

**TUGAS AKHIR**

**SKEMA SKRIPSI**

**IMPLEMENTASI DATA MINING  
UNTUK PREDIKSI PENYAKIT DIABETES  
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)**



**KURNIA FEBRIANTO**

**NIM : 215610033**

**PROGRAM STUDI SISTEM INFORMASI  
PROGRAM SARJANA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA  
YOGYAKARTA**

**2025**

**TUGAS AKHIR**

**SKEMA SKRIPSI**

**IMPLEMENTASI DATA MINING  
UNTUK PREDIKSI PENYAKIT DIABETES  
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)**

**Diajukan sebagai salah satu syarat untuk menyelesaikan studi pada**



**Program Sarjana  
Program Studi Sistem Informasi  
Fakultas Teknologi Informasi  
Universitas Teknologi Digital Indonesia  
Yogyakarta**

**Disusun Oleh**

**KURNIA FEBRIANTO**

**215610033**

**PROGRAM STUDI SISTEM INFORMASI  
PROGRAM SARJANA  
FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA  
YOGYAKARTA**

**2025**

## HALAMAN PERSETUJUAN UJIAN TUGAS AKHIR

Judul : Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Menggunakan Algoritma K-Nearest Neighbor (K-NN)

Nama : Kurnia Febrianto

NIM : 215610033

Program Studi : Sistem Informasi

Program : Sarjana

Semester : Ganjil

Tahun Akademik : 2024/2025



Telah diperiksa dan disetujui untuk diujikan di hadapan Dewan Penguji Skripsi

Yogyakarta, 04 Februari 2025

Dosen Pembimbing

A handwritten signature in blue ink, appearing to read 'Deborah Kurniawati', is written over the printed name.

Deborah Kurniawati, S.Kom., M.Cs.

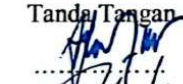

NIDN : 0511107301

**HALAMAN PENGESAHAN**  
**TUGAS AKHIR**  
**IMPLEMENTASI DATA MINING**  
**UNTUK PREDIKSI PENYAKIT DIABETES**  
**MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)**

Telah dipertahankan di depan Dewan Penguji Skripsi dan dinyatakan  
diterima sebagai salah satu persyaratan guna memperoleh

**Gelar Sarjana Komputer**  
**Program Studi Sistem Informasi**  
**Fakultas Teknologi Informasi**  
**Universitas Teknologi Digital Indonesia**  
**Yogyakarta**

Yogyakarta, 19 Februari 2025

| Dewan Penguji                         | NIDN       | Tanda Tangan  |
|---------------------------------------|------------|---|
| 1. Deborah Kurniawati, S.Kom., M.Cs.  | 0511107301 |  |
| 2. Y. Yohakim Marwanta, S.Kom., M.Cs. | 0026108101 |  |

Mengetahui,

Kepala Program Studi Sistem Informasi



Deborah Kurniawati, S.Kom., M.Cs.

NPP/NIP : 051149

### **PERNYATAAN KEASLIAN SKRIPSI**

Dengan ini saya menyatakan bahwa Laporan Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar Ahli Madya/kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 19 Februari 2025



Kurnia Febrianto

NIM : 215610033



## **HALAMAN PERSEMBAHAN**

Puji syukur Alhamdulillah saya panjatkan kepada Allah SWT, yang telah memberikan rahmat beserta karunia kepada saya sehingga dapat menyelesaikan skripsi ini yang saya persembahkan kepada:

1. Kedua orang tua saya, Bapak dan Ibu yang telah membesarkan saya dengan penuh cinta yang tidak mungkin dapat saya balas, dan selalu memberikan dorongan kepada saya dalam setiap langkah saya mencari ilmu sehingga saya dapat mengejar cita-cita saya kelak nanti.
2. Segenap keluarga besar saya yang telah memberikan bimbingan dan ilmunya agar saya dapat menjadi pribadi yang lebih baik dan mandiri serta dapat membanggakan mereka.
3. Salsabila Mufidati yang selalu ada disetiap situasi dan selalu mendukung saya disetiap waktu serta memberikan nasihat kepada saya.
4. Dan tidak lupa juga untuk semua pembaca skripsi ini, semoga dengan tulisan saya ini dapat membantu pembaca mendapatkan referensi yang bermanfaat.

**MOTTO**

مَنْ جَدَّ وَجَدَّ

**(Barang siapa yang bersungguh-sungguh, maka ia akan berhasil)**



## KATA PENGANTAR

Puji syukur saya panjatkan kepada Allah SWT, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan skripsi dengan mengangkat judul **“Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Menggunakan Algoritma K-Nearest Neighbor (K-NN)”**. Skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Program Studi Sistem Informasi Universitas Teknologi Digital Indonesia. Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi, sangatlah sulit bagi penulis untuk menyelesaikan skripsi ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kepada Allah SWT atas segala nikmat dan hidayah-nya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik.
2. Kedua orang tua saya, Bapak dan Ibu yang telah memberikan dorongan motivasi, pengorbanannya baik dari segi moril materi, serta doa yang selalu menyertai sehingga penulis dapat menyelesaikan skripsi ini.
3. Ibu Sri Redjeki, S.Si., M.Kom., Ph.D. selaku Rektor Universitas Teknologi Digital Indonesia.
4. Ibu Deborah Kurniawati, S.Kom., M.Cs. selaku Ketua Program Studi Sistem Informasi (S1) sekaligus menjadi dosen pembimbing peneliti ku yang selalu sabar dan tidak pernah bosan dalam membimbing peneliti dan memberikan saran, ilmu, motivasi, serta semangat nya setiap peneliti berada dalam kebingungan.
5. Salsabila Mufidati yang selalu ada disetiap situasi dan selalu mendukung saya disetiap saat serta memberikan nasihat kepada saya.
6. Teman teman dari Himpunan Mahasiswa Sistem Informasi yang senantiasa mendukung saya dalam keadaan apapun.

Penulis menyadari bahwa penelitian ini masih jauh dari kata sempurna dan masih banyak terdapat kekurangan serta kesalahan dalam penelitian ini. Penulis berharap

memberikan kritik dan saran dari semua pihak demi kesempurnaan Tugas Akhir peneliti ini.

Yang terakhir, penulis sangat berharap penelitian ini bisa berguna bagi semua pihak yang memiliki kepentingan, semoga penelitian ini bisa menjadi bermanfaat bagi yang membutuhkan dan menjadi ladang pahala bagi peneliti. Aamiin Ya Robbal Alamin.

## DAFTAR ISI

|  |       |
|--|-------|
| HALAMAN PERSETUJUAN.....                         | iii   |
| HALAMAN PENGESAHAN.....                          | iv    |
| PERNYATAAN KEASLIAN SKRIPSI.....                 | v     |
| HALAMAN PERSEMBAHAN .....                        | vii   |
| MOTTO .....                                      | viii  |
| KATA PENGANTAR .....                             | ix    |
| DAFTAR ISI .....                                 | x     |
| DAFTAR GAMBAR .....                              | xiii  |
| DAFTAR TABEL.....                                | xiv   |
| DAFTAR GRAFIK.....                               | xv    |
| DAFTAR LAMPIRAN.....                             | xv    |
| INTISARI.....                                    | xvii  |
| <i>ABSTRACT</i> .....                            | xviii |
| BAB 1 PENDAHULUAN .....                          | 1     |
| 1.1 Latar Belakang Masalah .....                 | 1     |
| 1.2 Rumusan Masalah.....                         | 3     |
| 1.3 Ruang Lingkup .....                          | 3     |
| 1.4 Tujuan .....                                 | 4     |
| 1.5 Manfaat .....                                | 4     |
| 1.6 Batasan Masalah .....                        | 5     |
| 1.7 Sistematika Penulisan .....                  | 6     |
| BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI ..... | 7     |
| 2.1 Tinjauan Pustaka .....                       | 7     |
| 2.2 Landasan Teori.....                          | 10    |
| 2.2.1 Penyakit Diabetes.....                     | 10    |
| 2.2.2 Data Mining .....                          | 16    |
| 2.2.3 Klasifikasi .....                          | 18    |
| 2.2.4 Algoritma K-Nearest Neighbor.....          | 18    |
| 2.2.5 Split Data.....                            | 20    |

|  |    |
|--|----|
| 2.2.6 Confusion Matrix .....               | 20 |
| 2.2.7 F1 Score .....                       | 22 |
| BAB III METODE PENELITIAN.....             | 23 |
| 3.1 Bahan Atau Data .....                  | 23 |
| 3.2 Peralatan.....                         | 24 |
| 3.2.1 Kebutuhan Perangkat Keras .....      | 24 |
| 3.2.2 Kebutuhan Perangkat Lunak .....      | 25 |
| 3.3 Prosedur Dan Pengumpulan Data .....    | 25 |
| 3.3.1 Teknik Pengumpulan Data .....        | 25 |
| 3.3.2 Alur Penelitian .....                | 27 |
| BAB IV IMPLEMENTASI DAN PEMBAHASAN .....   | 35 |
| 4.1 Analisis Deskripsi Data .....          | 36 |
| 4.2 Implementasi dan Uji Coba Sistem ..... | 49 |
| 4.2.1 Pre-Processing Data .....            | 49 |
| 4.2.2 Pembuatan Model .....                | 50 |
| 4.2.3 Interpretasi .....                   | 60 |
| 4.2.4 Penyajian Model .....                | 67 |
| BAB V PENUTUP.....                         | 70 |
| 5.1 Kesimpulan .....                       | 70 |
| 5.2 Saran .....                            | 70 |
| DAFTAR PUSTAKA .....                       | 71 |
| LAMPIRAN.....                              | 71 |

## DAFTAR GAMBAR

|   |     |
|---|-----|
| Gambar 3.1 Tahap Pengumpulan Data .....                                 | 26  |
| Gambar 3.2 Kerangka Penelitian.....                                     | 27  |
| Gambar 3.1 Tahap Proses Tranformasi.....                                | 29  |
| Gambar 3.5 Tahap Alur Pre-Pocessing.....                                | 35  |
| Gambar 3.4 Tahap Proses Tranformasi.....                                | 299 |
| Gambar 3.5 Tahap Proses Data Mining .....                               | 31  |
| Gambar 3.6 Tahap Proses Interpretasi .....                              | 33  |
| Gambar 4.1 Process Missing Value .....                                  | 45  |
| Gambar 4.2 Hasil Missing Value .....                                    | 46  |
| Gambar 4.3 Tahap Awal Proses Transformasi .....                         | 46  |
| Gambar 4.4 Pelabelan data.....  | 48  |
| Gambar 4.5 Perbandingan dengan skala 70:30 .....                        | 49  |
| Gambar 4.6 Perbandingan dengan skala 80:20 .....                        | 49  |
| Gambar 4.7 Proses Confusion Matrix.....                                 | 52  |
| Gambar 4.8 Train and test Model (K=5, Uji A).....                       | 53  |
| Gambar 4.1 Visualisasi Penyesuaian Formula Excel.....                   | 63  |
| Gambar 4.2 Visualisasi Akhir Model K-NN.....                            | 65  |
| Gambar 4.3 Visualisasi Proses Penyesuaian Model dengan rasio 80:20..... | 66  |
| Gambar 4.4 Visualisasi Proses Penyesuaian Model .....                   | 66  |
| Gambar 4.5 Confusion Matrix k=5 rasio 80:20.....                        | 67  |
| Gambar 4.6 Visualisasi Proses Penyesuaian Model.....                    | 68  |

## DAFTAR TABEL

|  |    |
|--|----|
| Tabel 2.1 Tinjauan Pustaka .....                                       | 7  |
| Tabel 2.1 Tinjauan Pustaka (Lanjutan) .....                            | 8  |
| Tabel 2.1 Tinjauan Pustaka (Lanjutan) .....                            | 8  |
| Tabel 2.1 Tinjauan Pustaka (Lanjutan) .....                            | 9  |
| Tabel 2.2 Keterangan Posisi Tabel <i>Confusion Matrix</i> .....        | 21 |
| Tabel 3.1 Rincian Variable Independen.....                             | 23 |
| Tabel 3.1 Rincian Variable Independen (Lanjutan).....                  | 35 |
| Tabel 4.1 Deskriptif Data .....  | 35 |
| Tabel 4.1 Deskriptif Data (Lanjutan).....                              | 36 |
| Tabel 4.2 Pelabelan Data.....  | 41 |
| Tabel 4.3 Akumulasi Hasil <i>Train And Test</i> Model Pada Uji A ..... | 53 |
| Tabel 4.4 Akumulasi Hasil <i>Train And Test</i> Model Pada Uji B.....  | 55 |
| Tabel 4.5 Hasil Evaluasi Model .....                                   | 64 |

## DAFTAR GRAFIK

|   |      |
|---|------|
| Grafik 4.1 Akumulasi Hasil <i>Class Recall</i> 1 .....    | 6457 |
| Grafik 4.2 Akumulasi Hasil <i>Class Recall</i> 0 .....    | 648  |
| Grafik 4.5 Akumulasi Hasil <i>Class Precision</i> 1 ..... | 649  |
| Grafik 4.5 Akumulasi Hasil <i>Class Precision</i> 0 ..... | 6460 |
| Grafik 4.5 Akumulasi Hasil <i>Accuracy</i> .....          | 6461 |

## DAFTAR LAMPIRAN

Lampiran 1. Dokumentasi Hasil Proses *Train And Test* Rasio 70:30

Lampiran 2. Dokumentasi Hasil Proses *Train And Test* Rasio 80:20

Lampiran 3. Ketentuan Kelulusan

Lampiran 4. Catatan Pendaran

Lampiran 5. Keputusan Hasil Uji

Lampiran 6. Biodata Penulis

Lampiran 7. Persetujuan Publikasi



## INTISARI

Gaya hidup manusia yang saat ini sudah modern membuat manusia serba instan dan gaya hidup yang kurang aktif menyebabkan berbagai macam penyakit salah satunya Diabetes Mellitus Tipe 2. Peningkatan kasus Diabetes Mellitus Tipe 2, terdapat sekitar 422 juta penderita Diabetes Mellitus Tipe 2 di dunia dan 1,5 juta kematian dihubungkan dengan Diabetes Mellitus setiap tahun (WHO, 2021). Salah satu faktor peningkatan kematian adalah karena penundaan diagnosis dan juga pencatatan. Teknik klasifikasi data mining adalah salah satu metode pencatatan. Penelitian ini bertujuan mengimplementasikan data mining serta mengetahui tingkat keakurasian untuk memprediksi diagnosis Penyakit Diabetes Mellitus Tipe 2. Data yang digunakan diperoleh dari website Kaggle yang mencakup variabel usia, BMI, kadar glukosa darah puasa, kadar insulin, dan tekanan darah dengan keseluruhan data berjumlah 2000 data. Model algoritma yang digunakan pada penelitian ini adalah algoritma K-Nearest Neighbor (K-NN) dengan menggunakan 2 pengujian yaitu Uji A dengan ratio 70% data training dan 30% data testing dan Uji B dengan ratio 80% data training dan 20% data testing. Hasil penelitian menunjukkan bahwa hasil akurasi terbaik dihasilkan oleh uji yang menggunakan perbandingan *split validation* 80:20 dengan hasil akurasi sebesar 93,25% dan *F1-Score Macro* sebesar 92.75%

Kata Kunci: Diabetes Mellitus Tipe 2, K-Nearest Neighbors, Kaggle, Rapid Miner, *Split Validation*

## ***ABSTRACT***

The modern human lifestyle makes humans instant and a less active lifestyle causes various diseases, one of which is Type 2 Diabetes Mellitus. The increase in cases of Type 2 Diabetes Mellitus, there are around 422 million people with Type 2 Diabetes Mellitus in the world and 1.5 million deaths are associated with Diabetes Mellitus each year (WHO, 2021). One of the factors for the increase in mortality is due to delays in diagnosis and recording. Data mining classification techniques are one of the recording methods. This study aims to implement data mining and determine the level of accuracy in predicting the diagnosis of Type 2 Diabetes Mellitus. The data used were obtained from the Kaggle website which includes variables of age, BMI, fasting blood glucose levels, insulin levels, and blood pressure with a total data of 2000 data. The algorithm model used in this study is the K-Nearest Neighbor (K-NN) algorithm using 2 testing. The results of the study showed that the best accuracy results were produced by a test using a split validation ratio of 80:20 with an accuracy result of 93.25% and a Macro F1-Score of 92.75%

Keywords: Diabetes Mellitus Type 2, K-Nearest Neighbors, Kaggle, Rapid Miner, Split Validation