 Preprocessing kalimat masukan..... | 119 |
| Tabel 4.9 Kode untuk tokenisasi..... | 120 |
| Tabel 4.10 hasil proses tokenisasi..... | 120 |
| Tabel 4.11 Kode untuk capital removal..... | 121 |
| Tabel 4.12 hasil proses capital removal..... | 121 |
| Tabel 4.13 Kode untuk remove line..... | 122 |
| Tabel 4.14 Hasil proses Punctuation Removal..... | 123 |
| Tabel 4.15 Kode untuk Query database..... | 124 |
| Tabel 4.16 Kode untuk mengambil kesamaan dokumen..... | 125 |
| Tabel 4.17 Kode untuk query database..... | 126 |
| Tabel 4.18 Kode set terms..... | 127 |
| Tabel 4.19 Kode pencarian set terms..... | 127 |
| Tabel 4.20 Kode untuk fungsi Term Frequency Inverse Document Frequency.. | 127 |
| Tabel 4.21 Kode menghitung vector space model..... | 129 |
| Tabel 4.22 Contoh hasil perhitungan..... | 130 |
| Tabel 4.23 Kode inisialisasi pelatihan algoritma Naive Bayes..... | 131 |

| | |
|--|-----|
| Tabel 4.24 Kode train algoritma Naive Bayes..... | 132 |
| Tabel 4.25 Kode klasifikasi algoritma Naive Bayes..... | 132 |
| Tabel 4.26 Kode prediksi algoritma Naive Bayes..... | 133 |
| Tabel 4.27 Kode input algoritma Random Forest..... | 135 |
| Tabel 4.28 Kode algoritma Random Forest..... | 136 |
| Tabel 4.29 Kode memulai algoritma Random Forest..... | 137 |
| Tabel 4.30 Kode memulai pohon keputusan algoritma Random Forest..... | 138 |
| Tabel 4.31 Kode pohon keputusan algoritma Random Forest..... | 139 |
| Tabel 4.32 Kode bootstrap algoritma Random Forest..... | 140 |
| Tabel 4.33 Kode pohon dalam algoritma Random Forest..... | 141 |
| Tabel 4.34 Kode evaluasi dalam algoritma Random Forest..... | 141 |
| Tabel 4.35 Deskripsi data Random Forest skenario 1..... | 144 |
| Tabel 4.36 Hasil confusion matrix Random Forest skenario 1..... | 144 |
| Tabel 4.37 Confusion matrix Random Forest skenario 1..... | 145 |
| Tabel 4.38 Penggunaan sumber daya Random Forest skenario 1..... | 147 |
| Tabel 4.39 Deskripsi data Random Forest skenario 2..... | 148 |
| Tabel 4.40 Hasil confusion matrix Random Forest skenario 2..... | 148 |
| Tabel 4.41 Confusion matrix Random Forest skenario 2..... | 149 |
| Tabel 4.42 penggunaan sumber daya Random Forest skenario 2..... | 151 |
| Tabel 4.43 Deskripsi data Random Forest skenario 3..... | 152 |
| Tabel 4.44 Hasil confusion matrix Random Forest skenario 3..... | 152 |
| Tabel 4.45 Confusion matrix Random Forest skenario 3..... | 153 |
| Tabel 4.46 Penggunaan sumber daya Random Forest skenario 3..... | 154 |
| Tabel 4.47 Deskripsi data Random Forest skenario 4..... | 155 |
| Tabel 4.48 Pengukuran confusion matrix Random Forest skenario 4..... | 156 |
| Tabel 4.49 Confusion matrix Random Forest skenario 4..... | 157 |
| Tabel 4.50 penggunaan sumber daya pada skenario 4..... | 158 |
| Tabel 4.51 deksripsi dataRandom Forest skenario 5..... | 159 |
| Tabel 4.52 Pengukuran confusion matrix Random Forest skenario 5..... | 159 |

| | |
|--|-----|
| Tabel 4.53 Confusion matrix Random Forest skenario 5..... | 160 |
| Tabel 4.54 Penggunaan sumber daya Random Forest skenario 5..... | 162 |
| Tabel 4.55 Deskripsi data Random Forest skenario 6..... | 163 |
| Tabel 4.56 Pengukuran confusion matrix Random Forest skenario 6..... | 163 |
| Tabel 4.57 confusion matrix skenario 6..... | 164 |
| Tabel 4.58 Penggunaan sumber daya Random Forest skenario 6..... | 165 |
| Tabel 4.59 Deskripsi data Naive Bayes skenario 1..... | 166 |
| Tabel 4.60 Pengukuran confusion matrix Naive Bayes skenario 1..... | 167 |
| Tabel 4.61 Confusion matrix Naive Bayes skenario 1..... | 167 |
| Tabel 4.62 penggunaan sumber daya skenario 1..... | 168 |
| Tabel 4.63 deskripsi data skenario 2..... | 169 |
| Tabel 4.64 Pengukuran confusion matrix Naive Bayes skenario 1..... | 169 |
| Tabel 4.65 Confusion matrix skenario 1..... | 169 |
| Tabel 4.66 Confusion matrix skenario 1..... | 170 |
| Tabel 4.67 Deskripsi data skenario 3..... | 171 |
| Tabel 4.68 Hasil pengukuran confusion matrix skenario 3..... | 171 |
| Tabel 4.69 Confusion matrix skenario 3..... | 171 |
| Tabel 4.70 Confusion matrix skenario 3..... | 172 |
| Tabel 4.71 Deskripsi data skenario 4..... | 173 |
| Tabel 4.72 Hasil pengukuran confusion matrix skenario 4..... | 173 |
| Tabel 4.73 Confusion matrix skenario 4..... | 174 |
| Tabel 4.74 Penggunaan sumber daya skenario 4..... | 174 |
| Tabel 4.75 Deskripsi data skenario 5..... | 175 |
| Tabel 4.76 Hasil pengukuran confusion matrix skenario 5..... | 175 |
| Tabel 4.77 Confusion matrix skenario 5..... | 176 |
| Tabel 4.78 Konsumsi daya skenario 5 naive bayes..... | 176 |
| Tabel 4.79 Deskripsi data Naive Bayes skenario 6..... | 177 |
| Tabel 4.80 Pengukuran confusion matrix Naive Bayes skenario 6..... | 177 |
| Tabel 4.81 Confusion matrix Naive Bayes skenario 6..... | 178 |

| | |
|--|-----|
| Tabel 4.82 Penggunaan sumber daya skenario 6..... | 178 |
| Tabel 4.83 hasil pengujian algoritma..... | 179 |
| Tabel 4.84 Deskripsi skenario pengingat..... | 183 |
| Tabel 4.85 Penggunaan sumber daya skenario pengingat..... | 188 |
| Tabel 4.86 Hasil skenario pengingat..... | 188 |
| Tabel 4.87 Deskripsi skenario informasi..... | 189 |
| Tabel 4.88 penggunaan sumber daya skenario informasi..... | 196 |
| Tabel 4.89 Hasil skenario informasi..... | 196 |
| Tabel 4.90 Deskripsi skenario klasifikasi..... | 197 |
| Tabel 4.91 Perhitungan Naive Bayes skenario lima..... | 204 |
| Tabel 4.92 Perhitungan Naive Bayes skenario enam..... | 206 |
| Tabel 4.93 Percobaan skenario tujuh klasifikasi diabetes..... | 207 |
| Tabel 4.94 Perhitungan Naive Bayes skenario delapan..... | 209 |
| Tabel 4.95 Penggunaan sumber daya skenario klasifikasi diabetes..... | 210 |
| Tabel 4.96 Hasil skenario pengujian klasifikasi diabetes..... | 211 |
| Tabel 4.97 Hasil skenario pengujian skenario chatbot..... | 212 |

DAFTAR GAMBAR

| | |
|--|-----|
| Gambar 2.2. Alur konsep pendekatan Machine Learning..... | 11 |
| Gambar 2.3. Tipe machine learning..... | 11 |
| Gambar 2.4. Konsep Supervised Learning..... | 12 |
| Gambar 2.5. Flowchart algoritma Naive Bayes..... | 18 |
| Gambar 2.6. Diagram Decission Tree..... | 22 |
| Gambar 2.7. Diagram Random Forest dengan kumpulan pohon keputusan..... | 24 |
| Gambar 2.8. Alur algoritma Random Forest..... | 25 |
| Gambar 2.10. Profiler Android Studio..... | 27 |
| Gambar 2.11. Diagram struktur Kotlin..... | 28 |
| Gambar 2.12. Diagram Model-View-Viewmodel..... | 29 |
| Gambar 2.14 Confusion Matrix..... | 31 |
| Gambar 2.15 Diagram pengujian metode dan algoritma..... | 33 |
| Gambar 2.16 Diagram profiler aplikasi..... | 34 |
| Gambar 3.1 Alur penelitian..... | 37 |
| Gambar 3.2 Alur preprocessing dataset..... | 41 |
| Gambar 3.3 Diagram alur sistem aplikasi..... | 43 |
| Gambar 3.4 Diagram pre-processing kalimat..... | 44 |
| Gambar 3.5 Alur proses Term Frequency Inverse Document Frequency..... | 47 |
| Gambar 3.6 Alur Term Frequency Inverse Document Frequency..... | 48 |
| Gambar 3.7 Alur diagram algoritma Naive Bayes..... | 56 |
| Gambar 3.8 Alur diagram algoritma Random Forest..... | 71 |
| Gambar 3.9 Diagram algoritma Random Forest..... | 100 |
| Gambar 3.10 Diagram pelatihan Random Forest..... | 101 |
| Gambar 3.11 Diagram alur data aplikasi..... | 102 |
| Gambar 3.12 Diagram Implementasi modul..... | 102 |
| Gambar 3.13 layout halaman Fitur Chat..... | 103 |
| Gambar 3.14 Modul Chat..... | 104 |
| Gambar 3.15 Diagram pengujian metode dan algoritma..... | 106 |

| | |
|---|-----|
| Gambar 3.16 Alur kerja aplikasi..... | 107 |
| Gambar 3.17 Alur kerja dalam arsitektur aplikasi..... | 108 |
| Gambar 3.18 Diagram profiler aplikasi..... | 108 |
| Gambar 4.1 Tampilan dataset yang telah diambil..... | 112 |
| Gambar 4.2 Informasi tipe dataset..... | 113 |
| Gambar 4.3 Hasil pencarian data duplikat..... | 114 |
| Gambar 4.4 Hasil pencarian data null..... | 115 |
| Gambar 4.5 Grafik persebaran data berat badan..... | 116 |
| Gambar 4.7 Bentuk data CSV dalam aplikasi..... | 117 |
| Gambar 4.8 Bentuk data kalimat awal dalam Excel..... | 118 |
| Gambar 4.9 Bentuk data kalimat awal dalam JSON..... | 118 |
| Gambar 4.10 Profiler Random Forest skenario 1..... | 146 |
| Gambar 4.11 Profile Random Forest skenario 2..... | 150 |
| Gambar 4.12 Profiler Random Forest skenario 3..... | 154 |
| Gambar 4.13 Profiler Random Forest skenario 4..... | 158 |
| Gambar 4.14 Profiler Random Forest skenario 5..... | 161 |
| Gambar 4.15 Profiler Random Forest skenario 6..... | 165 |
| Gambar 4.16 Profiler Naive Bayes skenario 1..... | 168 |
| Gambar 4.17 Profiler Naive Bayes skenario 2..... | 170 |
| Gambar 4.18 Profiler Naive Bayes skenario 3..... | 172 |
| Gambar 4.19 Profiler Naive Bayes skenario 4..... | 174 |
| Gambar 4.20 Profiler Naive Bayes skenario 5..... | 176 |
| Gambar 4.20 Profiler Naive Bayes skenario 6..... | 178 |
| Gambar 4.21 Percobaan memasukan skenario pengingat 1..... | 184 |
| Gambar 4.22 Percobaan memasukan skenario pengingat 2..... | 185 |
| Gambar 4.23 Percobaan memasukan skenario pengingat 3..... | 186 |
| Gambar 4.24 Percobaan memasukan skenario pengingat 4..... | 187 |
| Gambar 4.25 Percobaan skenario Informasi 1..... | 190 |
| Gambar 4.26 Percobaan memasukan skenario Informasi 2..... | 191 |

| | |
|--|-----|
| Gambar 4.27 Percobaan memasukan skenario Informasi 3..... | 192 |
| Gambar 4.28 Percobaan skenario informasi 4..... | 193 |
| Gambar 4.29 Percobaan skenario Informasi 5..... | 194 |
| Gambar 4.30 Percobaan skenario Informasi 6..... | 195 |
| Gambar 4.31 Percobaan skenario pertama klasifikasi diabetes..... | 199 |
| Gambar 4.32 Percobaan skenario dua klasifikasi diabetes..... | 200 |
| Gambar 4.33 Percobaan skenario ketiga klasifikasi diabetes..... | 201 |
| Gambar 4.34 Percobaan skenario empat klasifikasi diabetes..... | 202 |
| Gambar 4.35 Percobaan skenario lima klasifikasi diabetes..... | 204 |
| Gambar 4.36 Percobaan skenario enam klasifikasi diabetes..... | 205 |
| Gambar 4.37 Percobaan skenario tujuh klasifikasi diabetes..... | 207 |
| Gambar 4.38 Percobaan skenario delapan klasifikasi diabetes..... | 209 |