

**PERBANDINGAN METODE *RANDOM FOREST* DENGAN
METODE *K-NEAREST NEIGHBOR* UNTUK KLASIFIKASI
PENYAKIT DIABETES MELITUS**

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Infomatika



Oleh:

ANISA RIZKI YOLANDA
NIM : 09021181722077

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2022**

LEMBAR PENGESAHAN SKRIPSI


PERBANDINGAN METODE RANDOM FOREST DENGAN METODE K- NEAREST NEIGHBOR UNTUK KLASIFIKASI PENYAKIT DIABETES MELITUS

Oleh:


ANISA RIZKI YOLANDA
NIM : 09021181722077

Palembang, 21 Juli 2022

Pembimbing I,


Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005

Pembimbing II,


Rizki Kurniati, M.T.
NIP. 199107122019032016

Mengetahui,
Ketua Jurusan Teknik Informatika,


Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Rabu tanggal 06 Juli 2022 telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Anisa Rizki Yolanda
NIM : 09021181722077
Judul : Perbandingan Metode *Random Forest* dengan Metode *K-Nearest Neighbor* untuk Klasifikasi Penyakit Diabetes Melitus
dan dinyatakan **LULUS**.

1. Ketua Penguji

Mastura Diana Marieska, M.T.
NIP. 198603212018032001



2. Penguji I

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



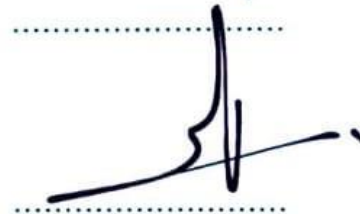
3. Penguji II

Annisa Darmawahyuni, M.Kom.
NIP. 1671147006900002



4. Pembimbing I

Dr. Abdiansah, S.Kom., M.Cs.
NIP. 198410012009121005



5. Pembimbing II

Rizki Kurniati, M.T.
NIP. 199107122019032016



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Anisa Rizki Yolanda
NIM : 09021181722077
Program Studi : Teknik Informatika Reguler
Judul Skripsi : Perbandingan Metode Random Forest dengan Metode K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus

Hasil Pengecekan Software iThenticate/Turnitin : 10%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya saya sendiri dan bukan hasil penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, Juli 2022



Anisa Rizki Yolanda

NIM. 09021181722077

HALAMAN MOTTO DAN PERSEMBAHAN

MOTTO :

“If we are close to Allah, whatever the most difficult thing, Allah will bring it closer and make it easy”

Kupersembahkan karya tulis ini kepada :

- ❖ Orang tuaku tercinta
- ❖ Saudaraku
- ❖ Sahabat dan teman seperjuanganku
- ❖ Jurusan Teknik Informatika
- ❖ Fakultas Ilmu Komputer
- ❖ Universitas Sriwijaya

KATA PENGANTAR

Puji syukur kepada Allah SWT atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini, banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun tidak secara langsung. Untuk itu, Penulis ingin menyampaikan rasa terima kasih kepada:

1. Bapak Jaidan Jauhari, S.Pd., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
2. Ibu Alvi Syahrini Utami, M.Kom. selaku Ketua Jurusan Teknik Informatika dan Ibu Mastura Diana Marieska, M.T. selaku Sekretaris Jurusan Teknik Informatika.
3. Ibu, bapak, dan saudara saya serta seluruh keluarga besar saya yang tidak pernah berhenti untuk selalu mendoakan, menasehati, dan memberikan motivasi dan dukungan yang luar biasa, baik moril maupun materil.
4. Bapak Dr. Abdiansah, S.Kom., M.Cs. selaku Dosen Pembimbing I dan Ibu Rizki Kurniati, M.T. selaku Dosen Pembimbing II yang telah membimbing, memotivasi dan memberi arahan dalam proses perkuliahan dan pengerjaan Tugas Akhir.
5. Ibu Alvi Syahrini Utami, M.Kom. selaku Dosen Penguji I sekaligus Dosen Pembimbing Akademik, yang telah membimbing, memotivasi dan memberi arahan dalam proses perkuliahan dan pengerjaan Tugas Akhir.
6. Ibu Annisa Darmawahyuni, M.Kom, selaku Dosen Penguji II yang telah memberikan arahan, masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
7. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

8. Kak Ricy serta seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
9. Kim Taehyung dan seluruh anggota BTS yang telah mewarnai serta membuat saya merasa tidak sendirian.
10. Binti Mutammimah, Fatimah Az-zahra, Herwantoro, M. Dwi Parlindungan, dan M.Ridhoi yang selalu ada dan mendukung serta memberikan doa dan semangat penulis untuk dapat menyelesaikan Tugas Akhir ini.
11. Amelia, Tita Dwi Yulian, dan Gina Damayanti yang telah mewarnai kehidupan perkuliahan dan mau membantu saat diperlukan.
12. Teman-teman TIREG A 2017 dan seluruh teman-teman Teknik Informatika Universitas Sriwijaya.
13. Untuk mereka yang sempat datang dan pergi, yang memberikan kekuatan kepada penulis serta pembelajaran yang sangat mendewasakan tentang bagaimana cara menerima dan mengikhlaskan.
14. Diri saya sendiri yang telah bertahan sejauh ini, yang mencoba untuk selalu tetap kuat dengan semua hal yang terjadi 2 tahun terakhir ini untuk dapat menyelesaikan tugas akhir dengan baik.
15. Semua pihak yang telah membantu dalam penyusunan dan penyempurnaan tugas akhir ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan yang disebabkan keterbatasan pengetahuan dan pengalaman. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan. Semoga Allah SWT selalu melimpahkan Rahmat dan Hidayah-Nya. Akhir kata, dengan segala kerendahan hati, semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, Juli 2022

Anisa Rizki Yolanda

ABSTRACT

Diabetes Mellitus is a disorder of the insulin system caused by excessive levels of glucose in the blood. Symptoms of diabetes mellitus so far have only been diagnosed by the general public based on known physical characteristics without being supported by facts and other medical considerations. The Basic Health Research (Riskesdas) shows a significant increase in the prevalence of diabetes, from 6.9% in 2013 to 8.5% in 2018. This study uses the Random Forest method and the K-Nearest Neighbor method where a comparison is made. to see which method is most appropriate in classifying Diabetes Mellitus. Both of these methods have the advantage that they can be used for large amounts of data. Determination of optimal parameter values is very important to support good accuracy results in both methods. Based on the results of the research, the Random Forest method produces an average accuracy of 80.91% accompanied by the average precision, recall, and f-measure values, respectively, which are 84.536%, 78.782%, and 81.422% with the best level of accuracy reaching 82.47%. While the K-Nearest Neighbor method produces an average accuracy of 74.062%, accompanied by the average value of precision, recall, and f-measure, respectively, which is 72.83%, 81.72%, and 76.992% with the best level of accuracy reaching 76.62%. This proves that the Random Forest method is the most suitable in classifying diabetes mellitus compared to the K-Nearest Neighbor method.

Keywords : Diabetes Mellitus, Classification, Random Forest, K-Nearest Neighbor

ABSTRAK

Penyakit Diabetes Melitus adalah kelainan sistem insulin yang diakibatkan berlebihnya kadar glukosa di dalam darah. Gejala dari penyakit diabetes melitus selama ini hanya didiagnosis masyarakat awam berdasarkan ciri-ciri fisik yang diketahui tanpa didukung oleh fakta dan pertimbangan medis lainnya. Menurut Riset Kesehatan Dasar (Riskesdas) memperlihatkan peningkatan angka prevalensi diabetes yang cukup signifikan, yaitu dari 6,9% di tahun 2013 menjadi 8,5% di tahun 2018. Pada penelitian ini menggunakan metode Random Forest dan metode K-Nearest Neighbor dimana dilakukan perbandingan untuk melihat metode mana yang paling sesuai untuk klasifikasi penyakit Diabetes Melitus. Kedua metode ini memiliki kelebihan yaitu dapat digunakan untuk data dengan jumlah yang besar. Penentuan nilai parameter yang optimal sangat penting untuk menunjang hasil akurasi yang baik pada kedua metode. Berdasarkan hasil penelitian metode Random Forest menghasilkan hasil akurasi rata-rata sebesar 80,91% disertai dengan nilai rata-rata presisi, recall, dan f-measure berturut-turut yaitu sebesar 84,536%, 78,782% dan 81,422% dengan tingkat akurasi terbaik mencapai 82,47%. Sedangkan pada metode K-Nearest Neighbor menghasilkan hasil akurasi rata-rata sebesar 74,062%, disertai dengan nilai rata-rata presisi, recall, dan f-measure berturut-turut yaitu sebesar 72,83%, 81,72%, dan 76,992% dengan tingkat akurasi terbaik mencapai 76,62%. Hal ini membuktikan bahwa metode Random Forest adalah metode yang paling sesuai dalam mengklasifikasikan penyakit diabetes melitus dibandingkan dengan metode K-Nearest Neighbor.

Kata Kunci : Diabetes Melitus, Klasifikasi, *Random Forest*, *K-Nearest Neighbor*

DAFTAR ISI

Halaman

HALAMAN JUDUL	i
LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	iii
HALAMAN PERNYATAAN	iv
HALAMAN MOTTO DAN PERSEMBAHAN	v
KATA PENGANTAR	vi
ABSTRACT.....	viii
ABSTRAK.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xvi
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan.....	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian	I-4
1.5 Manfaat Penelitian	I-4
1.6 Batasan Masalah	I-5
1.7 Sistematika Penulisan	I-5
1.8 Kesimpulan	I-6
BAB II KAJIAN LITERATUR	II-7
2.1 Pendahuluan.....	II-7
2.2 Landasan Teori.....	II-7
2.2.1 Diabetes Mellitus.....	II-7
2.2.2 Klasifikasi.....	II-8

2.2.3	<i>Decision Tree</i>	II-9
2.2.4	<i>Random Forest</i>	II-12
2.2.5	<i>K-Nearest Neighbor (K-NN)</i>	II-14
2.2.6	<i>Confusion Matrix</i>	II-16
2.3	Penelitian Lain Yang Relevan	II-18
2.4	Kesimpulan	II-19
BAB III METODOLOGI PENELITIAN		III-20
3.1	Pendahuluan	III-20
3.2	Pengumpulan Data	III-20
3.3	Tahapan Penelitian	III-21
3.3.1	Menetapkan Kerangka Kerja.....	III-22
3.3.2	Menetapkan Kriteria Pengujian.....	III-24
3.3.3	Menetapkan Format Data Pengujian	III-24
3.3.4	Menentukan Alat Bantu Penelitian	III-25
3.3.5	Melakukan Pengujian Penelitian	III-25
3.3.6	Melakukan Analisis Hasil Pengujian	III-26
3.3.7	Membuat Kesimpulan	III-26
3.4	Metode Pengembangan Perangkat Lunak.....	III-26
3.4.1	Rational Unified Process	III-26
3.4.1.1	Fase Insepsi	III-27
3.4.1.2	Fase Elaborasi	III-28
3.4.1.3	Fase Konstruksi.....	III-28
3.4.1.4	Fase Transisi	III-28
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-29
4.1	Pendahuluan	IV-29
4.2	Rational Unified Process.....	IV-29
4.2.1	Fase Insepsi	IV-29
4.2.1.1	Pemodelan Bisnis	IV-29
4.2.1.2	Kebutuhan	IV-30
4.2.1.3	Analisis dan Desain.....	IV-31

4.2.1.4	Desain Perangkat Lunak	IV-47
4.2.2	Fase Elaborasi	IV-54
4.2.2.1	Pemodelan Bisnis	IV-55
4.2.2.3	Diagram.....	IV-58
4.2.3.1	Kebutuhan Sistem	IV-69
4.2.3.2	Diagram Kelas.....	IV-69
4.2.3.3	Implementasi	IV-71
4.2.4	Fase Transisi.....	IV-77
4.2.4.1	Pemodelan Bisnis	IV-77
4.2.4.2	Kebutuhan Sistem	IV-77
4.2.4.3	Rencana Pengujian	IV-78
4.2.4.4	Implementasi	IV-81
4.3	Kesimpulan	IV-87
BAB V HASIL DAN ANALISIS PENELITIAN		V-88
5.1	Pendahuluan.....	V-88
5.2	Data Hasil Percobaan/Penelitian.....	V-88
5.1.1	Konfigurasi Percobaan	V-88
5.1.1.1	Data Hasil Percobaan.....	V-89
5.1.1.2	Hasil Pengujian Klasifikasi dengan Metode <i>Random Forest</i>	V-89
5.1.1.3	Hasil Pengujian Klasifikasi dengan Metode K-Nearest Neighbor	V-91
5.2	Analisis Hasil Pengujian	V-93
5.2.1	Analisis Hasil Pengujian Klasifikasi dengan Metode <i>Random Forest</i>	V-93
5.2.2	Analisis Hasil Pengujian Klasifikasi dengan Metode K-Nearest <i>Neighbor</i>	V-95
5.3	Kesimpulan	V-101
BAB VI KESIMPULAN DAN SARAN		VI-102
6.1	Pendahuluan.....	VI-102

6.2	Kesimpulan	VI-102
6.3	Saran	VI-103
DAFTAR PUSTAKA		xviii

DAFTAR TABEL

	Halaman
Tabel II- 2. Model Confusion Matrix	II-16
Tabel III- 1. Atribut Dataset Pima Indians Diabetes	III-21
Tabel III- 2. Rancangan Nilai Akurasi Hasil Pengujian Klasifikasi.....	III-24
Tabel III- 3. Rancangan Tabel Hasil Pengujian Klasifikasi	III-25
Tabel IV- 1. Tabel Kebutuhan Fungsional.....	IV-30
Tabel IV- 2. Tabel Kebutuhan Non-Fungsional.....	IV-31
Tabel IV- 3. Sampel Data Latih <i>Random Forest</i>	IV-34
Tabel IV- 4. Sampel Data Uji Random Forest.....	IV-43
Tabel IV- 5. Sampel Data Latih <i>K-Nearest Neighbor</i>	IV-44
Tabel IV- 6. Sampel Data Uji <i>K-Nearest Neighbor</i>	IV-45
Tabel IV- 7. Definisi Aktor	IV-48
Tabel IV- 8. Definisi Use Case	IV-48
Tabel IV- 9. Skenario Use Case Memuat Data	IV-50
Tabel IV- 10. Skenario Use Case Melakukan Pelatihan Random Forest.....	IV-51
Tabel IV- 11. Skenario Use Case Melakukan Pengujian Random Forest	IV-52
Tabel IV- 12. Skenario Use Case Melakukan Pelatihan <i>K-Nearest Neighbor</i>	IV-53
.....	VI-54
Tabel IV- 13. Skenario Use Case Melakukan Pengujian <i>K-Nearest Neighbor</i>	VI-54
.....	IV-71
Tabel IV- 14. Implementasi Kelas	IV-71
Tabel IV- 15. Rencana Pengujian Use Case Memuat Data.....	IV-78
Tabel IV- 16. Rencana Pengujian Use Case Melakukan Pelatihan dengan Metode Random Forest	IV-78
Tabel IV- 17. Rencana Pengujian Use Case Melakukan Pengujian dengan Metode Random Forest	IV-79
Tabel IV- 18. Rencana Pengujian <i>Use Case</i> Melakukan Pelatihan dengan Metode <i>K-Nearest Neighbor</i>	IV-79

Tabel IV- 19. Rencana Pengujian <i>Use Case</i> Melakukan Pengujian dengan Metode <i>K-Nearest Neighbor</i>	IV-80
Tabel IV- 20. Pengujian <i>Use Case</i> Memuat Data	IV-81
Tabel IV- 21. Pengujian <i>Use Case</i> Melakukan Pelatihan dengan Metode Random Forest.....	IV-82
Tabel IV- 22. Pengujian <i>Use Case</i> Melakukan Pengujian dengan Metode Random Forest	IV-83
Tabel IV- 23. Pengujian <i>Use Case</i> Melakukan Pelatihan dengan Metode <i>K-Nearest Neighbor</i>	IV-84
Tabel IV- 24. Pengujian <i>Use Case</i> Melakukan Pengujian dengan Metode <i>K-Nearest Neighbor</i>	IV-86
Tabel V- 1. Hasil Pengujian metode Random Forest (tree = 10)	V-89
Tabel V- 2. Hasil Pengujian metode Random Forest (tree = 15)	V-90
Tabel V- 3. Hasil Pengujian metode Random Forest (tree = 25)	V-90
Tabel V- 4. Hasil Pengujian metode Random Forest (tree = 50)	V-90
Tabel V- 5. Hasil Pengujian metode Random Forest (tree = 100)	V-91
Tabel V- 6. Hasil Pengujian Metode <i>K-Nearest Neighbor</i> (K = 5).....	V-91
Tabel V- 7. Hasil Pengujian Metode <i>K-Nearest Neighbor</i> (K = 25).....	V-92
Tabel V- 8. Hasil Pengujian Metode <i>K-Nearest Neighbor</i> (K = 35).....	V-92
Tabel V- 9. Hasil Pengujian Metode <i>K-Nearest Neighbor</i> (K = 45).....	V-92
Tabel V- 10. Hasil Pengujian Metode <i>K-Nearest Neighbor</i> (K = 49).....	V-93
Tabel V- 11. Tabel Confusion Matrix Random Forest.....	V-94
Tabel V- 12. Tabel Confusion Matrix <i>K-Nearest Neighbor</i>	V-95
Tabel V- 13. Perbandingan Hasil Akurasi Klasifikasi	V-97
Tabel V- 14. Perbandingan Hasil Klasifikasi.....	V-97

DAFTAR GAMBAR

Halaman

Gambar I- 1. Model Susunan decision tree	I-10
Gambar II- 2. Proses random forest	II-14s
Gambar III- 1. Digaram Tahapan Penelitian	III-22
Gambar III- 2. Diagram Kerangka Kerja.....	III-23
Gambar III- 3. Diagram Proses <i>Rational Unified Process</i> (RUP).....	III-27
Gambar IV- 1. Diagram <i>Use Case</i>	IV-47
Gambar IV- 2. Rancangan Antarmuka Panel Utama Perangkat Lunak.....	IV-56s
Gambar IV- 3. Rancangan Antarmuka Panel Pelatihan <i>Random Forest</i>	IV-56
Gambar IV- 4. Rancangan Antarmuka Panel Pelatihan <i>K-Nearest Neighbor</i>	IV-57
.....	IV-57
Gambar IV- 5. Rancangan Antarmuka Panel Pengujian <i>Random Forest</i> dan	IV-57
.....	IV-57
Gambar IV- 6. Diagram Aktivitas Memuat Data	IV-59
Gambar IV- 7. Diagram Aktivitas Pelatihan Metode <i>Random Forest</i>	IV-60
Gambar IV- 8. Diagram Aktivitas Pengujian Metode <i>Random Forest</i>	IV-61
Gambar IV- 9. Diagram Aktivitas Pelatihan Metode <i>K-Nearest Neighbor</i> ...IV-62	IV-62
Gambar IV- 10. Diagram Aktivitas Pengujian Metode <i>K-Nearest Neighbor</i>	IV-63
.....	IV-63
Gambar IV- 11. Diagram Alur Memuat Data	IV-64
Gambar IV- 12. Diagram Alur Pelatihan Metode <i>Random Forest</i>	IV-65
Gambar IV- 13. Diagram Alur Pengujian Metode <i>Random Forest</i>	IV-66
Gambar IV- 14. Diagram Alur Pelatihan Metode <i>K-Nearest Neighbor</i>	IV-67
Gambar IV- 15. Diagram Alur Pengujian Metode <i>K-Nearest Neighbor</i>	IV-68
Gambar IV- 16. Diagram Kelas	IV-70
Gambar IV- 17. Antarmuka Halaman Awal Perangkat Lunak	IV-75
Gambar IV- 18. Antarmuka Pelatihan Metode <i>Random Forest</i>	IV-75
Gambar IV- 19. Antarmuka Pelatihan Metode <i>K-Nearest Neighbor</i>	IV-76

Gambar IV- 20. Antarmuka Pengujian Metode <i>Random Forest</i> dan <i>K-Nearest Neighbor</i>	IV-76
Gambar V- 1. Grafik Hasil Klasifikasi Metode <i>Random Forest</i>	V-94
Gambar V- 2. Grafik Hasil Klasifikasi Metode <i>K-Nearest Neighbor</i>	V-96
Gambar V- 3. Grafik Perbandingan Nilai Akurasi.....	V-98
Gambar V- 4. Grafik Perbandingan Nilai Presisi.....	V-98
Gambar V- 5. Grafik Perbandingan Nilai <i>Recall</i>	V-99
Gambar V- 6. Grafik Perbandingan Nilai <i>F-Measure</i>	V-99
Gambar V- 7. Grafik Perbandingan Rata-rata Hasil Klasifikasi	V-100

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab pendahuluan akan membahas latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan. Pendahuluan dimulai dengan menjelaskan alasan melaksanakan penelitian ini dan perencanaan awal penelitian, serta susunan dan bentuk hasil penelitian.

1.2 Latar Belakang

Penyakit Diabetes Melitus adalah kelainan sistem insulin yang diakibatkan berlebihnya kadar glukosa di dalam darah. Insulin merupakan hormon yang membantu metabolisme karbohidrat. Ada beberapa faktor penyebab berlebihnya kadar glukosa dalam darah, salah satunya faktor genetika. Diabetes termasuk kedalam Penyakit Tidak Menular (PTM) yang saat ini telah menjadi ancaman serius kesehatan global.

Menurut data Organisasi Kesehatan Dunia (WHO) pada tahun 2016, 70% dari total kematian disebabkan oleh penyakit diabetes. Data terbaru yang dipublikasikan dalam *International Diabetes Federation (IDF) Atlas* edisi ke-9 menunjukkan bahwa 463 juta orang dewasa saat ini hidup dengan diabetes. Sehubungan dengan hal tersebut, Riset Kesehatan Dasar (Riskesdas) memperlihatkan peningkatan angka prevalensi diabetes yang cukup signifikan, yaitu dari 6,9% di tahun 2013 menjadi 8,5% di tahun 2018.

Gejala penyakit diabetes melitus selama ini hanya didiagnosis masyarakat awam berdasarkan ciri-ciri fisik yang diketahui tanpa didukung oleh fakta dan pertimbangan medis lainnya. Akibatnya tidak jarang penyakit-penyakit tersebut ditangani dengan cara yang salah dan kesembuhan pun tidak kunjung diraih. Alternatif yang dapat dilakukan untuk mendapatkan hasil diagnosis yang lebih baik adalah pemeriksaan laboratorium, akan tetapi cara ini relatif mahal dan butuh waktu lama untuk mengetahui hasilnya, selain itu tidak semua daerah di Indonesia memiliki laboratorium diagnosis dengan fasilitas yang memadai.

Seiring dengan kemajuan teknologi informasi terutama dalam bidang kecerdasan buatan, teknik *machine learning* diperkenalkan untuk membantu meningkatkan kemampuan pengambilan keputusan (Wahyuni, 2016). Pada penelitian ini, metode klasifikasi digunakan untuk mengelompokkan data apakah pasien terkena diabetes atau tidak, yang tujuannya untuk pencegahan awal agar tidak salah dalam penanganannya. Klasifikasi merupakan teknik pengumpulan data berdasarkan ciri atau karakteristik suatu data (Wibawa et al., 2018). Metode klasifikasi yang umum digunakan antara lain *Naive Bayes*, *K-Nearest Neighbor*, *Decision Tree*, *Random Forest*, dan *Support Vector Machines*.

Pada penelitian sebelumnya, banyak metode ciri yang telah digunakan antara lain penelitian yang dilakukan oleh Yunita (2016) yang mengklasifikasikan penyakit diabetes melitus menggunakan metode *K-Nearest Network* (K-NN). Dari hasil penelitian ini, didapatkan nilai akurasi yang cukup tinggi yakni 96%. Adapun penelitian lainnya yang menggunakan metode klasifikasi yakni penelitian yang dilakukan oleh Benbelkacem & Atmani (2019).

Dari penelitian Benbelkacem & Atmani (2019), menggunakan metode *Random Forest* untuk mendiagnosis penyakit diabetes melitus, dengan memastikan apakah metode ini efektif untuk diagnosis diabetes. Peneliti mengevaluasi performa metode *Random Forest* dengan membandingkannya dengan teknik *machine learning* lainnya yaitu C4.5, RepTree, SimpleCart, BFTree, dan SVM. Hasil yang didapatkan menunjukkan bahwa metode *Random Forest* memberikan kinerja yang lebih baik dibandingkan dengan metode lain dengan tingkat kesalahan yang paling rendah.

Selain itu juga, terdapat penelitian lain yang dilakukan oleh Singh et al., (2017) yang mengevaluasi pengaruh tiga tipe data yang berbeda terhadap kinerja pengklasifikasi data yakni *Random Forest*, *K-Nearest Neighbor*, dan *Naïve Bayes*. Dari 8 jenis data yang di klasifikasi salah satunya adalah data *Pima Indians Diabetes Data Set*, dari data tersebut didapatkan hasil bahwa kinerja *Random Forest* dan *K-Nearest Neighbor* berperilaku identik dalam hal akurasi, yakni 64,868% pada *Random Forest* dan 64.86% pada *K-Nearest Neighbor*. Sedangkan *Naïve Bayes* berkinerja lebih rendah yakni dengan hasil akurasi 50.00%.

Berdasarkan latar belakang diatas, penelitian ini akan membandingkan metode *Random Forest* dengan metode *K-Nearest Neighbor* untuk klasifikasi penyakit diabetes melitus, hasil yang diharapkan pada penelitian ini dapat mengetahui metode yang paling sesuai dalam mengklasifikasi.

1.3 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana cara mengklasifikasikan penyakit diabetes melitus menggunakan metode *Random Forest* dan metode *K-Nearest Neighbor*?
2. Berapa besar tingkat akurasi yang dihasilkan dari kedua metode tersebut dalam mengklasifikasikan penyakit diabetes melitus?

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini adalah:

1. Mengembangkan perangkat lunak menggunakan metode *Random Forest* dan *K-Nearest Neighbor* untuk mengklasifikasikan penyakit diabetes melitus.
2. Mengukur seberapa besar tingkat akurasi yang dihasilkan untuk mengetahui metode manakah yang paling sesuai dalam mengklasifikasikan penyakit diabetes melitus.

1.5 Manfaat Penelitian

Manfaat dalam pelaksanaan penelitian ini adalah:

1. Perangkat lunak yang dihasilkan dapat digunakan untuk mengklasifikasikan penyakit diabetes melitus.
2. Dapat mengetahui metode yang paling sesuai untuk klasifikasi penyakit diabetes melitus dengan menghitung nilai akurasi keduanya.
3. Hasil dan akurasi dari penelitian ini dapat dibandingkan ke penelitian atau metode lain yang sejenis.

4. Hasil penelitian dapat dijadikan sebagai sumber atau referensi bagi peneliti yang ingin membahas tentang klasifikasi penyakit diabetes melitus menggunakan metode *Random Forest* dan *K-Nearest Neighbor*.

1.6 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Penelitian ini menggunakan data penyakit pasien dari suku Indian Pima yang diperoleh dari *National Institute of Diabetes and Digestive and Kidney Diseases*¹.
2. Formula perhitungan jarak yang digunakan pada *K-Nearest Neighbor* adalah *Euclidean Distance*.

1.7 Sistematika Penulisan

Sistematika pada penulisan tugas akhir ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini akan menjelaskan pokok-pokok pikiran rencana penulisan skripsi seperti latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, serta batasan masalah.

BAB II. KAJIAN LITERATUR

Pada bab ini akan menjelaskan dasar-dasar teori yang digunakan dalam penelitian serta menguraikan hasil dari penelitian terdahulu yang relevan dengan penelitian ini.

¹ <https://www.kaggle.com/uciml/pima-indians-diabetes-database/home>

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan menjelaskan deskripsi data penelitian, tahapan-tahapan yang akan dilaksanakan pada penelitian, metode pengembangan perangkat lunak, serta manajemen dari penelitian ini.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan diuraikan tahapan pada proses pengembangan perangkat lunak dengan metode pemrograman berorientasi objek berdasarkan panduan *Rational Unified Process* (RUP).

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini akan menjelaskan hasil dan analisa hasil dari pengujian perangkat lunak. Analisis diberikan sebagai basis dari kesimpulan yang akan diambil dalam penelitian.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini akan menjelaskan mengenai kesimpulan penelitian dan saran yang diharapkan dapat berguna untuk pengembangan penelitian lain.

1.8 Kesimpulan

Pada bab ini telah dibahas mengenai penelitian yang akan dilaksanakan yaitu perbandingan metode *Random Forest* dan *K-Nearest Neighbor* dalam mengklasifikasi penyakit diabetes melitus.

DAFTAR PUSTAKA

- Benbelkacem, S., & Atmani, B. (2019, April). Random forests for diabetes diagnosis. In *2019 International Conference on Computer and Information Sciences (ICCIS)* (pp. 1-4). IEEE.
- Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.
- Byeon, H. (2021). Predicting the depression of the South Korean elderly using SMOTE and an imbalanced binary dataset. *International Journal of Advanced Computer Science and Applications*, *12*(1).
- Data, P. (2014). Informasi Kementerian Kesehatan Republik Indonesia. *Situasi dan Analisis Diabetes*. Jakarta: Kementerian Kesehatan RI.
- Kandhasamy, J. P., & Balamurali, S. J. P. C. S. (2015). Performance analysis of classifier models to predict diabetes mellitus. *Procedia Computer Science*, *47*, 45-51.
- Mujawar, I. K., Jadhav, B. T., & Patil, K. (2018). Web-based Fuzzy Expert System for Symptomatic Risk Assessment of Diabetes Mellitus. *International Journal of Computer Applications*, *182*(3), 5-12.
- Murugan, A., Nair, S. A. H., & Kumar, K. S. (2019). Detection of skin cancer using SVM, random forest and kNN classifiers. *Journal of medical systems*, *43*(8), 1-9.
- Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification

- and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), 128-138.
- Perwitasari, R., Afawani, R., & Anjarwani, S. E. (2020). Penerapan Metode Rational Unified Process (RUP) Dalam Pengembangan Sistem Informasi Medical Check Up Pada Citra Medical Centre. *Jurnal Teknologi Informasi, Komputer, dan Aplikasinya (JTika)*, 2(1), 76-88.
- Shah, K., Patel, H., Sanghvi, D., & Shah, M. (2020). A comparative analysis of logistic regression, random forest and KNN models for the text classification. *Augmented Human Research*, 5(1), 1-16.
- Singh, A., & Lakshmiganthan, R. (2018). Impact of different data types on classifier performance of random forest, naive bayes, and k-nearest neighbors algorithms.
- Tigga, N. P., & Garg, S. (2020). Prediction of type 2 diabetes using machine learning classification methods. *Procedia Computer Science*, 167, 706-716.
- Vijayarani, S., & Dhayanand, S. (2015). Data mining classification algorithms for kidney disease prediction. *Int J Cybernetics Inform*, 4(4), 13-25.
- Wahyuni, E. S. (2016). Penerapan metode seleksi fitur untuk meningkatkan hasil diagnosis kanker payudara. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 7(1), 283-294.
- Wibawa, A. P., Purnama, M. G. A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. In *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi Vol* (Vol. 3, No. 1).

- Yunita, F. (2016). Sistem Klasifikasi Penyakit Diabetes Mellitus Menggunakan Metode K-Nearest Neighbor (K-NN). *Selodang Mayang: Jurnal Ilmiah Badan Perencanaan Pembangunan Daerah Kabupaten Indragiri Hilir*, 2(1).
- Yustanti, W. (2018). Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah. *Jurnal Matematika, Statistika dan Komputasi*, 9(1), 57-68.
- Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting diabetes mellitus with machine learning techniques. *Frontiers in genetics*, 9, 515.