MODEL KLASIFIKASI POTENSI PENYAKIT *DIABETES MELLITUS* MENGGUNAKAN METODE *K-NEAREST NEIGHBOR* (K-NN)

SKRIPSI

OLEH
CANDRANI SRI MURTONO
NIM. 18610069

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022

MODEL KLASIFIKASI POTENSI PENYAKIT DIABETES MELLITUS MENGGUNAKAN METODE *K-NEAREST NEIGHBOR* (K-NN)

SKRIPSI

Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)

Oleh
Candrani Sri Murtono
NIM. 18610069

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022

MODEL KLASIFIKASI POTENSI PENYAKIT DIABETES MELLITUS MENGGUNAKAN METODE *K-NEAREST NEIGHBOR* (K-NN)

SKRIPSI

Oleh
Candrani Sri Murtono
NIM. 18610069

Telah Diperiksa dan Disetujui Untuk Diuji

Malang, 13 Desember 2022

Dosen Pembimbing I	Dosen Pembimbing II
Hisyam Fahmi, M.Kom.	Erna Herawati, M.Pd.
NIP. 19890727 201903 1 018	NIDT. 19760723 20180201 2 222

Mengetahui,
Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

MODEL KLASIFIKASI POTENSI PENYAKIT DIABETES MELLITUS MENGGUNAKAN METODE *K-NEAREST NEIGHBOR* (K-NN)

SKRIPSI

Oleh
Candrani Sri Murtono
NIM. 18610069

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal 21 Desember 2022

Ketua Penguji	: Dr. Usman Pagalay, M.Si	.....
Anggota Penguji I	: Dr. H. Imam Sujarwo, M.Pd	.....
Anggota Penguji II	: Hisyam Fahmi, M.Kom	.....
Anggota Penguji III	: Erna Herawati, M.Pd	.....

Mengetahui,
Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: Candrani Sri Murtono
NIM	: 18610069
Program Studi	: Matematika
Fakultas	: Sains dan Teknologi
Judul Skripsi	: Model Klasifikasi Potensi Penyakit Diabetes Mellitus menggunakan Metode *K-Nearest Neighbor* (K-NN)

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perilaku tersebut.

Malang, 22 November 2022

Yang membuat pernyataan,

Candrani Sri Murtono

NIM. 18610069

MOTO

Love anything you do, live only once, do your best, take every chance, jadilah muda yang bersemangat

PERSEMBAHAN

Skripsi ini penulis persembahkan untuk:

Diri saya sendiri yang telah berjuang menyelesaikan skripsi ini hingga tuntas dengan segala suka duka yang telah di lalui, Kedua orang tua saya tercinta yang sudah ada di surga, yaitu
Alm Drs. Sri Murtono, M.Pd C.D dan Almh Ir. RR. Sri Redjeki Wahjoe Astoeti, serta kedua adik saya Raditya Sri Murtono dan Dewi Sri Murtono yang dengan kasih sayangnya senantiasa mendukung penuh penulis dalam segala hal.

Seluruh keluarga besar R. Yoyok Subagyo yang senantiasa memberikan dukungan moril maupun materiil kepada penulis. Seluruh Om dan Tante para donatur dari teman-teman almh Mama dari arsitektur Universitas Brawijaya yang selalu memberikan dukungan moril dan materiil kepada penulis.

Partner terbaik saya yang senantiasa mendukung penuh kapanpun dan dimanapun saat penulis membutuhkan bantuan maupun dukungan dalam hal apapun. Seluruh teman-teman dari AKSIOMA dan GenBI yang telah memberikan semangat pada penulis. Terimakasih kepada seluruh pihak yang telah memberikan segala dukungan pada penulis, sehingga penulis dapat menyelesaikan skripsi ini dengan lancar.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dengan memanjatkan rasa syukur kehadirat Allah SWT yang telah melimpahkan rahmat, taufik, serta hidayah-Nya sehingga penulis dapat menyelesaikan pengerjaan skripsi yang berjudul "Model Klasifikasi Potensi Penyakit Diabetes Mellitus dengan menggunakan Metode *K-Nearest Neighbor* (k-NN)", sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) serta mendapatkan gelar sarjana matematika di Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Selama proses penyusunan skripsi ini, penulis menyampaikan banyak terima kasih kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim.
2. Dr. Sri Harini, M.Si, selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim.
3. Dr. Elly Susanti, S.Pd., M.Sc, selaku ketua Program Studi Matematika, Universitas Islam Negeri Maulana Malik Ibrahim.
4. Hisyam Fahmi, M.Kom., selaku dosen pembimbing I.
5. Erna Herawati, M.Pd., selaku dosen pembimbing II.
6. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim.

7. Kedua orang tua tercinta yang telah di surga, Almarhum Drs. Sri Murtono, M.Pd, C.D dan Almarhumah Ir. Rr. Sri Redjeki Wahjoe Astoeti yang telah melahirkan serta mendidik saya.
8. Kedua adik kandung saya Raditya Sri Murtono dan Dewi Sri Murtono yang selalu memberikan dukungan.
9. Tante Yayuk, seluruh donatur Arsitektur UB (teman Almarhumah Ibu), serta Om dan Tante keluarga Yoyok Subagyo yang telah memberikan bantuan baik segi materi maupun perhatian lainnya sehingga penulis dapat berkuliah hingga akhir.
10. Seluruh mahasiswa angkatan 2018 yang senantiasa mendukung serta selalu berbagi informasi.

Semoga skripsi yang berjudul "Model Klasifikasi Potensi Penyakit Diabetes Mellitus Menggunakan Metode *K-Nearest Neighbor* (k-NN)" yang telah penulis buat dapat membawa banyak manfaat bagi penulis, pembaca, serta masyarakat. Aamiin

Wassalamu'alaikum Wr. Wb

Malang, 19 September 2022

Penulis

DAFTAR ISI

HALAMAN JUDUL i
HALAMAN PENGAJUAN ii
HALAMAN PERSETUJUAN iii
PERNYATAAN KEASLIAN TULISAN .....
Error! Bookmark not defined.
MOTO vi
PERSEMBAHAN vii
KATA PENGANTAR viii
DAFTAR ISI x
DAFTAR TABEL xii
DAFTAR GAMBAR xiii
DAFTAR LAMPIRAN xiv
ABSTRAK .....
Error! Bookmark not defined.
ABSTRACT .....
Error! Bookmark not defined.
مستخلص البحث .....
Error! Bookmark not defined.
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Tujuan Penelitian 4
1.4 Manfaat Penelitian 4
1.5 Batasan Masalah 4
1.6 Definisi Istilah 5
BAB II TINJAUAN PUSTAKA 7
2.1 Teori Pendukung 7
2.1.1. Data Mining 7
2.1.2. *K-fold Cross Validation* 8
2.1.3. Klasifikasi 9
2.1.4. Regresi Logistik Biner 9
2.1.5. Algoritma *k-Nearest Neighbor* (k-NN) 11
2.1.6. *Confusion Matrix* 13
2.1.7. Kurva ROC 14
2.1.8. Diabetes Mellitus 15
2.1 Kajian Integrasi Penyakit Diabetes Mellitus dengan Al-Qur'an ... 20
2.2 Kajian Diabetes Mellitus dengan Metode *K-Nearest Neighbor* ... 21
BAB III METODE PENELITIAN 23
3.1 Jenis Penelitian 23
3.2 Data dan Sumber Data 23
3.3 Lokasi Penelitian 24
3.4 Teknik Pengumpulan Data 24
3.5 Instrumen Penelitian 25
3.6 Variabel Penelitian 25
3.7 Teknik Analisis Data 26
3.7.1. *Preprocessing* 27

3.7.2. *Data Mining* 28
3.7.3. Evaluasi 29

BAB IV HASIL DAN PEMBAHASAN 30
4.1 Hasil Penelitian 30
4.1.1 Analisis Deskriptif 30
4.1.2 *Pre-processing* 34
4.1.2.1 *Data Cleaning* 35
4.1.2.2 *Data Selection* 35
4.1.3 *K-Nearest Neighbor* (k-NN) 50
4.1.3.1 Data RST tk.II dr. Soepraoen Malang Sebelum Uji Regresi Logistik Biner 50
4.1.3.2 Data RST tk.II dr. Soepraoen Malang Setelah Uji Regresi Logistik Biner 57
4.1.4 Evaluasi 63
4.1.4.1 Data RST tk.II dr.Soepraoen Malang sebelum Uji Regresi Logistik Biner 63
4.1.4.2 Data RST tk.II dr.Soepraoen Malang setelah Uji Regresi Logistik Biner 66
4.1.5 Grafik ROC 68
4.2 Pembahasan 69
4.3 Kajian Keislaman dengan Hasil Penelitian 72
BAB V PENUTUP 74
5.1. Kesimpulan 74
5.2. Saran untuk Penelitian Lanjutan 74
DAFTAR PUSTAKA 76

DAFTAR TABEL

Tabel 2.1	*Confusion Matrix*	13
Tabel 3.1	Atribut Data RST	24
Tabel 4.1	Atribut Data RST tk. II dr. Soepraoen Malang	30
Tabel 4.2	Tahap *Selection* 1	36
Tabel 4.3	*Variables not in the Equation*	38
Tabel 4.4	Statistik Uji Data RST tk. II dr. Soepraoen Malang	40
Tabel 4.5	Tahap *Selection* 2	41
Tabel 4.6	Atribut Data RST tk. II dr. Soepraoen Malang	42
Tabel 4.7	Tabulasi BMI terhadap Diabetes Mellitus Data RST tk. II dr. Soepraoen Malang	43
Tabel 4.8	Tabulasi Kolesterol terhadap Diabetes Mellitus RST tk.II dr.Soepraoen Malang	44
Tabel 4.9	Tabulasi LDL Kolesterol terhadap Diabetes Mellitus	45
Tabel 4.10	Tabulasi Trigliserida terhadap Diabetes Mellitus	46
Tabel 4.11	Tabulasi Gula Darah Puasa terhadap Diabetes Mellitus	47
Tabel 4.12	Tabulasi Gula Darah 2 jam terhadap Diabetes Mellitus	48
Tabel 4.13	Tabulasi HBA1C terhadap diabetes mellitus	49
Tabel 4.14	*Euclidean Distance*	57
Tabel 4.15	*Euclidean Distance* Data RST tk. II dr. Soepraoen Malang	59
Tabel 4.16	*Confusion Matrix* Data RST tk. II dr. Soepraoen Malang	64
Tabel 4.17	Tingkat Akurasi	65
Tabel 4.18	Tingkat Akurasi berdasarkan nilai k	66
Tabel 4.19	*Confusion Matrix*	66
Tabel 4.20	Tingkat Akurasi RST tk.II dr.Soepraoen Malang Setelah	68
Tabel 4.21	Akurasi berdasarkan parameter	68
Tabel 4.22	Keterangan Atribut	71

DAFTAR GAMBAR

Gambar 3.1	Alur Penelitian	27
Gambar 3.2	*Flowchart* k-NN	29
Gambar 4.1	*Euclidean Distance* dengan Excel (sebelum)	52
Gambar 4.2	*Euclidean Distance* dengan Excel (setelah)	59
Gambar 4.3	Tahap 1 Uji Validasi	62
Gambar 4.4	*Cross Validation*	63
Gambar 4.5	Grafik AUC	69

DAFTAR LAMPIRAN

Lampiran 1 Dataset Diabetes Mellitus RST
Lampiran 2 Tingkat *Accuracy*
Lampiran 3 Tingkat *Precission*
Lampiran 4 Tingkat *Recall*
Lampiran 5 Statistik Uji
Lampiran 6 Statistik Uji

ABSTRAK

Murtono, Candrani Sri. 2022. **Model Klasifikasi Potensi Penyakit Diabetes Mellitus Menggunakan Metode *K-Nearest Neighbor* (k-NN).** Skripsi. Program Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim, Malang. Pembimbing: (I) Hisyam Fahmi, M.Kom. (II) Erna Herawati M.Pd.

Kata Kunci: Diabetes Mellitus, *k-Nearest Neighbor*, Klasifikasi, *RapidMiner*

Diabetes mellitus merupakan penyakit tidak menular yang terjadi di seluruh dunia. Banyak faktor yang menyebabkan seseorang dapat terkena diabetes mellitus. Selain faktor genetik, pola hidup juga sangat berpengaruh. Oleh karena itu diperlukan sebuah penelitian untuk mengetahui faktor yang paling berpengaruh terhadap seseorang dapat terkena diabetes mellitus. Penelitian ini membahas model klasifikasi potensi penyakit diabetes mellitus menggunakan metode *k-Nearest Neighbor* (k-NN). Tujuan penelitian ini adalah untuk mengetahui model dan hasil klasifikasi potensi penyakit diabetes mellitus agar masyarakat dapat melakukan tindakan preventif terhadap penyakit diabetes mellitus berdasarkan hasil yang didapatkan. Data yang digunakan merupakan data diabetes mellitus dari RST tk.II dr.Soepraoen Malang. Klasifikasi k-NN dilakukan dengan cara menentukan ketetanggaan terdekat. Didapatkan sembilan faktor yang paling berpengaruh terhadap seseorang dapat terkena diabetes mellitus dari hasil klasifikasi data pertama dari RST tk.II dr. Soepraoen Malang, yaitu usia, BMI, kolesterol total, LDL kolesterol, Trigliserida, gula darah sewaktu, gula darah puasa, gula darah 2 jam, dan HBA1C. Hasil yang diperoleh dari klasifikasi pada data ini adalah 98,08%, *precision* sebesar 96,30%, dan *recall* sebesar 100.00%. Telah dilakukan uji menggunakan data baru yang telah mendapatkan diagnosa dari dokter, dan saat data tersebut diproses menggunakan model yang telah terbentuk, hasil diagnosa yang diperoleh adalah sama dengan diagnosa yang di tegakkan oleh dokter. Sehingga penelitian ini dapat dikatakan berhasil.

ABSTRACT

Murtono, Candrani Sri. 2022. **Potential Classification Model for Diabetes Mellitus Using the K-Nearest Neighbor (K-NN) Method.** Thesis. Department of Mathematics, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Hisyam Fahmi, M.Kom. (II) Erna Herawati M.Pd.

Keywords: Classification, Diabetes Mellitus, k-Nearest Neighbor, RapidMiner

Diabetes mellitus is a non-communicable disease that occurs throughout the world. There are many factors that cause a person to develop diabetes mellitus. Not only genetic factors, lifestyle also has big impact. Therefore, a research is needed to find out the main factors that cause a person to get diabetes mellitus. This study discusses a potential classification model for diabetes mellitus using the k-Nearest Neighbor (k-NN) method. The purpose of this study was to find out the model and result of the potential classification of diabetes mellitus so that people can take preventive actions against diabetes mellitus based on the results obtained. The data that used in this research is diabetes mellitus data from RST tk.II Dr. Soepraoen Malang. K-NN classification is done by determining the nearest neighbor. There are nine factors that really effect on a person getting diabetes mellitus from the results of the first data classification from RST tk.II dr. Soepraoen Malang, namely age, BMI, total cholesterol, LDL cholesterol, triglycerides, random blood sugar, fasting blood sugar, 2 hour blood sugar, and HBA1C. The results obtained from the classification of this data are the accuracy is 98.08%, precision is 96.30%, and recall is 100.00%. A test has been carried out using new data that has received a diagnosis from a doctor, and when the data is processed using the model that has been formed, the diagnosis results obtained are the same as the diagnosis established by the doctor. So, this research has been successful.

مستخلص البحث

مورتونو، جندراني سري. ٢٠٢٢. **غموذج التصنيف المحتمل لمرض السكري باستخدام طريقة (*K-Nearest Neighbor*) الجار الأقرب**. البحث الجامعي. قسم الرّياضيات، كلية العلوم والتكنولوجيا. جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأوّل (١) هشام فهمي، الماجستير. المشرف الثانية (٢) إيرنا هيراواتي، الماجستير.

الكلمات الأساسية: التصنيف، السكري، الجار الأقرب، (*RapidMiner*)

مرض السكري مرض غير معدي يحدث في جميع أنحاء العالم. هنك كثير من العوامل التي تسبب أصابة الشخص بمرض السكري. ليس فقط العوامل الوراثية، نمط الحياة له أيضًا تأثير كبير. لذلك، هنك حاجة إلى إجراء بحث لمعرفة العوامل الرئيسية التي تسبب إصابة الشخص بمرض السكري. تناقش هذه الدراسة نموذجًا محتملاً للتصنيف كان الغرض من هذه الدراسة هو (*K-Nearest Neighbor*) الجار الأقرب. لمرض السكري باستخدام طريقة معرفة نموذج ونتيجة التصنيف المحتمل لمرض السكري حتى يتمكن الناس من اتخاذ إجراءات وقائية ضد مرض السكري بناءً على النتا ئج التي تم الحصول عليها. البيانات المستخدمة في هذا البحث هي بيانات مرض السكري من *RST tk.II dr. Soepraoen Malang*. عن طريق تحديد (*K-Nearest Neighbor*) الجار الأقرب يتم تصنيف. هناك تسعة عوامل تؤثر حقًا على إصابة الشخص بمرض السكري من نتائج تصنيف البيانات الأول من *RST tk.II dr. Soepraoen Malang* أي العمر، مؤشر كتلة الجسم، الكوليسترول الكلي، *LDL* كوليسترول، والدهون الثلاثية، وسكر الدم العشوائي، وسكر الدم الصائم وسكر الدم لمدة ٢ ساعة، و *HBA1C*. التي تم الحصول عليها من تصنيف هذه البيانات هي الدقة 98,08% والذقة 96,30% والاستدعاء 100,00% تم إجراء اختبار باستخدام بيانات جديدة تم تشخيصها من الطبيب، وعندما تتم معالجة البيانات باستخدام النموذج، الذي تم تشكيله، فإن نتائج التشخيص التي تم الحصول عليها هي نفس التشخيص الذي حدده الطبيب. لذلك، كان هذا البحث ناجحًا.

BAB I
PENDAHULUAN

1.1 Latar Belakang

Diabetes Mellitus termasuk ke dalam salah satu penyakit metabolik kronik yang jika tidak dilakukan tindakan perawatan maupun pengobatan yang tepat maka dapat mengakibatkan keadaan yang membahayakan hingga menyebabkan komplikasi (Aris & Benyamin, 2019). Diabetes merupakan masalah kesehatan masyarakat yang umum dan terus meningkat di seluruh dunia serta di seluruh negara maju atau berkembang. Diabetes merupakan kumpulan gangguan metabolisme yang ditandai dengan hiperglikemia. Karena sekresi insulin secara tidak normal, fungsi insulin, atau keduanya. (Hermayudi dkk, 2017). Diabetes Mellitus telah meningkat sebanyak 4 kali lipat pada tahun 1980 sebanyak 108 juta penderita diabetes usia dewasa menjadi 422 juta orang yang mengalami Diabetes Mellitus pada tahun 2014 (WHO, 2016). Ini juga dikuatkan dengan informasi dari (IDF, 2017) bahwa sebanyak 424,9 juta orang dewasa mengalami diabetes yang diperkirakan akan meningkat pada 2045 hingga sebanyak 628,6 juta penderita. Dikarenakan oleh semakin meningkatnya jumlah penderita penyakit Dabetes maka peneliti ingin melakukan penelitian terkait bagaimana klasifikasi potensi penyakit Diabetes Mellitus.

Dalam Islam sudah ada aturan tentang bagaimana tata cara maupun takaran seorang umat dalam hal dalam mengonsumsi makanan dan minuman. Seperti yang tertuang pada Hadis berikut ini (Al-Jauziyah, 2000):

الْمِقْدَامَ بن معدي كرب الكندي قال سمعت رَسُولَ اللَّهِ صلى الله عليه وسلم يقول ما مَلأَ بن آدَمَ وِعَاءً شَرًّا من بَطْنٍ حَسْبُ بن آدَمَ أُكُلاَتٌ يُقِمْنَ صُلْبَهُ فان كان لاَ مَحَالَةَ فَثُلُثٌ طَعَامٍ وَثُلُثٌ شَرَابٍ وَثُلُثٌ لِنَفَسِهِ رواه أحمد والترمذي وصححه الألباني

"Tidaklah seorang anak Adam (manusia) mengisi bejana (kantong) yang lebih buruk daripada perutnya. Cukuplah baginya beberapa suap yang bisa menegakkan tulang sulbinya. Jikalau memang harus berbuat, maka sepertiga untuk makanannya, sepertiga untuk minumannya dan sepertiga untuk nafasnya." (HR. Imam Ahmad, at-Tirmidzi dan rahimahumullah selainnya).

Sesuai dengan yang telah diketahui bahwa sebagai umat muslim yang baik hendaknya berperilaku sesuai dengan anjuran Islam untuk berbagai aspek kehidupan kita, seperti bagaimana kita menjaga kesehatan. Seperti dengan tidak melakukan sesuatu atau makan secara berlebihan.

Konsep yang digunakan pada *K-Nearest Neighbor* (k-NN) digunakan untuk melakukan klasifikasi pada objek berdasarkan nilai k yang ada di ruang fitur. Ukuran jarak dibutuhkan pada metode ini untuk proses penentuan kedekatan objek tertentu. k-NN juga merupakan mesin atau metode yang paling populer dan juga paling sederhana untuk memprediksi penyakit menggunakan data kesehatan. Model berbasis k-NN hanya terdiri dari penyimpanan set data pelatihan. Pada *machine learning* dan *data science*, klasifikasi merupakan salah satu teknik yang paling penting untuk menentukan prediksi. Metode yang berbeda digunakan untuk mendapatkan hasil yang berbeda pula, sehingga dapat memeriksa data dari berbagai perspektif sehingga kumpulan informasi yang diperoleh tersebut dapat diringkas menjadi suatu informasi yang dapat meningkatkan keakuratan hasil yang didapatkan. Objek yang berada di data uji akan diklasifikasikan dengan tetangga terdekat. Dalam proses yang dilakukan, metode ini mencari tetangga terdekat lalu memilih mayoritas kelas yang terdapat pada *cluster*. Metode *k-Nearest Neighbor*

akan menghasilkan keputusan untuk dapat mengklasifikasi data dari data latih dan mendapatkan hasil yang baik jika menggunakan data dalam jumlah besar(Saxena dkk, 2018)

Alasan penulis menggunakan metode *k-Nearest Neighbor* adalah, jika ditinjau berdasarkan dua jurnal yang membandingkan beberapa metode untuk meneliti tentang diabetes ini, metode *k-Nearest Neighbor* merupakan metode terbaik karena memunculkan hasil dengan tingkat akurasi yang paling tinggi jika dibandingkan dengan metode lain. Selain itu, metode k-NN ini merupakan metode paling sederhana untuk memprediksi penyakit menggunakan data kesehatan. Berdasarkan penelitian terdahulu terkait Diabetes Mellitus ini, metode *k-Nearest Neighbor* merupakan metode terbaik untuk masalah diabetes mellitus.

Sudah banyak dilakukan penelitian terkait diabetes mellitus pada penelitian sebelumnya, namun yang menjadikan pembeda pada penelitian kali ini dengan penelitian terdahulu adalah, pada penelitian kali ini juga dapat diketahui faktor apa saja yang paling berpengaruh diantara faktor-faktor penyakit diabetes mellitus. Hal itu dikuatkan dengan dilakukannya penghitungan terhadap dua jenis data yang berbeda serta memiliki atribut yang berbeda pula. Sehingga diharapkan masyarakat dapat melakukan tindakan preventif dari penyakit diabetes mellitus. Oleh karena itu, berdasarkan fakta dan informasi dari data yang telah didapat, pada penelitian kali ini penulis ingin mengetahui klasifikasi potensi penyakit Diabetes Mellitus menggunakan metode *k-Nearest Neighbor*.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang, maka rumusan masalah yang diambil dalam penelitian ini adalah bagaimana klasifikasi potensi penyakit Diabetes Mellitus menggunakan metode *k-Nearest Neighbor* (k-NN)?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, maka tujuan dari penelitian ini adalah untuk mengetahui hasil klasifikasi potensi penyakit Diabetes Mellitus menggunakan metode *k-Nearest Neighbor* (k-NN).

1.4 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut:

1. Menambah informasi terkait hal-hal yang dapat menjadi faktor peningkatan risiko terkena penyakit Diabetes Mellitus. Sehingga diharapkan, masyarakat dapat melakukan tindakan preventif sejak dini.
2. Mengetahui fakta, tanda gejala, faktor penyebab, cara mengantisipasi, serta tipe dari penyakit Diabetes Mellitus.
3. Mengetahui faktor apa saja yang paling berpengaruh terhadap penyakit Diabetes Mellitus.

1.5 Batasan Masalah

Mengingat luasnya pembahasan, maka penulis menentukan batasan masalah:

1. Data yang digunakan merupakan data Diabetes Mellitus yang didapatkan dari RST tk. II dr. Soepraoen Malang.

2. Batasan usia pasien yang digunakan pada penelitian kali ini adalah usia muda dan usia paruh baya menurut Orgaisasi Kesehatan Dunia (WHO) pada tahun 2015, yakni usia 25 tahun sampai dengan 60 tahun.
3. Metode yang digunakan dalam proses seleksi atribut adalah Uji Regresi Logistik Biner.
4. Metode yang digunakan dalam proses klasifikasi adalah metode *k-Nearest Neighbor*.
5. *Output* yang dihasilkan berupa atribut kategorikal dengan *output* YES dan NO.
6. Aplikasi yang digunakan untuk melakukan seleksi atribut adalah aplikasi SPSS.
7. Aplikasi yang digunakan untuk melakukan proses klasifikasi adalah *RapidMiner*.
8. Nilai *k* yang digunakan pada *k-Nearest Neighbor* adalah 9.
9. Nilai *k* yang digunakan pada *k-fold Cross Validation* adalah 10.

1.6 Definisi Istilah

Anomali	:	Keadaan penyimpangan atau keanehan yang terjadi atau dengan kata lain tidak seperti biasanya
Atribut	:	Bagian data yang mewakili karakteristik atau *feature* dari objek data
Database	:	Istilah teknologi jaringan komputer yang berfungsi sebagai penyimpanan data

Data *Training*	:	Data yang digunakan untuk menemukan model/pola yang nantinya akan diterapkan pada data *testing*
Data *Testing*	:	Data yang akan di uji dengan menggunakan pola/model yang telah ditemukan pada data *testing.* Digunakan untuk mengklasifikasikan data yang belum diketahui kelasnya.
Domain	:	Nama unik yang merupakan alamat dari suatu *website*
Glukosa	:	Kadar gula darah
Hiperglikemia	:	Kondisi ketika kadar gula di dalam darah melebihi batas normal
IMT	:	Indeks Massa Tubuh
Insulin	:	Insulin adalah hormon alami yang diproduksi oleh pankreas. Hormon insulin membantu mengontrol kadar gula darah (glukosa) dalam tubuh
Kolesterol	:	Lemak yang terkandung dalam aliran darah yang dibutuhkan untuk pembentukan dinding sel dan sebagai bahan baku beberapa hormon
Kronis	:	Kondisi atau sifat penyakit yang telah lama terjadi
Preventif	:	Tindakan pencegahan
Rapidminer	:	Aplikasi yang digunakan dalam proses penghitungan penelitian ini
Sekresi	:	Suatu proses pengeluaran zat yang dilakukan oleh kelenjar

BAB II
TINJAUAN PUSTAKA

2.1 Teori Pendukung

2.1.1. *Data Mining*

Data mining merupakan proses pengambilan data dari data *warehouse* berdasarkan prediksi variabel. Prediksi juga digunakan untuk menemukan data dari sekumpulan data yang ditentukan sebagai domain, seperti kecerdasan buatan, basis data, dan juga lainnya. Penggunaan data mining memiliki dampak di berbagai aspek kehidupan, sebagai contoh digunakan di bidang kesehatan. Data mining pada sistem medis berfungsi untuk mengekstrak informasi dari database sehingga dapat melakukan diagnosis penyakit (Aris & Benyamin, 2019).

Data mining juga disebut sebagai *knowledge discovery in database* yang merupakan proses pengumpulan serta pemakaian data untuk menemukan keteraturan, pola hubungan dalam set data berukuran besar. Hasil yang didapatkan melalui proses *data mining* ini dapat digunakan untuk pengambilan keputusan di masa depan. Tahapan yang dilakukan dalam proses *data mining* adalah sebagai berikut: Pembersihan data (*data cleaning*), proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan. Integrasi data (*data integration*), penggabungan data dari berbagai *database* ke dalam satu *database* baru. Seleksi data (*Data Selection*), tidak semua data yang ada pada *database* dipakai semuanya, maka dari itu data yang diambil dari *database* hanyalah data yang sesuai untuk dianalisis. Transformasi data (*Data Transformation*), yaitu proses saat data diubah atau digabung ke dalam format yang sesuai untuk dilakukan proses dalam *data mining*. Proses *mining*, merupakan suatu proses utama saat metode diterapkan

untuk menemukan pengetahuan berharga dan tersembunyi dari data. Evaluasi pola (*pattern evaluation*), proses untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.

Data mining adalah bidang yang terintegrasi dengan berbagai macam teknik dari beberapa bidang. Ini adalah kombinasi dari pembelajaran mesin, statistik, pengenalan pola, dan sistem kecerdasan buatan untuk analisis sejumlah besar data untuk menemukan pola tersembunyi dalam data. Saat ini teknik *data mining* diterapkan dalam ilmu kedokteran untuk pengambilan keputusan. Hal ini juga memegang peran penting dalam *dataset* Diabetes untuk mengekspos dan mengungkap pengetahuan tersembunyi dari sejumlah besar data diabetes yang tidak terpakai yang secara signifikan akan membantu untuk kemajuan pengobatan berkualitas untuk pasien yang menderita diabetes (Das dkk, 2018)

2.1.2. *K-fold Cross Validation*

K-fold Cross Validation merupakan metode yang dapat digunakan untuk melakukan validasi ulang guna melakukan cek keakuratan sebuah sistem. Pada pendekatan kali ini, seluruh data yang digunakan dibagi menjadi k partisi secara acak $(P1, P2, \dots, P_k)$. Jika menggunakan $k = 3$ atau disebut juga *3-fold cross validation* maka berarti sebanyak 2/3 dari seluruh data digunakan sebagai data latih dan 1/3 sisanya digunakan sebagai data uji(Mangande, 2020). Dalam tahapan yang dilakukan pada proses penelitian kali ini, *k-fold Cross Validation* dilakukan percobaan sebanyak *k* kali dengan menggunakan perbandingan data uji dan data latih yang berbeda pada setiap percobaannya guna mendapatkan perbandingan antara data uji dan data latih yang memiliki tingkat akurasi yang paling tinggi. Lalu

dilakukan pembagian data latih dan data uji dengan perbandingan yang telah didapatkan.

2.1.3. Klasifikasi

Metode klasifikasi merupakan proses yang bertujuan guna menemukan model atau fungsi yang menjelaskan konsep atau kelas data agar dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Hal ini juga dapat dikatakan sebagai pembelajaran (klasifikasi) yang memetakan sebuah item data ke dalam salah satu dari beberapa kelas yang sudah didefinisikan (Aris & Benyamin, 2019). Model tersebut biasanya berupa aturan "jika-maka", berupa *decision tree*, formula matematis maupun *neural network.* Metode yang ada pada klasifikasi diantaranya ada C4.5, *RainForest, Naïve Bayes, Neural Network, Genetic Algorithm, Fuzzy, Case-Based Reasoning,* dan *k-Nearest Neighbor* (Arriawati dkk, 2011).

2.1.4. Regresi Logistik Biner

Regresi logistik merupakan metode analisis statistika guna mendeskripsikan hubungan antara peubah respon yang memiliki dua atau lebih kategori dengan satu atau lebih peubah penjelas yang berskala kategori atau interval (Hosmer & Lemeshow, 2000). Analisis regresi logistik adalah suatu metode analisis untuk mengetahui hubungan atau pengaruh, di mana variabel responnya berbentuk kategori yang memiliki satu atau lebih variabel bebas (Suniantara & Rusli, 2017).

Regresi logistik biner merupakan salah satu metode yang dapat digunakan untuk mencari hubungan antara variabel dependen yang bersifat *dichotomous* (berskala nominal atau ordinal dengan dua kategori) yang memiliki satu atau lebih variabel *independent* yang bersifat kontinu atau kategorik (Agresti, 2002). Regresi

logistik biner terdiri dari beberapa tahapan. Tahap pertama adalah uji independensi yang bertujuan untuk mengetahui apakah di antara variabel-variabel yang akan diteliti memiliki hubungan atau tidak. Model regresi logistik biner digunakan jika variabel responnya menghasilkan dua kategori bernilai Yes dan No. Model regresi logistiknya adalah sebagai berikut (Anggraeni & Zain, 2015):

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p}}$$

dimana p = banyaknya variabel predictor x_i

Estimasi Parameter dilakukan dengan menggunakan metode MLE. Pada dasarnya metode maksimum likelihood memberikan nilai estimasi β dengan memaksimumkan fungsi likelihood nya (Agresti, 2002) secara matematis fungsi likelihood dapat dinyatakan

$$l(\beta) = \prod_{i=1}^{n} \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$$

dimana:

$$(x_i) = \frac{\exp(\sum_{j=0}^{p} \beta_i x_{ij})}{1 + \exp(\sum_{j=0}^{p} \beta_i x_{ij})}$$

$$\mathrm{L}(\beta) = ln[l(\beta)]$$

$$\mathrm{L}(\beta) = \sum_{j=0}^{p} \left[\sum_{i=1}^{n} y_i x_{ij} \right] \beta_j - \sum_{i=1}^{n} n_i \ln \left[1 + \exp(\sum_{j=0}^{p} \beta_j x_{ij}) \right]$$

Langkah selanjutnya adalah melakukan pengujian parameter model regresi logistik, dilakukan untuk menguji variabel predictor berpengaruh atau tidak terhadap variabel respon. Adapun pengujian parameter yang dilakukan adalah uji

signifikansi secara serentak dan uji signifikansi secara individu (Hosmer & Lemeshow, 2000).

Bentuk logit nya dapat dituliskan sebagai berikut:

$$\text{Logit}\left(\pi(x)\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

Definisi 1. Fungsi kepadatan peluang (fkp) bagi variabel *random* x yang berdistribusi logistik adalah:

$$f(x) = \frac{\exp\left\{\frac{x-\mu}{\tau}\right\}}{\tau\left[1+\exp\left\{\frac{x-\mu}{\tau}\right\}\right]^2}; \; -\infty \leq x \leq \infty; \tau > 0$$

Dengan nilai tengah μ dan ragam $\sigma^2 = \pi^2 \, \tau^{2}/_{3}$

2.1.5. Algoritma *k-Nearest Neighbor* (k-NN)

Algoritma k-NN adalah algoritma yang menentukan nilai jarak pada pengujian data uji dengan data latih berdasarkan nilai terkecil dari nilai ketetanggaan terdekat. Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan *training samples*(Krisandi dkk, 2013). Metode *k-Nearest Neighbor* (k-NN) adalah teknik klasifikasi pembelajaran mesin non-parametrik yaitu, tidak ada asumsi untuk distribusi data yang mendasarinya. Dengan kata lain, struktur model ditentukan dari *dataset* dengan menyimpan semua kasus yang tersedia dan memprediksi kasus baru berdasarkan ukuran kesamaan (Sarker dkk, 2018). Tujuan dari algoritma k-NN adalah guna mengklasifikasi objek baru berdasarkan atribut dan data latih, di mana hasil dari data latih yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada k-NN. Algoritma k-NN

menggunakan klasifikasi ketetanggaan yang digunakan sebagai nilai prediksi dari sampel uji yang baru (Daniel T & Larose Ph.D, 2005).

Menurut (Karegowda dkk, 2012) langkah yang dapat dilakukan untuk mengklasifikasikan data dengan menggunakan metode *k-Nearest Neighbor* adalah sebagai berikut :

1. Menyiapkan data latih dan data uji.
2. Mendefinisikan nilai *k.*
3. Melakukan perhitungan nilai jarak antara data latih dengan data uji. Dengan rumus penghitung jarak *euclidean* sebagai berikut (Alfina dkk, 2012):

$$d(x_i, x_j) = \sqrt{\sum_{i,j=1}^{n} (x_i - x_j)^2} \qquad (2.1)$$

Keterangan:

$d(x_i, x_j)$ = Jarak

i = Variabel Data

n = Dimensi Data

X_i = Data latih (*data training*)

X_j = Data uji (*data testing*)

4. Mengelompokkan data berdasarkan perhitungan jarak.
5. Mengelompokkan data berdasarkan nilai tetangga terdekat.
6. Memilih nilai yang sering muncul dari tetangga terdekat sebagai acuan prediksi data selanjutnya.

2.1.6. *Confusion Matrix*

Confusion Matrix merupakan metode perhitungan guna melakukan analisis kualitas model klasifikasi dalam mengenali *tuple* dari kelas yang ada (Suyanto, 2018). *Confusion Matrix* merupakan tabel yang berisikan hasil dari proses klasifikasi. Sebagai contoh:

Tabel 2. 1 *Confusion Matrix*

	Positif	Negatif
Positif	TP (*True Positive*)	FP (*False Positive*)
Negatif	FN (*False Negative*)	TN (*True Negative*)

Setelah didapatkan nilai *Confusion Matrix* maka nilai yang telah didapatkan akan digunakan untuk mendapatkan nilai *accuracy, precission, recall* dan *Receiver Operation Characteristic* (ROC). Persamaan menurut (Mangande, 2020) untuk menghitung Nilai Akurasi:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.2)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2.3)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2.4)$$

Keterangan:

TP (*True Positive*) : Jumlah positif yang diklasifikasikan sebagai positif

TN (*True Negative*) : Jumlah negatif yang diklasifikasikan sebagai negatif

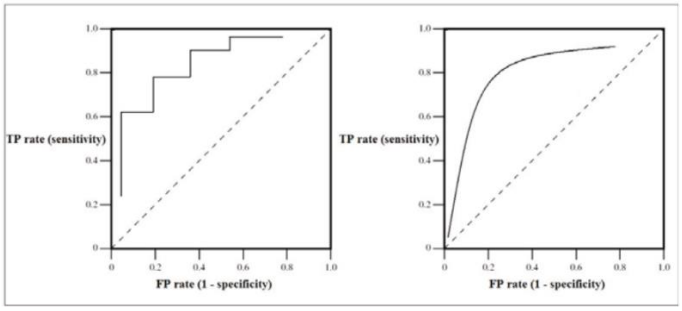
FP (*False Positive*) : Jumlah negatif yang diklasifikasikan sebagai positif

FN (*False Negative*) : Jumlah positif yang diklasifikasikan sebagai negatif

2.1.7. Kurva ROC

Grafik kurva ROC (*Receiver Operating Characteristic*) digunakan untuk mengevaluasi akurasi *classifier* dan untuk membandingkan klasifikasi yang berbeda model. Sebuah grafik ROC adalah grafik dua dimensi dengan proporsi negatif pada sumbu horisontal dan proporsi positif yang benar di sumbu vertikal (Vercellis, 2008). Kegunaan kurva ROC (*Receiver Operating Characteristic*) adalah untuk radar selama Perang Dunia II untuk mendeteksi benda-benda musuh di medan perang, teori deteksi sinyal, dalam psikologi ke rekening untuk deteksi sinyal persepsi, penelitian medis dan dalam mesin pembelajaran dan penelitian data mining, serta masalah klasifikasi (Gorunescu, 2011).

Dalam masalah klasifikasi menggunakan kelas keputusan dua (klasifikasi biner), masing-masing objek dikelompokkan dalam (P, N), yaitu positif atau negatif. Sementara beberapa model klasifikasi menghasilkan label kelas diskrit (menunjukkan hanya kelas diprediksi objek), pengklasifikasi lainnya (misalnya, *naive bayes*, jaringan saraf) menghasilkan output yang berkesinambungan, yang ambang batas yang berbeda mungkin diterapkan untuk memprediksi keanggotaan kelas. Secara teknis, *ROC curve* juga dikenal sebagai grafik ROC, dua-dimensi grafik dimana tingkat TP diplot pada sumbu-Y dan tingkat FP diplot pada sumbu-X (Gorunescu, 2011).



Gambar 2.1 Grafik ROC

Pada gambar 2.1 di atas garis diagonal membagi ruang ROC, yaitu:

1. Pada gambar pertama, terdapat garis diagonal yang merupakan hasil klasifikasi yang baik.
2. Pada gambar kedua, terdapat garis diagonal yang merupakan hasil klasifikasi yang buruk.

Digunakan metode yang menghitung luas daerah di bawah kurva ROC yang disebut AUC (*Area Under the ROC Curve*) yang diartikan sebagai probabilitas. AUC merupakan perhitungan untuk mengukur perbedaan performansi metode yang digunakan untuk mengekspresikan data *confusion matrix*. Garis horizontal mewakili nilai *false positives* (FP) dan garis vertikal mewakili nilai *true positives* (TP). AUC mengukur kinerja diskriminatif dengan memperkirakan probabilitas *output* dari sampel yang dipilih secara acak dari populasi positif atau negatif, semakin besar AUC, semakin kuat klasifikasi yang digunakan. Karena AUC adalah bagian dari daerah unit persegi, nilainya akan selalu antara 0,0 dan 1,0. Untuk keakurasian nilai AUC dalam klasifikasi *data mining* dibagi menjadi lima kelompok (Gorunescu, 2011), yaitu:

1. 0.90 – 1.00 = klasifikasi sangat baik (*excellent classification*)
2. 0.80 – 0.90 = klasifikasi baik (*good classification*)
3. 0.70 – 0.80 = klasifikasi cukup (*fair classification*)
4. 0.60 – 0.70 = klasifikasi buruk (*poor classification*)
5. 0.50 – 0.60 = klasifikasi salah (*failure*)

2.1.8. Diabetes Mellitus

Diabetes Mellitus merupakan penyakit yang memiliki tingkat risiko kematian yang tinggi. Penyakit Diabetes Mellitus terjadi ketika produksi insulin

dalam tubuh tidak memadai atau tubuh tidak dapat memproduksi insulin dengan tepat. Pada umumnya penyakit Diabetes Mellitus terjadi ketika kadar gula darah berada di atas normal (Iyer dkk., 2015). Diabetes Mellitus tergolong pada penyakit yang tidak dapat disembuhkan, akan tetapi kita dapat mempertahankan secara optimal terkait kualitas pertumbuhan dan perkembangan pasien, dengan cara melakukan kontrol metabolik pada pasien. Hal yang dapat dilakukan untuk mempertahankan keseimbangan metabolik yaitu dengan pemberian injeksi insulin secara berkelanjutan, melakukan diet olah raga secara rutin, pemeliharaan kesehatan, edukasi diri sendiri dan lingkungan keluarga (Asli et al., 2007).

Diabetes Mellitus termasuk ke dalam salah satu penyakit yang tidak menular yang paling sering terjadi secara global. Penyakit ini berada pada urutan keempat penyebab kematian di sebagian besar negara berkembang. Diabetes Mellitus juga dikenal sebagai penyakit yang heterogen yang biasanya ditandai dengan kadar gula darah yang tinggi serta toleransi glukosa terganggu, serta kekurangan insulin, kelemahan keefektifan peran insulin, ataupun karena kedua alasan tersebut (Nor et al., 2020). Diabetes Mellitus adalah gangguan metabolisme kronis yang mengakibatkan regulasi glukosa darah abnormal. Tingkat glukosa darah sebaiknya dipertahankan mendekati normal melalui praktik manajemen diri, yang melibatkan pelacakan kadar glukosa darah secara aktif dan mengambil tindakan yang tepat termasuk menyesuaikan diet dan obat insulin. Anomali glukosa darah dapat didefinisikan sebagai pembacaan yang tidak diinginkan baik karena alasan yang diketahui secara tepat (variasi penyebab normal) atau alasan yang tidak diketahui (variasi penyebab khusus) pada pasien (Sarker et al., 2018).

A. Tipe Diabetes Mellitus

1. Diabetes Mellitus Tipe 1

   Diabetes Mellitus tipe 1 juga disebut sebagai insulin *dependent.* Penyakit ini bermanifestasi sebagai penyakit auto imun yang menyerang pada usia muda, bahkan usia di bawah 20 tahun. Sering kali diabetes tipe ini juga disebut dengan *juvenile-onset diabetes*. Pada diabetes tipe 1 ini, sel pankreas yang berfungsi untuk memproduksi insulin telah dirusak oleh sistem pertahanan tubuh. Injeksi insulin serta melakukan pengecekan kadar gula darah, dan juga melakukan diet harus dilakukan oleh penderita Diabetes Mellitus tipe 1 ini (Iyer dkk., 2015).

2. Diabetes Mellitus Tipe 2

   Diabetes tipe ini menyumbang hampir 90% dari total keseluruhan kasus diabetes, biasa disebut *adult-onset* diabetes atau diabetes yang tidak tergantung pada insulin (non-insulin dependent) (Sa'di dkk., 2015). Dalam kondisi ini, seluruh organ tubuh menjadi resisten insulin dan hal ini menyebabkan peningkatan permintaan insulin. Pada titik ini, organ pankreas tidak memproduksi insulin sesuai yang dibutuhkan. Tindakan yang dapat dilakukan agar terhindar dari diabetes tipe 2 ini, pasien harus melakukan diet ketat, olahraga secara rutin, dan selalu memantau kadar glukosa pada darah secara rutin dan berkala. Sebagian besar penderita diabetes tipe 2 mengalami suatu kondisi di mana kadar glukosa darah lebih tinggi dari batas normal akan tetapi tidak setinggi kadar pada pasien diabetes (Iyer dkk., 2015).

3. Diabetes Mellitus Gestasional (DMG)

   Merupakan penyakit dengan berbagai gejala yang muncul pada ibu hamil yang disebabkan oleh adanya peningkatan kadar gula darah akibat terjadinya penurunan sekresi insulin yang progresif. Diabetes Mellitus tipe ini ditegakkan atas dasar pemeriksaan kadar gula darah secara enzimatik dengan bahan darah plasma vena. Pemeriksaan glukosa puasa ≥126 mg/dl atau pemeriksaan glukosa plasma ≥ 200 mg/dl 2 jam setelah Tes Toleransi Glukosa Oral (TTGO) atau pemeriksaan glukosa sewaktu ≥ 200 mg/dl. Biasanya keadaan ini terjadi pada saat usia kehamilan memasuki usia 24 minggu. Sebagian penderita akan kembali normal setelah melahirkan. Hampir setengah dari angka kejadiannya, diabetes dapat muncul kembali (Fitriani, 2017).

B. Tanda Gejala Diabetes Mellitus

Tanda gejala awal seseorang terserang Diabetes Mellitus akan ditandai dengan kebiasaan mengonsumsi makanan serta minum yang berlebih, sering melakukan buang air kecil, mudah mengantuk dan sering tidur. Namun pada usia muda meskipun mengalami rasa kantuk dan suka tidur tidak dapat dijadikan patokan jika orang tersebut terkena Diabetes Mellitus. Mereka bisa saja terkena penyakit kurang darah atau anemia. Jika memiliki ciri tersebut di atas, sebaiknya memeriksakan diri ke dokter untuk memastikan mengidap Diabetes Mellitus atau tidak serta dapat melakukan tindakan preventif agar terhindar dari penyakit Diabetes Mellitus. Sangat diperlukan untuk menjalani pola hidup yang sehat juga. Diabetes ini tidak semata berlatar belakang keluarga alias terpaut gen, tetapi pola makan dan aktivitas juga sangat berpengaruh (Wistiani, 2011).

C. Faktor Penyebab Diabetes Mellitus

Berikut adalah beberapa faktor penyebab terjadinya Diabetes Mellitus menurut beberapa sumber:

1. Faktor Risiko Diabetes Mellitus menurut P2PTM Kemenkes RI

a. Berat badan berlebih

b. Kurang aktifitas fisik

c. Kadar kolesterol HDL (kolesterol baik) ≤35 mg/dl

d. Kadar Trigliserida ≥250 mg/dl

e. Riwayat penyakit jantung

f. Tekanan darah tinggi

g. Diet tidak seimbang (tinggi gula, garam, lemak, dan rendah serat)

2. Faktor Risiko Diabetes Mellitus menurut WHO

Informasi dikutip dari laman resmi *Centers for Disease Control and Prevention* pada sub-Diabetes. Berikut akan ditampilkan faktor risiko diabetes mellitus tipe 1:

a) Riwayat keturunan keluarga

Memiliki orang tua, saudara laki-laki, atau saudara perempuan dengan diabetes tipe 1

b) Usia

Diabetes tipe 1 dapat menyerang berbagai usia, akan tetapi biasanya berkembang pada anak-anak, remaja, atau dewasa muda.

Informasi dikutip dari laman resmi *Centers for Disease Control and Prevention* pada sub-Diabetes. Berikut akan ditampilkan faktor risiko diabetes mellitus tipe 2:

a) Kelebihan berat badan
b) Usia 45 tahun atau lebih
c) Memiliki orang tua, saudara laki-laki, atau saudara perempuan dengan diabetes tipe 2
d) Aktifitas fisik kurang dari 3 kali dalam seminggu
e) Mempunyai Riwayat penyakit hati berlemak non-alkohol
f) Pernah mengalami diabetes gestasional (diabetes mellitus saat hamil) atau pernah melahirkan bayi dengan berat badan lebih dari 9 kg
g) Penduduk Amerika Afrika, Hispanik atau Latin, Indian Amerika, beberapa orang di Kepulauan Pasifik, Asia Amerika, atau penduduk asli Alaska beresiko lebih tinggi terkena diabetes mellitus tipe 2

2.2 Kajian Integrasi Penyakit Diabetes Mellitus dengan Al-Qur'an

Manusia telah dianjurkan untuk senantiasa memelihara kesehatannya, salah satunya dengan cara menjaga pola makannya. Karena kesehatan tubuh manusia sangat berkaitan dengan makanan dan minuman yang dikonsumsi. Makanan serta minuman yang berlebihan dapat mengakibatkan terjadinya berbagai gangguan kesehatan. Oleh karena itu Allah SWT melarang manusia jika berlebihan dalam makan maupun minum. Perbuatan berlebihan yang melebihi batas selain merusak, juga akan merugikan, juga Allah tidak menyukainya. Sesuatu jika tidak disukai oleh Allah SWT, jika tetap dikerjakan juga maka akan mendatangkan kerugian. Hal ini juga dikuatkan oleh Hadis berikut (Al-Jauziyah, 2000):

الْمِقْدَامَ بن معدي كرب الكندي قال سمعت رَسُولَ اللَّهِ صلى الله عليه وسلم يقول ما مَلأَ بن آدَمَ وِعَاءً شَرًّا من بَطْنٍ حَسْبُ بن آدَمَ أُكُلاَتٌ يُقِمْنَ صُلْبَهُ فان كان لاَ مَحَالَةَ فَثُلُثٌ طَعَامٍ وَثُلُثٌ شَرَابٍ وَثُلُثٌ لِنَفَسِهِ رواه أحمد والترمذي وصححه الألباني

"Tidaklah seorang anak Adam (manusia) mengisi bejana (kantong) yang lebih buruk daripada perutnya. Cukuplah baginya beberapa suap yang bisa menegakkan tulang sulbinya. Jikalau memang harus berbuat, maka sepertiga untuk makanannya, sepertiga untuk minumannya dan sepertiga untuk nafasnya." (HR. Imam Ahmad, at-Tirmidzi dan rahimahumullah selainnya).

Dari Hadis di atas, hendaknya kita sebagai manusia yang beriman untuk dapat mengupayakan agar dapat melaksanakan perintah Allah SWT dalam setiap urusan kita di dunia. Hendaknya kita tidak makan atau minum berlebih, cukup sesuai dengan porsi yang dibutuhkan dengan mempertimbangkan gizi nya. Hal tersebut juga dikuatkan dengan Hadis di bawah ini (Al-Jauziyah, 2000):

نعمتان مغبون فيهما كثير من الناس الصحة والفراغ

"Dua kenikmatan yang sering dilupakan oleh kebanyakan manusia adalah kesehatan dan waktu luang." (HR. Al-Bukhari: 6412, at-Tirmidzi: 2304, Ibnu Majah: 4170).

Ibnu Bathal menjelaskan bahwa makna hadis ini adalah seseorang tidak dikatakan memiliki waktu luang hingga ia juga memiliki badan yang sehat. Barang siapa yang memiliki hal tersebut (waktu luang dan badan yang sehat) hendaknya ia bersemangat agar jangan sampai ia tertipu dengan meninggalkan syukur kepada Allah atas segala nikmat yang diberikan-Nya. Termasuk bersyukur kepada Allah adalah dengan melaksanakan segala perintah-Nya dan menjauhi segala larangan-Nya. Barang siapa yang tidak bersyukur seperti itu maka ialah orang yang tertipu. (Fathul Bari bi Syarhi Shahihil Bukhari: 14).

2.3 Kajian Diabetes Mellitus dengan Metode *K-Nearest Neighbor*

Penelitian yang dilakukan kali ini berkaitan dengan proses klasifikasi penyakit Diabetes Mellitus dengan menggunakan metode *k-Nearest Neighbor* (k-NN). Metode k-NN merupakan metode *data mining* yang paling sederhana dalam memprediksi atau melakukan klasifikasi terhadap penyakit menggunakan data

kesehatan. Metode k-NN adalah metode yang menentukan nilai jarak pada pengujian data uji dengan data latih berdasarkan nilai terkecil dari nilai ketetanggaan terdekat.

Tahapan yang dilakukan dalam penelitian kali ini adalah menentukan atribut apa saja yang akan digunakan dalam pemrosesan data, setelah itu ditentukan *output* apa yang diinginkan untuk dimunculkan sebagai *output* dari aplikasi. Proses pengklasifikasian yang dilakukan yaitu dengan memasukkan data pasien Diabetes Mellitus yang telah didapat dari RS. Tk II dr. Soepraoen. Data yang didapatkan harus di masukkan ke aplikasi Ms. Excel, agar menjadi format .xlsx agar dapat dibaca dan di proses menggunakan aplikasi *RapidMiner*. Sebelum dilakukannya proses klasifikasi menggunakan metode *k-Nearest Neighbor* (k-NN), dilakukan *preprocessing* berupa normalisasi yang bertujuan agar tidak terdapat jarak yang terlalu jauh antara data dari atribut satu dengan yang lainnya. Data yang telah di normalisasi di proses menggunakan metode k-NN dengan bantuan aplikasi *RapidMiner*. Sehingga didapatkan hasil klasifikasi yang diinginkan dengan *output* berupa “YES” jika terkonfirmasi Diabetes Mellitus, dan *output* berupa “NO” tidak terkonfirmasi Diabetes Mellitus.

BAB III
METODE PENELITIAN

3.1 Jenis Penelitian

Jenis penelitian yang dilakukan adalah penelitian kuantitatif dengan memanfaatkan data kuantitatif dengan cara melakukan uji terhadap data sekunder berupa data Diabetes Mellitus dari rekam medik pasien yang peneliti dapatkan dari RS Tk. II dr. Soepraoen Malang. Penelitian dilakukan secara sistematis dengan menggunakan model dan teori yang berkaitan. Penelitian yang dilakukan pada kali ini merupakan proses *data mining* dengan menggunakan metode *k-Nearest Neighbo*r (k-NN).

3.2 Data dan Sumber Data

Penelitian yang dilakukan pada kali ini menggunakan 2 jenis data sekunder yang didapatkan dari 2 sumber yang berbeda pula. Kedua data memiliki beberapa atribut yang berbeda yang akan diolah dengan menggunakan metode yang sama. Data pertama merupakan data sekunder berupa data Diabetes Mellitus yang diambil langsung dari Rumah Sakit Tentara tk. II dr. Soepraoen Malang. Data yang di dapatkan berupa data *hardfile* rekam medik pasien yang disortir berdasarkan kebutuhan peneliti, diantaranya adalah pasien Diabetes Mellitus yang berkunjung pada RST tk. II dr. Soepraoen Malang pada bulan Januari 2021 sampai dengan bulan Maret 2022. Data tersebut berjumlah 52 data pasien, terdiri dari 9 *independent variable* dan 1 *dependent variable* yang terdapat pada data rekam medik penyakit Diabetes Mellitus. Berikut merupakan tabel terkait atribut yang digunakan pada data ini:

Tabel 3. 1 Atribut Data RST tk. II dr. Soepraoen Malang

No	Atribut	Tipe
1	Usia	Rasio
2	BMI	Rasio
3	Kolesterol Total	Rasio
4	LDL Kolesterol	Rasio
5	Trigliserida	Rasio
6	Gula darah sewaktu	Rasio
7	Gula darah puasa	Rasio
8	Gula darah 2 jam	Rasio
9	HBA1C	Rasio
10	Diagnosa	Kategorikal

Acuan yang digunakan dalam mendapatkan satuan data dan kadar pada setiap atribut adalah informasi dari WHO.

3.3 Lokasi Penelitian

Penelitian yang dilakukan dalam penyusunan skripsi "Model Klasifikasi Prediksi Penyakit Diabetes Mellitus dengan Menggunakan Metode *k-Nearest Neighbo*r (k-NN)" ini dilaksanakan di Rumah Sakit Tk. II dr. Soepraoen, Kota Malang, Jawa Timur.

3.4 Teknik Pengumpulan Data

Proses pengumpulan data yang dilakukan pada penelitian kali ini adalah sebagai berikut:

1. Mendapatkan data pertama dari pihak Rumah Sakit tk. II dr. Soepraoen Malang berupa *hardfile* Rekam Medik pasien yang mengidap Diabetes Mellitus pada usia 25 sampai dengan 60 tahun. Data tersebut diperoleh dengan terdapat 46 atribut awal yang selanjutnya dilakukan tahap seleksi

sehingga didapatkan 10 atribut utama yang digunakan pada penelitian kali ini.

2. Memasukkan data *hardfile* dari rumah sakit, ke dalam aplikasi Ms. Excel sesuai dengan atribut yang dibutuhkan untuk dilakukan tahap selanjutnya.

3.5 Instrumen Penelitian

Dikarenakan data yang digunakan pada penelitian kali ini merupakan data sekunder yang telah tersedia, sehingga peneliti tidak terjun langsung ke lapangan guna mendapatkan data pasien satu persatu. Sehingga instrumen yang digunakan dalam proses penulisan skripsi ini serta mempermudah penulis dalam melakukan proses penelitian adalah sebagai berikut:

1. Data

   Merupakan data sekunder berupa data Diabetes Mellitus dari RST tk. II dr. Soepraoen Malang.

2. Laptop / PC

   Laptop yang digunakan sebaiknya menggunakan RAM 4, guna menjalankan aplikasi yang dibutuhkan dalam proses penelitian serta dapat menjalankan *multitasking* beberapa aplikasi dengan lancar.

3.6 Variabel Penelitian

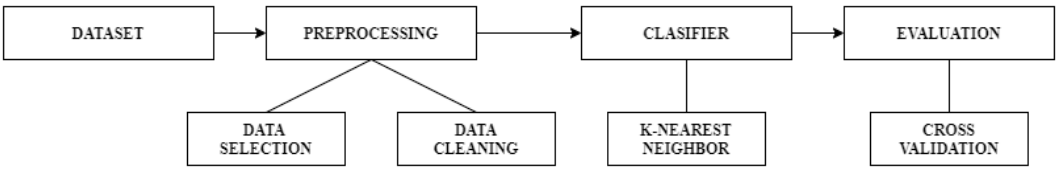
Penelitian kali ini menggunakan dua jenis data dengan variabel yang berbeda. Pada data pertama yang diperoleh dari RST tk. II dr. Soepraoen Malang menggunakan dua jenis variabel yaitu variabel bebas (*independent variable*) yang terdapat 9 (sembilan) atribut yaitu usia, BMI, kolesterol total, LDL kolesterol, Trigliserida, gula darah sewaktu, gula darah puasa, gula darah 2 jam, dan HBA1C.

Serta 1 (satu) variabel terikat (*dependent variable*) yaitu berupa target (*output*) pada penelitian kali ini yaitu atribut “Diagnosa”.

3.7 Teknik Analisis Data

Pada penelitian kali ini terdapat 3 langkah utama yang dilakukan, yaitu tahap *pre-processing, data mining,* dan *evaluation.* Setiap tahapnya mempunyai fungsi masing-masing. Dimulai dari *pre-processing,* pada tahap ini akan dilakukan tahapan *data selection* dan *data cleaning. Data selection* dilakukan guna memilih data berdasarkan data serta atribut yang sesuai dengan penelitian kali ini, pada tahap ini data dilakukan sebanyak dua kali pemilihan data, yang pertama adalah seleksi data berdasarkan informasi dari P2PTM Kemenkes Republik Indonesia, yang selanjutnya dilakukan tahap pemilihan atribut berdasarkan hasil Uji Regresi Logistik biner. Tahap *preprocessing* selanjutnya adalah tahap *data cleaning* yang dilakukan guna membersihkan data dari data yang tidak digunakan, maupun data yang terdapat *noise* atau *missing value* yang terlalu banyak.

Setelah data melalui tahap *preprocessing* maka data siap untuk dilakukan tahap selanjutnya yaitu tahap *data mining.* Tahap *mining* yang dilakukan pada penelitian kali ini menggunakan metode *k-Nearest Neighbor.* Data akan diklasifikasikan berdasarkan atribut yang paling berpengaruh. Tahap uji data dilakukan dengan membagi data menjadi dua bagian yaitu *data training* dan *data testing*. Setelah data dibagi maka dapat dilakukan tahap klasifikasi dengan mengimplementasikan metode *k-Nearest Neighbor*. Langkah terakhir yang dilakukan adalah tahapan evaluasi dengan menggunakan metode *k-fold Cross Validation* guna mendapatkan hasil klasifikasi terbaik. Berikut merupakan penjabaran tahapan yang dilalui dalam proses penelitian kali ini:



Gambar 3. 1 Alur Penelitian

3.7.1. Preprocessing

Preprocessing merupakan suatu teknik awal yang digunakan pada data agar lebih mudah saat digunakan dalam proses *data mining*. Tujuan dilakukannya *preprocessing* ini yaitu menjadikan kualitas data lebih baik dalam hal konsistensi, ketepatan, kelengkapan serta meningkatkan hasil akurasi (de Souza Lima dkk., 2017). Berikut merupakan tahapan *preprocessing* yang dilakukan pada kedua *dataset* diabetes dari RST tk.II dr. Soepraoen Malang.

1. ***Data Cleaning***

   Data cleaning merupakan tahap yang dilakukan guna menghilangkan *noise* serta menghilangkan data yang tidak konsisten maupun data yang tidak relevan. Data yang diperoleh dari suatu instansi maupun dari *website* pada umumnya terdapat isi yang tidak sempurna seperti adanya *missing value*, data yang tidak valid maupun data yang salah ketik. Selain itu biasanya terdapat atribut data yang tidak sesuai dengan kaidah *data mining*. Data yang tidak relevan sebaiknya dihilangkan (Nofriansyah, 2016).

2. ***Data Selection***

   Tahap *selection* dilakukan guna melakukan seleksi pada data dan atribut agar mendapatkan data yang sesuai dengan tujuan penelitian yang dilakukan. Setelah dilakukannya tahap seleksi ini, diharapkan hasil akurasi yang diperoleh akan lebih baik. Selain itu dengan dilakukannya seleksi data,

peneliti dapat mengetahui benar atribut apa saja yang paling berpengaruh dalam penentuan Diabetes Mellitus.

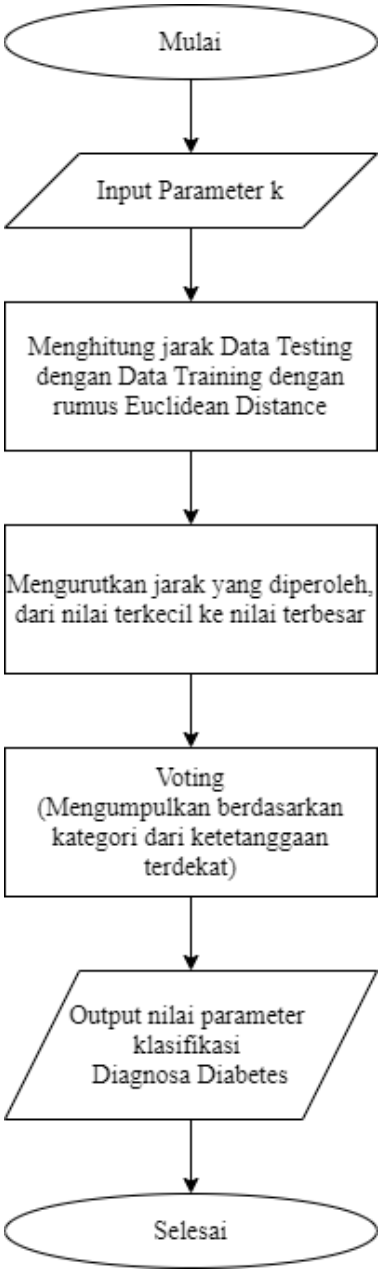
Pada tahap yang dilakukan pada penelitian kali ini data dilakukan sebanyak dua kali pemilihan data (*data selection*), yang pertama adalah seleksi data berdasarkan informasi dari P2PTM Kemenkes Republik Indonesia dan tahap selanjutnya adalah melakukan tahap pemilihan atribut berdasarkan hasil dari uji regresi logistik biner yang dilakukan. Metode uji regresi logistik biner merupakan metode yang paling tepat untuk menentukan atribut yang paling berpengaruh jika variabel *dependent* nya merupakan label kategorikal.

3.7.2. Data Mining

Mining dilakukan sebagai tahap utama dalam proses klasifikasi pada penelitian kali ini. Pada proses ini akan dilakukan pengujian terhadap data uji dan data latih guna menemukan model klasifikasi yang menghasilkan akurasi paling tinggi. Tahap *mining* pada penelitian kali ini akan dilakukan dengan menggunakan metode k-*Nearest Neighbor* (k-NN). Dalam proses pengolahan data peneliti akan dilakukan dengan melakukan perhitungan secara matematis dengan cara manual dan penghitungan menggunakan aplikasi *RapidMiner* sebagai alat bantu klasifikasi. Dengan menggunakan metode k-NN akan didapatkan *euclidean distance* yang nantinya akan diurutkan berdasarkan jarak ketetangga-an terdekat untuk mendapatkan model. Pada tahap ini akan didapatkan model klasifikasi dari data yang telah diolah berupa tabel *Confusion Matrix* yang akan menampilkan hasil pengujian yang dihasilkan oleh aplikasi dan akan disesuaikan dengan hasil pada data asli. Berikut tahapan *data mining* yang dilakukan pada penelitian kali ini.

3.7.3. Evaluasi

Proses evaluasi adalah proses penerjemahan pola-pola yang dihasilkan dari data mining untuk menguji atau mengevaluasi keakuratan dan *performance* dari metode yang digunakan. Pada proses ini menggunakan *cross validation* dengan 10ji. *k-fold cross validation* merupakan metode uji untuk mengevaluasi kinerja dari algoritma



Gambar 3. 2 Flowchart k-NN

BAB IV
HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Pada bab 4 ini akan dibahas mengenai hasil penelitian yang telah dilakukan, dimulai dari analisis deskriptif yang berhubungan dengan penjelasan mengenai data yang digunakan, hasil pengujian, dan pembahasan terhadap hasil yang didapatkan dari pengujian yang telah dilakukan.

4.1.1 Analisis Deskriptif

Pada sub bab ini akan dipaparkan secara rinci terkait data yang akan digunakan pada penelitian kali ini. Analisis deskriptif bertujuan untuk mengetahui karakteristik data yang akan diteliti, akan dibahas mengenai deskripsi untuk masing-masing atribut yang digunakan. Berikut merupakan atribut-atribut dari data awal yang didapatkan dari RST tk. II dr. Soepraoen Malang:

Tabel 4.1 Atribut Data RST tk. II dr. Soepraoen Malang

No	Atribut	Tipe	Keterangan
1	Nomor rekam medik	Rasio	Nomor rekam medik
2	Tempat tanggal lahir	Rasio	Tempat tanggal lahir
3	Tanggal masuk rumah sakit	Rasio	Tanggal pasien masuk rumah sakit
4	Nama	Rasio	Nama pasien
5	Usia	Rasio	Usia pasien saat dilakukan pemeriksaan
6	Keluhan	Rasio	Keluhan yang dirasakan pasien

No	Atribut	Tipe	Keterangan
7	HB	Rasio	Kadar hemoglobin dalam darah. Hemoglobin adalah protein pada sel darah merah
8	Leukosit	Rasio	Sel darah putih
9	LED (Laju Endap Darah)	Rasio	Pemeriksaan untuk mengukur berapa lama sel darah merah menggumpal atau mengendap pada alat pemeriksaan
10	Trombosit	Rasio	Sel darah yang berperan penting dalam proses pembekuan darah
11	PCV (*packed-cell volume*)	Rasio	Tes volume darah, salah satu pemeriksaan pada tes darah lengkap. Menganalisis dan mengetahui jumlah sel darah merah dalam volume darah
12	Tanggal pemeriksaan	Rasio	Tanggal pemeriksaan
13	Tekanan darah	Rasio	Gaya atau dorongan darah ke dinding arteri saat darah dipompa keluar dari jantung keseluruh tubuh (Palmer, 2007)
14	RR (*Respiratory rate*)	Rasio	Jumlah siklus pernafasan (inspirasi dan ekspirasi penuh) yang dihitung dalam waktu 1 menit atau 60 detik (Perry & Potter, 2005)
15	Nadi	Rasio	Pemeriksaan jumlah denyut jantung yang dihitung dalam waktu 1 menit atau 60 detik
16	Suhu	Rasio	Pemeriksaan suhu
17	SPO2 (*oximeter*)	Rasio	Pemeriksaan kadar saturasi oksigen

No	Atribut	Tipe	Keterangan
18	Berat Badan	Rasio	Berat badan pasien saat dilakukan pemeriksaan
19	Tinggi Badan	Rasio	Tinggi badan pasien saat dilakukan pemeriksaan
20	BMI	Rasio	*Body Mass Index* / Indeks Massa Tubuh
21	Ureum	Rasio	Kadar ureum dalam darah sebagai indikator fungsi ginjal
22	Kreatinin	Rasio	Kadar kreatinin dalam darah sebagai indikator fungsi ginjal
23	Uric Acid	Rasio	Kadar asam urat
24	Natrium	Rasio	Mineral penghasil ion bermuatan listrik
25	Kalium	Rasio	Mineral dalam tubuh yang mengendalikan fungsi sel saraf dan otot, terutama otot jantung
26	Klorida	Rasio	Suatu elektrolit yang memiliki peranan penting dalam menjaga keseimbangan cairan di dalam dan di luar sel-sel tubuh, serta mempertahankan volume darah normal, tekanan darah, dan pH cairan tubuh
27	SGOT (Serum Glutamic Oxaloacetic Transaminase)	Rasio	Enzim yang biasanya ditemukan pada hati (liver), jantung, otot, ginjal, hingga otak yang berfungsi untuk mencerna protein dalam tubuh
28	SGPT	Rasio	Enzim yang biasanya ditemukan pada hati (liver), jantung, otot, ginjal, hingga otak yang berfungsi

No	Atribut	Tipe	Keterangan
	(Serum Glutamic Pyruvic Transaminase)		untuk mencerna protein dalam tubuh
29	Bilirubin total	Rasio	Bentuk larut air yang dikeluarkan dari hati
30	Bilirubin direk	Rasio	
31	Duke	Rasio	Pemeriksaan *bleeding*, jari dilukai dengan lanset untuk diambil darahnya
32	Lee & white	Rasio	Mengukur waktu yang diperlukan oleh darah lengkap untuk membeku di dalam tabung
33	PH	Rasio	Pemeriksaan yang dilakukan guna melihat tingkat asam dan basa dalam cairan urine
34	BJ (Berat Jenis)	Rasio	Ukuran konsentrasi solut dalam urin. Berat jenis urin memberi informasi tentang kemampuan ginjal dalam mengonsentrasikan urin
35	Leukosit	Rasio	Salah satu pemeriksaan penunjang pada penyakit ginjal
36	Eritrosit	Rasio	Menunjukkan adanya infeksi bakteri
37	Epitel Squamous	Rasio	Salah satu parameter yang ada pada pemeriksaan urin untuk mengetahui adanya sel epitel squamous
38	Kolesterol Total	Rasio	Pemeriksaan kadar HDL dan LDL
39	LDL Kolesterol	Rasio	*Low-density lipoprotein* (kadar Kolesterol jahat)
40	Trigliserida	Rasio	Kadar Trigliserida dalam darah dan lemak

No	Atribut	Tipe	Keterangan
41	Gula darah sewaktu	Rasio	Kadar gula darah sewaktu
42	Gula darah puasa	Rasio	Kadar gula darah puasa
43	Gula darah 2 jam	Rasio	Kadar gula darah 2 jam
44	HBA1C	Rasio	Kadar hemoglobin terglikasi. Bentuk hemoglobin yang secara kimiawi terkait dengan gula
45	TSH	Rasio	Untuk mengukur konsentrasi thyroid-stimulating hormone (TSH) dalam darah
46	Diagnosa	Kategorikal	Diagnosa yang telah ditetapkan, diabetes atau tidak diabetes

Tabel 4.1 di atas menampilkan atribut dari seluruh data awal yang didapatkan dari RST tk II dr. Soepraoen. Seluruh data tersebut di atas akan dilakukan tahap *pre-processing* yakni penghapusan data *missing value.* Data tersebut akan dilakukan tiga tahapan utama, yaitu tahap *preprocessing*, *data mining*, dan *evaluation* seperti halnya yang telah dibahas pada bab 3. Berikut akan dijabarkan secara lebih rinci terkait tahapan penelitian yang dilakukan:

4.1.2 *Pre-processing*

Pada tahap ini terdapat dua proses yang dilakukan yaitu tahap *data selection* dan tahap *data cleaning.* Tahap *data selection* melakukan seleksi data serta atribut sebelum dilakukan *mining* dan tahap kedua adalah *data cleaning* dilakukan guna membersihkan data dari *missing value* agar tingkat akurasi hasil yang didapatkan lebih maksimal.

4.1.2.1 *Data Cleaning*

Data yang didapatkan dari pihak Rumah Sakit berupa data *hardfile* rekam medik pasien, data harus dilakukan pembersihan pada data yang didapatkan. Proses pembersihan yang dilakukan di antaranya adalah melakukan pemilihan data sesuai atribut yang telah ditetapkan guna mendapatkan data dengan atribut yang sesuai, melakukan proses pengelompokan data berdasarkan kelas (label) yaitu kelompok pasien diabetes dan non-diabetes, selain itu pada tahap ini juga dilakukan penghilangan data yang memiliki banyak *missing value*, dan yang terakhir pada tahap *cleaning* kali ini, dilakukan pengurangan jumlah atribut dari yang semula terdapat 46 atribut menjadi 10 atribut utama yang akan digunakan, yang terdiri dari 9 atribut *independent* dan 1 *output* (target) yang merupakan atribut *dependent* berupa Diagnosa. Pengurangan atribut dilakukan denngan cara mengeliminasi data yang tidak sesuai dengan faktor penyebab Diabetes Mellitus menurut P2PTM Kemenkes Republik Indonesia. Setelah dilakukannya pengurangan jumlah atribut berdasarkan faktor penyebab, dilakukan pula penentuan atribut yang paling berpengaruh dan pengurangan atribut yang tidak berpengaruh melalui tahap uji regresi logistik biner yang dilakukan menggunakan bantuan aplikasi SPSS. Sehingga didapatkan 10 atribut yang digunakan pada penelitian kali ini, dapat dilihat pada Tabel 4.3.

4.1.2.2 *Data Selection*

Pada tahap ini dilakukan proses seleksi untuk menentukan data dan atribut yang akan digunakan pada penelitian kali ini. Proses pemilihan atribut yang digunakan yaitu berdasarkan hal-hal yang mempengaruhi seseorang dapat

terkena penyakit Diabetes Mellitus. Akan dilakukan beberapa tahapan seleksi atribut pada penelitian kali ini. Pada tabel di bawah ini akan ditampilkan seluruh atribut awal yang didapatkan sebelum dilakukannya tahapan seleksi atribut, dan akan dilakukan tahap seleksi yang pertama.

1. Tahap *Selection* 1 (P2PTM Kemenkes RI dan WHO)

Pada tahap ini akan dilakukan proses seleksi atribut berdasarkan informasi terkait faktor penyebab Diabetes Mellitus menurut P2PTM Kemenkes Republik Indonesia dan WHO. Berikut akan ditampilkan tabel yang berisikan keputusan terhadap atribut yang akan digunakan atau dieliminasi.

Tabel 4.2 Tahap *Selection* 1

No	Atribut	Keterangan
1	Nomor rekam medik	Tidak
2	Tempat tanggal lahir	Tidak
3	Tanggal masuk rumah sakit	Tidak
4	Nama	Digunakan
5	Usia	Tidak
6	Keluhan	Tidak
7	Hemoglobin	Tidak
8	Lekosit	Tidak
9	LED	Tidak
10	Trombosit	Tidak
11	PCV	Tidak
12	Tanggal pemeriksaan	Tidak
13	Tekanan darah	Digunakan
14	RR	Digunakan
15	Nadi	Digunakan
16	Suhu	Digunakan
17	SPO2	Digunakan
18	Berat Badan	Digunakan

No	Atribut	Keterangan
19	Tinggi Badan	Digunakan
20	BMI	Digunakan
21	Ureum	Tidak
22	Kreatinin	Tidak
23	Uric Acid	Tidak
24	Natrium	Tidak
25	Kalium	Tidak
26	Klorida	Tidak
27	SGOT	Tidak
28	SGPT	Tidak
29	Bilirubin total	Tidak
30	Bilirubin direk	Tidak
31	Duke	Tidak
32	Lee & white	Tidak
33	PH	Tidak
34	BJ	Tidak
35	Lekosit	Tidak
36	Eritrosit	Tidak
37	Epitel Squamous	Tidak
38	Kolesterol Total	Digunakan
39	LDL Kolesterol	Digunakan
40	Trigliserida	Digunakan
41	Gula darah sewaktu	Digunakan
42	Gula darah puasa	Digunakan
43	Gula darah 2 jam	Digunakan
44	HBA1C	Digunakan
45	TSH	Tidak
46	Diagnosa	Digunakan (label)

Tabel di atas menampilkan atribut yang didapatkan setelah melalui tahap seleksi awal yang telah dilakukan, menunjukkan atribut-atribut yang digunakan dan tidak digunakan. Penetapan atribut di atas adalah berdasarkan faktor penyebab Diabetes Mellitus menurut P2PTM Kemenkes Republik Indonesia dan WHO. Selanjutnya akan dilakukan seleksi atribut dengan menggunakan uji regresi logistik biner, guna mendapatkan dan mengetahui atribut yang paling berpengaruh dan yang paling tidak berpengaruh.

2. Tahap *selection* 2 (Regresi Logistik Biner)

Pada tahap ini akan dilakukan proses seleksi atribut dengan dilakukannya uji regresi logistik biner. Tahap ini dilakukan guna mengetahui atribut-atribut yang paling berpengaruh serta mengetahui atribut yang dapat di eliminasi terhadap penyakit Diabetes Mellitus, guna mendapatkan hasil akurasi yang maksimal. Uji regresi logistik biner yang dilakukan pada penelitian kali ini dilakukan dengan dua cara, yakni penghitungan secara manual serta dilakukan juga proses penghitungan dengan menggunakan bantuan aplikasi SPSS. Berikut akan ditampilkan hasil uji yang telah dilakukan.

Tabel 4.3 *Variables not in the Equation*

Variables	***Score***	***df***	***Sig.***
Usia	31.294	1	0.000
Sistole	7.792	1	0.005
Diastole	0.190	1	0.663
RR	1.224	1	0.269
Nadi	2.398	1	0.122

Variables	*Score*	*df*	*Sig.*
Suhu	1.615	1	0.204
SPO2	0.773	1	0.379
BB	2.443	1	0.118
TB	5.485	1	0.019
BMI	26.642	1	0.000
Kolesterol	12.068	1	0.001
Ldl kolesterol	20.126	1	0.000
Trigliserida	20.560	1	0.000
Gula darah sewaktu	37.867	1	0.000
Gula darah puasa	16.103	1	0.000
Gula darah 2 jam	29.498	1	0.000
HBA1C	44.913	1	0.000

Dapat dilihat pada Tabel 4.3 di atas, ditunjukkan bahwa terdapat beberapa atribut yang memiliki tingkat signifikan >0,05 yang berarti bahwa atribut-atribut tersebut tidak berpengaruh secara signifikan. Sehingga atribut yang tidak berpengaruh signifikan dapat dilakukan eliminasi. Sehingga tersisa 10 (sepuluh) atribut yang akan digunakan dari data Diabetes Mellitus RST tk. II dr. Soepraoen Malang.

I. Uji Hipotesis

H_0 = Atribut tidak berpengaruh signifikan

H_1 = Atribut berpengaruh signifikan

II. Daerah Kritis

$\propto = 5\% = 0{,}05$

H_0 ditolak jika *Sig*. < 0,05

III. Statistik Uji

Tabel 4.4 Statistik Uji Data RST tk. II dr. Soepraoen Malang

Variabel	Sig		Kesimpulan
Sistole	Sig. 0,005	> 0,05	tidak berpengaruh signifikan
Diastole	Sig. 0,663	> 0,05	tidak berpengaruh signifikan
RR	Sig. 0,269	> 0,05	tidak berpengaruh signifikan
Nadi	Sig. 0,122	> 0,05	tidak berpengaruh signifikan
Suhu	Sig. 0,204	> 0,05	tidak berpengaruh signifikan
SPO2	Sig. 0,379	> 0,05	tidak berpengaruh signifikan
BB	Sig. 0,118	> 0,05	tidak berpengaruh signifikan
TB	Sig. 0,019	> 0,05	tidak berpengaruh signifikan
Usia	Sig. 0,000	< 0,05	berpengaruh signifikan
BMI	Sig. 0,000	< 0,05	berpengaruh signifikan
Kolesterol	Sig. 0,001	< 0,05	berpengaruh signifikan
LDL Kolesterol	Sig. 0,000	< 0,05	berpengaruh signifikan
Trigliserida	Sig. 0,000	< 0,05	berpengaruh signifikan
Gd Sewaktu	Sig. 0,000	< 0,05	berpengaruh signifikan
Gd Puasa	Sig. 0,000	< 0,05	berpengaruh signifikan
Gd 2 Jam	Sig. 0,000	< 0,05	berpengaruh signifikan
HBA1C	Sig. 0,000	< 0,05	berpengaruh signifikan

Berdasarkan hasil pengujian yang telah dilakukan, atribut yang tidak berpengaruh signifikan yang ditunjukkan dengan nilai sig. lebih dari 0,05 akan dilakukan proses eliminasi, sehingga atribut yang akan diolah hanyalah atribut yang berpengaruh signifikan setelah melalui uji regresi logistik biner. Atribut-atribut tersebut merupakan representasi dari faktor yang paling berpengaruh pada seseorang dapat terkena diabetes mellitus, sehingga masyarakat dapat melakukan tindakan preventif dengan menghindari

makanan yang dapat meningkatkan kadar gula darah dan kolesterol, serta dapat melaksanakan pola hidup sehat sejak dini.

IV. Model Regresi Logistik Biner

$$
\begin{aligned}
(\pi(x)) &= \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p \\
&= 0.296 + 0.073\,X_1 + 0.94\,X_2 + 0.458\,X_3 - 0.323 + \\
&\quad 0.458\,X_4 + 2.687\,X_5 - 148.720\,X_6 - 5.135\,X_7 + \\
&\quad 3.395\,X_8 + 16.631\,X_9 + 0.011\,X_{10} + 0.040\,X_{11} + \\
&\quad 0.044\,X_{12} + 0.028\,X_{13} + 0.30\,X_{14} + 0.18\,X_{15} + 3.070\,X_{16}
\end{aligned}
$$

Berdasarkan hasil uji yang telah dilakukan maka terdapat satu atau lebih variabel prediktor yang berpengaruh signifikan terhadap variabel respon. Dari Tabel 4.5 di atas dapat kita ketahui bahwa atribut usia, BMI, Kolesterol, LDL Kolesterol, Trigliserida, Gula Darah Sewaktu, Gula Darah Puasa, Gula Darah 2 Jam, HBA1C, berpengaruh signifikan. Sehingga atribut yang tidak berpengaruh signifikan dieliminasi. Didapatkan atribut-atribut berikut yang selanjutnya akan dilakukan tahap klasifikasi.

Berikut akan ditampilkan tabel yang berisi atribut yang digunakan dan tidak digunakan setelah dilakukan uji Regresi Logistik Biner yang telah dilakukan.

Tabel 4.5 Tahap *Selection* 2

No	Atribut	Keterangan
1	Sistole	Tidak
2	RR	Tidak
3	Nadi	Tidak
4	Suhu	Tidak
5	SPO2	Tidak

No	Atribut	Keterangan
6	Berat Badan	Tidak
7	Tinggi Badan	Tidak
8	Usia	Digunakan
9	BMI	Digunakan
10	Kolesterol Total	Digunakan
11	LDL Kolesterol	Digunakan
12	Trigliserida	Digunakan
13	Gula darah sewaktu	Digunakan
14	Gula darah puasa	Digunakan
15	Gula darah 2 jam	Digunakan
16	HBA1C	Digunakan
17	Diagnosa	Digunakan (label)

Setelah dilakukan tahap uji regresi logistik biner, maka didapatkan atribut-atribut yang digunakan serta atribut yang dapat dilakukan eliminasi. Sehingga didapatkan *list* atribut yang akan digunakan pada data Diabetes Mellitus dari RST tk. II dr. Soepraoen Malang sebagai berikut:

Tabel 4.6 Atribut Data RST tk. II dr. Soepraoen Malang

No	Atribut
1	Usia
2	BMI
3	Kolesterol Total
4	LDL Kolesterol
5	Trigliserida
6	Gula darah sewaktu
7	Gula darah puasa
8	Gula darah 2 jam
9	HBA1C
10	Diagnosa (label)

Data yang digunakan merupakan data sekunder berupa data Diabetes Mellitus yang berasal dari 2 (dua) sumber yang berbeda. Data pertama merupakan data Diabetes Mellitus yang didapatkan dari RST tk. II dr. Soepraoen Malang, data terdiri dari 9 *independent variable* dan 1 *dependent variable.* Atribut-atribut tersebut adalah Usia BMI, Kolesterol Total, LDL Kolesterol, Trigliserida, Gula darah sewaktu, Gula darah puasa, Gula darah 2 jam, HBA1C, dan Diagnosa. Data berjumlah 52 data yang terdiri dari 50% pasien dengan diagnosa Diabetes mellitus dan 50% sisa nya merupakan pasien dengan diagnosa penyakit selain Diabetes Mellitus. Data yang digunakan merupakan data pasien dengan usia 25 sampai dengan 60 tahun. Berikut merupakan tabulasi dari atribut-atribut yang digunakan berdasarkan data Diabetes Mellitus dari RST tk.II dr. Soepraoen Malang yang akan digunakan pada penelitian kali ini:

Tabel 4. 7 Tabulasi BMI terhadap Diabetes Mellitus Data RST tk. II dr. Soepraoen Malang

Diabetes Mellitus	**BMI**		**Jumlah**
	18.6-24.9	**25-29.9**	
	Normal	**Pra-Obesitas**	
Yes	**24**	**2**	**26**
	46,15%	3,85%	50%
No	**26**	**0**	**26**
	50%	0%	50%
Jumlah	**50**	**2**	**52**
	96,15%	3,85%	100%

Berdasarkan Tabel 4.7 di atas, dapat diketahui bahwa dari seluruh data pasien yang diambil, terdapat 96,15% yakni 50 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar BMI normal dengan rincian sebagai berikut. Terdapat 46,15% yakni 24 pasien ter diagnosa diabetes mellitus dengan kadar BMI

normal. Terdapat 50% yakni 26 pasien tidak ter diagnosa diabetes mellitus dengan kadar BMI normal.

Terdapat 3,85% yakni 2 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar BMI pra-obesitas dengan rincian sebagai berikut. Terdapat 3,85% yakni 2 pasien ter diagnosa diabetes dengan kadar BMI pra-obesitas. Terdapat 0% yakni tidak ada pasien tidak ter diagnosa diabetes.

Tabel 4. 8 Tabulasi Kolesterol terhadap Diabetes Mellitus RST tk.II dr.Soepraoen Malang

Diabetes Mellitus	**Kolesterol**			**Jumlah**
	< 200 mg/dl	**200-239 mg/dl**	**> 240 mg/dl**	
	Normal	**Batas Tinggi**	**Tinggi**	
Yes	**3**	**9**	**14**	**26**
	5,78%	17,30%	26,92%	50%
No	**24**	**2**	**0**	**26**
	46,15%	3,85%	0%	50%
Jumlah	**27**	**11**	**14**	**52**
	51,93%	21,15%	26,92%	100%

Berdasarkan Tabel 4.8 di atas, dapat diketahui bahwa dari seluruh data pasien yang diambil, terdapat 51,93% yakni 27 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar Kolesterol normal dengan rincian sebagai berikut. Terdapat 5,78% yakni 3 pasien ter diagnosa diabetes mellitus dengan kadar Kolesterol normal. Terdapat 46,15% yakni 24 pasien tidak ter diagnosa diabetes mellitus dengan kadar Kolesterol normal.

Terdapat 21,15% yakni 11 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar Kolesterol Batas Tinggi dengan rincian sebagai berikut. Terdapat 17,30% yakni 9 pasien ter diagnosa diabetes dengan kadar Kolesterol batas tinggi. Terdapat 3,85% yakni 2 pasien tidak ter diagnosa diabetes dengan kadar Kolesterol batas tinggi.

Terdapat 26,92% yakni 14 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar Kolesterol Tinggi dengan rincian sebagai berikut. Terdapat 26,92% yakni 14 pasien ter diagnosa diabetes dengan kadar Kolesterol tinggi. Terdapat 0% yakni tidak ada pasien non-diabetes dengan kadar Kolesterol tinggi.

Tabel 4.9 Tabulasi LDL Kolesterol terhadap Diabetes Mellitus

Diabetes Mellitus	**LDL Kolesterol**					**Jumlah**
	Optimal	**Hampir Optimal**	**Batas Tinggi**	**Tinggi**	**Sangat Tinggi**	
	>100 mg/dl	100 - 129 mg/dl	130 - 159 mg/dl	160 - 189 mg/dl	190 mg/dl	
Yes	**1**	**0**	**11**	**9**	**5**	**26**
	1,92%	0%	21,15%	17,31%	9,61%	50%
No	**1**	**21**	**4**	**0**	**0**	**26**
	1,92%	40,38%	7,70%	0%	0%	50%
Jumlah	**2**	**21**	**15**	**9**	**5**	**52**
	3,84%	40,38%	28,85%	17,31%	9,61%	100%

Berdasarkan Tabel 4.9 di atas, dapat diketahui bahwa dari seluruh data pasien yang diambil, terdapat 3,84% yakni 2 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar LDL Kolesterol optimal dengan rincian sebagai berikut. Terdapat 1,92% yakni 1 pasien ter diagnosa diabetes mellitus dengan kadar LDL Kolesterol optimal. Terdapat 1,92% yakni 1 pasien tidak ter diagnosa diabetes mellitus dengan kadar LDL Kolesterol optimal.

Terdapat 40,38% yakni 21 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar LDL Kolesterol hampir optimal dengan rincian sebagai berikut. Terdapat 0% yakni tidak ada pasien ter diagnosa diabetes mellitus dengan kadar LDL Kolesterol hampir optimal. Terdapat 40,38% yakni 21 pasien tidak ter diagnosa diabetes mellitus dengan kadar LDL Kolesterol hampir optimal.

Terdapat 28,85% yakni 15 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar LDL Kolesterol batas tinggi dengan rincian sebagai berikut. Terdapat 21,15% yakni 11 pasien ter diagnosa diabetes mellitus dengan kadar LDL Kolesterol batas tinggi. Terdapat 7,70% yakni 4 pasien tidak ter diagnosa diabetes mellitus dengan kadar LDL Kolesterol batas tinggi.

Terdapat 17,31% yakni 9 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar LDL Kolesterol tinggi dengan rincian sebagai berikut. Terdapat 17,31% yakni 9 pasien ter diagnosa diabetes mellitus dengan kadar LDL Kolesterol tinggi. Terdapat 0% yakni tidak ada pasien tidak ter diagnosa diabetes mellitus dengan kadar LDL Kolesterol tinggi.

Terdapat 9,61% yakni 5 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar LDL Kolesterol sangat tinggi dengan rincian sebagai berikut. Terdapat 9,61% yakni 5 pasien ter diagnosa diabetes mellitus dengan kadar LDL Kolesterol sangat tinggi. Terdapat 0% yakni tidak ada pasien tidak ter diagnosa diabetes mellitus dengan kadar LDL Kolesterol sangat tinggi.

Tabel 4.10 Tabulasi Trigliserida terhadap Diabetes Mellitus

Diabetes Mellitus	**Trigliserida**				**Jumlah**
	Normal	**Batas Tinggi**	**Tinggi**	**Sangat Tinggi**	
	< 150 mg/dl	150 - 199 mg/dl	200 - 499 mg/dl	> 500 mg/dl	
Yes	**1**	**8**	**16**	**0**	**25**
	1,92%	15,38%	30,77%	0%	48,07%
No	**25**	**2**	**0**	**0**	**27**
	48,08%	3,84%	0%	0%	51,92%
Jumlah	**26**	**10**	**16**	**0**	**52**
	50%	15,38%	30,77%	0%	100%

Berdasarkan Tabel 4.10 di atas, dapat diketahui bahwa dari seluruh data pasien yang diambil, terdapat 50% yakni 26 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar trigliserida normal dengan rincian sebagai berikut. Terdapat 1,92% yakni 1 pasien ter diagnosa diabetes mellitus dengan kadar trigliserida normal. Terdapat 48,08% yakni 25 pasien tidak ter diagnosa diabetes mellitus dengan kadar trigliserida normal.

Terdapat 15,38% yakni 10 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar trigliserida batas tinggi dengan rincian sebagai berikut. Terdapat 15,38% yakni 8 pasien ter diagnosa diabetes mellitus dengan kadar Trigliserida batas tinggi. Terdapat 3,84 % yakni 2 pasien tidak ter diagnosa diabetes mellitus dengan kadar Trigliserida batas tinggi.

Terdapat 30,77% yakni 16 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar trigliserida tinggi dengan rincian sebagai berikut. Terdapat 30,77% yakni 16 pasien ter diagnosa diabetes mellitus dengan kadar Trigliserida tinggi. Terdapat 0% yakni tidak ada pasien tidak ter diagnosa diabetes mellitus dengan kadar Trigliserida tinggi. Terdapat 0% yakni tidak pasien diabetes mellitus dan non-diabetes mellitus dengan kadar trigliserida sangat tinggi.

Tabel 4.11 Tabulasi Gula Darah Puasa terhadap Diabetes Mellitus

Diabetes Mellitus	**Gula Darah Puasa**			**Jumlah**
	< 90 mg/dl	**100-125 mg/dl**	**> 126 mg/dl**	
	Normal	**Pra-diabetes**	**Diabetes**	
Yes	**1**	**1**	**23**	**25**
	1,92%	1,92%	44,23%	48,08%
No	**11**	**16**	**0**	**27**
	21,15%	30,77%	0%	51,92%
Jumlah	**12**	**17**	**23**	**52**
	23,07%	32,69%	44,23%	100%

Berdasarkan Tabel 4.11 di atas, dapat diketahui bahwa dari seluruh data pasien yang diambil, terdapat 23,07% yakni 12 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar gula darah puasa normal dengan rincian sebagai berikut. Terdapat 1,92% yakni 1 pasien ter diagnosa diabetes mellitus dengan kadar gula darah puasa normal. Terdapat 21,15% yakni 11 pasien tidak ter diagnosa diabetes mellitus dengan kadar gula darah puasa normal.

Terdapat 32,69% yakni 17 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar gula darah puasa kategori pra-diabetes dengan rincian sebagai berikut. Terdapat 1,92% yakni 1 pasien ter diagnosa diabetes mellitus dengan kadar gula darah puasa kategori pra-diabetes. Terdapat 30,77% yakni 16 pasien non-diabetes mellitus dengan kadar gula darah puasa kategori pra-diabetes.

Terdapat 44,23% yakni 23 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar gula darah puasa kategori diabetes dengan rincian sebagai berikut. Terdapat 44,23% yakni 23 pasien ter diagnosa diabetes mellitus dengan kadar gula darah puasa kategori diabetes. Terdapat 0% yakni tidak ada pasien non-diabetes mellitus dengan kadar gula darah puasa kategori diabetes.

Tabel 4.12 Tabulasi Gula Darah 2 jam terhadap Diabetes Mellitus

Diabetes Mellitus	**Gula Darah 2 Jam**			**Jumlah**
	< 140 mg/dl	**140-199 mg/dl**	**> 200 mg/dl**	
	Normal	**Pra Diabetes**	**Diabetes**	
Yes	**0**	**13**	**15**	**28**
	0%	25%	28,85%	53,85%
No	**22**	**2**	**0**	**24**
	42,30%	3,85%	0%	46,15%
Jumlah	**22**	**15**	**15**	**52**
	42,30%	28,85%	28,85%	100%

Berdasarkan Tabel 4.12 di atas, dapat diketahui bahwa dari seluruh data pasien yang diambil, terdapat 42,30% yakni 22 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar gula darah 2 jam normal dengan rincian sebagai berikut. Terdapat 0% yakni tidak ada pasien ter diagnosa diabetes mellitus dengan kadar gula darah 2 jam normal. Terdapat 42,30% yakni 22 pasien tidak ter diagnosa diabetes mellitus dengan kadar gula darah 2 jam normal.

Terdapat 28,85% yakni 15 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar gula darah 2 jam kategori pra-diabetes dengan rincian sebagai berikut. Terdapat 25% yakni 13 pasien ter diagnosa diabetes mellitus dengan kadar gula darah 2 jam kategori pra-diabetes. Terdapat 3,85% yakni 2 pasien non-diabetes mellitus dengan kadar gula darah 2 jam kategori pra-diabetes.

Terdapat 28,85% yakni 15 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar gula darah 2 jam kategori diabetes dengan rincian sebagai berikut. Terdapat 28,85% yakni 15 pasien ter diagnosa diabetes mellitus dengan kadar gula darah 2 jam kategori diabetes. Terdapat 0% yakni tidak ada pasien non-diabetes mellitus dengan kadar gula darah 2 jam kategori diabetes.

Tabel 4.13 Tabulasi HBA1C terhadap diabetes mellitus

Diabetes Mellitus	**HBA1C**			**Jumlah**
	< 5,7%	**5,7% – 6,4%**	**> 6,5%**	
	Normal	**Pra-diabetes**	**Diabetes**	
Yes	**0**	**0**	**26**	**26**
	0%	0%	50%	50%
No	**19**	**7**	**0**	**26**
	36,54%	13,46%	0%	50%
Jumlah	**19**	**7**	**26**	**52**
	36,54%	13,46%	50%	100%

Berdasarkan Tabel 4.13 di atas, dapat diketahui bahwa dari seluruh data pasien yang diambil, terdapat 36,54% yakni 19 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar HBA1C normal dengan rincian sebagai berikut. Terdapat 0% yakni tidak ada pasien ter diagnosa diabetes mellitus dengan kadar HBA1C normal. Terdapat 36,54% yakni 19 pasien tidak ter diagnosa diabetes mellitus dengan kadar HBA1C normal.

Terdapat 13,46% yakni 7 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar HBA1C kategori pra-diabetes dengan rincian sebagai berikut. Terdapat 0% yakni tidak ada pasien ter diagnosa diabetes mellitus dengan kadar HBA1C kategori pra-diabetes. Terdapat 13,46% yakni 7 pasien tidak ter diagnosa diabetes mellitus dengan kadar HBA1C kategori pra-diabetes.

Terdapat 50% yakni 26 pasien diabetes mellitus dan non-diabetes mellitus dengan kadar HBA1C kategori diabetes dengan rincian sebagai berikut. Terdapat 50% yakni 26 pasien ter diagnosa diabetes mellitus dengan kadar HBA1C kategori diabetes. Terdapat 0% yakni tidak ada pasien tidak ter diagnosa diabetes mellitus dengan kadar HBA1C kategori diabetes.

4.1.3 K-Nearest Neighbor (k-NN)

4.1.3.1 Data RST tk.II dr. Soepraoen Malang Sebelum Uji Regresi Logistik Biner

Pada tahap ini penghitungan dilakukan dengan dua cara yaitu menggunakan perhitungan manual secara matematis serta penghitungan dengan menggunakan bantuan aplikasi Excel. Rumus yang digunakan untuk mencari euclidean distance adalah seperti rumus (2.1). Berikut adalah perhitungan manual secara matematis yang dilakukan untuk mengetahui jarak antara data *training* dan data *testing* yang

digunakan pada penelitian kali ini. Diambil salah satu data untuk dilakukan perhitungan *euclidean distance* secara matematis. Data yang diambil merupakan data acak dari data diabetes mellitus RST tk. II dr. Soepraoen Malang yang digunakan pada penelitian kali ini. Berikut merupakan perhitungan untuk x_{1i} hingga x_{10i}

$$d(x_i, x_j) = \sqrt{\sum_{i,j=1}^{n} (x_i - x_j)^2}$$

$$d(x_i, x_j) = \sqrt{\begin{pmatrix} (54-40)^2 + (23{,}438 - 22{,}567)^2 + (223 - 150)^2 + \\ (152-152)^2 + (191-134)^2 + (265-200)^2 + \\ (165-180)^2 + (185-175)^2 + (11{,}9-4{,}4)^2 \end{pmatrix}}$$

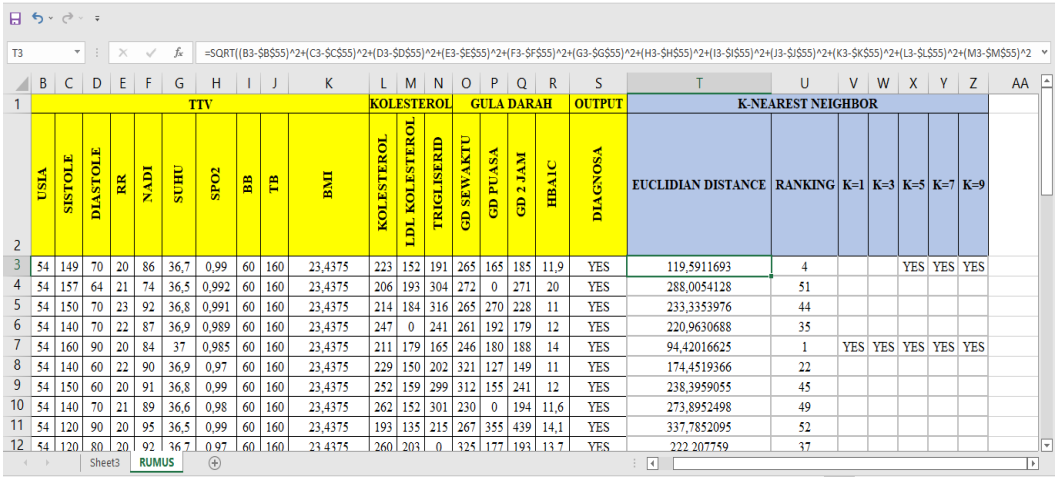
$$d(x_i, x_j) = \sqrt{14^2 + 0{,}871^2 + 73^2 + 0^2 + 57^2 + 65^2 + (-15)^2 + 10^2 + 7{,}5^2}$$

$$d(x_i, x_j) = \sqrt{\begin{matrix} 196 + 0{,}759 + 5329 + 0 + 3349 + 4225 + 225 \\ + 100 + 56{,}25 \end{matrix}}$$

$$d(x_i, x_j) = \sqrt{13.380{,}937}$$

$$d(x_i, x_j) = 115{,}676$$

Perhitungan jarak *euclidian* dilakukan dengan menggunakan bantuan aplikasi Excel, dengan rumus sebagai berikut =SQRT((B3-B55)^2+(C3-C55)^2+(D3-D55)^2+(E3-E55)^2+(F3-F55)^2+(G3-G55)^2+(H3-H55)^2+(I3-I55)^2+(J3-J55)^2+(K3-K55)^2+(L3-L55)^2+(M3-M55)^2+(N3-N55)^2+(O3-O55)^2+(P3-P55)^2+(Q3-Q55)^2+(R3-R55)^2). Berikut akan ditampilkan proses penghitungan *euclidean distance* menggunakan bantuan aplikasi Ms. Excel untuk mendapatkan nilai *euclidean distance* pada data lainnya.



=SQRT((B3-B55)^2+(C3-C55)^2+(D3-D55)^2+(E3-E55)^2+(F3-F55)^2+(G3-G55)^2+(H3-H55)^2+(I3-I55)^2+(J3-J55)^2+(K3-K55)^2+(L3-L55)^2+(M3-M55)^2

	TTV									KOLESTEROL			GULA DARAH				OUTPUT	K-NEAREST NEIGHBOR						
	USIA	SISTOLE	DIASTOLE	RR	NADI	SUHU	SPO2	BB	TB	BMI	KOLESTEROL	LDL KOLESTEROL	TRIGLISERID	GD SEWAKTU	GD PUASA	GD 2 JAM	HBA1C	DIAGNOSA	EUCLIDIAN DISTANCE	RANKING	K=1	K=3	K=5	K=7	K=9
3	54	149	70	20	86	36,7	0,99	60	160	23,4375	223	152	191	265	165	185	11,9	YES	119,5911693	4			YES	YES	YES
4	54	157	64	21	74	36,5	0,992	60	160	23,4375	206	193	304	272	0	271	20	YES	288,0054128	51					
5	54	150	70	23	92	36,8	0,991	60	160	23,4375	214	184	316	265	270	228	11	YES	233,3353976	44					
6	54	140	70	22	87	36,9	0,989	60	160	23,4375	247	0	241	261	192	179	12	YES	220,9630688	35					
7	54	160	90	20	84	37	0,985	60	160	23,4375	211	179	165	246	180	188	14	YES	94,42016625	1	YES	YES	YES	YES	YES
8	54	140	60	22	90	36,9	0,97	60	160	23,4375	229	150	202	321	127	149	11	YES	174,4519366	22					
9	54	150	60	20	91	36,8	0,99	60	160	23,4375	252	159	299	312	155	241	12	YES	238,3959055	45					
10	54	140	70	21	89	36,6	0,98	60	160	23,4375	262	152	301	230	0	194	11,6	YES	273,8952498	49					
11	54	120	90	20	95	36,5	0,99	60	160	23,4375	193	135	215	267	355	439	14,1	YES	337,7852095	52					
12	54	120	80	20	92	36,7	0,97	60	160	23,4375	260	203	0	325	177	193	13,7	YES	222,207759	37					

Sheet3 RUMUS

Gambar 4.1 *Euclidean Distance* dengan Excel (sebelum)

Gambar di atas menunjukkan proses penghitungan jarak Euclidean yang dilakukan dengan menggunakan bantuan aplikasi Ms. Excel pada data Diabetes Mellitus RST tk. II dr. Soepraoen Malang dengan jumlah 17 atribut awal. Tabel berikut ini merupakan tampilan keseluruhan data yang telah dilakukan penghitungan *Euclidean distance.*

	USIA	SISTOLE	DIASTOLE	RR	NADI	SUHU	SPO2	BB	TB	BMI	KOLESTEROL	LDL KOLESTEROL	TRIGLISERIDA	GD SEWAKTU	GD PUASA	GD 2 JAM	HBA1C	DIAGNOSA	EUCLIDIAN DISTANCE
1	54	149	70	20	86	36,7	0,99	60	160	23,438	223	152	191	265	165	185	11,9	YES	0,000
2	54	160	90	20	84	37	0,985	60	160	23,438	211	179	165	246	180	188	14	YES	51,706
3	54	160	90	22	87	36,6	0,985	60	160	23,438	255	154	181	304	192	179	12,9	YES	62,777
4	54	140	60	22	90	36,9	0,97	60	160	23,438	229	150	202	321	127	149	11	YES	78,986
5	49	130	90	26	88	36,5	0,99	75	175	24,490	287	191	181	258	180	210	13,2	YES	88,853
6	54	150	60	23	89	36,5	0,99	60	160	23,438	217	153	259	311	192	210	11,5	YES	90,831
7	54	140	70	21	93	36,6	0,993	60	160	23,438	254	173	271	217	177	193	12	YES	102,201
8	57	130	80	23	93	36,6	0,98	65	168	23,030	225	172	270	315	196	184	11	YES	103,562
9	54	130	80	22	90	36,8	0,985	60	160	23,438	207	135	195	300	270	228	9,8	YES	123,067
10	54	158	62	19	88	37,1	0,985	60	160	23,438	246	175	291	230	200	227	13	YES	124,191
11	54	150	60	20	91	36,8	0,99	60	160	23,438	252	159	299	312	155	241	12	YES	134,629
12	57	200	80	22	90	36,7	0,99	65	168	23,030	284	174	292	268	230	220	11,8	YES	150,613
13	49	132	94	20	80	36,5	0,98	80	175	26,122	248	181	183	338	210	300	11	YES	153,844
14	54	140	70	22	87	36,9	0,989	60	160	23,438	247	0	241	261	192	179	12	YES	164,460

	USIA	SISTOLE	DIASTOLE	RR	NADI	SUHU	SPO2	BB	TB	BMI	KOLESTEROL	LDL KOLESTEROL	TRIGLISERIDA	GD SEWAKTU	GD PUASA	GD 2 JAM	HBA1C	DIAGNOSA	EUCLIDIAN DISTANCE
15	54	150	70	23	92	36,8	0,991	60	160	23,438	214	184	316	265	270	228	11	YES	172,194
16	49	134	85	20	88	36,5	0,96	80	175	26,122	242	152	266	289	250	310	13	YES	174,713
17	47	130	80	20	90	36,9	0,98	65	170	22,491	176	114	123	140	111	129	4	NO	175,124
18	29	120	70	20	80	36,5	0,98	56	170	19,377	179	130	139	127	117	133	5,1	NO	175,661
19	47	130	90	20	97	36,2	0,993	65	170	22,491	187	110	132	150	88	111	6	NO	179,619
20	47	120	70	22	90	37,7	0,993	50	157	20,285	193	105	135	143	82	118	5,9	NO	183,229
21	35	125	80	21	88	36,6	0,99	59	169	20,658	167	119	131	139	81	140	5,9	NO	184,255
22	30	160	80	21	85	36,7	0,98	60	165	22,039	197	111	137	112	120	109	4,6	NO	193,293
23	32	120	70	20	92	37,8	0,99	63	167	22,590	165	121	132	115	105	138	4,7	NO	193,876
24	54	140	80	20	86	36,8	0,99	60	160	23,438	254	133	0	265	170	200	14,2	YES	195,549
25	57	170	80	22	70	36,4	0,97	65	168	23,030	229	200	163	236	0	264	10,4	YES	195,787
26	29	130	70	25	70	36,5	0,98	60	170	20,761	169	103	133	118	86	123	6,1	NO	204,409
27	30	156	101	20	87	36,6	0,99	60	165	22,039	170	130	131	105	124	107	4,5	NO	204,760
28	54	140	70	21	89	36,6	0,98	60	160	23,438	262	152	301	230	0	194	11,6	YES	205,531

	USIA	SISTOLE	DIASTOLE	RR	NADI	SUHU	SPO2	BB	TB	BMI	KOLESTEROL	LDL KOLESTEROL	TRIGLISERIDA	GD SEWAKTU	GD PUASA	GD 2 JAM	HBA1C	DIAGNOSA	EUCLIDIAN DISTANCE
29	47	120	70	20	87	36,9	0,99	65	170	22,491	189	134	124	90	101	131	3	NO	211,429
30	54	146	85	22	90	36,7	0,993	60	160	23,438	284	178	258	300	277	334	10,2	YES	212,398
31	54	120	80	20	92	36,7	0,97	60	160	23,438	260	203	0	325	177	193	13,7	YES	212,695
32	29	130	70	23	79	36,6	0,99	56	157	22,719	183	117	131	89	92	139	4,9	NO	214,393
33	47	120	80	20	74	36,5	0,99	65	170	22,491	167	103	124	110	89	107	5,2	NO	217,292
34	26	120	70	21	75	36,7	0,99	65	170	22,491	156	122	131	92	102	130	5,3	NO	218,697
35	29	80	40	19	88	40	0,98	56	157	22,719	151	105	132	106	98	140	4,8	NO	221,465
36	47	130	90	21	89	36,6	0,98	65	170	22,491	201	102	137	95	82	104	6	NO	221,880
37	54	157	64	21	74	36,5	0,992	60	160	23,438	206	193	304	272	0	271	20	YES	222,979
38	29	80	70	20	90	38	0,98	56	157	22,719	151	132	120	89	118	130	5,2	NO	228,712
39	57	146	85	21	95	36,9	0,99	65	168	23,030	0	140	284	248	125	165	13,7	YES	247,444
40	31	120	80	22	85	36,6	0,99	60	165	22,039	102	118	134	87	84	131	4,7	NO	248,517
41	47	120	70	20	91	36,8	0,99	65	170	22,491	200	0	131	97	103	134	5,1	NO	250,976
42	54	120	80	21	93	36,9	0,99	60	160	23,438	0	183	171	319	195	271	12,1	YES	251,581

	USIA	SISTOLE	DIASTOLE	RR	NADI	SUHU	SPO2	BB	TB	BMI	KOLESTEROL	LDL KOLESTEROL	TRIGLISERIDA	GD SEWAKTU	GD PUASA	GD 2 JAM	HBA1C	DIAGNOSA	EUCLIDIAN DISTANCE
43	30	159	86	20	90	36,5	0,985	60	165	22,039	166	125	0	127	101	104	4,8	NO	266,806
44	30	152	86	20	90	36,6	0,985	60	165	22,039	166	105	126	111	102	0	5,5	NO	269,292
45	47	130	80	22	92	36,8	0,99	65	170	22,491	175	119	121	105	89	0	5,9	NO	273,141
46	54	130	80	20	86	36,5	0,98	60	160	23,438	243	202	251	0	165	185	11,9	YES	277,824
47	58	140	70	20	80	36,3	0,993	56	165	20,569	0	104	139	108	117	128	5,7	NO	291,822
48	30	160	100	22	92	36,6	0,985	60	165	22,039	185	123	120	0	121	117	4	NO	292,980
49	47	130	90	21	84	37	0,99	65	170	22,491	159	110	130	0	82	132	4,9	NO	300,822
50	30	148	98	22	80	36,6	0,97	60	165	22,039	0	107	128	102	116	110	4,1	NO	302,996
51	45	100	80	20	65	36,5	0,99	60	170	20,761	0	106	130	96	117	104	3,7	NO	310,144
52	54	120	90	20	95	36,5	0,99	60	160	23,438	193	135	215	267	355	439	14,1	YES	322,043

Setelah didapatkan *euclidean distance* yang dibutuhkan, lalu diurutkan berdasarkan ketetanggan terdekat dan akan diambil 9 tetangga terdekat yang akan ditampilkan pada tabel di bawah ini:

Tabel 4.14 *Euclidean Distance*

Nomor	Diagnosa	Euclidian Distance	Ranking	K=1	K=3	K=5	K=7	K=9
1	Yes	0,000	1	Yes				
2	Yes	51,706	2					
3	Yes	62,777	3	Yes	Yes			
4	Yes	78,986	4					
5	Yes	88,853	5	Yes	Yes	Yes		
6	Yes	90,831	6					
7	Yes	102,201	7	Yes	Yes	Yes	Yes	
8	Yes	103,562	8					
9	Yes	123,067	9	Yes	Yes	Yes	Yes	Yes
10	Yes	124,191	10					
Data baru				Yes	Yes	Yes	Yes	Yes

Dari tabel di atas dapat dilihat bahwa mayoritas *output* yang dihasilkan dari 9 ketetanggaan terdekat adalah “yes” sehingga data baru yang dimasukkan termasuk ke dalam kategori diabetes mellitus.

4.1.3.2 Data RST tk.II dr. Soepraoen Malang Setelah Uji Regresi Logistik Biner

Pada tahap ini penghitungan dilakukan dengan dua cara yaitu menggunakan perhitungan manual secara matematis serta penghitungan dengan menggunakan

bantuan aplikasi Excel. Rumus yang digunakan untuk mencari euclidean distance adalah seperti rumus (2.1). Berikut adalah perhitungan manual secara matematis yang dilakukan untuk mengetahui jarak antara data *training* dan data *testing* yang digunakan pada penelitian kali ini. Diambil salah satu data untuk dilakukan perhitungan *euclidean distance* secara matematis. Data yang diambil merupakan data acak dari data diabetes mellitus RST tk. II dr. Soepraoen Malang yang digunakan pada penelitian kali ini.

$$d(x_i, x_j) = \sqrt{\sum_{i,j=1}^{n} (x_i - x_j)^2}$$

$$d(x_i, x_j) = \sqrt{\begin{matrix}(54-54)^2 + (23{,}438 - 23{,}438)^2 + (223 - 246)^2 \\ +(152-175)^2 + (191-291)^2 + (265-230)^2 + (165-200)^2 \\ + (185-227)^2 + (11{,}9-13)^2\end{matrix}}$$

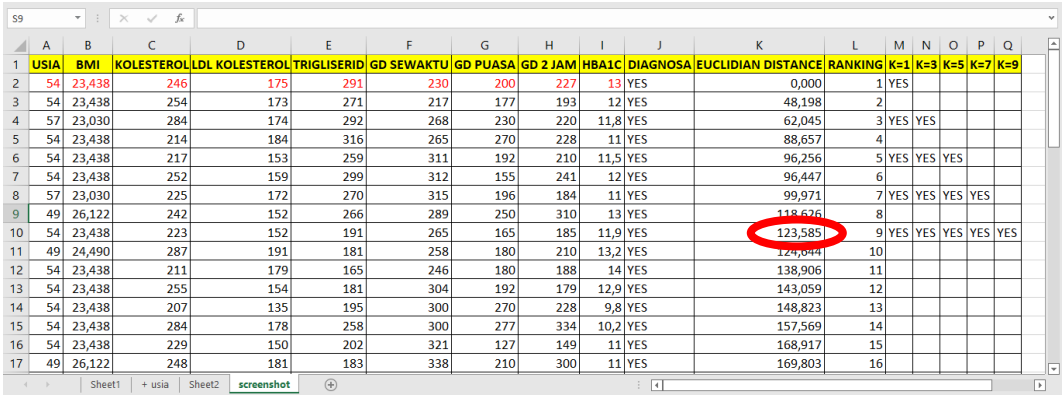
$$d(x_i, x_j) = \sqrt{0^2 + 0^2 + (-23)^2 + (-23)^2 + (-100)^2 + 35^2 + (-35)^2 + (-42)^2 + (-1{,}1)^2}$$

$$d(x_i, x_j) = \sqrt{0 + 0 + 529 + 529 + 10000 + 1225 + 1225 + 1764 + 1{,}21}$$

$$d(x_i, x_j) = \sqrt{15.273{,}21}$$

$$d(x_i, x_j) = 123{,}585$$

Setelah dilakukan perhitungan secara matematis, dilakukan pula perhitungan jarak Euclidean menggunakan bantuan aplikasi Ms. Excel guna mendapatkan nilai jarak antara data *training* dan *testing* lainnya dengan menggunakan rumus sebagai berikut =SQRT((A3-A55)^2+(B3-B55)^2+(C3-C55)^2+(D3-D55)^2+(E3-E55)^2+(F3-F55)^2+(G3-G55)^2+(H3-H55)^2). Lalu *euclidean distance* yang telah didapatkan di urutkan sesuai jarak, dimulai dari jarak ketetanggaan terdekat. Di bawah ini merupakan tampilan proses perhitungan yang dilakukan dengan menggunakan aplikasi Excel dengan jarak ketetanggaan k = 1, 3, 5, 7, 9.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	USIA	BMI	KOLESTEROL	LDL KOLESTEROL	TRIGLISERID	GD SEWAKTU	GD PUASA	GD 2 JAM	HBA1C	DIAGNOSA	EUCLIDIAN DISTANCE	RANKING	K=1	K=3	K=5	K=7	K=9
2	54	23,438	246	175	291	230	200	227	13	YES	0,000	1	YES				
3	54	23,438	254	173	271	217	177	193	12	YES	48,198	2					
4	57	23,030	284	174	292	268	230	220	11,8	YES	62,045	3	YES	YES			
5	54	23,438	214	184	316	265	270	228	11	YES	88,657	4					
6	54	23,438	217	153	259	311	192	210	11,5	YES	96,256	5	YES	YES	YES		
7	54	23,438	252	159	299	312	155	241	12	YES	96,447	6					
8	57	23,030	225	172	270	315	196	184	11	YES	99,971	7	YES	YES	YES	YES	
9	49	26,122	242	152	266	289	250	310	13	YES	118,626	8					
10	54	23,438	223	152	191	265	165	185	11,9	YES	123,585	9	YES	YES	YES	YES	YES
11	49	24,490	287	191	181	258	180	210	13,2	YES	124,644	10					
12	54	23,438	211	179	165	246	180	188	14	YES	138,906	11					
13	54	23,438	255	154	181	304	192	179	12,9	YES	143,059	12					
14	54	23,438	207	135	195	300	270	228	9,8	YES	148,823	13					
15	54	23,438	284	178	258	300	277	334	10,2	YES	157,569	14					
16	54	23,438	229	150	202	321	127	149	11	YES	168,917	15					
17	49	26,122	248	181	183	338	210	300	11	YES	169,803	16					

Gambar 4.2 *Euclidean Distance* dengan Excel (setelah)

Gambar di atas menunjukkan proses penghitungan jarak euclidean yang dilakukan dengan menggunakan bantuan aplikasi Ms. Excel pada data Diabetes Mellitus RST tk. II dr. Soepraoen Malang dengan jumlah 10 atribut. Hasil yang diperoleh dari perhitungan melalui aplikasi MS. Excel sesuai dengan perhitungan matematis yang telah dilakukan secara manual. Setelah didapatkan *euclidean distance* yang dibutuhkan, lalu diurutkan berdasarkan ketetanggan terdekat yang akan ditampilkan pada tabel di bawah ini:

Tabel 4.15 *Euclidean D*istance Data RST tk. II dr. Soepraoen Malang

USIA	BMI	KOLESTEROL	LDL KOLESTEROL	TRIGLISERIDA	GD SEWAKTU	GD PUASA	GD 2 JAM	HBA1C	DIAGNOSA	EUCLIDIAN DISTANCE	RANKING	K=1	K=3	K=5	K=7	K=9
54	23,438	246	175	291	230	200	227	13	YES	0,000	1	YES				
54	23,438	254	173	271	217	177	193	12	YES	48,198	2					
57	23,030	284	174	292	268	230	220	11,8	YES	62,045	3	YES	YES			
54	23,438	214	184	316	265	270	228	11	YES	88,657	4					
54	23,438	217	153	259	311	192	210	11,5	YES	96,256	5	YES	YES	YES		
54	23,438	252	159	299	312	155	241	12	YES	96,447	6					
57	23,030	225	172	270	315	196	184	11	YES	99,971	7	YES	YES	YES	YES	
49	26,122	242	152	266	289	250	310	13	YES	118,626	8					

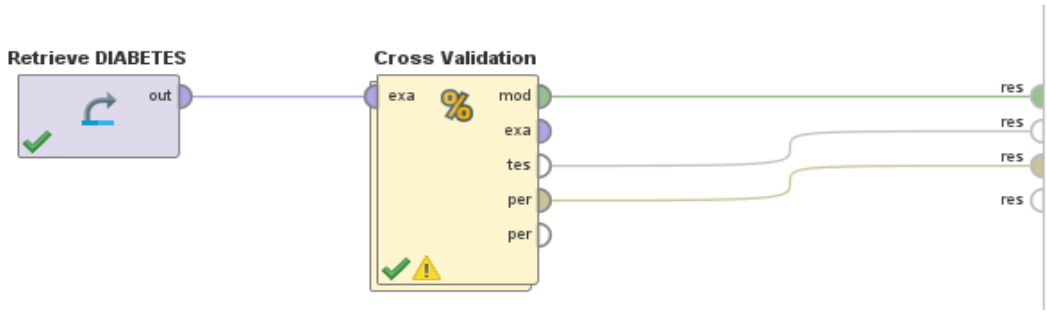
USIA	BMI	KOLESTEROL	LDL KOLESTEROL	TRIGLISERIDA	GD SEWAKTU	GD PUASA	GD 2 JAM	HBA1C	DIAGNOSA	EUCLIDIAN DISTANCE	RANKING	K=1	K=3	K=5	K=7	K=9
54	23,438	223	152	191	265	165	185	11,9	YES	**123,585**	9	YES	YES	YES	YES	YES
49	24,490	287	191	181	258	180	210	13,2	YES	124,644	10					
54	23,438	211	179	165	246	180	188	14	YES	138,906	11					
54	23,438	255	154	181	304	192	179	12,9	YES	143,059	12					
54	23,438	207	135	195	300	270	228	9,8	YES	148,823	13					
54	23,438	284	178	258	300	277	334	10,2	YES	157,569	14					
54	23,438	229	150	202	321	127	149	11	YES	168,917	15					
49	26,122	248	181	183	338	210	300	11	YES	169,803	16					
54	23,438	247	0	241	261	192	179	12	YES	190,935	17					
54	23,438	262	152	301	230	0	194	11,6	YES	204,880	18					
54	23,438	206	193	304	272	0	271	20	YES	214,107	19					
29	19,377	179	130	139	127	117	133	5,1	NO	238,025	20					
54	23,438	243	202	251	0	165	185	11,9	YES	241,305	21					
57	23,030	229	200	163	236	0	264	10,4	YES	242,320	22					
47	22,491	176	114	123	140	111	129	4	NO	250,202	23					
30	22,039	197	111	137	112	120	109	4,6	NO	255,166	24					
35	20,658	167	119	131	139	81	140	5,9	NO	255,748	25					
32	22,590	165	121	132	115	105	138	4,7	NO	255,896	26					
47	22,491	187	110	132	150	88	111	6	NO	255,902	27					
47	20,285	193	105	135	143	82	118	5,9	NO	255,985	28					
30	22,039	170	130	131	105	124	107	4,5	NO	264,296	29					
47	22,491	189	134	124	90	101	131	3	NO	267,555	30					
57	23,030	0	140	284	248	125	165	13,7	YES	267,568	31					
29	22,719	183	117	131	89	92	139	4,9	NO	270,024	32					
29	20,761	169	103	133	118	86	123	6,1	NO	270,394	33					
29	22,719	151	105	132	106	98	140	4,8	NO	270,643	34					

USIA	BMI	KOLESTEROL	LDL KOLESTEROL	TRIGLISERIDA	GD SEWAKTU	GD PUASA	GD 2 JAM	HBA1C	DIAGNOSA	EUCLIDIAN DISTANCE	RANKING	K=1	K=3	K=5	K=7	K=9
26	22,491	156	122	131	92	102	130	5,3	NO	274,609	35					
29	22,719	151	132	120	89	118	130	5,2	NO	277,156	36					
47	22,491	201	102	137	95	82	104	6	NO	280,084	37					
54	23,438	193	135	215	267	355	439	14,1	YES	283,768	38					
47	22,491	167	103	124	110	89	107	5,2	NO	283,806	39					
54	23,438	0	183	171	319	195	271	12,1	YES	291,312	40					
54	23,438	254	133	0	265	170	200	14,2	YES	298,939	41					
31	22,039	102	118	134	87	84	131	4,7	NO	303,899	42					
47	22,491	200	0	131	97	103	134	5,1	NO	306,921	43					
54	23,438	260	203	0	325	177	193	13,7	YES	310,438	44					
30	22,039	185	123	120	0	121	117	4	NO	327,972	45					
47	22,491	159	110	130	0	82	132	4,9	NO	337,164	46					
30	22,039	166	105	126	111	102	0	5,5	NO	338,309	47					
47	22,491	175	119	121	105	89	0	5,9	NO	341,544	48					
58	20,569	0	104	139	108	117	128	5,7	NO	346,861	49					
30	22,039	0	107	128	102	116	110	4,1	NO	359,854	50					
30	22,039	166	125	0	127	101	104	4,8	NO	360,229	51					
45	20,761	0	106	130	96	117	104	3,7	NO	362,418	52					
54	23,438	246	175	291	230	200	227	13	DATA BARU	0,000		YES	YES	YES	YES	YES

Hasil perhitungan jarak Euclid dengan data training sejumlah 51 data diurutkan mulai dari terkecil hingga terbesar. Selanjutnya dilihat mayoritas klasifikasi yang muncul dari perhitungan jarak pertama. Dari tabel di atas dapat dilihat bahwa dari 9 ketetanggaan terdekat, mayoritas *output* yang diperoleh adalah

"YES" sehingga data baru yang dimasukkan sebagai data *testing* adalah ter klasifikasi "YES" berdasarkan model yang terbentuk dari metode *k-Nearest Neighbor* dengan *k*=9, yang berarti data baru yang dimasukkan termasuk dalam kategori diabetes mellitus. Hal ini sesuai dengan diagnosa yang telah ditegakkan oleh dokter yang memeriksa, karena data baru yang digunakan sebagai data testing merupakan data telah mendapatkan diagnosa dari dokter. Sehingga perhitungan yang dilakukan berhasil. Hasil *output* (diagnosa) yang diperoleh melalui perhitungan *k-Nearest Neighbor* pada penelitian ini sesuai dengan diagnosa yang diberikan oleh dokter.

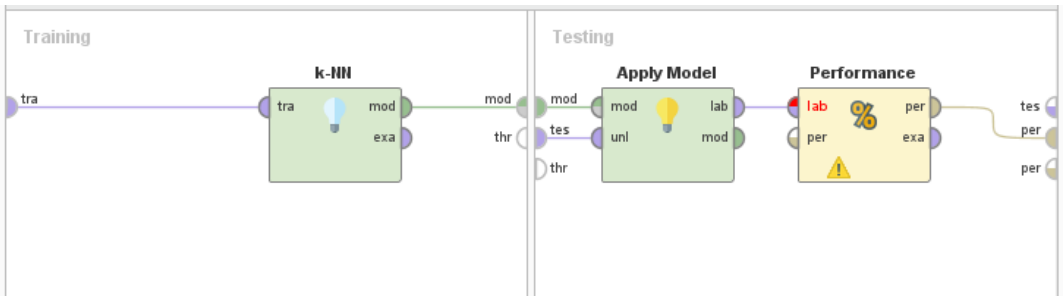
Akan dilakukan perhitungan menggunakan aplikasi *RapidMiner* guna meningkatkan ketepatan pada perhitungan yang telah dilakukan. Pada tahap ini dilakukan proses klasifikasi dengan menggunakan metode yang telah dipilih yaitu metode *k-Nearest Neighbor* (k-NN). Pada tahap ini, dilakukan proses klasifikasi dengan metode akan dilakukan dengan bantuan aplikasi *RapidMiner*. *Output* yang dihasilkan berupa tabel *confusion matrix* yang dapat dilihat pada Tabel 4.15



Gambar 4.3 Tahap 1 Uji Validasi

Pada Gambar 4.2 ditunjukkan tahap awal uji validasi menggunakan aplikasi *RapidMiner*. Pada tahap ini dilakukan *import* data diabetes RST tk. II dr. Soepraoen Malang dalam format .CSV, setelah proses import selesai, maka data di *drag* ke bagian halaman *process*, lalu dimasukkan proses *cross validation*, lalu hubungkan.

Pada bagian cross validation lakukan *double click*, maka akan muncul halaman berikutnya. Dapat dilihat pada Gambar 4.4 di bawah ini.



Gambar 4.4 *Cross Validation*

Pada halaman yang ditunjukkan pada Gambar 4.4 akan muncul bagian *training* dan *testing*, maka masukkan metode k-NN pada bagian *training* lalu hubungkan ke bagian *testing* dan masukkan *Apply Model* dan *Performance* pada bagian *testing*. Maka akan menampilkan hasil *accucarcy*, *precision*, *recall*, *AUC (optimistic)*, *dan AUC (pessimistic)*. Di bawah ini merupakan tampilan hasil yang diperoleh dari pengolahan data Diabetes Mellitus RST tk II dr. Soepraoen Malang dengan menggunakan aplikasi *RapidMiner* dengan menggunakan metode k-NN. Kedua data diolah dengan menggunakan metode k-NN dengan k=9. Akan ditunjukkan hasil perhitungan dengan menggunakan bantuan aplikasi *RapidMiner* dari dua data yang berbeda, yaitu data yang didapatkan dari RST tk. II dr. Soepraoen Malang.

4.1.4 Evaluasi

4.1.4.1 Data RST tk.II dr.Soepraoen Malang sebelum Uji Regresi Logistik Biner

Pada sub bab ini akan dilakukan pengolahan data pertama yang merupakan data Diabetes Mellitus yang didapat dari RST tk. II dr. Soepraoen Malang. Di bawah ini merupakan tampilan hasil klasifikasi yang diperoleh dari pengolahan data

Diabetes Mellitus yang didapatkan dari data diabetes mellitus RST sebelum dilakukannya uji regresi logistik biner dengan menggunakan aplikasi *RapidMiner* dengan menggunakan metode k-NN yang akan menjadi acuan untuk melakukan tahapan evaluasi. Kedua data diolah dengan menggunakan *k*=9 metode k-NN.

Tabel 4.16 *Confusion Matrix* Data RST tk. II dr. Soepraoen Malang

	True Yes (Diabetes)	**True No (Non-Diabetes)**	**Class precision**
Pred. Yes (Diabetes)	25	0	100.00%
Pred. No (Non-Diabetes)	1	26	96.30%
Class recall	96.15%	100.00%	

Tabel 4.16 menunjukkan hasil dari perhitungan *confusion matrix* yang didapatkan melalui tahap *cross validation* dari data yang didapatkan dari RST tk. II dr. Soepraoen Malang seblum dilakukannya uji regresi logistik biner dengan menggunakan 17 atribut, pengolahan dilakukan dengan menggunakan *RapidMiner.* Ditunjukkan bahwa *True Yes* (TY) berjumlah 25 data, *False Yes* (FY) sebanyak 0 data, *False Negative* (FN) sebanyak 1 data, dan jumlah *True Negative* (TN) sebanyak 26 data. Setelah didapatkan nilai pada *confusion matrix*, maka akan dilakukan tahap evaluasi yang dilakukan secara manual untuk menentukan tingkat *accuracy, precission,* dan *recall* dengan menggunakan rumus sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$= \frac{25 + 26}{25 + 26 + 1 + 0} \times 100\%$$

$$= \frac{51}{52} \times 100\%$$

$$= \ 98{,}08\%$$

$$Precision \ = \frac{TN}{TN \ + \ FN} \times \ 100\%$$

$$= \frac{26}{26 \ + \ 1} \times \ 100\%$$

$$= \frac{26}{27} \times \ 100\%$$

$$= \ 96.30\%$$

$$Recall \quad = \frac{TN}{TN \ + \ FY} \times \ 100\%$$

$$= \frac{26}{26 \ + \ 0} \times \ 100\%$$

$$= \frac{26}{26} \times \ 100\%$$

$$= 100\%$$

Berikut adalah tabel tingkat akurasi yang didapat dari proses klasifikasi menggunakan metode *k-Nearest Neighbor* menggunakan aplikasi *RapidMiner*.

Tabel 4.17 Tingkat Akurasi

Uji	***Accuracy***	***Precision***	***Recall***
Tingkat Akurasi	98.08%	96.30%	100.00%

Hasil akurasi klasifikasi menggunakan metode k-NN yang ditinjau dari parameter *k* = 1, 3, 5, 7, 9 untuk data RST tk.II dr.Soepraoen Malang sebelum dilakukan uji Regresi Logistik Biner memiliki 52 data, dengan 17 atribut ditampilkan pada tabel berikut.

Tabel 4.18 Tingkat Akurasi berdasarkan nilai *k*

Parameter	Akurasi (%)
k = 3	96.15%
k = 5	96.15%
k = 7	96.15%
k = 9	98.08%

4.1.4.2 Data RST tk.II dr.Soepraoen Malang setelah Uji Regresi Logistik Biner

Setelah dilakukan uji regresi logistik biner, data yang akan diolah mengalami pengurangan jumlah atribut, sehingga perlu dilakukan proses klasifikasi ulang agar mendapatkan hasil yang dapat menjadi perbandingan antara data sebelum dilakukan uji dan setelah dilakukan uji. Berikut merupakan hasil klasifikasi yang didapatkan:

Tabel 4.19 *Confusion Matrix*

	True Yes (Diabetes)	***True No (Non-Diabetes)***	***Class precision***
Pred. Yes (Diabetes)	25	0	100.00%
Pred. No (Non-Diabetes)	1	26	96.30%
Class recall	96.15%	100.00%	

Pada tabel 4.19 menunjukkan hasil dari klasifikasi yang dilakukan dengan menggunakan metode k-NN dengan bantuan aplikasi *RapidMiner*. Ditunjukkan bahwa *True Yes* (TY) berjumlah 25 data, *False Yes* (FY) sebanyak 0 data, *False Negative* (FN) sebanyak 1 data, dan jumlah *True Negatve* (TN) sebanyak 26 data. Diperoleh hasil akurasi sebesar 98,08%.

Setelah didapatkan hasil klasifikasi, maka akan dilakukan tahap evaluasi yang dilakukan secara manual untuk menentukan tingkat *accuracy*, *precission*, dan *recall* dengan menggunakan rumus sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$= \frac{25 + 26}{25 + 26 + 1 + 0} \times 100\%$$

$$= \frac{51}{52} \times 100\%$$

$$= 98{,}08\%$$

$$Precision = \frac{TN}{TN + FN} \times 100\%$$

$$= \frac{26}{26 + 1} \times 100\%$$

$$= \frac{26}{27} \times 100\%$$

$$= 96.30\%$$

$$Recall = \frac{TN}{TN + FY} \times 100\%$$

$$= \frac{26}{26 + 0} \times 100\%$$

$$= \frac{26}{26} \times 100\%$$

$$= 100\%$$

Setelah dilakukannya tahap evaluasi dengan cara penghitungan manual, maka akan dilakukan tahap evaluasi dengan menggunakan aplikasi *RapidMiner* guna mendapatkan tingkat *accuracy*, *precission*, dan *recall* dengan menggunakan *k*=9 pada algoritma k-NN. Berikut adalah tabel tingkat akurasi yang didapat dari perhitungan menggunakan aplikasi *RapidMiner*.

Tabel 4.20 Tingkat Akurasi RST tk.II dr.Soepraoen Malang Setelah

Uji	***Accuracy***	***Precision***	***Recall***
Tingkat Akurasi	98,08%	96,30%	100.00%

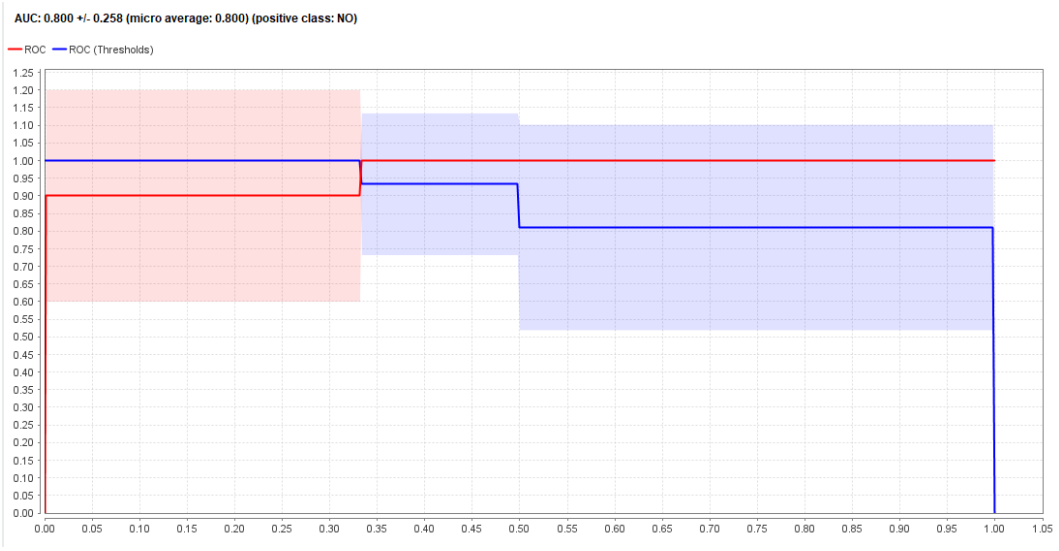
Algoritma k-NN dilakukan dengan cara menentukan nilai *k* dengan data *training* yang digunakan. Hasil akurasi klasifikasi yang ditinjau dari penghitungan menggunakan parameter *k* = 3, 5, 7, 9 untuk data sejumlah 52 data dengan 10 atribut yang paling berpengaruh. Tingkat akurasi berdasarkan nilai *k* yang berbeda akan ditampilkan pada tabel berikut.

Tabel 4.21 Akurasi berdasarkan parameter

Parameter	**Akurasi (%)**
k = 3	96.15%
k = 5	96.15%
k = 7	96.15%
k = 9	98.08%

4.1.5 Grafik ROC

Area Under Curve (AUC) menampilkan hasil perhitungan pengukuran perbedaan performasi metode yang digunakan untuk menghasilkan nilai yang sesuai dengan kriteria yang ditentukan. Bentuk dari hasil proses AUC tersebut di bawah ini:



Gambar 4. 5 Grafik AUC

Hasil AUC yang didapatkan dari aplikasi *RapidMiner* sebesar 0.800, dengan positive class = No. Sehingga didapatkan hasil bahwa klasifikasi yang dilakukan masuk ke kategori baik (*good classification*).

4.2 Pembahasan

Setelah mendapatkan hasil dari penelitian yang telah dilakukan, akan dibahas lebih lanjut terkait hasil yang didapatkan dan akan dikaitkan berdasarkan teori yang ada. Tingkat akurasi yang diperoleh dari proses klasifikasi menggunakan metode *k-Nearest Neighbor* dengan sembilan ketetanggaan terdekat (k = 9) adalah sama antara data RST tk.II dr.Soepraoen Malang sebelum dilakukan seleksi atribut dan setelah dilakukan seleksi atribut, yaitu sebesar 98,08% sehingga dapat dikatakan bahwa proses seleksi atribut yang dilakukan berhasil, karena meskipun terjadi pengurangan jumlah atribut dengan mengeliminasi data yang tidak berpengaruh, tingkat akurasi yang didapatkan adalah sama, sehingga atribut yang dieliminasi merupakan atribut yang tidak berpengaruh signifikan. Didapatkan hasil dari uji regresi logistik biner yang telah didapatkan bahwa faktor yang paling

berpengaruh terhadap seseorang dapat terkena diabetes mellitus adalah atribut yang memiliki nilai sig. < 0,05, yaitu atribut usia, *Body Mass Index* (BMI), kolesterol total, LDL kolesterol, trigliserida, gula darah sewaktu, gula darah puasa, gula darah 2 jam, dan HBA1C, dimana atribut-atribut tersebut merupakan representasi dari faktor yang mempengaruhi seseorang dapat terkena diabetes. Di antara seluruh atribut utama yang didapatkan, nilai *score* tertinggi adalah pada 10 atribut yang tergolong pada kategori gula darah dan kolesterol. Hal tersebut sesuai dengan pendapat dari Kemenkes RI dan WHO. Sehingga untuk dapat melakukan tindakan preventif pencegahan diabetes mellitus adalah dengan cara menghindari makanan yang dapat meningkatkan kadar gula darah dan kolesterol dalam tubuh. Hal tersebut juga dapat diimbangi dengan melakukan pola hidup sehat dan aktivitas fisik secara rutin.

Berdasarkan hasil pengujian dari uji regresi logistik biner, hipotesis yang didapatkan menunjukkan bahwa atribut yang berpengaruh signifikan adalah atribut yang memiliki hasil *Sig*.<0,05. Sehingga didapatkan bahwa usia, BMI, kolesterol total, LDL kolesterol, trigliserida, gula darah sewaktu, gula darah puasa, gula darah 2 jam, dan HBA1C berpengaruh signifikan terhadap faktor sesorang dapat terklasifikasi menjadi penderita diabetes mellitus, sehingga jika pasien tersebut mengalami *abnormal* pada atribut tersebut di atas, maka dapat dipastikan bahwa pasien tersebut merupakan pasien diabetes mellitus.

Selanjutnya akan ditampilkan keterangan lebih lanjut mengenai kadar normal masing-masing atribut yang digunakan pada data Diabetes Mellitus dari RST tk. II dr. Soepraoen Malang. Kadar normal yang dicantumkan merupakan kadar normal yang telah ditetapkan oleh WHO.

Tabel 4.22 Keterangan Atribut

No	Kategori	Atribut	Kadar	
1	TTV	Usia	-	
2		BMI	Kurang	< 18,5
			Normal	18,5 – 24,9
			Pra-obesitas	25,0 – 29,9
			Obesitas kelas I	30,0 – 34,9
			Obesitas kelas II	35,0 – 39,9
			Obesitas kelas III	> 40
3	KOLESTEROL	Kolesterol Total	Normal	< 200 mg/dl
			Batas Tinggi	200 – 239 mg/dl
			Tinggi	> 240 mg/dl
4		LDL Kolesterol	Optimal	>100 mg/dl
			Hampir Optimal	100 – 129 mg/dl
			Batas Tinggi	130 -159 mg/dl
			Tinggi	160-189 mg/dl
			Sangat Tinggi	190 mg/dl
5		Trigliserida	Normal	< 150 mg/dl
			Batas Tinggi	150 hingga 199 mg/dl
			Tinggi	200 hingga 499 mg/dl
			Sangat Tinggi	> 500 mg/dl
6	GULA DARAH	Gula darah sewaktu	Normal	N/A
			Pra-diabetes	N/A
			Diabetes	> 200 mg/dl
7		Gula darah puasa	Normal	< 99 mg/dl
			Pra-diabetes	100 -125 mg/dl
			Diabetes	> 126 mg/dl
8		Gula darah 2 jam	Normal	< 140 mg/dl
			Pradiabetes	140 - 199
			Diabetes	> 200
9		HBA1C	Normal	< 5,7%
			Pradiabetes	5,7% - 6,4%
			Diabetes	> 6,5%
	OUTPUT	Diagnosa	Yes	Ter diagnosa Diabetes
			No	Tidak ter diagnosa Diabetes

Berdasarkan informasi tersebut di atas, dapat diketahui bahwa hasil klasifikasi yang didapatkan sesuai dengan informasi dan teori dari WHO bahwa

pasien yang mengalami peningkatan kadar pada gula darah dan kolesterol, merupakan pasien diabetes mellitus.

4.3 Kajian Keislaman dengan Hasil Penelitian

Berdasarkan hasil yang diperoleh dari penelitian ini, faktor penyebab utama diabetes mellitus adalah meningkatnya kadar gula darah dan kolesterol, dimana hal tersebut disebabkan karena pola makan yang tidak baik. Dalam Al-Qur'an telah disebutkan untuk tidak makan atau minum secara berlebihan. Hal ini tertuang pada Surat Al-Araf ayat 31 yang artinya (Kemenag, 2002):

يَا بَنِي آدَمَ خُذُوا زِينَتَكُمْ عِنْدَ كُلِّ مَسْجِدٍ وَكُلُوا وَاشْرَبُوا وَلَا تُسْرِفُوا ۚ إِنَّهُ لَا يُحِبُّ الْمُسْرِفِينَ

"Wahai anak cucu Adam! Pakailah pakaianmu yang bagus pada setiap (memasuki) masjid, makan dan minumlah, tetapi jangan berlebihan. Sungguh, Allah tidak menyukai orang yang berlebih-lebihan"

Pada ayat ini Allah memerintahkan agar manusia berlaku adil dalam semua urusan, maka pada ayat ini Allah memerintahkan manusia untuk makan dan minum secukupnya tanpa berlebih-lebihan. Wahai anak cucu Adam! Pakailah pakaianmu yang bagus yaitu pakaian yang dapat menutupi aurat kalian atau bahkan yang lebih dari itu ketika kalian beribadah, sehingga kalian bisa melakukan salat dan tawaf dengan nyaman, dan lakukanlah itu pada setiap memasuki dan berada di dalam masjid atau tempat lainnya di muka bumi ini. Dalam rangka beribadah, Kami telah menyediakan makanan dan minuman, maka makan dan minumlah apa saja yang kamu sukai dari makanan dan minuman yang halal, baik dan bergizi, tetapi jangan berlebihan dalam segala hal, baik dalam beribadah dengan menambah cara atau kadarnya, ataupun dalam makan dan minum. Karena sungguh, Allah tidak

menyukai, yakni tidak melimpahkan rahmat dan ganjaran-Nya kepada orang yang berlebih-lebihan dalam hal apa pun (Kemenag, 2002).

Hal tersebut di atas jika dikaitkan dengan diabetes mellitus sesuai dengan topik yang diangkat, maka hendaknya kita sebagai manusia dapat membatasi diri untuk tidak makan ataupun minum secara berlebihan, karena faktor yang paling berpengaruh terhadap diabetes mellitus adalah peningkatan kadar glukosa dan kolesterol.

BAB V
PENUTUP

5.1. Kesimpulan

Kesimpulan yang dapat diambil setelah dilakukan penelitian ini, maka dapat diketahui bahwa hasil *accuracy*, *precision,* dan *recall* yang didapatkan antara data dengan 17 atribut dengan data yang telah dilakukan uji regresi logistik biner yang memiliki jumlah atribut 10 adalah sama yaitu *accuracy* sebesar 98,08%, *precision* sebesar 96,30%, dan *recall* sebesar 100.00%. Sehingga dapat diambil kesimpulan bahwa atribut yang digunakan merupakan atribut yang paling berpengaruh signifikan terhadap faktor terjadinya Diabetes Mellitus. Atribut yang paling berpengaruh merupakan representasi dari faktor yang paling berpengaruh terhadap seseorang dapat terkena penyakit diabetes mellitus yaitu usia, *Body Mass Index* (BMI), kolesterol total, LDL kolesterol, trigliserida, gula darah sewaktu, gula darah puasa, gula darah 2 jam, dan HBA1C. Di antara 10 atribut utama yang didapatkan, nilai *score* tertinggi adalah pada atribut-atribut yang tergolong pada kategori Gula Darah dan Kolesterol, sehingga untuk melakukan tindakan preventif agar terhindar dari Diabetes Mellitus, hendaknya kita dapat menghindari makanan maupun kebiasaan yang dapat meningkatkan kadar gula darah dan kolesterol.

5.2. Saran untuk Penelitian Lanjutan

Pada penelitian kali ini telah dibahas terkait klasifikasi potensi penyakit diabetes mellitus dengan menggunakan metode *k-Nearest Neighbor* sehingga, saran peneliti untuk peneliti selanjutnya adalah:

1. Melakukan penelitian serupa secara lebih detail dan komprehensif serta dapat melakukan deteksi dini penyakit Diabetes Mellitus yang berfokus pada generasi usia muda, khususnya pada Generasi Z.
2. Menambahkan perhitungan matematis secara lebih rinci terkait Regresi Logistik Biner.

DAFTAR PUSTAKA

Agresti, A. (2002). *Categorical Data Analysis*.

Alfina, T., Santosa, B., & Barakbah, A. R. (2012). *Analisa Perbandingan Metode Hierarchical - K means - gabungan, ITS*.

Al-Jauziyah, I. Q. (2000). *ZAADUL-MA'AD BEKAL PERJALANAN KE AKHIRAT* (R. Huda, Ed.; 1st ed.). PUSTAKA AZZAM.

Anggraeni, Y., & Zain, I. (2015). *Pemodelan Regresi Logistik Biner terhadap Peminat ITS di Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) 2014*. https://media.neliti.com/media/publications/15584-ID-pemodelan-regresi-logistik-biner-terhadap-peminat-its-di-seleksi-bersama-masuk-p.pdf

Aris, F., & Benyamin. (2019). *Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi - SRM Publishing*. *1*(1), 01–06.

Arriawati, A. J., Santoso, I., & Christyono, Y. (2011). *KLASIFIKASI CITRA TEKSTUR MENGGUNAKAN k-NEAREST NEIGHBOUR BERDASARKAN EKSTRAKSI CIRI METODE MATRIKS KOOKURENSI*.

Asli, A., Rahmawati, L., Gunardi, H., Sekartini, R., Batubara, J. R., & BPulungan, A. (2007). *Gangguan Perilaku Pasien Diabetes Melitus tipe-1 di Poliklinik Endokrinologi Anak Rumah Sakit Cipto Mangunkusumo* (Vol. 9, Issue 4).

Daniel T, & Larose Ph.D. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*.

Das, H., Naik, B., & Behera, H. S. (2018). Classification of diabetes mellitus disease (DMD): A data mining (DM) approach. *Advances in Intelligent Systems and Computing*, *710*, 539–549. https://doi.org/10.1007/978-981-10-7871-2_52

de Souza Lima, E., de Souza, Z. M., Montanari, R., de Medeiros Oliveira, S. R., Lovera, L. H., & Vieira Farhate, C. V. (2017). CLASSIFICATION OF THE INITIAL DEVELOPMENT OF EUCALIPTUS USING DATA MINING TECHNIQUES. *Cerne*, *23*(2), 201–208. https://doi.org/10.1590/01047760201723022296

Fitriani, R. (2017). *ANALISIS FAKTOR RISIKO KEJADIAN DIABETES MELITUS GESTASIONAL*. *10*.

Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques*. Springer.

Hermayudi dkk. (2017). *Metabolik Endokrin*. 32–32.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression (Second Edition)*.

Iyer, A., S, J., & Sumbaly, R. (2015). Diagnosis of Diabetes Using Classification Mining Techniques. *International Journal of Data Mining & Knowledge Management Process*, *5*(1), 01–14. https://doi.org/10.5121/ijdkp.2015.5101

Karegowda, A. G., Jayaram, M. A., & Manjunath, A. S. (2012). *Cascading K-means Clustering and K-Nearest Neighbor Classifier for Categorization of Diabetic Patients 148*. http://www.galaxy.gmu.edu/interface/I01/I2001Proceedings/Jbreault

Kemenag. (2002). *Al-Qur'an Kemenag* (1st ed.). LPMQ.

Krisandi, N., Helmi, B., & Prihandono, I. (2013). *ALGORITMA k-NEAREST NEIGHBOR DALAM KLASIFIKASI DATA HASIL PRODUKSI KELAPA SAWIT PADA PT. MINAMAS KECAMATAN PARINDU* (Vol. 02, Issue 1).

Mangande, P. (2020). *CLASSIFICATION OF COSTUMER DATA THAT POTENTIALS TO OPEN DEPOSITS USING NAIVE BAYES ALGORITHM*.

Nofriansyah, D. (2016). *Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi)*.

Nor, A., 1∗, F., Dyah, Y., Santik, P., & Artikel, I. (2020). *33 HIGEIA 4 (1) (2020) HIGEIA JOURNAL OF PUBLIC HEALTH RESEARCH AND DEVELOPMENT Kejadian Diabetes Melitus Tipe I pada Usia 10-30 Tahun*. https://doi.org/10.15294/higeia/v4i1/31763

Sa'di, S., Maleki, A., Hashemi, R., Panbechi, Z., & Chalabi, K. (2015). Comparison of Data Mining Algorithms in the Diagnosis of Type Ii Diabetes. *International Journal on Computational Science & Applications*, *5*(5), 1–12. https://doi.org/10.5121/ijcsa.2015.5501

Sarker, I., Faruque, Md., Alqahtani, H., & Kalim, A. (2018). K-Nearest Neighbor Learning based Diabetes Mellitus Prediction and Analysis for eHealth Services. *ICST Transactions on Scalable Information Systems*, *0*(0), 162737. https://doi.org/10.4108/eai.13-7-2018.162737

Saxena, K., Khan, Z., Singh, S., Research, M.-T., & 1&3, S. (n.d.). Diagnosis of Diabetes Mellitus using K Nearest Neighbor Algorithm. *International Journal of Computer Science Trends and Technology*, *2*. www.ijcstjournal.org

Suniantara, K. P., & Rusli, M. (2017). *KLASIFIKASI WAKTU KELULUSAN MAHASISWA STIKOM BALI MENGGUNAKAN CHAID REGRESSION-TREES DAN REGRESI LOGISTIK BINER*. *5*, 27–32.

Suyanto. (2018). *Machine Learning Tingkat Dasar dan Lanjut*. INFORMATIKA.

Vercellis, Carlo. (2008). *Business intelligence : data mining and optimization for decision making*. Wiley.

WHO. (2016, April 21). *Global report on diabetes*.

Wistiani. (2011). *Manifestasi Klinis Beberapa Penyakit dengan Konfirmasi diagnostik lupus eritematosus*.

LAMPIRAN

Data RST tk. II dr. Soepraoen Malang

Lampiran 1 Dataset Diabetes Mellitus RST tk.II dr.Soepraoen Malang Setelah Uji Regresi

USIA	BMI	KOLESTEROL	LDL KOLESTEROL	TRIGLISERIDA	GD SEWAKTU	GD PUASA	GD 2 JAM	HBA1C	DIAGNOSA
54	23,438	246	175	291	230	200	227	13	YES
54	23,438	254	173	271	217	177	193	12	YES
57	23,030	284	174	292	268	230	220	11,8	YES
54	23,438	214	184	316	265	270	228	11	YES
54	23,438	217	153	259	311	192	210	11,5	YES
54	23,438	252	159	299	312	155	241	12	YES
57	23,030	225	172	270	315	196	184	11	YES
49	26,122	242	152	266	289	250	310	13	YES
54	23,438	223	152	191	265	165	185	11,9	YES
49	24,490	287	191	181	258	180	210	13,2	YES
54	23,438	211	179	165	246	180	188	14	YES
54	23,438	255	154	181	304	192	179	12,9	YES
54	23,438	207	135	195	300	270	228	9,8	YES
54	23,438	284	178	258	300	277	334	10,2	YES
54	23,438	229	150	202	321	127	149	11	YES
49	26,122	248	181	183	338	210	300	11	YES
54	23,438	247	0	241	261	192	179	12	YES
54	23,438	262	152	301	230	0	194	11,6	YES
54	23,438	206	193	304	272	0	271	20	YES
29	19,377	179	130	139	127	117	133	5,1	NO
54	23,438	243	202	251	0	165	185	11,9	YES
57	23,030	229	200	163	236	0	264	10,4	YES
47	22,491	176	114	123	140	111	129	4	NO
30	22,039	197	111	137	112	120	109	4,6	NO
35	20,658	167	119	131	139	81	140	5,9	NO
32	22,590	165	121	132	115	105	138	4,7	NO
47	22,491	187	110	132	150	88	111	6	NO
47	20,285	193	105	135	143	82	118	5,9	NO
30	22,039	170	130	131	105	124	107	4,5	NO
47	22,491	189	134	124	90	101	131	3	NO
57	23,030	0	140	284	248	125	165	13,7	YES

29	22,719	183	117	131	89	92	139	4,9	NO
29	20,761	169	103	133	118	86	123	6,1	NO
29	22,719	151	105	132	106	98	140	4,8	NO
26	22,491	156	122	131	92	102	130	5,3	NO
29	22,719	151	132	120	89	118	130	5,2	NO
47	22,491	201	102	137	95	82	104	6	NO
54	23,438	193	135	215	267	355	439	14,1	YES
47	22,491	167	103	124	110	89	107	5,2	NO
54	23,438	0	183	171	319	195	271	12,1	YES
54	23,438	254	133	0	265	170	200	14,2	YES
31	22,039	102	118	134	87	84	131	4,7	NO
47	22,491	200	0	131	97	103	134	5,1	NO
54	23,438	260	203	0	325	177	193	13,7	YES
30	22,039	185	123	120	0	121	117	4	NO
47	22,491	159	110	130	0	82	132	4,9	NO
30	22,039	166	105	126	111	102	0	5,5	NO
47	22,491	175	119	121	105	89	0	5,9	NO
58	20,569	0	104	139	108	117	128	5,7	NO
30	22,039	0	107	128	102	116	110	4,1	NO
30	22,039	166	125	0	127	101	104	4,8	NO
45	20,761	0	106	130	96	117	104	3,7	NO
54	23,438	246	175	291	230	200	227	13	DATA BARU

Lampiran 2 Tingkat *Accuracy*

accuracy: 98.33% +/- 5.27% (micro average: 98.08%)

	true YES	true NO	class precision
pred. YES	25	0	100.00%
pred. NO	1	26	96.30%
class recall	96.15%	100.00%	

Lampiran 3 Tingkat *Precission*

precision: 97.50% +/- 7.91% (micro average: 96.30%) (positive class: NO)

	true YES	true NO	class precision
pred. YES	25	0	100.00%
pred. NO	1	26	96.30%
class recall	96.15%	100.00%	

Lampiran 4 Tingkat *Recall*

recall: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: NO)

	true YES	true NO	class precision
pred. YES	25	0	100.00%
pred. NO	1	26	96.30%
class recall	96.15%	100.00%	

Lampiran 5 Statistik Uji

Variables not in the Equation[a]

			Score	df	Sig.
Step 0	Variables	USIA	31.294	1	.000
		SISTOLE	7.792	1	.005
		DIASTOLE	.190	1	.663
		RR	1.224	1	.269
		NADI	2.398	1	.122
		SUHU	1.615	1	.204
		SPO2	.773	1	.379
		BB	2.443	1	.118
		TB	5.485	1	.019
		BMI	26.642	1	.000
		KOLESTEROL	12.068	1	.001
		LDL KOLESTEROL	20.126	1	.000
		TRIGLISERID	20.560	1	.000
		GD SEWAKTU	37.867	1	.000
		GD PUASA	16.103	1	.000
		GD 2 JAM	29.498	1	.000
		HBA1C	44.913	1	.000

a. Residual Chi-Squares are not computed because of redundancies.

Lampiran 6 Statistik Uji

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	USIA	.296	1314.731	.000	1	1.000	1.344	.000	.
	SISTOLE	.073	699.048	.000	1	1.000	1.076	.000	.
	DIASTOLE	.094	1590.979	.000	1	1.000	1.098	.000	.
	RR	.458	5884.166	.000	1	1.000	1.582	.000	.
	NADI	-.323	1994.206	.000	1	1.000	.724	.000	.
	SUHU	2.687	22428.263	.000	1	1.000	14.682	.000	.
	SPO2	-148.720	1265394.206	.000	1	1.000	.000	.000	.
	BB	-5.135	35267.437	.000	1	1.000	.006	.000	.
	TB	3.395	27865.332	.000	1	1.000	29.802	.000	.
	BMI	16.631	105553.105	.000	1	1.000	16693407.84	.000	.
	KOLESTEROL	.011	128.922	.000	1	1.000	1.011	.000	5.542E+109
	LDL KOLESTEROL	.040	167.749	.000	1	1.000	1.040	.000	6.389E+142
	TRIGLISERID	.044	141.595	.000	1	1.000	1.045	.000	3.505E+120
	GD SEWAKTU	.028	124.231	.000	1	1.000	1.028	.000	5.728E+105
	GD PUASA	.030	113.610	.000	1	1.000	1.031	.000	5.229E+96
	GD 2 JAM	.018	330.090	.000	1	1.000	1.019	.000	9.566E+280
	HBA1C	3.070	4051.958	.000	1	.999	21.541	.000	.

RIWAYAT HIDUP

Candrani Sri Murtono, lahir di kota Jakarta pada 23 Maret 2000. Penulis lahir dari **Drs. Sri Murtono, M.Pd, C.D** dan **Rr. Ir. Sri Redjeki Wahjoe Astoeti** sebagai anak pertama dari tiga bersaudara. Jenjang pendidikan pertama yang ditempuh adalah PAUD dan TK Cemara Dua, lalu melanjutkan Sekolah Dasar pada tiga sekolah yang berbeda, SD plus Al-Firdaus Surakarta, SDN Kebonsari 1, dan lulus dari SDN Sumberagung 03 yang lulus pada tahun 2012. Penulis melanjutkan studi pada Sekolah Menengah Pertama di SMPN 1 Selorejo yang lulus pada tahun 2015. Pendidikan selanjutnya adalah Sekolah Menengah Kejuruhan pada SMKN 2 Malang yang lulus pada tahun 2018. Penulis melanjutkan pendidikan dengan berkuliah pada UIN Maulana Malik Ibrahim Malang, dengan memilih jurusan Matematika murni di Fakultas Sains dan Teknologi.

Selama menempuh Pendidikan di Kampus UIN Maulana Malik Ibrahim Malang, selain menyelesaikan tugasnya sebagai mahasiswi, penulis juga merupakan penerima Beasiswa dari Bank Indonesia yang tergabung pada Generasi Baru Indonesia (GenBI Malang) selama dua periode, penulis juga merupakan Ketua Mathematics English Club periode 2020/2022. Selain itu penulis juga merupakan pengurus dari HMJ “Integral” Matematika. Tidak hanya itu, penulis juga merupakan Duta Felari. Menurut penulis, hal-hal tersebut dapat meningkatkan *soft-skill* dan kemampuan managemen lainnya.

KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

BUKTI KONSULTASI SKRIPSI

Nama	: Candrani Sri Murtono
NIM	: 18610069
Fakultas / Program Studi	: Sains dan Teknologi / Matematika
Judul Skripsi	: Model Klasifikasi Potensi Penyakit Diabetes Mellitus menggunakan Metode *K-Nearest Neighbor* (K-NN)
Pembimbing I	: Hisyam Fahmi, M.Kom
Pembimbing II	: Erna Herawati, M.Pd

No	Tanggal	Hal	Tanda Tangan
1.	25 Februari 2022	Konsultasi BAB I	1.
2.	17 Maret 2022	Konsultasi Revisi BAB I	2.
3.	25 Maret 2022	Konsultasi BAB II dan III	3.
4.	12 April 2022	Konsultasi Revisi BAB II dan III	4.
5.	25 April 2022	ACC Seminar Proposal	5.
6.	20 Mei 2022	Konsultasi Keagamaan	6.
7.	25 Mei 2022	Konsultasi Revisi Keagamaan	7.
8.	29 Mei 2022	Konsultasi BAB IV dan V	8.
9.	18 Juli 2022	Konsultasi Revisi BAB IV dan V	9.
10.	02 Agustus 2022	Konsultasi BAB IV dan V	10.
11.	23 Agustus 2022	Konsultasi Revisi BAB IV dan V	11.
12.	29 Agustus 2022	Konsultasi Kajian Agama	12.
13.	09 September 2022	Konsultasi Revisi Kajian Agama TTD Untuk Seminar Hasil	13.

KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

No	Tanggal	Hal	Tanda Tangan
14.	23 September 2022	ACC Seminar Hasil	14.
15.	01 November 2022	Konsultasi Revisi Seminar Hasil	15.
16.	16 November 2022	Konsultasi Revisi Seminar Hasil	16.
17.	21 November 2022	Konsultasi Kajian Agama	17.
18.	02 Desember 2022	ACC Sidang Skripsi	18.
19.	21 Desember 2022	ACC Skripsi untuk Syarat Yudisium	19.

Malang, 21 Desember 2022
Mengetahui,
Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005