

**PENINGKATAN PERFORMA *DECISION TREE* DENGAN  
MENGUNAKAN *ADABOOST* UNTUK KLASIFIKASI PENYAKIT  
DIABETES**

**(Skripsi)**

**Oleh**

**ASER YOSUA RUMBRAWER  
NPM 1915061031**



**JURUSAN TEKNIK ELEKTRO  
FAKULTAS TEKNIK  
UNIVERSITAS LAMPUNG  
2025**

**PENINGKATAN PERFORMA *DECISION TREE* DENGAN  
MENGUNAKAN *ADABOOST* UNTUK KLASIFIKASI PENYAKIT  
DIABETES**

**Skripsi**

**Sebagai Salah Satu Syarat untuk  
memperoleh gelar Sarjana S1 dari  
Universitas Lampung**

**Oleh**

**ASER YOSUA RUMBRAWER  
NPM: 1915061031  
Program Studi Teknik Informatika**



**JURUSAN TEKNIK ELEKTRO  
FAKULTAS TEKNIK  
UNIVERSITAS LAMPUNG  
2025  
ABSTRAK**

**PENINGKATAN PERFORMA *DECISION TREE* DENGAN  
MENGUNAKAN *ADABOOST* UNTUK KLASIFIKASI PENYAKIT  
DIABETES**

**Oleh**

**ASER YOSUA RUMBRAWER**

Penyakit diabetes merupakan salah satu penyakit kronis yang paling umum di dunia, penyakit ini memiliki ciri khas yaitu terjadi peningkatan kadar glukosa dalam darah dan mengakibatkan gangguan produksi atau fungsi insulin tidak bekerja dengan efektif, penyakit ini memiliki beberapa tipe yaitu tipe 1 dan tipe 2. Akibat dari tidak di tangani dengan cepat dan tepat dapat membuat komplikasi serius seperti penyakit jantung, gagal ginjal, neuropati dan gangguan lain nya, sehingga di perlukan deteksi dini melalu pemeriksaan. Oleh karena itu melalui penelitian ini melakukan uji coba pengembangan prediksi diabetes dengan menggunakan algoritma pada machine learning yaitu *decision tree*. Menggunakan decision tree memiliki kelemahan dalam menghadapi overfitting dan kurang generalisasi pada dataset yang kompleks. Sehingga diperlukan adaboost (Adaptive Boosting) yang dapat menggabungkan sejumlah decision tree sederhana menjadi model yang lebih kuat dengan memberikan bobot lebih tinggi pada data yang sulit di klasifikasi. Pada penelitian ini melakukan perbandingan performa decision tree yang menunjukkan hasil akurasi 94%, meningkat menjadi 96% setelah decision tree digabungkan bersama adaboost, terjadi peningkatan juga pada data testing precision, F1-score tetapi terjadi penurunan pada recall diakibat data yang tidak seimbang. Tetapi menunjukkan program berjalan dengan baik karena menunjukkan kinerja algoritma yang baik untuk melakukan klasifikasi penyakit diabetes. Kombinasi decision tree dan adaboost juga menghasilkan peningkatan kinerja pada metrik se area under the curve (AUC), dengan tetap menjaga efisiensi komputasi

Kata kunci : Adaboost, Decision tree, Klasifikasi diabetes, Penyakit diabetes.

**ABSTRACT****IMPROVING THE PERFORMANCE OF DECISION TREE USING  
ADABOOST FOR DIABETES CLASSIFICATION****By****ASER YOSUA RUMBRAWER**

Diabetes is one of the most common chronic diseases worldwide, characterized by an increase in blood glucose levels and impaired insulin production or function, rendering it ineffective. This disease has several types, including type 1 and type 2 diabetes. If left untreated promptly and appropriately, it can lead to serious complications such as heart disease, kidney failure, neuropathy, and other health issues, highlighting the need for early detection through proper examination. This study explores the development of diabetes prediction using machine learning algorithms, specifically the Decision Tree. While Decision Tree is intuitive and effective, it struggles with overfitting and lacks generalization on complex datasets. To address these issues, Adaptive Boosting (AdaBoost) is employed, which combines multiple simple Decision Trees into a stronger model by assigning higher weights to data points that are difficult to classify. This study compares the performance of Decision Tree, which achieved an accuracy of 94%, with the combined Decision Tree and AdaBoost model, which improved to 96%. The results also showed improvements in testing precision and F1-score, although a decrease in recall was observed due to imbalanced data. Despite this, the program performed well and demonstrated the effectiveness of the algorithm for diabetes classification. The combination of Decision Tree and AdaBoost also improved performance metrics, such as the area under the curve (AUC), while maintaining computational efficiency.

Keywords: AdaBoost, Diabetes, Diabetes Classification, Decision Tree.

Judul Skripsi : **PENINGKATAN PERFORMA *DECISION TREE* DENGAN MENGGUNAKAN *ADABOOST* UNTUK KLASIFIKASI PENYAKIT DIABETES**

Nama Mahasiswa : **Aser Yosua Rumbrawer**

Nomor Pokok Mahasiswa : **1915061031**

Jurusan : **Teknik Elektro**

Fakultas : **Teknik**

**MENYETUJUI**

**1. Komisi Pembimbing**

Pembimbing Pendamping

Pembimbing Utama

  
**Ir. Titin Yulianti S.T., M.Eng.**

**NIP. 198807092019032015**

  
**Ir. Gigih Forda Nama, S.T., M.T., IPM.**

**NIP. 198307122008121003**

**2. Mengetahui**

Ketua Jurusan  
Teknik Elektro

  
**Herlinawati, S.T., M.T.**

**NIP. 197103141999032001**

Ketua Program Studi  
Teknik Informatika

  
**Yessi Mulyani, S.T., M.T.**

**NIP. 197312262000122001**

**MENGESAHKAN**

**Tim Penguji**

**Ketua : Ir. Gigih Forda Nama, S.T., M.T., IPM.**



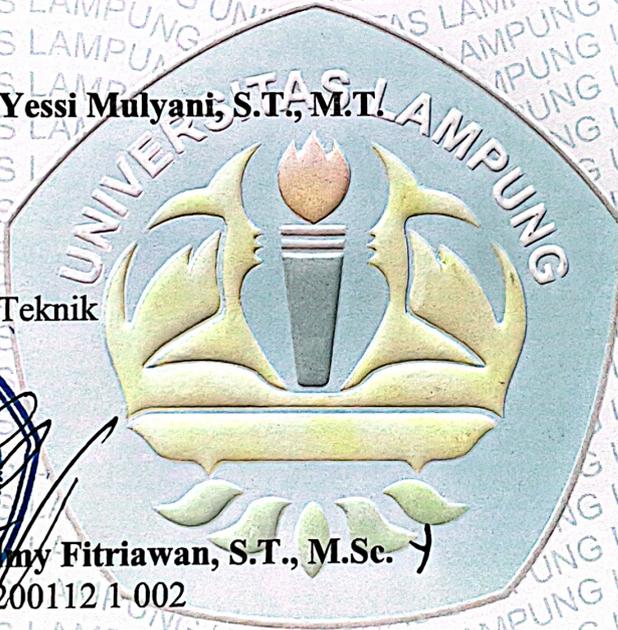
**Sekretaris : Ir. Titin Yulianti S.T., M.Eng.**



**Penguji : Yessi Mulyani, S.T., M.T.**



**Dekan Fakultas Teknik**



**Dr. Eng. Ir. Helmy Fitriawan, S.T., M.Sc.**

**NIP: 196509282001121002**

**Tanggal Lulus Ujian Skripsi : 06 Maret 2025**

## SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi saya dengan judul “Peningkatan Performa Decision Tree Menggunakan Adaboost untuk Klasifikasi Penyakit Diabetes” dibuat oleh saya sendiri. Semua hasil yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti bahwa skripsi ini merupakan salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan hukum atau akademik yang berlaku.

Bandar Lampung, 21 Maret 2025

Pembuat pernyataan,



Aser Yosua Rumbrawer

NPM. 1915061031

## RIWAYAT HIDUP



Penulis dilahirkan di Timika, pada tanggal 16 Maret 2001, sebagai anak keempat dari enam bersaudara, dari Keliopas Rumbrawer dan Ibu Regina Rejauw. Riwayat pendidikan penulis dimulai dari SD INPRES Koperapoka 1 pada tahun 2007 hingga 2013, SMP N 2 Mimika pada tahun 2013 hingga 2016, dan SMA N 1 Mimika pada tahun 2016 hingga 2019.

Tahun 2019 penulis terdaftar sebagai mahasiswa Universitas Lampung pada Program Studi Teknik Informatika, Jurusan Teknik Elektro, Fakultas Teknik melalui jalur Afirmasi Dikti. Selama menjadi mahasiswa, penulis mengikuti beberapa kegiatan, antara lain:

1. Menjadi anggota divisi Pengembangan Keteknikan dan Pengabdian Masyarakat Himpunan Mahasiswa Elektro (HIMATRO) Universitas Lampung pada tahun 2020 dan 2021.
2. Mengikuti program Kampus Merderka Studi Independen Bersertifikat di PT Cipta Konsultan Internasional dengan tema Metaverse 3D world creation + commercialization pada tahun 2022.
3. Melaksanakan Kuliah Kerja Nyata di Desa Margasari, Kecamatan Labuhan Maringgai, Kabupaten Lampung Timur, Provinsi Lampung pada tahun 2022.
4. Melaksanakan Kerja Praktik di PT Freeport Indonesia dengan posisi sebagai Data Analis pada tahun 2023.

**MOTTO**

Yin&Yang

**(Filosofi Tionghoa)**

For God everything is good

**(Penulis)**

Just keep swimming, just keep swimming, swimming swimming and swimming

**(Dory, Finding Nemo)**

There comes a day when you're gonna look around and realize happiness is where  
you are

**(Chief Tui, Moana)**

Crying help me slow down and obsess over the weight of life's problems

**(Sadness, Inside Out)**

Kita tidak bisa mengatur angin, tapi kita bisa mengatur layar

**(Sri Mulyani)**

**Puji Tuhan**

**Dengan Rahmatnya dan Kuasanya yang Besar**

**Kupersembahkan karya saya ini untuk**

*Diri Sendiri*

**ASER YOSUA RUMBRAWER**

*Orang Tua Tercinta*

**REGINA REJAUW**

**Dan**

**ELSIE RUMBRAWER**

*Keluarga Terkasih*

**RULAND RUMBRAWER**

**ESTER SALAY**

**WEMPI KROBO**

**YANA RUMBRAWER**

**EKA RUMBRAWER**

**SILA RUMBRAWER**

**NANSI RUMBRAWER**

**MARTINA RUMBRAWER**

## SANWACANA

Segala puji bagi Tuhan Yang Maha Esa, atas limpahan rahmat dan nikmat-Nya yang diberikan kepada penulis sehingga penulis dapat menyelesaikan skripsi ini. Skripsi dengan judul “Peningkatan Performa Decision Tree dengan Menggunakan Adaboost untuk Klasifikasi Penyakit Diabetes” ini merupakan salah satu syarat untuk memperoleh gelar Sarjana Teknik pada Program Studi Teknik Informatika Jurusan Teknik Elektro, Fakultas Teknik, Universitas Lampung. Pada pengerjaan skripsi ini, penulis ingin mengucapkan terimakasih kepada:

1. Bapak Dr. Eng. Ir. Helmy Fitriawan, S.T., M.Sc. selaku Dekan Fakultas Teknik Universitas Lampung.
2. Ibu Herlinawati, S.T., M.T. selaku Kepala Jurusan Teknik Elektro Universitas Lampung.
3. Ibu Yessi Mulyani, S.T.,M.T. selaku Ketua Program Studi Teknik Informatika Universitas Lampung dan telah membantu proses kelancaran pengerjaan penelitian. Dan juga sebagai dosen Penguji yang telah memberika kritik dan saran yang membangun kepada penulis bisa menyelesaikan skripsi
4. Bapak Gigih Forda Nama, S.T., M.T., IPM. selaku dosen pembimbing utama yang telah meluangkan waktunya untuk membimbing penulis dalam penyelesaian skripsi serta memberi masukan dan nasihat yang bermanfaat.
5. Ibu Titin Yulianti, S.T., M.Eng. selaku dosen pembimbing pendamping yang telah bersedia membimbing sekaligus memberikan arahan bagi penulis sehingga penulis bisa menyelesaikan skripsi.
6. Segenap Dosen di Jurusan Teknik Elektro yang telah memberikan ilmu yang bermanfaat, wawasan, dan pengalaman semasa kuliah.

7. Mbak Rika dan segenap Staff di Jurusan Teknik Elektro dan Fakultas Teknik yang telah sangat membantu penulis baik dalam hal administrasi dan hal-hal lainnya.
8. Keluarga besar yang sangat penulis cintai Keluarga Besar Rumbrawer di Timika yang selalu mendoakan yang terbaik untuk penulis serta mendukung penuh penulis selama ini.
9. Bagus Stevanus, Anggi Dwi serta teman-teman Teknik Informatika Kelas A yang telah menemani serta membantu penulis dengan memberikan saran, masukan, dan hiburan kepada penulis sejak mahasiswa baru sampai penulis menyelesaikan skripsi.
10. Angkatan ETERNITY 2019 yang telah menjadi teman seperjuangan sejak mahasiswa baru. Terimakasih telah mewarnai masa perkuliahan penulis dan menulis banyak cerita bersama dari botak hingga gondrong.
11. Rex Orange Country yang telah membantu dan membangun semangat kepada penulis untuk menyelesaikan skripsi melalui lagu-lagu yang dinyanyikan
12. Teman-teman badminton yang mendukung penulis dengan berolahraga Bersama bermain badminton.
13. Afiriasi Lampung 2019 yang paling heboh yang menghibur dan mendukung penulis.

Penulis menyadari bahwa terdapat banyak kekurangan dalam penulisan skripsi ini. Penulis mengharap kritik dan saran yang membangun dari semua pihak. Penulis berharap skripsi ini dapat bermanfaat bagi kita semua.

Bandar Lampung, 21 Maret 2025

Aser Yosua Rumbrawer

## DAFTAR ISI

BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Sistematika Penulisan Skripsi .....	3
BAB II TINJAUAN PUSTAKA .....	5
2.1 Penyakit Diabetes .....	5
2.2 <i>Machine Learning</i> .....	7
2.3 Decesion Tree.....	9
2.4 Adactive Boosting ( <i>Adaboost</i> ) .....	11
2.5 Confusion Matrix .....	13
2.7 Phyton.....	14
2.9 Penelitian Terdahulu.....	16
BAB III METODE PENELITIAN .....	21
3.1 Tempat dan Waktu Penelitian .....	21
3.1.1 Tempat Penelitian .....	21
3.1.2 Waktu Penelitian.....	21
3.2 Alat dan Bahan.....	22
3.2.1 Bahan.....	22
3.2.2 Alat.....	23
3.3 Tahapan Penelitian .....	24
3.4 Desain Sistem.....	26
BAB IV HASIL DAN PEMBAHASAN .....	32
4.1 Pemahaman Bisnis.....	32
4.2 Selection Data .....	33
4.2.1 Persiapan Data .....	33

4.3 Preprocessing Data .....	34
4.3.1 Persiapan data .....	39
4.4 Transformasi Data .....	40
4.4 Persiapan Data Mining.....	49
4.4.1 Splitting data .....	49
4.4.2 Klasifikasi.....	50
4.5 Evaluation .....	56
4.5.1 hasil klasifikasi Testing menggunakan Decision Tree.....	56
4.5.2 Hasil Testing klasifikasi menggunakan Decision tree Adaabost .....	58
4.5.3 Perbandingan Kinerja Algoritma Decision Tree dan Decision Tree Adaboost .....	62
4.7 Pembahasan.....	63
Bab V Kesimpulan dan saran .....	66
5.1 Kesimpulan .....	66
5.2 Saran .....	66
Daftar Pustaka .....	67

## DAFTAR GAMBAR

	Halaman
Gambar 1. 1 Data risdiskes .....	2
Gambar 2. 1 Machine Learning.....	7
Gambar 2. 2 Decision tree.....	9
Gambar 3. 1 Tahapan penelitian.....	26
Gambar 3. 2 Desain penelitian KDD .....	26
Gambar 3. 3 flowchart decision tree.....	29
Gambar 3. 4 flowchart decision tree adaboost .....	30
Gambar 4. 1 Tampilan data .....	34
Gambar 4. 2 Tampilan data bersih.....	35
Gambar 4. 3 Jumlah dan konsistensi data .....	35
Gambar 4. 4 Penghapusan data other.....	35
Gambar 4. 5 Contoh data duplikat .....	36
Gambar 4. 6 Data duplikat .....	37
Gambar 4. 7 Pengehapusan data duplikat .....	37
Gambar 4. 8 Data hilang .....	38
Gambar 4. 9 Encode Gender .....	38
Gambar 4. 10 Data categorical .....	38
Gambar 4. 11 Nilai encode.....	39
Gambar 4. 12 Distribusi data diabetes berdasarkan jumlah pasien .....	41
Gambar 4. 13 Distribusi data hypertension dan BGL.....	41
Gambar 4. 14 Distribusi usia .....	42
Gambar 4. 15 Distribusi riwayat merokok.....	43
Gambar 4. 16 Distribusi data riwayat merokok terhadap gender .....	44
Gambar 4. 17 Distribusi HbA1c_level dan blood glucose level .....	45
Gambar 4. 18 Distribusi data BMI dan Blood Glucose Level .....	46
Gambar 4. 19 Distribusi usia berdasarkan hypertension.....	47
Gambar 4. 20 Distribusi hypertension dan hearth disease .....	48
Gambar 4. 21 Distribusi data gender dan blood glucose, serta BMI.....	49
Gambar 4. 22 Library pandas Numpy.....	50
Gambar 4. 23 Proses klasifikasi menggunakan decision tree .....	50
Gambar 4. 24 Model data training .....	51
Gambar 4. 25 Visualisasi decision tree.....	52
Gambar 4. 26 Proses klasifikasi pemodelan adaboost .....	53
Gambar 4. 27 Tampilan data training decision tree.....	54
Gambar 4. 28 Visualisai decision tree adaboost.....	55
Gambar 4. 29 program decision tree.....	56

Gambar 4. 30 Grafik ROC .....	58
Gambar 4. 31 program decision tree dengan adaboost .....	59
Gambar 4. 32 Grafik ROC adaboost.....	61
Gambar 4. 33 diagram perbandingan decision tree .....	63

## DAFTAR TABEL

	Halaman
Tabel 2. 1 Confusion matrix.....	13
Tabel 2. 2 Penelitian terdahulu.....	17
Tabel 3. 1 Timeline.....	21
Tabel 3. 2 Atribut.....	23
Tabel 4. 1 Atribut dataset rekaman pasien.....	33
Tabel 4. 2 Data atribut yang digunakan.....	39
Tabel 4. 3 Proporsi data training dan testing.....	50
Tabel 4. 4 Confusion matrix data training decision tree.....	50
Tabel 4. 5 Confusion matrix dari testing decision tree.....	57
Tabel 4. 6 data testing decision Tree.....	57
Tabel 4. 7 Confusion matrix testing decision tree adaboost.....	59
Tabel 4. 8 Hasil data testing decision tree adaboost.....	60
Tabel 4. 9 Hasil perbandingan kerja model.....	62

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Diabetes merupakan penyakit medis yang ditandai dengan tingginya kadar gula darah pada penderitanya, penyakit diabetes merupakan salah satu faktor utama penyebab penyakit jantung koroner dan stroke, [1] Diabetes melitus adalah penyakit gangguan metabolik yang disebabkan oleh gagalnya organ pankreas dalam memproduksi hormon insulin secara memadai. Penyakit ini bisa dikatakan sebagai penyakit kronis karena dapat terjadi secara menahun. Berdasarkan penyebabnya diabetes melitus di golongan menjadi tiga jenis, diantaranya diabetes melitus tipe 1, tipe 2 dan diabetes melitus gestasional

Diabetes melitus didefinisikan sebagai penyakit atau kelainan metabolisme kronis dengan berbagai etiologi yang ditandai dengan tingginya kadar gula darah disertai gangguan metabolisme karbohidrat, lipid, dan protein sebagai akibat dari insufisiensi fungsi insulin. Insufisiensi fungsi insulin dapat disebabkan oleh gangguan atau kekurangan produksi insulin oleh sel beta Langerhans kelenjar pankreas, atau disebabkan oleh kurangnya respon sel-sel tubuh terhadap insulin [2].

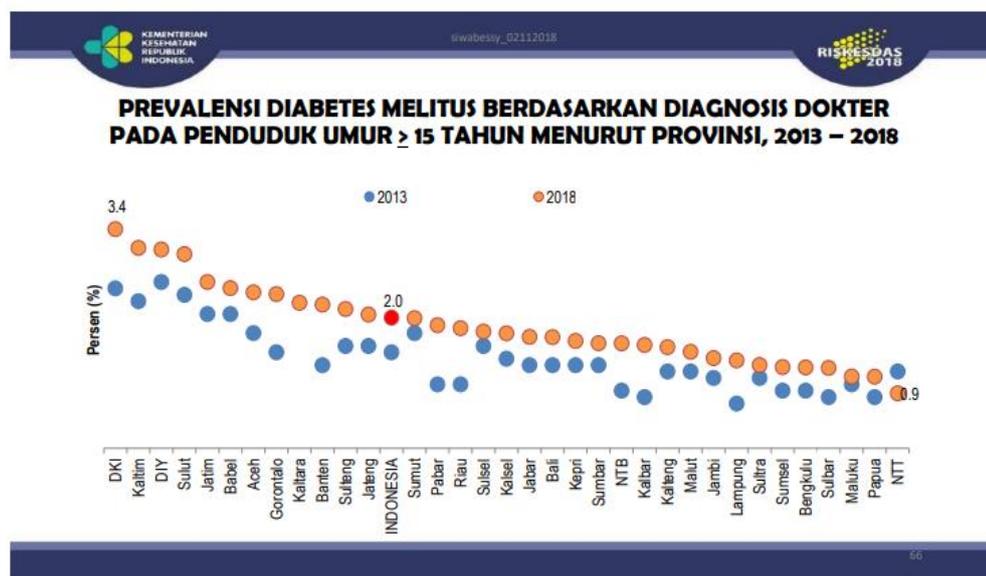
Peningkatan angka prevalensi Diabetes melitus yang cukup signifikan, yaitu dari 6,9% di tahun 2013 menjadi 8,5% di tahun 2018; sehingga estimasi jumlah penderita di Indonesia mencapai lebih dari 16 juta orang yang kemudian berisiko terkena penyakit lain, seperti: serangan jantung, stroke, kebutaan dan gagal ginjal bahkan dapat menyebabkan kelumpuhan dan kematian. [3]

Saat ini penderita kencing manis penyakit adalah satu jenis penyakit diabetes mematikan ketiga di Indonesia setelah stroke dan jantung sekitar 10 juta orang.

Jumlahnya sekitar 10 tahun mendatang dapat meningkat dua sampai tiga kali lipat, [4]

Jumlah penderita diabetes terus meningkat dan tercatat saat ini mencapai 422 juta orang di dunia empat kali lebih banyak dari pada 30 tahun lalu, menurut Organisasi Kesehatan Dunia (World Health Organization).

Menurut Riskesdas pada tahun 2018 diabetes melitus dengan usia penduduk lebih dari 15 tahun yang telah diagnosis oleh dokter memiliki kenaikan grafik dari tahun 2013-2018 disetiap provinsi di Indonesia.



Gambar 1. 1 Data risdiskes

Salah satu masalah utama diabetes di Indonesia adalah keterlambatan diagnosis, sebanyak 73,5 persen penderita diabetes tidak menyadari dirinya sakit [5]. Pengetahuan yang kurang mengenai gejala dan cara menangani penyakit diabetes melitus, serta penggunaan teknologi yang kurang memadai dapat juga mempengaruhi penanganan pada pasien yang gejalanya menunjukkan seseorang mengidap penyakit diabetes, sehingga keterlambatan penanganan mengakibatkan seseorang dapat dengan cepat mengidap komplikasi penyakit seperti serangan jantung dan stroke yang dapat berujung pada kematian,

Pada penelitian ini menguji performa algoritma pada klasifikasi penyakit diabetes dengan menggunakan *Decision tree* yang di kolaborasikan bersama *Adaboost* untuk menguatkan hasil akurasi terhadap data diabetes tersebut. Sehingga diharapkan menghasilkan keputusan yang dapat membantu memprediksi seseorang yang memiliki gejala diabetes dengan cepat.

## **1.2 Rumusan Masalah**

Adapun Rumusan Masalah sebagai berikut.

1. Bagaimana menggunakan *Decision tree* dalam mengklasifikasi Penyakit Diabetes Melitus
2. Bagaimana meningkatkan performa akurasi *Decision tree* dengan menggunakan *Adaboost*
3. Bagaimana hasil evaluasi akurasi terhadap *Decision tree* yang telah menggunakan *Adaboost*

## **1.3 Tujuan Penelitian**

Adapun tujuan pada penelitian ini adalah sebagai berikut.

1. Mengoptimasi *Decision tree* menggunakan Metode *Adaboost* untuk peningkatan hasil akurasi.
2. Mengevaluasi hasil kinerja klasifikasi *Decision tree* menggunakan metode *Adaboost* pada penderita penyakit Diabetes Melitus.

## **1.4 Batasan Masalah**

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Algoritma yang digunakan pada penelitian ini hanya menggunakan *Decision tree* dan *Adaboost*
2. Tidak melakukan pengembangan interface

## **1.5 Sistematika Penulisan Skripsi**

Adapun sistematika Penulisan Skripsi yang terdiri dari 5 (lima) bab yang dijabarkan sebagai berikut.

## BAB I PENDAHULUAN

Memuat latar belakang yang menjelaskan mengenai presentasi penderita diabetes di dunia dan di Indonesia, serta permasalahan penelitian yaitu, bagaimana memanfaatkan machine learning untuk membantu mengambil keputusan cepat untuk diagnosa diabetes. Perumusan masalah, menjelaskan bagaimana performa *Decision tree* menggunakan metode *Adaboost*. diharapkan menghasilkan akurasi yang tepat tujuan penelitian, untuk membantu pengambilan keputusan dengan akurasi yang tepat dan manfaat penelitian juga terdapat di bab ini.

## BAB II TINJAUAN PUSTAKA

Memuat dasar teori mengenai penyakit diabetes, machine learning, *Decision tree*, Adaptive Boosting(*Adaboost*), Interface dan Processing serta penelitian yang telah dilakukan dahulu sebagai referensi dalam penelitian ini.

## BAB III METODOLOGI PENELITIAN

Memuat waktu dan tempat penelitian, waktu penelitian Januari-Maret yang akan dilakukan dilaboratorium komputer Jurusan Teknik Elektro. Tahapan penelitian secara umum studi literasi, Pengumpulan data, pra processing, pemodelan *Decision tree* klasifikasi, *Decision tree* menggunakan *Adaboost* klasifikasi dan tahapan penelitian secara terperinci.

## BAB IV PEMBAHASAN

Memuat hasil dari setiap tahapan klasifikasi data diabetes menggunakan *Adaboost* pada decision tree seperti Accuracy, Precision, Recall, F1-score, Spesifisitas

## BAB V KESIMPULAN DAN SARAN

Memuat kesimpulan dari kegiatan yang dilakukan pada penelitian peningkatan performa *Decision tree* menggunakan metode Adabost. Saran-saran mengenai hasil evaluasi nilai akurasi untuk melakukan perbaikan dan pengembangan lebih lanjut.

## **BAB II TINJAUAN PUSTAKA**

### **2.1 Penyakit Diabetes**

Diabetes Melitus merupakan sekelompok penyakit metabolik yang ditandai dengan kadar gula darah melebihi nilai normal. Diabetes adalah suatu penyakit. Dimana tubuh penderitanya secara otomatis tidak dapat mengontrol kadar gula darah. Dalam tubuh yang sehat, pankreas melepaskan hormon insulin yang bertugas mengangkut gula melalui darah ke otot dan jaringan lain untuk memasok energi. Penderita diabetes tidak dapat memproduksi insulin dalam jumlah yang cukup atau tubuh tidak dapat menggunakan insulin secara efektif sehingga mengakibatkan kelebihan gula dalam darah. Kelebihan gula kronis dalam darah bersifat racun bagi tubuh [6]. Diabetes melitus adalah kondisi di mana tubuh tidak dapat mengatur kadar glukosa dalam darah dengan baik. Hal ini bisa disebabkan oleh:

Kurangnya produksi insulin: Pankreas tidak memproduksi cukup insulin.

Resistensi insulin: Tubuh tidak dapat menggunakan insulin secara efektif.

Gangguan hormonal: Hormon lain dapat menghambat fungsi insulin.

Kondisi ini menyebabkan hiperglikemia, yang jika dibiarkan dapat menimbulkan berbagai komplikasi serius seperti kerusakan pada organ tubuh, termasuk jantung, ginjal, dan mata.

#### **Ciri-Ciri dan Gejala Diabetes**

Gejala diabetes sering kali tidak terlihat pada tahap awal, terutama pada diabetes tipe 2. Beberapa ciri-ciri gejala pada diabetes

Sering buang air kecil (poliuria): Penderita diabetes cenderung lebih sering buang air kecil, terutama di malam hari.

Gampang haus (polidipsia): Rasa haus yang berlebihan yang tidak hilang meskipun sudah minum cukup air.

Cepat lapar (polifagia): Meskipun makan dengan cukup, penderita merasa lapar terus-menerus.

Berat badan turun drastis: Penurunan berat badan yang tidak diinginkan meskipun pola makan tetap sama.

Kelelahan: Merasa lelah dan lesu tanpa alasan yang jelas.

Perubahan penglihatan: Pandangan kabur atau perubahan lainnya dalam penglihatan.

Kulit kering atau gatal: Kulit mungkin menjadi kering dan gatal akibat dehidrasi.

Luka yang sulit sembuh: Luka atau infeksi yang memerlukan waktu lama untuk sembuh.

Gejala-gejala ini dapat bervariasi antara diabetes tipe 1 dan tipe 2. Pada diabetes tipe 1, gejala biasanya muncul lebih cepat dalam beberapa minggu, sedangkan pada tipe 2, gejala mungkin berkembang secara perlahan dan sering kali tidak terdeteksi selama bertahun-tahun

Tipe Diabetes mellitus menurut American Diabetes Association, 2010 adalah sebagai berikut.

#### a. Diabetes Melitus tipe I

Remaja lebih mungkin terkena diabetes melitus tipe I, yang sering dikenal sebagai diabetes tergantung insulin. Lebih dari 90% sel penghasil insulin di pankreas mengalami kerusakan permanen. Akibatnya, sedikit atau bahkan tidak ada insulin yang diproduksi dalam waktu singkat. Hanya sekitar 10% dari seluruh penderita Diabetes Melitus yang menderita diabetes tipe I. Mayoritas penderita diabetes tipe I berusia di bawah 30 tahun.

#### b. Diabetes Melitus tipe II

Diabetes melitus tipe II ( Diabetes Non Insulin Dependent) ini tidak ada kerusakan pada pankreasnya dan dapat terus menghasilkan insulin, bahkan kadang-kadang insulin pada tingkat tinggi dari normal. Akan tetapi, tubuh manusia resisten terhadap efek insulin, sehingga tidak ada insulin yang cukup untuk memenuhi

kebutuhan tubuh. Diabetes mellitus tipe ini sering terjadi pada dewasa yang berumur lebih dari 30 tahun dan menjadi lebih umum dengan peningkatan usia.

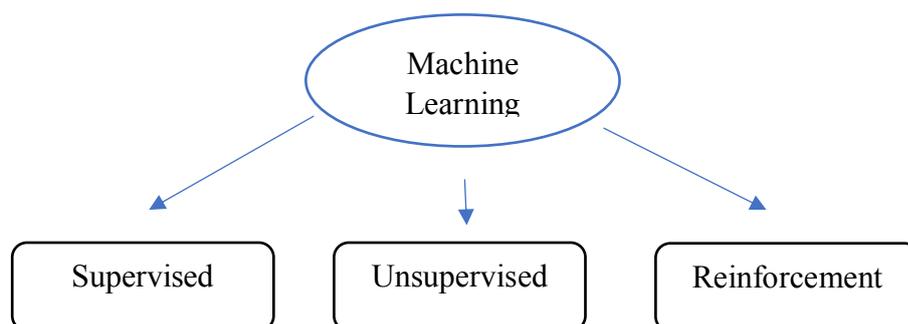
### c. Diabetes Mellitus Gestasional

Diabetes mellitus gestasional adalah diabetes yang timbul selama kehamilan. Ini meliputi 2-5% daripada seluruh diabetes. Jenis ini sangat penting diketahui karena dampaknya pada janin kurang baik bila tidak ditangani dengan benar[4]. Suatu kondisi yang memiliki tingkat toleransi glukosa berbeda-beda yang awalnya diidentifikasi selama kehamilan disebut diabetes gestasional. Meskipun kadar glukosa dapat kembali normal setelah melahirkan, sebagian besar wanita hamil dengan diabetes mellitus gestasional memiliki homeostatis glukosa yang relatif normal selama lima bulan pertama kehamilannya dan mungkin juga mengalami kekurangan insulin relatif selama trimester kedua [4].

## 2.2 Machine Learning

Teknik komputer berdasarkan pengalaman untuk meningkatkan kinerja atau menghasilkan prediksi yang tepat dikenal sebagai *Machine Learning*. Pengalaman dalam konteks ini diartikan sebagai pengetahuan sebelumnya yang dapat diakses dan diterapkan pada data siswa. [7] berikut adalah gambar klasifikasi pada machine learning yang membagi data berwarna merah dan data berwarna biru.

Berikut adalah contoh situasi dalam pembelajaran mesin::



Gambar 2. 1 Machine Learning

1. *Supervised Learning* Penggunaan skenario supervised learning, pembelajaran menggunakan masukan data pembelajaran yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label. Model mempelajari hubungan

antara input (variabel independen  $X$ ) dan output (variabel target  $Y$ ) untuk memprediksi hasil dari data baru yang belum dilihat sebelumnya.

### Komponen Utama Supervised Learning

Data yang digunakan berisi pasangan fitur (input)  $x_i$  dan label (output)  $y_i$ . Label dapat berupa nilai kontinu untuk regresi atau kategori diskrit untuk klasifikasi.

Model adalah representasi matematis yang memetakan  $X$  ke  $Y$ . Contoh model meliputi regresi linier, Decision Tree, dan Support Vector Machines (SVM). Model dilatih untuk meminimalkan fungsi kesalahan (loss function), seperti mean squared error (MSE) atau cross-entropy loss, yang mengukur seberapa baik prediksi model terhadap data pelatihan.[8]

2. *Unsupervised Learning* Penggunaan skenario unsupervised learning, pembelajaran menggunakan masukan data pembelajaran yang tidak diberi label. Setelah itu mencoba untuk mengelompokkan data berdasarkan karakteristik-karakteristik yang ditemui.

3. *Reinforcement learning* Pada skenario reinforcement learning fase pembelajaran dan tes saling dicampur. Untuk mengumpulkan informasi pembelajar secara aktif dengan berinteraksi ke lingkungan sehingga untuk mendapatkan balasan untuk setiap aksi dari pembelajar.

*Decision tree* adalah model Machine Learning yang digunakan untuk pengambilan keputusan. Ini adalah struktur pohon dengan simpul yang mewakili keputusan atau pemisahan data berdasarkan fitur-fitur tertentu. *Decision tree* digunakan untuk menggambarkan dan memecah masalah keputusan menjadi serangkaian keputusan yang lebih kecil dan lebih mudah diinterpretasikan. *Adaboost* adalah algoritma ensemble learning yang digunakan untuk meningkatkan kinerja model Machine Learning. *Adaboost* bekerja dengan menggabungkan beberapa model lemah (weak learners), seperti *Decision tree* yang dangkal, dan memberikan bobot (weight) lebih pada data yang salah diklasifikasikan pada iterasi sebelumnya.

Dalam penelitian ini *Decision tree* dan *Adaboost* masuk kedalam skenario Supervised Learning, yang memiliki label pada data sebelum di klasifikasikan menjadi sebuah hasil keputusan.

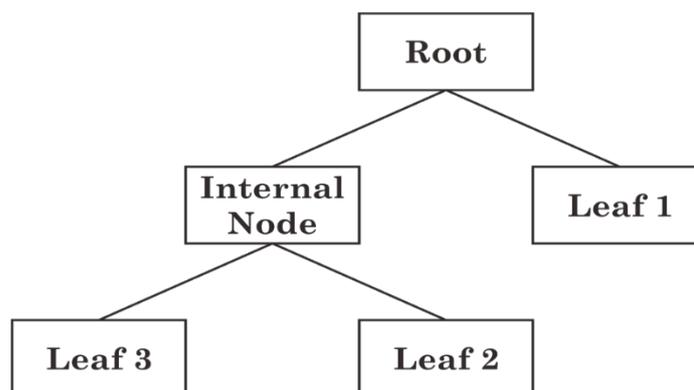
### 2.3 Decision Tree

*Decision tree* adalah algoritma pembelajaran mesin terawasi yang kuat dan banyak digunakan untuk tugas klasifikasi dan regresi. Ini mewakili struktur seperti diagram alur di mana setiap node internal mewakili keputusan pada suatu atribut, setiap cabang mewakili hasil dari keputusan tersebut, dan setiap node daun mewakili label kelas (dalam kasus klasifikasi) atau nilai numerik (dalam kasus regresi). Pohon keputusan bersifat intuitif dan mudah dipahami, menjadikannya populer untuk memvisualisasikan dan menafsirkan proses pengambilan keputusan yang kompleks. Alat prediksi dan klasifikasi yang awalnya digunakan dalam perhitungan manual sebelum era digital. Diagram berbentuk pohon ini memberikan hasil yang sangat intuitif dan mudah dipahami.

Struktur Pohon: Data dibagi secara rekursif berdasarkan atribut tertentu untuk menghasilkan partisi (node).

Root node: Node awal yang mewakili data secara keseluruhan.

Leaf node: Node akhir yang mengandung data yang homogen dalam atribut target. Pembentukan pohon memilih atribut terbaik untuk membagi data pada setiap level berdasarkan kekuatan partisi. Melakukan partisi secara rekursif hingga kondisi penghentian terpenuhi, seperti minimnya keuntungan statistik atau jumlah data. Berdasarkan algoritma seperti ID3, C4.5, atau Random Forests. Pendekatan resampling seperti bootstrap dan boosting digunakan untuk meningkatkan akurasi.[8]



Gambar 2. 2 Decision tree

a) Simpul Akar (Root)

Titik awal percabangan berupa simpul pengujian pertama yang membuat cabang keputusan.

b) Simpul Perantara (Internal Node)

Simpul pengujian Selanjutnya yang membuat cabang keputusan lanjutan.

c) Simpul Daun (Leaf)

Keputusan akhir yang didapatkan setelah melewati simpul akar dan perantara.

Gini Impurity (Untuk CART - Classification and Regression Trees): Gini Impurity mengukur seberapa seragam atau murni sebuah node dalam konteks klasifikasi. Ini dihitung dengan rumus berikut:

$$Gini(D) = 1 - \sum_{i=1}^k p_i^2 \quad (1)$$

Keterangan:

$p_i$  adalah proporsi data yang termasuk dalam kelas ke-I dalam data set D.

K adalah jumlah kelas yang ada.

Information Gain (IG)

Information Gain adalah selisih antara ketidakpastian (entropy atau impurity) sebelum dan setelah pembagian data pada sebuah atribut.

$$Information\ Gain(D,A) = Gini(D) - \sum_{u \in Values(A)} \frac{|D_u|}{|D|} \cdot Gini(D_u) \quad (2)$$

Keterangan ;

D adalah data set yang akan dibagi

A adalah atribut yang sedang dipertimbangkan untuk pembagian.

Values (A) adalah semua nilai u untuk atribut A

$D_u$  adalah subset data yang memiliki nilai u untuk atribut A

$|D_u|$  adalah ukuran subset  $D_u$  dan  $|D|$  adalah ukuran total dari dataset D.

Entropy

Entropy (Untuk C4.5): Entropy mengukur ketidakpastian atau kebingungan dalam node dalam konteks klasifikasi. Ini dihitung dengan rumus berikut:

Rumus Entropy untuk suatu set data  $D$  yang memiliki  $k$  kelas adalah sebagai berikut:

$$\text{Entropy}(D) = - \sum_{i=1}^k p_i \log_2(p_i) \quad (3)$$

Keterangan:

$p_i$  adalah proporsi data yang termasuk dalam kelas ke  $i$  dalam set data  $D$ .

$k$  adalah jumlah kelas yang ada.

## 2.4 Adaptive Boosting (*Adaboost*)

Adaptive Boosting adalah algoritma pembelajaran mesin yang termasuk dalam keluarga algoritma boosting. Algoritma ini dirancang untuk meningkatkan kinerja model lemah (weak learner) menjadi model yang kuat (strong learner) dengan cara mengatasi kesalahan yang dibuat oleh model sebelumnya.

Metode Classification and Regression Trees (CART) merupakan salah satu metode yang dapat digunakan sebagai base learning pada metode Adaboost. CART merupakan metode non parametrik untuk melakukan analisis regresi dan klasifikasi, melalui pembentukan pohon keputusan. CART dikenal sebagai metode pengurutan rekursif biner [9], artinya sekelompok data dalam simpul pohon dipilah menjadi dua simpul anak kemudian masing-masing simpul anak dipilah lagi menjadi dua simpul anak, dan seterusnya sampai berhenti karena kriteria tertentu.

*Adaboost* dan variannya telah sukses diterapkan pada beberapa bidang (domain) karena dasar teorinya yang kuat, prediksi yang akurat, dan kesederhanaan yang besar[10]. Langkah-langkah algoritma *Adaboost* adalah sebagai berikut :

Input : suatu kumpulan sample penelitian dengan label  $\{(x_i, y_i), \dots, (x_N, y_N)\}$ , suatu component learn algoritma, jumlah perputaran  $T$ .

Intialize : bobot suatu sample pelatihan  $W_i^1 = 1/N$ , untuk semua  $i = 1, \dots, N$ .

For  $t = 1, \dots, T$ .

Gunakan component learn algoritma untuk melatih suatu komponen klasifikasi  $h_t$ , pada sample bobot pelatihan.

Hitung kesalahan pelatihannya pada

$$h_t, \varepsilon_t = \sum_{i=1}^N W_i^t, y_i \neq h_t(x_i).$$

Tetapkan bobot untuk component classifier

$$h_t = \alpha_t \frac{1}{2} \ln \left( \frac{1 - \varepsilon_t}{\varepsilon_t} \right).$$

Update bobot sample pelatihan

$$W_i^{t+1} = \frac{W_i^t \exp\{-a_t y_i h_t(x_i)\}}{C_t}, \quad i = 1, \dots, N \quad C_t \text{ adalah salah satu konstanta normalisasi.}$$

$$\text{Output } f(x) = \text{sign}\left(\sum_{t=1}^T a_t h_t(x)\right).$$

Berikut adalah gambaran umum tentang cara *Adaboost* bekerja:

a. Model Lemah (Weak Learner)

*Adaboost* dimulai dengan menggunakan model lemah sebagai model dasar. Model lemah ini bisa menjadi model sederhana yang memiliki kinerja sedikit di atas tingkat acak, seperti pohon keputusan dangkal (decision stump).

b. Bobot Data

Setiap contoh data dalam set pelatihan diberi bobot awal. Bobot ini digunakan untuk memberikan penekanan lebih pada contoh yang salah diklasifikasikan oleh model lemah sebelumnya.

c. Pelatihan Berulang

*Adaboost* melatih model lemah berulang kali pada set data yang sama. Pada setiap iterasi, bobot diperbarui berdasarkan kinerja model sebelumnya. Contoh yang salah diklasifikasikan diberi bobot lebih besar, sehingga fokus lebih besar pada contoh tersebut pada iterasi berikutnya.

d. Gabungan Model

Setiap model lemah yang dihasilkan pada setiap iterasi diberi bobot sesuai dengan kinerjanya. Bobot ini mencerminkan seberapa baik model tersebut dapat mengklasifikasikan contoh-contoh dalam set pelatihan.

#### e. Voting Weighted

Pada tahap pengujian (testing), model-model lemah diuji pada contoh-contoh baru. Prediksi dari setiap model lemah dijumlahkan dengan bobotnya, dan hasilnya diambil berdasarkan hasil voting yang terponderasi.

## 2.5 Confusion Matrix

Salah satu teknik yang sering digunakan untuk menentukan kebenaran konsep data mining adalah confusion matrix. Tabel yang mencantumkan jumlah hasil tes yang teridentifikasi dengan benar dan salah berfungsi sebagai deskripsi confusion matrix. [11].

Tabel 2. 1 Confusion matrix

		Nilai Aktual	
		Positive	Negative
Nilai Prediksi	Positive	TP	FP
	Negative	FN	TN

Tabel diatas memiliki penjelasan sebagai berikut

True Positive (TP) adalah jumlah data dengan nilai sebenarnya positif dan nilai prediksi Positif

False Positive (FP) adalah jumlah data dengan nilai sebenarnya negatif dan nilai prediksi positif

False Negative (FN) adalah jumlah data dengan nilai sebenarnya positif dan nilai prediksi negatif

True Negative (TN) adalah jumlah data dengan nilai sebenarnya negatif dan nilai prediksi negatif

Accuracy yaitu presentase seberapa akuratnya suatu model dalam mengklasifikasikan (prediksi) secara benar oleh algoritma. Fungsi dari accuracy dapat dilihat dari persamaan

$$\text{Accuracy} = \frac{TP+TN}{(TP+FP+TN+FN)} \quad (1)$$

Recall (sensitivitas) mengukur seberapa banyak dari total instans positif yang berhasil ditemukan oleh model.

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (2)$$

Precision (presisi) mengukur seberapa banyak dari instans yang diprediksi sebagai positif oleh model yang benar-benar positif.

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (3)$$

F-Score adalah ukuran komposit yang menggabungkan precision dan recall. F-Score sering digunakan ketika kita ingin mencapai keseimbangan antara precision dan recall.

$$\text{F-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{(\text{Precision} + \text{Recall})} \quad (4)$$

Spesifisitas mengukur seberapa baik model dapat memprediksi negatif sejati.

$$\text{Spesifisitas} = \frac{TN}{(TN+FP)} \quad (5)$$

## 2.7 Python

Python adalah bahasa pemrograman interpretatif yang dianggap mudah dipelajari serta berfokus pada keterbacaan kode. Dengan kata lain python diklaim sebagai bahasa pemrograman yang memiliki kode-kode pemrograman yang sangat jelas dan mudah dipahami, lengkap, dan mudah. Python secara umum berbentuk pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman

fungsional. Python juga dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi. [12] library yang akan digunakan pada penelitian dijelaskan sebagai berikut.

- a. Jupyter Notebook. Jupyter Notebook adalah lingkungan pengembangan interaktif yang memungkinkan Anda membuat dan berbagi dokumen berisi kode, teks naratif, visualisasi, dan hasil eksekusi kode. Nama "Jupyter" berasal dari gabungan tiga bahasa pemrograman yang didukung oleh lingkungan ini, yaitu Julia, Python, dan R.
- b. Scikit-learn adalah proyek sumber terbuka, artinya bebas untuk digunakan dan didistribusikan, dan siapa pun dapat dengan mudah memperoleh kode sumber untuk melihat apa yang terjadi di balik adegan [13] beberapa parameter yang dapat digunakan untuk klasifikasi *Decision tree*
  - criterion: Kriteria pemilihan fitur. 'gini' untuk indeks Gini atau 'entropy' untuk gain informasi.
  - splitter: Strategi pemilihan pemisah. 'best' untuk memilih pemisah terbaik atau 'random' untuk memilih pemisah secara acak.
  - max\_depth: Kedalaman maksimum pohon. Jika None, pohon akan diperluas sampai semua simpul daun berisi kurang dari min\_samples\_split sampel.
  - min\_samples\_split: Jumlah sampel minimum yang diperlukan untuk membagi simpul internal.
  - min\_samples\_leaf: Jumlah sampel minimum yang diperlukan di simpul daun.
  - max\_features: Jumlah fitur yang harus dipertimbangkan saat mencari pemisah terbaik. Jika None, semua fitur akan dipertimbangkan.
  - random\_state: Digunakan untuk mengontrol reproduksibilitas hasil. Jika diatur, hasil akan tetap sama pada setiap pelatihan.
- c. pandas adalah pustaka Python untuk perselisihan dan analisis data. Itu dibangun di sekitar struktur data yang disebut DataFrame yang dimodelkan setelah R DataFrame.[13]. Pandas digunakan untuk membantu pra-pemrosesan data dan manipulasi fitur serta target. Model *Decision tree* dengan *Adaboost* kemudian dibuat dan dilatih menggunakan data latih, dan akurasi model diukur menggunakan data uji. dapat menyesuaikan parameter

seperti *max\_depth* pada model *Decision tree* dan *n\_estimators* pada model *Adaboost*.

## 2.9 Penelitian Terdahulu

Beberapa penelitian yang telah dilakukan mengenai machine learning ya itu tentang algortima *Decision tree* dan *Adaboost* dengan beberapa jenis data yang berbeda-beda.

Ginanjari Abdurrahman melakukan penelitian menggunakan metode *Adaboost* classifier dan mendapatkan hasil akurasi sebesar 80,09%[14]. Adapun Lidia Pebrianti, Fitrihuda Aulia, Halimatun Nisa dan Kana Saputra S melakukan penelitian dengan menggunakan data berupa tabular dari kondisi kesehatan pasien yang terindikasi diabetes maupun tidak, dengan metode *Adaboost* dan algoritma naïve bayes mendapatkan hasil akurasi optimasi *Adaboost* dan naïve bayes senilai 0.7694[15]. Hasbi Firmansyah, dan Zainal Abidin melakukan penelitian menggunakan data podes untuk mengklasifikasi status sebuah desa penelitian ini menggunakan metode *Adaboost* dengan algoritma *Decision tree* dan mendapatkan hasil akurasi 88,91%[16]. Pada penelitian yang di lakukan oleh Zico Karya Saputra Domas, dan Roby Rakhmad yang menambahkan algoritma decision terhadap *Adaboost* dengan dataset kurangtransparanan pengungkapan informasi anti-korupsi sebagai data untuk di klasifikasi mendapatkan hasil akurasi meningkat menjadi 0,768[17].

Data kelulusan Mahasiswa Program Studi Statistika Universitas Tanjungpura dengan jumlah data 140, Yuveinsiana Crismayella, Neva Satyahadewi, Hendra Perdana melakukan penelitian dengan menggunakan metode algortima C5.0 dan *Adaboost* akurasi meningkat sebesar 12,14% menjadi 82,14% [18]. Nia Novianti, Muhammad Zarlis, Poltak Sihombing juga melakukan penelitian mengenai penyakit diabetes, dengan metode K-NN dan algoritma *Adaboost* sebagai klasifikasi nya dan dalam 5 pengujian dengan nilai  $K = 7, 13, 19, 25$  dan 31 terdapat peningkatan hasil akurasi[19]. Penelitian yang dilakukan oleh Ankara Science University yang menjelaskan tentang tingkat akurasi penggunaan *Decision tree* dengan metode *Adaboost* meningkat menjadi 0,95 dengan dataset yang digunakan yaitu data covid 19[20].

Pada penelitian yang dilakukan oleh Triando Hamonangan Saragih, Muliadi, M. Reza Faisal dan Muhammad Al Ichsan Nur Rizqi Said terhadap tanam jarak pagar dengan menggunakan *Adaboost* classifier untuk mengklasifikasi penyakit yang terdapat pada tanaman tersebut. Dengan hasil akurasi rata-rata terbaik 50% dan maksimal terbaik sebesar 53,01% fold sebanyak 2[21].

Dewi Anggraeni, Ramadhani mengambil data diabetes diRumah Sakit M. Djamil Padang. Untuk melakukan pengujian terhadap *Decision tree* algoritma C4.5 dan mendapatkan hasil Akurasi gain tertinggi adalah atribut glukosa yaitu sebesar 0.613305[22]. Primandani Arsi, dan Oman Somantri melakukan penelitian terhadap penyakit diabetes, menggunakan Neurol Network berbasis algoritma genetika dan mendapatkan. Hasil eksperimen menunjukkan terjadinya penurunan nilai RMSE yang berarti terjadi peningkatan pada nilai akurasi prediksi, yaitu dari 0,402 +/- 0,035 menjadi 0,396 +/-0,032[23].

berikut ini adalah table penelitian sebelum nya

Tabel 2. 2 Penelitian terdahulu

No	Peneliti	Data	Metode	Hasil
1	Ginanjari Abdurrahman (2022)	Pima Indian diabetes dataset.	algoritma <i>Adaboost</i> Classifier	<i>Adaboost</i> Classifier pada dataset hasil imputing mean menghasilkan akurasi sebesar 80.09 %
2	Lidia Pebrianti, Fitrihuda Aulia,	Dataset berupa data tabular dari kondisi kesehatan	<i>Adaboost</i> menggunakan algoritma Naïve Bayes	nilai akurasi optimasi <i>Adaboost</i> dan Naïve

	Halimatun Nisa dan, Kana Saputra S (2022)	pasien yang terindikasi diabetes maupun tidak.		Bayes senilai 0.7694
3	Hasbi Firmansyah, dan Zainal Abidin (2022)	Data Podes 2014	<i>Adaboost</i> menggunakan algoritma <i>Decision tree</i>	Akurasi 88.91%
4	Zico Karya Saputra Domas, Roby Rakhmad (2022)	dataset kurangtransparana n pengungkapan informasi anti-korupsi	<i>Decision tree</i> yang dikombinasika n dengan metode <i>Adaboost</i>	hasil yang lebih objektif, jika algoritma decision tree berbasis <i>Adaboost</i> meningkat menjadi 0,768
5	Yuveinsiana Crismayella, Neva Satyahadewi, Hendra Perdana (2023)	Mahasiswa Program Studi Statistika Universitas Tanjungpura dengan jumlah data 140	Algoritma C5.0 dan <i>Adaboost</i>	boosting menggunakan Algoritma <i>Adaboost</i> , akurasi meningkat sebesar 12,14% menjadi 82,14%

6	Nia Novianti, Muhammad Zarlis, Poltak Sihombing (2022)	dataset penyakit diabetes yang bersumber dari UCI Repository Machine Learning dimana dataset memiliki 1001	K-NN dan algoritma <i>Adaboost</i> .	5 pengujian dengan nilai K = 7, 13, 19, 25 dan 31 terdapat peningkatan hasil akurasi
7	Ankara Science University, Ankara, Turkey	Covid 19 dataset	Algoritma <i>Adaboost-Decision tree</i>	Perfomance model naik menjadi nilai akurasi 0,95
8	Triando Hamonangan Saragih, Muliadi, M. Reza Faisal dan Muhammad Al Ichsan Nur Rizqi Said (2021)	Dataset Tanaman jarak pagar	<i>Adaboost Classifier</i>	hasil akurasi rata-rata terbaik 50% dan maksimal terbaik sebesar 53,01% fold 2
9	Dewi Anggraeni, Ramadhani (2018)	Rumah Sakit M. Djamil Padang.	<i>Decision tree</i> algoritma C4.5	Akurasi gain tertinggi adalah atribut glukosa yaitu sebesar 0.613305.

10	Primandani Arsi, Oman Somantri (2018)	Pima Diabetes India yang berjumlah 768 record	Neurol Network berbasis algoritma genetika	nilai RMSE yang berarti terjadi peningkatan pada nilai akurasi prediksi, yaitu dari 0,402 +/- 0,035 menjadi 0,396 +/- 0,032
----	--	---	--	---

Pada penelitian ini akan dilakukan klasifikasi penyakit diabetes menggunakan *Decision tree* sebagai pemodelannya dan akan ditambahkan menggunakan metode *Adaboost* untuk meningkatkan hasil akurasi klasifikasi tersebut.

## BAB III METODE PENELITIAN

### 3.1 Tempat dan Waktu Penelitian

#### 3.1.1 Tempat Penelitian

Penelitian ini dilaksanakan di Laboratorium Komputer Jurusan Teknik Elektro Program studi Teknik Informatika Fakultas Teknik Universitas Lampung yang beralamatkan di Jl. Soemantri Brojonegoro No. 1 Gedong Meneng, Bandar Lampung.

#### 3.1.2 Waktu Penelitian

Penelitian ini dilaksanakan mulai dari minggu ketiga bulan Januari 2024 sampai dengan akhir bulan Maret 2024.

Adapun timeline waktu pelaksanaan penelitian dijabarkan pada Tabel berikut

Tabel 3. 1 Timeline

No. Kegiatan		Waktu Pelaksanaan Penelitian 2024								
		Feb	Mar	Apr	Mei	Juni	Ags	Sept	Okt	Nov
1.	Studi Literasi									
2	Pengumpulan Data									

3.	Preprocessing Data									
4.	Pemodelan Decision tree									
5.	Decision Tree klasifikasi Adaboost									
6.	Evaluasi dan Analisis									
7.	Menyusun hasil analisis dan Pembahasan									

### 3.2 Alat dan Bahan

#### 3.2.1 Bahan

Data yang digunakan berasal dari Institut Nasional Diabetes dan Penyakit Pencernaan dan Ginjal, semua pasien di sini adalah hasil rekam medis yang dilakukan oleh rumah sakit, data pasien tersebut merupakan keturunan India Pima. Suku Indian Pima adalah kelompok masyarakat adat yang berasal dari daerah barat daya Amerika Serikat, khususnya di negara bagian Arizona, serta wilayah utara Meksiko, di negara bagian Sonora. Nama asli mereka dalam bahasa Pima adalah Akimel O'odham, yang berarti Orang Sungai (River People), mencerminkan hubungan erat mereka dengan sungai-sungai di wilayah tersebut, seperti Sungai

Studi tentang diabetes di suku Pima dimulai pada tahun 1960-an. Mereka memiliki salah satu tingkat diabetes tipe 2 tertinggi di dunia, yang sebagian besar dihubungkan dengan perubahan pola makan tradisional menjadi makanan modern. <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset> Data yang dipakai di dalam penelitian ini terdiri dari 100.000 record dengan 9 variabel atribut

Berikut deskripsi mengenai atribut penelitian yang digunakan terdapat pada Tabel

Tabel 3. 2 Atribut

<b>Nama Atribut</b>	<b>Deskripsi</b>
Gender	Menjelaskan perbedaan jenis kelamin
Hypertension	Menjelaskan tekanan darah dari batas normal
heart_disease	Menjelaskan pasien yang memiliki penyakit jantung
Smoking_History	Mengacu pada riwayat pasien perokok
blood_glucose_level	Menjelaskan tentang jumlah kadar glukosa dalam darah
BMI (Body Mass Index)	Metode yang digunakan untuk mengukur berat badan dan tinggi badan seseorang
HbA1c_level	Menjelaskan pemeriksaan yang akurat dan tepat dalam mengukur kadar gula darah yang kronis serta berkorelasi positif dengan terjadinya risiko komplikasi diabetes.
Age	Menjelaskan tentang rentang usia
Outcome	Mengindikasikan apakah seseorang memiliki diabetes atau tidak (nilai 1 untuk diabetes, 0 untuk tidak diabetes).

### 3.2.2 Alat

Untuk penelitian ini dibutuhkan Perangkat Hardware dan beberapa Software

#### a. Hardware

Laptop dengan spesifikasi Processor Intel(R) Celeron® N4000 CPU @ 1.10GHz (2 CPUs), -1.1GHz dengan 4096 RAM dan Flash Disch

#### b. Software

Operating System Windows 11 Home 64-bit (10.0, build 22621)

Microsoft Excel 2013 untuk Menampilkan data

Python untuk mengolah data dan memvisualkan

Figma untuk pengembangan interface

### 3.3 Tahapan Penelitian

Berikut ini adalah tahap penelitian secara umum yang akan dilakukan pada penelitian ini.

#### 1 Studi Literasi

Studi literasi yang dilakukan secara tulisan melalui buku, jurnal maupun artikel guna mencari dan menambah wawasan mengenai masa masalah dan solusi yang dapat membantu pengembangan penelitian ini sehingga, studi literasi dilakukan selama penelitian ini dijalankan untuk menambah referensi dan peningkatan dalam penelitian sebelumnya.

#### 2 Pengumpulan data

Data di kumpulkan dapat melalui sebuah lembaga maupun platform open source yang dapat digunakan menjadi bahan data yang dapat dikelola atau diteliti untuk memecahkan sebuah masalah. Dalam hal ini pengumpulan data diambil pada platform media online open source.

#### 3. Preprocessing Data

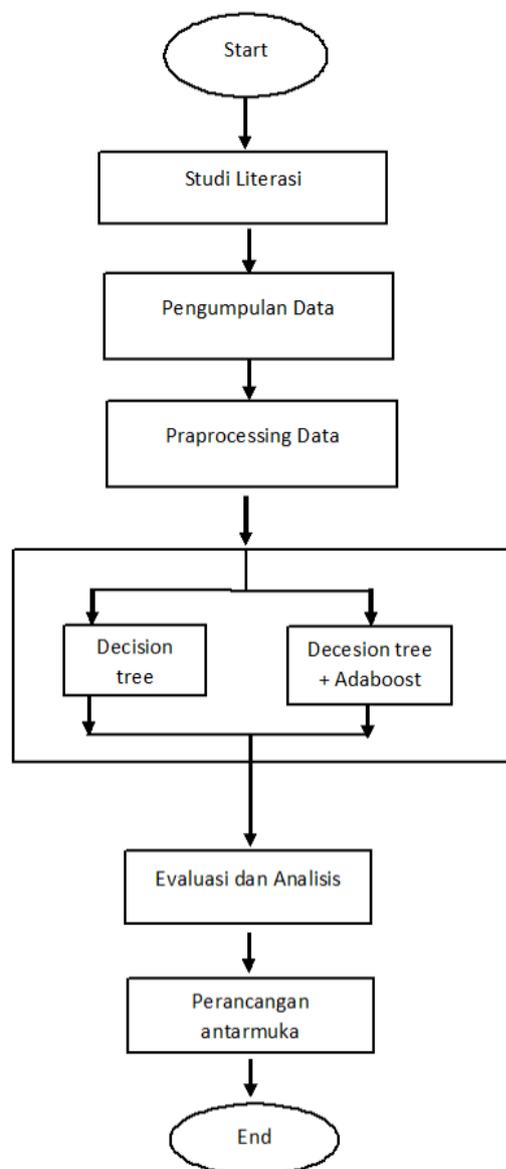
Setelah mengumpulkan data selanjutnya yaitu tahap preprocessing, dimana pada tahap ini data dibersihkan untuk memisahkan data yang gagal dan noise, serta tidak terbaca, lalu data tersebut di transformasi untuk mengambil data yang perlu digunakan untuk pengembangan selanjutnya.

#### 4. *Decision tree + Adaboost*

Tahap selanjutnya yaitu memvisualkan data ke dalam *Decision tree* untuk dapat melihat hasil akurasi dari data tersebut. Lalu setelah itu hasil yang telah menggunakan algoritma *Decision tree* ditambahkan lagi menggunakan metode *Adaboost* untuk menguatkan hasil akurasi dari *Decision tree* tersebut.

## 5. Evaluasi dan Analisis

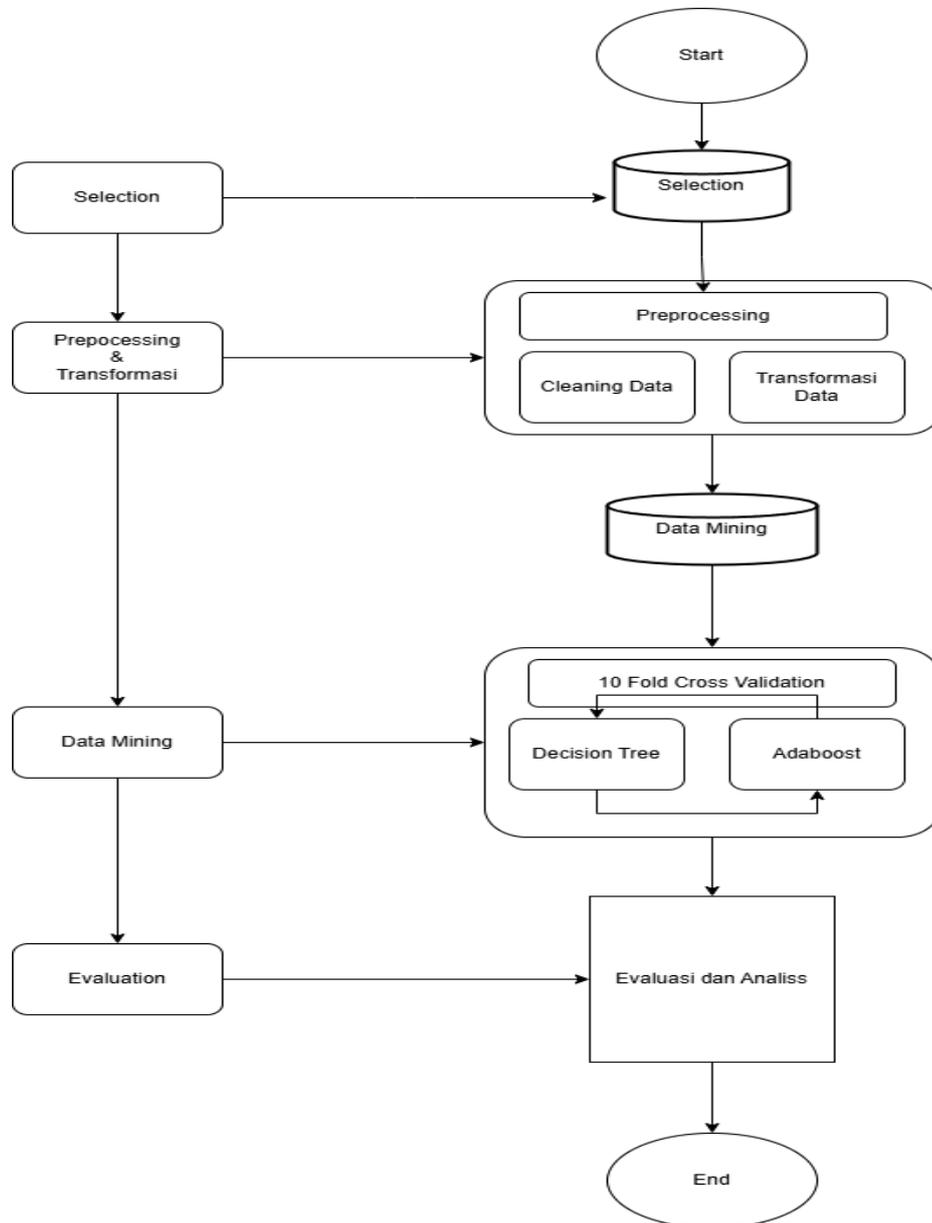
Tahap ini dilakukan evaluasi pada hasil akurasi *Decision tree* yang telah ditambahkan bersama *Adaboost*, untuk di analisis menjadi sebuah data baru yang yang dapat menjadi sebuah hasil kesimpulan yang dipakai untuk mengambil sebuah keputusan yang akurat. Dan di visualkan kedalam interface untuk menjadi sebuah informasi yang dapat digunakan untuk bahan pertimbangan mengambil sebuah keputusan.



Gambar 3. 1 Tahapan penelitian

### 3.4 Desain Sistem

Berikut adalah desain sistem penelitian yang akan dilakukan



Gambar 3. 2 Desain penelitian KDD

Penjelasan dari Desain sistem Penelitian pada gambar diatas adalah sebagai berikut

Pada desain sistem untuk penelitian ini menggunakan model KDD Metodologi pada penelitian ini menggunakan proses KDD (*Knowledge Discovery in Database*).

Proses ini menjelaskan secara sistematis dalam mencari suatu hubungan baru

di dalam Market Basket Analysis menggunakan beberapa tahap pengolahan data [24]. di dalam penelitiannya menjelaskan bahwa terdapat beberapa langkah di dalam proses KDD (*Knowledge Discovery in Database*) diantaranya secara berurutan selection, preprocessing, transformation, data mining, dan Interpretation/evaluation.

### 1. Selection

Selection Dataset berupa data tabular dari kondisi kesehatan pasien yang terindikasi diabetes maupun tidak yang memiliki bentuk gejala dan pemeriksaan diabetes pada data ini, jumlah data 100.000 dengan dataset ini memiliki 9 variabel yaitu Gender, Hypertension, Heart\_Disease, Smoking\_History, Blood\_Glukose\_Level, BMI, HbA1C\_Level Age, Outcome. Dengan dua kelas yaitu positif dan negatif

### 2. Preprocessing data dan Transformasi Data

Preprocessing data adalah salah satu proses yang penting dilakukan untuk menghasilkan data dengan kualitas yang baik dengan cara di antaranya validasi, integrasi dan transformasi. Preprocessing / Cleaning proses preprocessing mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data, seperti kesalahan cetak(tipografi) [25]. Adapun fitur yang diperbaiki yaitu Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Diabetes PedigreeFunction, BMI dan Age. Preprocessing data dilakukan dengan 2 tahap yaitu, dengan cara penghapusan data / Cleaning Data dan Transformation Data / Data Transformasi. Setelah data terkumpul, dilakukan pembersihan data dengan cara menghapus data-data yang tidak sesuai atau tidak diperlukan dalam penelitian, seperti menghapus data-data kosong yang tidak mempunyai nilai dalam data penelitian dan mengambil data-data penting yang akan diteliti secara manual.

### 3. Data Mining

#### a. Ten-Fold cross Validation

Proses selanjutnya yaitu melakukan k-fold cross validation dengan membagi data penelitian menjadi data testing dan data training. Pembagian data meliputi 3/10 data

sebagai data testing, dan 7/10 data sebagai data training. Uji validitas menggunakan 10-fold cross validation yang membagi data secara acak menjadi 10 iterasi. Kemudian sesuaikan model menggunakan iterasi  $k - 1$  ( $k$  minus 1) dan validasikan model menggunakan iterasi ke- $k$  yang tersisa, lalu catat skor/kesalahan (error). Ulangi proses ini sampai setiap iterasi  $k$  berfungsi sebagai set tes dan mengambil rata-rata dari skor yang tercatat. Rata-rata skor yang dicatat akan menjadi metrik kinerja model.[26]

#### b. Model Decision tree

*Decision tree* membagi ruang fitur (feature space) dengan membuat keputusan berdasarkan nilai-nilai fitur di setiap simpul (node), setelah itu membuat model decision tree untuk mengklasifikasi data lalu memvisualkan model dari data tersebut.

#### c. *Decision tree* menggunakan *Adaboost*

Mencoba decision tree (weak learner) pada data pelatihan dengan mempertimbangkan bobot pada setiap sampel. Menghitung kesalahan klasifikasi model pada data pelatihan. menghitung bobot baru untuk setiap sampel data berdasarkan seberapa baik atau buruk model dapat mengklasifikasikan sampel tersebut. Perbarui bobot data. Penggabungan model Decision tree yang dihasilkan pada setiap iterasi dengan memberikan bobot pada masing-masing model. Model gabungan (ensemble model) akan menjadi hasil akhir dari proses *Adaboost*. Model gabungan untuk melakukan prediksi pada data pengujian atau data baru. menghasilkan dengan mengambil mayoritas suara dari hasil klasifikasi model pohon keputusan pada setiap iterasi. Persamaan nya sebagai berikut.

Model Gabungan ( $H(x)$ ) dihasilkn dengan menggabungkan model Decision tree dengan bobot.  $H(x) = \text{Sign} \{ \sum_{t=1}^T a_t h_t(x) \}$

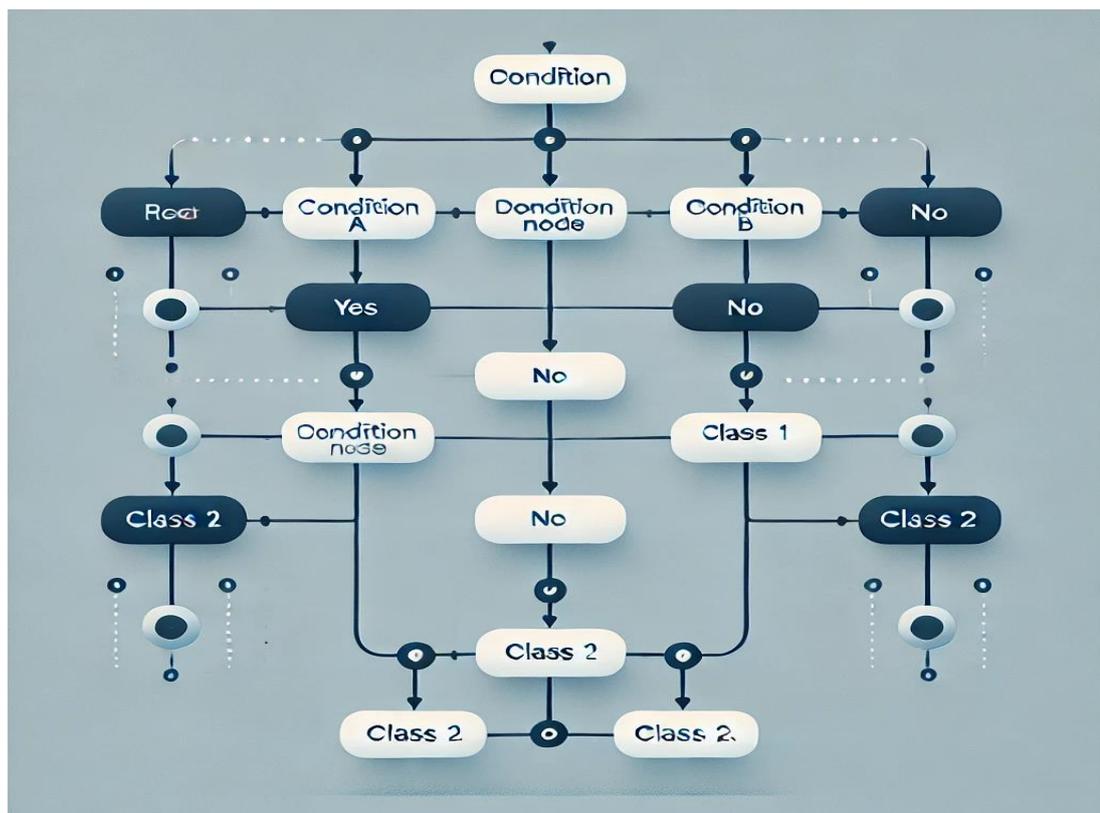
$(h_t)(x)$  adalah output dari model *decision tree* pada iterasi ke- $t$ .

### 4. Evaluasi

#### a. Evaluasi dan Analisis

Tahap terakhir pada penilitan ini dengan selesai melakukan pengujian pada model decesion yang digabungkan menggunakan metode *Adaboost*, Pengujian yang digunakan menggunakan confusion matrix untuk mendapatkan nilai accuracy, recall, precision, dan f-score

Flowchart decision tree



Gambar 3. 3 flowchart *decision tree*

Root Node (Akar Pohon)

Merupakan titik awal dari decision tree. Biasanya, root node memuat atribut utama yang menjadi dasar pembagian awal dataset.

Condition Nodes (Node Kondisi) node ini adalah titik pengambilan keputusan berdasarkan nilai atribut tertentu. Setiap node kondisi memiliki aturan yang memisahkan data menjadi dua atau lebih cabang berdasarkan nilai atribut.

Branches (Cabang)

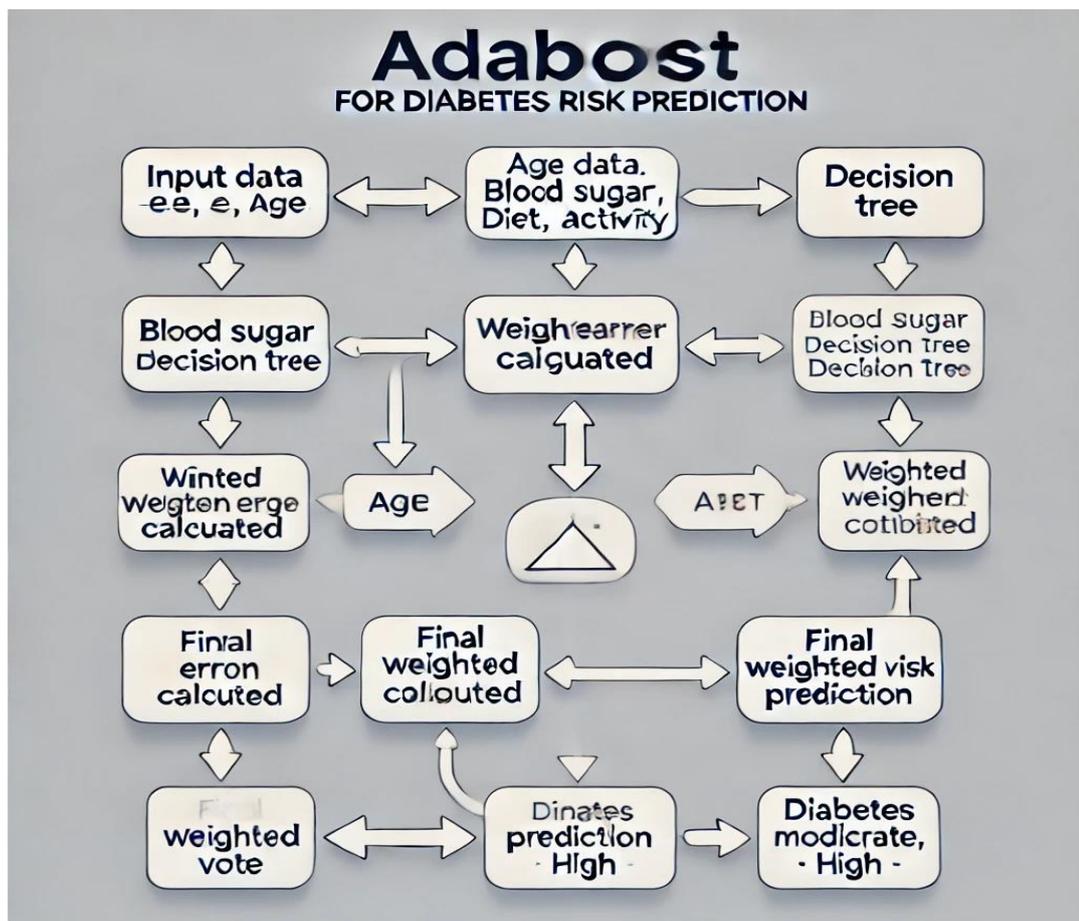
Cabang menunjukkan hasil keputusan di node kondisi. Misalnya, "Yes" atau "No" yang melanjutkan ke node atau kondisi berikutnya.

#### Leaf Nodes (Node Daun)

Node ini adalah titik akhir dari *decision tree* biasanya, leaf node merepresentasikan kelas tertentu (misalnya, Class 1 atau Class 2) yang menjadi hasil akhir dari proses klasifikasi.

Dataset melewati root node dan dinilai berdasarkan aturan tertentu. Dataset akan mengalir melalui cabang ke node kondisi berikutnya sesuai dengan hasil keputusan. Proses ini berlanjut hingga semua dataset diklasifikasikan ke leaf node, yang memberikan hasil akhir berupa label kelas.

#### Flowchart Adaboost



Gambar 3. 4 flowchart decision tree adaboost

Algoritma *adaboost* dimulai dengan membangun weak classifiers (decision trees) yang secara iteratif memperbaiki kesalahan dengan menyesuaikan bobot data.

Setiap weak classifier berkontribusi pada prediksi akhir melalui sistem weighted voting, di mana model dengan performa lebih baik memiliki pengaruh lebih besar.

Pada akhirnya, prediksi akhir risiko diabetes dihitung dengan menggabungkan seluruh kontribusi weak classifiers, menghasilkan klasifikasi risiko diabetes yang lebih akurat dan andal.

## **BAB V KESIMPULAN DAN SARAN**

### **5.1 Kesimpulan**

Adapun kesimpulan dari penelitian yang telah dilakukan yaitu menggunakan decision tree pada data diabetes yang di klasifikasikan menggunakan adaboost

1. Optimasi algoritma Decision Tree dengan AdaBoost mampu meningkatkan akurasi dan performa klasifikasi dibandingkan model Decision Tree biasa. Dengan akurasi sebelumnya 94% naik menjadi 96% setelah menggunakan adaboost pada algoritma decision tree
2. Hasil evaluasi kinerja cukup baik terlihat terjadi peningkatan pada setiap nilai seperti precision f1-score dan AUC hanya terjadi penurunan pada nilai recall yang awalnya 73% turun menjadi 68%

### **5.2 Saran**

Adapun saran yang dapat diberikan untuk pengembangan penelitian selanjutnya

1. Penelitian lanjutan dapat mempertimbangkan pendekatan lain seperti Gradient Boosting atau Random Forest untuk membandingkan efektivitas terhadap AdaBoost.
2. Penyeimbangan data (oversampling atau undersampling) pada kelas minoritas agar recall untuk kelas positif meningkat.
3. Mengintegrasikan model ke dalam sistem kesehatan berbasis web atau aplikasi untuk mendukung diagnosa awal pasien secara cepat dan efisien.
4. Menguji model menggunakan data dari sumber berbeda untuk mengevaluasi generalisasi model terhadap populasi yang lebih luas.
5. Mengembangkan antarmuka pengguna untuk mempermudah dokter atau pengguna dalam memahami hasil prediksi dan variabel yang relevan.

## DAFTAR PUSTAKA

- [1] KEPUTUSAN MENTERI KESEHATAN REPUBLIK INDONESIA NOMOR HK.01.07/MENKES/603/2020, “Diabetes Kemenses.”
- [2] B. Berglund, T. Lindvall, and D. H. Schwela, "Definition, diagnosis and classification of diabetes Mellitus and its complication“\_ 1999.
- [3] Badan Penelitian dan Pengembangan Kesehatan RI, “Laporan Riskesdas 2018 Nasional.pdf,” *Lembaga Penerbit Balitbangkes*. p. 156, 2018.
- [4] N. Wahidah and S. R. Rahayu, “Determinan Diabetes Melitus pada Usia Dewasa Muda,” *Higeia*, vol. 6, no. 1, pp. 114–125, 2022, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/higeia>
- [5] D. Di *et al.*, “INSIDEN DAN FAKTOR RISIKO DIABETES MELITUS PADA ORANG PENYAKIT TIDAK MENULAR ( Incident and Risk Factor of Diabetes Mellitus in Adults at Bogor . Prospective Cohort Study Risk Factors Non Communicable Diseases ),” pp. 151–160, 2015.
- [6] Ameet Talwalkar and A. T. Mehryar Mohri, Afshin Rostamizadeh, *Machine learning*, vol. 0, no. 9783319056050. 2014. doi: 10.1007/978-3-319-05606-7\_6.
- [7] S. Tiwari, “Mathematics for Machine Learning,” *2024 3rd Int. Conf. Innov. Technol. INOCON 2024*, 2024, doi: 10.1109/INOCON60754.2024.10511475.
- [8] A. Irma Prianti, “Pebandingan Metode K-Nearest Neighbor dan Adaptive Boosting pada Kasus Klasifikasi Multi Kelas,” *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 13, no. 1, pp. 39–47, 2020, doi: 10.36456/jstat.vol13.no1.a3269.

- [9] E. Listiana and M. A. Muslim, "Peningkatan Akurasi Pada Algoritma C4.5 Menggunakan Adaboost Untuk Meminimalkan Resiko Kredit," *Pros. SNATIF*, no. 2015, pp. 875–881, 2017.
- [10] M. F. Rahman, D. Alamsah, M. I. Darmawidjadja, and I. Nurma, "Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN)," *J. Inform.*, vol. 11, no. 1, p. 36, 2017, doi: 10.26555/jifo.v11i1.a5452.
- [11] K. Umam, *Algoritma Dan Pemrograman Komputer Dengan Python*. 2021. [Online]. Available: <http://repository.iainmadura.ac.id/id/eprint/455>
- [12] M. Dr. Budi Raharjo, S.Kom., M.Kom., *Pembelajaran Mesin*. 2016. [Online]. Available: <https://www.codepolitan.com/mengenal-teknologi-machine-learning-pembelajaran-mesin>
- [13] G. Abdurrahman, "Jurnal Sistem dan Teknologi Informasi Klasifikasi Penyakit Diabetes Melitus Menggunakan Adaboost Classifier," *JUSTINDO (Jurnal Sist. dan Teknol. Informasi)*, vol. 7, no. 1, pp. 59–66, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JUSTINDO>
- [14] L. Pebrianti, F. Aulia, H. Nisa, and K. Saputra, "Informasi Implementasi Metode Adaboost untuk Mengoptimasi Klasifikasi Penyakit Diabetes dengan Algoritma Naïve Bayes," *J. Sist. dan Teknol.*, vol. 7, no. 2, pp. 122–127, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JUSTINDO/article/view/8627%0Ahttp://jurnal.unmuhjember.ac.id/index.php/JUSTINDO/article/download/8627/4296>
- [15] H. Firmansyah and Z. Abidin, "Penerapan Algoritma Gradient Boosted Decision Trees Pada Adaboost Untuk Klasifikasi Status Desa," *J. Inform. UPS*, vol. 1, no. 1, pp. 27–35, 2022.
- [18] Z. K. S. Domas and R. Rakhmadi, "Peningkatan Performa Decision Tree dengan AdaBoost untuk Klasifikasi Kekurangtransparanan Informasi Anti-Korupsi," *Appl. Inf. Syst. Manag.*, vol. 5, no. 2, pp. 75–82, 2022, doi: 10.15408/aism.v5i2.24887.

- [19] Y. Crismayella, N. Satyahadewi, and H. Perdana, "Algoritma Adaboost pada Metode Decision Tree untuk Klasifikasi Kelulusan Mahasiswa," *Jambura J. Math.*, vol. 5, no. 2, pp. 278–288, 2023, doi: 10.34312/jjom.v5i2.18790.
- [20] N. Novianti, M. Zarlis, and P. Sihombing, "Penerapan Algoritma Adaboost Untuk Peningkatan Kinerja Klasifikasi Data Mining Pada Imbalance Dataset Diabetes," *J. Media Inform. Budidarma*, vol. 6, no. 2, p. 1200, 2022, doi: 10.30865/mib.v6i2.4017.
- [21] E. Sevinç, "An empowered AdaBoost algorithm implementation: A COVID-19 dataset study," *Comput. Ind. Eng.*, vol. 165, no. June 2021, p. 107912, 2022, doi: 10.1016/j.cie.2021.107912.
- [22] T. H. Saragih, M. R. Faisal, D. M. Al, I. Nur, and R. Said, "AdaBoost Classifier untuk Klasifikasi Tanaman Jarak Pagar," *J. Komputasi*, vol. 9, no. 2, pp. 60–66, 2021.
- [23] D. Anggraeni and Ramadhani, "Analisa Decision Tree Untuk Prediksi Diagnosa Diabetes Mellitus," *online) Kisaran, Asahan*, vol. 9986, no. September, pp. 153–158, 2018.
- [24] P. Arsi and O. Somantri, "Deteksi Dini Penyakit Diabetes Menggunakan Algoritma Neural Network Berbasis Algoritma Genetika," *J. Inform. J. Pengemb. IT*, vol. 3, no. 3, pp. 290–294, 2018, doi: 10.30591/jpit.v3i3.1008.
- [25] Tj. Favvad, D. Haussler, and P. Stolorz, "KDD for Science Data Analysis: Issues and Examples," *Proc. - 2nd Int. Conf. Knowl. Discov. Data Mining, KDD 1996*, pp. 50–56, 1996.
- [26] N. Purwati and A. Dwi Januanti, "Aplikasi Data Mining Dengan Algoritma Naive Bayes Untuk Memprediksi Tingkat Kelulusan Mahasiswa," *J. Pepadun*, vol. 2, no. 1, pp. 123–137, 2021, doi: 10.23960/pepadun.v2i1.38.
- [27] J. M. Gorriz, F. Segovia, J. Ramirez, A. Ortiz, and J. Suckling, "Is K-fold cross validation the best model selection method for Machine Learning?,"

2024, [Online]. Available: <http://arxiv.org/abs/2401.16407>