**IMPLEMENTASI METODE *NAIVE BAYES-ADABOOST* UNTUK DIAGNOSIS PENYAKIT DIABETES MELITUS**

# PROPOSAL SKRIPSI



**Oleh:**

**ANDREAN FIRMAN PRADANA**

**06.2018.1.07024**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK ELEKTRO DAN TEKNOLOGI INFORMASI**

**INSTITUT TEKNOLOGI ADHI TAMA SURABAYA**

**2022**

# KATA PENGANTAR

Puji syukur marilah panjatkan kehadirat Allah SWT, yang telah memberikan kami karunia nikmat dan kesehatan, sehingga kami dapat menyelesaikan proposal skripsi ini yang berjudul “IMPLEMENTASI METODE *NAÏVE BAYES-ADABOOST* UNTUK DIAGNOSIS PENYAKIT DIABETES MELITUS”. Proposal skripsi ini disusun untuk memenuhi Sebagian persyaratan dalam memperoleh gelar studi strata satu (S1) jurusan Teknik Informatika, Fakultas Teknik Elektro dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya.

Dalam penulisan proposal skripsi ini tidak mungkin akan terwujud apabila tidak ada bantuan dari berbagai pihak, melalui kesempatan ini izinkan penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Allah Subhanahu wa ta’ala.
2. Kepada keluarga yang mendoakan dan memberikan dukungan.
3. Bapak Hendro Nugroho ST.,M.kom. selaku dosen pembimbing 1 yang telah meluangkan waktu, dan pikiran didalam memberikan bimbingan dalam pengerjaan proposal skripsi.
4. Ibu Gusti Eka Yuliastuti S.Kom.,M.Kom. selaku dosen pembimbing 2 yang telah memberikan bimbingan, pikiran, dan motivasi dalam penyelesaian proposal skripsi.
5. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih atas segala ilmu dan pengetahuan yang telah diberikan kepada penulis.

Penulis menyadari bahwa dalam pengerjaan proposal skripsi ini masih jauh dari kata sempurna, karena terbatasnya kemampuan dan pengalaman penulis, oleh karena itu, segala kritik dan saran baik secara langsung maupun tidak langsung yang dapat membangun kedepannya. Semoga proposal skripsi ini dapat memberikan manfaat bagi semua pihak.

Sidoarjo, …/../ 2022

Penulis

# DAFTAR ISI

[**PROPOSAL SKRIPSI i**](#_Toc107319685)

[**KATA PENGANTAR ii**](#_Toc107319686)

[**DAFTAR ISI iii**](#_Toc107319687)

[**DAFTAR GAMBAR v**](#_Toc107319688)

[**DAFTAR TABEL vi**](#_Toc107319689)

[**BAB I 1**](#_Toc107319690)

[**PENDAHULUAN 1**](#_Toc107319691)

[**1.1 Latar Belakang 1**](#_Toc107319692)

[**1.2 Perumusan Masalah 3**](#_Toc107319693)

[**1.3 Tujuan Penilitian 3**](#_Toc107319694)

[**1.4 Batasan Masalah 3**](#_Toc107319695)

[**1.5 Manfaat Penelitian 4**](#_Toc107319696)

[**1.6 Metodologi Penelitian 4**](#_Toc107319697)

[**1.7 Sistematika Penulisan 5**](#_Toc107319698)

[**BAB II 6**](#_Toc107319699)

[**LANDASAN TEORI 6**](#_Toc107319700)

[**2.1 Peneliti Terdahulu 6**](#_Toc107319701)

[**2.2 Diabetes 7**](#_Toc107319702)

[**2.2.1 Gejala dan Faktor Diabetes 8**](#_Toc107319703)

[**2.2.2 Klasifikasi Diabetes Melitus 9**](#_Toc107319704)

[**2.3 Kecerdasan Buatan 10**](#_Toc107319705)

[**2.4 Data Mining 10**](#_Toc107319706)

[**2.5 Klasifikasi 12**](#_Toc107319707)

[**2.6 Machine Learning 12**](#_Toc107319708)

[**2.7 Confusion Matrix 13**](#_Toc107319709)

[**2.8 Naïve Bayes 14**](#_Toc107319710)

[**2.9 Adaboost 17**](#_Toc107319711)

[**3.0 Python 19**](#_Toc107319712)

[**BAB III 20**](#_Toc107319713)

[**METODOLOGI PENELITIAN 20**](#_Toc107319714)

[**3.1 Gambaran Umum 20**](#_Toc107319715)

[**3.2 Analisa Kebutuhan 20**](#_Toc107319716)

[**3.2.1 Kebutuhan Software 21**](#_Toc107319717)

[**3.2.2 Kebutuhan Hardware 21**](#_Toc107319718)

[**3.3.3 Data yang dibutuhkan 21**](#_Toc107319719)

[**3.3 Pengumpulan Data 21**](#_Toc107319720)

[**3.4 Preprocessing Dataset 22**](#_Toc107319721)

[**3.5 Klasifikasi Naïve Bayes - Adaboost 23**](#_Toc107319722)

[**3.5.1 Flowchart Klasifikasi *Naïve Bayes* 24**](#_Toc107319723)

[**3.5.2 Flowchart Klasifikasi *Adaboost* 25**](#_Toc107319724)

[**3.5.3 Preprocessing Data 26**](#_Toc107319725)

[**3.5.4 Perhitungan Manual Naïve Bayes – Adaboost 27**](#_Toc107319726)

[**3.6 Pengujian Sistem 44**](#_Toc107319727)

[**DAFTAR PUSTAKA 45**](#_Toc107319728)

# DAFTAR GAMBAR

[**Gambar 2. 1 Langkah-langkah proses data mining 11**](#_Toc107319220)

[**Gambar 3. 1 Gambaran Umum 20**](#_Toc107319221)

[**Gambar 3. 2 Dataset Diabetes Melitus 21**](#_Toc107319222)

[**Gambar 3. 3 Flowchart Algoritma Sistem 23**](#_Toc107319223)

[**Gambar 3. 4 Flowchart Naïve Bayes 24**](#_Toc107319224)

[**Gambar 3. 5 Flowchart Adaboost 25**](#_Toc107319225)

[**Gambar 3. 6 Gambar Data Training 30**](#_Toc107319226)

# DAFTAR TABEL

[**Tabel 2. 1 Model Confusin Matrix 13**](#_Toc107319394)

[**Tabel 3. 1 Variabel Dataset Diabetes 22**](#_Toc107319395)

[**Tabel 3. 2 Variabel BMI Sebelum di Preprocessing 26**](#_Toc107319396)

[**Tabel 3. 3 Variabel BMI Setelah di Preprocessing 27**](#_Toc107319397)

[**Tabel 3. 4 Nilai Prior 30**](#_Toc107319398)

[**Tabel 3. 5 Nilai Likehood Kelas Diabetes 31**](#_Toc107319399)

[**Tabel 3. 6 Nilai Likehood Kelas Tidak Diabetes 32**](#_Toc107319400)

[**Tabel 3. 7 Probabilitas Variabel Numerik 37**](#_Toc107319401)

[**Tabel 3. 8 Hasil Probabilitas Akhir Setiap Kelas 41**](#_Toc107319402)

[**Tabel 3. 9 Tabel Probabilitas Akhir 42**](#_Toc107319403)

[**Tabel 3. 10 Confusion Matrix 42**](#_Toc107319404)

# 

# BAB I

# PENDAHULUAN

## 1.1 Latar Belakang

Diabetes melitus adalah penyakit kronis serius yang disebabkan oleh pankreas yang tidak dapat menghasilkan cukup insulin (hormon yang mengatur gula darah atau glukosa), atau ketika tubuh tidak dapat secara efektif menggunakan insulin yang dibutuhkannya. Diabetes melitus tidak termasuk ke dalam penyakit menular, namun penyakit ini menjadi salah satu penyebab utama dari kecacatan dan kematian di dunia. Ini telah menjadi penyebab masalah kesehatan masyarakat yang sangat penting dan telah menjadi salah satu dari empat penyakit tidak menular prioritas yang harus di tindaklanjuti oleh pemerintah di seluruh dunia (Khairani, 2019). Menurut *World Health Organization* (WHO), di Indonesia jumlah penderita diabetes mengalami peningkatan dari 10,7 juta pada tahun 2019 menjadi 19,5 juta di tahun 2021. Peningkatan ini merupakan hal yang sangat memprihatikan dan membuat Indonesia menduduki peringkat kelima dengan penderita penyakit diabetes terbanyak di dunia (Reza Pahlevi, 2021).

Diabetes melitus memiliki komplikasi yang sering terjadi adalah penyakit mikrovaskular dan neuropati. Gangguan kesehatan komplikasi diabetes melitus antara lain seperti gangguan mata (retinopati), gangguan ginjal (nefropati), gangguan pembuluh darah (vaskulopati) dan gangguan pada kaki. Komplikasi pada kaki merupakan yang paling umum pada diabetes, keseluruhan sekitar 20%-40% memiliki penyakit vascular (PVD). Sekitar 5% pasien penyakit diabetes mengembangkan luka kaki pada setiap tahun, dan tingkat amputasi sekitar 0,5% pertahun (*NICE*, 2004).

Menurut Dalimartha, 2005 dalam Jilao, (2017) menjelaskan peningkatan pada penderita penyakit diabetes melitus salah satunya yaitu dapat disebabkan karena pola makan penderita yang tidak seimbang. Pola makan yang berlebihan atau tidak seimbang akan menyebabkan obesitas. Kelebihan berat badan atau obesitas adalah kondisi dimana lemak berlebih baik diseluruh tubuh atau terlokalissi pada bagian tertentu. Obesitas ditemukan apabila kelebihan berat badan >20% pada pria dan >25% pada wanita karena lemak (Ganong W.F, 2013).

Prevelensi diabetes biasanya terjadi dikarenakan beberapa faktor, yaitu seperti umur, jenis kelamin, keturunan, tekanan darah, BMI, status perokok, alkohol, dan waktu tidur. Sebagian banyak orang yang menderita diabetes melitus masih menganggap gejala-gejala ringan yang disebabkan oleh penyakit diabetes merupakan hal yang wajar. Sehingga penanganan dini sering kali diabaikan.

Melihat kondisi tersebut langkah pencegahan dan penanganan dini harus segera dilakukan, data mining menjadi hal yang cukup penting didalam melakukan prediksi. Prediksi yang dilakukan kepada pasien penderita diabetes dapat diperoleh dari kumpulan data pasien penderita diabetes, lalu data tersebut diolah membentuk suatu pola tertentu sehingga mendapatkan hasil yang bisa digunakan untuk diagnosa awal diabetes. Beberapa penelitian telah sudah menerapkan teknik data mining dan metode klasifikasi untuk prediksi berbagai penyakit. Klasifikasi yang umum digunakan adalah *Naïve Bayesian*, Pohon Keputusan, SVM (*Support Vector Machine*), dan Jaringan Saraf Tiruan, K-NN (Pristyanto 2019).

Dari penjelasan yang diuraikan diatas, maka untuk menangani masalah tersebut dibutuhkan metode untuk mendiagnosa awal penyakit diabetes agar seseorang dapat segera mungkin mendapatkan penanganan yang tepat. Selain itu, sistem ini bertujuan untuk menghindari komplikasi penyakit yang dapat terjadi apabila terlambat dalam penanganan dini. Pemilihan metode yang tepat untuk melakukan diagnosa awal penyakit diabetes memiliki pengaruh terhadap hasil akhir yang akan didapatkan.

Pada penelitian ini menggunakan data sekunder yang telah diambil dari situs *web Kagg*el 2021 *version* 1 yaitu Diabetes Dataset 2019. Penelitian ini dapat dilakukan dengan kombinasi dua metode yaitu metode *Naïve Bayes* dan *Adaboost*. Pada penelitian sebelumnya, menunjukkan bahwa kombinasi antara 2 metode (*Naïve Bayes* dan *Adaboost*) telah memperbaiki tingkat kinerja metode *naïve bayes* dengan mendapatkan hasil pengujian yang pertama dengan data split 80/20 menggunakan dataset penyakit stroke memberikan nilai akurasi 0,976 memiliki diagnosis yang cukup baik. Setelah di optimasi menggunakan algoritma *Adaboost* dengan split data 70/30 memberikan nilai akurasi 0,981. Dari pengujian sebelum dan setelah opstimasi memiliki selisih sebesar 0,005 (Byna & Basit, 2020). Penelitian selanjutnya yaitu prediksi awal diabetes melitus yang juga menggunakan algoritma *naïve bayes*. Dari total jumlah data yang digunakan sebanyak 500 data, nilai akurasi tertinggi didapatkan pada pengujian ketiga dengan akurasi 89% (Nasruddin, 2021).

Berdasarkan permasalahan diatas, penelitian ini dilakukan untuk diagnosa terhadap penyakit diabetes melitus menggunakan kombinasi 2 metode (*Naïve Bayes* dan *Adaboost*). Dari penelitian tersebut telah dilakukan dengan menggunakan metode *Naïve Bayes* dan *Adaboost* medapatkan nilai akurasi yang cukup baik, diharapkan dari penelitian ini yang menggunakan kombinasi dua metode *Naïve Bayes* dan *Adaboost* dapat diterapkan dengan baik untuk membantu mendiagnosa penyakit diabetes melitus dengan akurat.

## 1.2 Perumusan Masalah

Berdasarkan penjelasan dari latar belakang yang telah dijelaskan maka dapat di uraikan rumusan masalah seperti berikut:

1. Bagaimana membuat sistem untuk mendiagnosa penyakit diabetes melitus?
2. Bagaimana cara mengetahui akurasi keberhasilan dari sistem diagnosa penyakit diabetes melitus?

## 1.3 Tujuan Penilitian

Berdasarkan penjelasan dari perumusan masalah yang telah diuraikan diatas maka tujuan dari penelitian ini sebagai berikut:

1. Membuat sistem untuk mendiagnosa penyakit diabetes melitus dengan menggunakan metode *Naïve Bayes* – *Adaboost*.
2. Mengetahui akurasi keberhasilan dari sistem diagnosa penyakit diabetes melitus dengan cara membandingkan data sebenarnya.

## 1.4 Batasan Masalah

Agar permasalahan dari penelitian ini tidak meluas, maka diperlukan batasan masalah yang akan dibahas, diantaranya:

1. Dataset yang digunakan untuk penelitian ini adalah *Diabetes Dataset 2019* dari situs *web Kaggle* 2021 version 1.

|  |  |
| --- | --- |
| 1 | <https://www.kaggle.com/datasets/tigganeha4/diabetes-dataset-2019?resource=download> |

1. Penelitian ini menggunakan Jupyter Notebook 6.4.6.
2. Bahasa pemrograman yang digunakan adalah Python.
3. Penelitian ini hanya untuk penanganan awal dan bukan untuk menggantikan posisi ahli medis bila penyakit yang diderita memerlukan penanganan lebih lanjut.

## 1.5 Manfaat Penelitian

1. Secara Akademis
2. Dapat dijadikan sarana menambah wawasan bagi pembaca mengenai metode *Naïve Bayes* dan metode *Adaboost*.
3. Dapat membantu para peneliti selanjutnya sebagai bahan refrensi.
4. Secara Praktis

Diharapkan dapat membantu para ahli medis dalam mendiagnosa awal agar tidak terjadi peningkatan kasus penyakit diabetes melitus yang lebih tinggi.

## 1.6 Metodologi Penelitian

Penelitian ini dapat bekerja dengan baik apabila dilakukan secara konsisten dan sesuai alur yang telah ditetapkan, adapun metodologi yang akan dilakukan yaitu sebagai berikut:

1. Studi literatur

Pada tahap ini yaitu mencari paper dari berbagai sumber terkait dengan penelitian yang akan dilakukan sebagai bahan refrensi. Sumber refrensi yang dapat digunakan dalam bentuk jurnal, ebook, yang sesuai dengan penelitian.

1. Bimbingan dosen

Melakukan konsultasi ke dosen pembimbing terkait pengerjaan yang akan dilakukan

1. Pengolahan data

Pada tahap ini yaitu melakukan proses preprocessing data untuk mengubah data mentah menjadi data yang dapat digunakan untuk proses data mining.

1. Analisa data dan kesimpulan

Pada tahap yang terakhir yaitu melakukan analisa dan menuliskan kesimpulan dari hasil yang percobaan yang telah dilakukan.

## 1.7 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir meliputi:

**BAB I: PENDAHULUAN**

Dalam bab ini terdapat latar belakang, perumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, dan sistematika penulisan.

**BAB II: LANDASAN TEORI**

Bab ini memuat tentang teori penunjang penelitian yang akan dilakukan oleh peneliti, serta menjelaskan pemahaman seperti data mining, teori algoritma naïve bayes, teori adaboost dan berbagai teori yang menunjang penelitian.

**BAB III: PERANCANGAN SISTEM**

Pada bab ini memuat tentang bagaimana tahapan dalam merancang sistem untuk diagnosa dini penyakit diabetes melitus.

# BAB II

# LANDASAN TEORI

## 2.1 Peneliti Terdahulu

Terdapat beberapa penelitian terdahulu yang berhubungan dengan penelitian yang dilakukan antara lain:

Penelitian yang berjudul “*Prediction of Type 2 Diabetes using Machine Learning Classification Method*”, Neha Prerna Tigga, Shruti Garg. Penelitian tersebut bertujuan melakukan klasifikasi untuk memprediksi penyakit diabetes dari beberapa metode seperti *Logistic Regression, K Nearest Neighbour, Support Vector Machine, Naïve Bayes, Decission Tree, Random Forest.* Dari beberapa teknik klasifikasi tersebut menunjukkan bahwa algoritman *random forest* memiliki tingkat akurasi paling tinggi dengan nilai 90% (Tigga & Garg, 2020).

Penelitian yang berjudul “Penerapan metode *Adaboost* untuk mengoptimasi Prediksi penyakit Stoke Dengan Algoritma *Naïve Bayes*“. Agus Byna, Muhammad Basit. Penelitian tersebut bertujuan untuk memprediksi penyakit stroke dengan memanfaatkan algoritma *adaboost* yang memiliki kelebihan dapat digabung dengan algoritma *naïve bayes*, hasil pengujian pertama menggunakan data split 80/20 memberikan akurasi 0,976 memiliki tingkat diagnosa klasifikasi yang sangat baik. Setelah dilakukan opstimasi menggunakan algoritma *adaboost* dengan split data 70/30 dengan hasil akurasi 0,981. Hasil dari klasifikasi setelah dilakukan optimasi menunjukkan bahwa akurasi mengalami peningkatan dengan selisih sebesar 0,005 (Byna & Basit, 2020).

Penelitian yang berjudul “Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi *Random Forest”*, Widya Apriliah, Ilham Kurniawan, Muhamad Baydhowi, Tri Haryanti. Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, 2020. Metode yang digunakan dalam penelitian ini adalah tiga algoritma klasifikasi machine learning yaitu Support *Vector Machine*, *Naïve Bayes*, dan *Random Forest*. Tujuan dari penelitian ini yaitu mendeteksi penyakit diabetes tahap awal menggunakan tiga algoritma klasifikasi machine learning yaitu *Suport Vector Machine*, *Naive Bayes* dan *Random Forest* dan mencari algoritma dengan hasil nilai akurasi paling tinggi. Hasil dari penelitian ini menentukan kecukupan sistem yang dirancang dengan akurasi sebesar 97,88% (Ahsana et al., 2021).

Prediksi Diabetes Menggunakan Algoritma *Naïve Bayes* dan *Greedy Forward Selection*, Fitriyani, Universitas ARS, 2021. Metode yang digunakan dala penelitian ini adalah algoritma *Naïve Bayes* dan *Greedy Forward* untuk meemprediksi prediksi penyakit diabetes. Algoritma Forward digunakan untuk seleksi fitur-fitur tidak relevan yang dapat mengurangi kinerja dari model yang digunakan. Hasil penelitian menunjukkan bahwa model *Naïve Bayes* dan *Greedy Forward* mendapatkan nilai akurasi tertinggi sebesar 97,73% dan model *Naïve Bayes* tanpa seleksi fitur *Greedy Forward* mendapatkan nilai akurasi hanya sebesar 87,69% (Fitriyani, 2021).

## 2.2 Diabetes

Diabetes Melitus atau yang sering kita dengar dengan sebutan kencing manis adalah sekelompok penyakit metabolik yang ditandai dengan tingginya kadar gula darah (hiperglikemia), yang diakibatkan oleh gangguan sekresi insulin dan defisiensi insulin (Anggeria & Siregar, 2019).

Secara umum, diabetes didefinisikan sebagai penyakit endemik yang telah menyerang hampir setiap negara didunia. Berdasarkan data *International Diabetes Federation* (IDF) pada tahun 2015, diperkirakan hingga 415 juta orang menderita diabetes. Jumlah ini juga diperkirakan akan meningkat menjadi 640 juta orang pada tahun 2040. Selain dari itu, penyakit diabetes memiliki presentase kematian angka usia sebelum 70 tahun yang lebih tinggi terjadi di negara-negara yang memiliki penghasilan rendah dan menengah disbanding dengan negara-negara berpenghasilan tinggi (Ahsana et al., 2021).

Efek dari diabetes mellitus termasuk lama-kerusakan jangka panjang, disfungsi dan kegagalan berbagai organ. Diabetes mellitus dapat muncul dengan gejala yang khas seperti haus, poliuria, pandangan kabur, dan penurunan berat badan. Di dalam bentuk yang paling parah, ketoasidosis atau non-ketotik keadaan hiperosmolar dapat berkembang dan menyebabkan pingsan, koma dan, jika tidak ada pengobatan yang efektif maka kemungkinan terburuk yang akan terjadi yaitu kematian. Seringkali gejala tidak parah, atau mungkin tidak ada, dan akibatnya hiperglikemia yang cukup untuk menyebabkan patologis dan perubahan fungsional mungkin ada untuk waktu yang lama sebelum diagnosis dibuat. Efek jangka panjang dari diabetes melitus termasuk perkembangan progresif dari spesifik komplikasi retinopati dengan potensi kebutaan nefropati yang dapat menyebabkan gagal ginjal, dan/atau neuropati dengan risiko ulkus kaki, amputasi, sendi Charcot, dan fitur disfungsi otonom, termasuk seksual penyelewengan fungsi. Orang dengan diabetes berada pada peningkatan risiko kardiovaskular, vaskular perifer, dan serebrovaskular penyakit (Lenhard et al., 1998).

### 2.2.1 Gejala dan Faktor Diabetes

Beberapa gejala atau faktor yang disebabkan oleh diabetes di kehidupan sehari hari yang dapat menyebabkan penyakit diabetes:

1. Usia

Dari penelitian sebelumnya menunjukkan bahwa Sebagian besar orang yang menderita diabetes terjadi pada usia 40-49 tahun, tetapi hal tersebut tidak menutup kemungkinan bahwa seseorang dengan usia kurang dari 45 tahun dapat terserang penyakit ini (Yosmar et al., 2018).

1. BMI

*Body Mass Index* (BMI) merupakan salah satu faktor yang sangat segnifikan terhadap penyakit diabetes melitus, seseorang yang memiliki rentang BMI 30-40 yang dimana tergolong kategori obesitas. Seseorang yang mengalami obesitas sangat rentan terkena pemyakit diabetes dikarenakan sesuai dengan literatur bahwa obesitas memiliki hubungan erat dengan resistensi insulin sehingga menghambat absorbsi glukosa ke dalam otot dan sel lemak sehingga glukosa di dalam darah meningkat (Yosmar et al., 2018).

1. Jenis Kelamin

Mayoritas penderita diabetes yaitu berjenis kelamin perempuan, seorang perempuan beresiko terkena diabetes melitus karena fisik perempuan memiliki kesempatan untuk terjadinya peningkatan indeks masa tubuh yang lebih besar. Sindrom siklus (*Premenstual Syndrome*), pasca *menopouse* yang dapat membuat distribusi lemak tubuh mudah terakumulasi sehingga wanita lebih beresiko terkena diabetes (Imelda, 2019).

1. Hipertensi

Hipertensi adalah penyakit yang sering menyertai pengidap diabetes melitus. Tingginya prevalensi hipertensi terhadap penderita diabetes beresiko 4-5 kali lebih besar sebagai penyebab kematian jantung koroner dan stroke. Menurut penelitian sebelumnya ditemukan bahwa riwayat hipertensi memiliki keterkaitan erat dengan kasus diabetes melitus tipe 2, resikonya lebih tinggi 2,629 kali disbanding bukan pengidap hipertensi (Nuraisyah, 2018).

Selain faktor dari diabetes melitus, gejala dari diabetes juga harus diperhatikan, seseorang yang mengidap DM memiliki gejala seperti *polyuria* (sering buang air kecil), polydipsia (sering merasa haus), dan juga polyfagia (sering merasa lapar), serta turunnya berat badan tanpa diketahui penyebabnya. Selain hal tersebut, gejala lain yang sering dirasakan oleh penderita DM adalah seperti keluhan tubuh terasa lemas, kurang bertenaga, kaki dan tangan seringkali terasa kesemutan, penyembuhan luka yang berkepanjagan, dan mata kabur. Namun beberapa kasus juga penderita tidak mengalami gejala yang telah disebutkan diatas (Febrinasari et al., 2020).

### 2.2.2 Klasifikasi Diabetes Melitus

Berdasarkan penelitian terdahulu, klasifikasi diabetes melitus dibedakan menjadi tiga macam :

1. Diabetes Melitus Tipe 1

Diabetes tipe I merupakan gangguan sistemik dampak terjadinya kelainan metabolisme glukosa biasanya ditandai oleh tingginya gula darah melebihi batas normal dampak dari kerusakan destruksi pankreas baik oleh karena proses autoimun atau idiopatik sebagai dampaknya produksi insulin berkurang atau bahkan berhenti (Adelita et al., 2020).

1. Diabetes Melitus Tipe 2

Diabetes melitus tipe II adalah penyakit hiperglikemi dampak dari insentivitas sel terhadap insulin. Kadar insulin yang mungkin sedikit menurun atau berada pada rentang normal. Diabetes melitus tipe II ini sering disebut dengan non insulin dependent diabetes melitus dikarenakan insulin tetap dihasilkan oleh sel-sel beta pankreas.

1. Diabetes Pada Kehamilan

Diabetes melitus pada kehamilan atu disebut dengan diabetes melitus gestasional (DMG) merupakan gangguan toleransi glukosa yang pertama kali ditemukan pada saat wanita mengalami kehamilan. DMG adalah keadaan dimana saat wanita yang sebelumya tidak pernah menderita diabetes melitus lalu saat kehamilan menunjukkan kadar gula darah tinggi (hiperglikemia) yang terjadi selama kehamilan (Kurniawan, 2016).

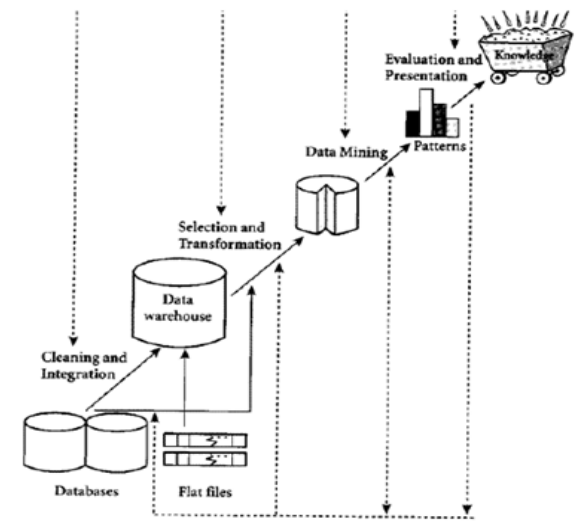
## 2.3 Kecerdasan Buatan

Kecerdasan buatan (*Artifical Intellegence* atau AI) merupakan suatu kecerdasan yang ditambahkan didalam suatu sistem atau bisa disebut kemampuan untuk mentafsirkan data eksternal dengan benar serta mengolah data tersebut yang dimaksudkan untuk tujuan tertentu. Pemanfaatan AI oleh industri tidak hanya terbatasi di sektor industri telekomunikasi, namun juga disektor pemerintahan, manufaktur, jasa serta disektor perbankan (Ririh et al., 2020). Kecerdasan buatan diciptakan lalu ditanamkan pada mesin atau komputer dengan maksud dan tujuan agar mesin dapat bekerja seperti yang dilakukan oleh manusia.

## 2.4 Data Mining

Ilmu data mining telah menjadi sangat diminati oleh masyarakat di berbagai dunia dalam beberapa tahun terakhir, karena didalam dunia sistem informasi ilmu data mining memiliki fungsi untuk alat mengukur data yang memiliki jumlah besar dan dapat mengolah data menjadi informasi yang berguna. Data mining, juga disebut *Knowledge Discovery in Databases* (KDD) suatu kegiatan yang meliputi pengumpulan data, pemakaian data historis berguna untuk menentukan pola keteraturan, pola hubungan didalam set data yang berukuran besar (Septiani, 2017).

Didalam data mining memiliki langkah-langkah proses yang sangat penting dan harus dilakukan saat mengukur data sebelum menjadi informasi yang berguna. Berikut merupakan gambar dalam proses data mining.



Gambar 2. 1 Langkah-langkah proses data mining

**(Abdillah 2015)**

Berikut adalah penjelasan dari Langkah-langkah proses data mining (Carolina and Ade, 2020) :

1. *Data Cleaning:* Pada proses ini yaitu pembersihan data agar tidak ada noise didalam data, data-data yang tidak sesuai, dan data yang telah hilang saat proses klasifikasi. Apabila terdapat data yang memiliki atribut yang tidak konsisten maka harus dibersihkan agar tidak mempengaruhi atau mengurangi nilai akurasi.
2. *Data integration*: Mengenali atribut-atribut data untuk dilakukan proses integrasi data agar hasil tidakmenyimpang dari yang diinginkan. Sesudah melakukan integrasi data, pembersihan dan transformasi data diulang kembali untuk menjaga data, hal ini dilakukan bertujuan untuk menjaga konsistensi data.
3. *Data selection*: Proses penyeleksian data harus dilakukan karena tidak semua data didalam database digunakan. Data yang digunakan hanya yang sesuai dengan analisis yang akan dilakukan.
4. *Data transformation*: Didalam proses ini dilakukan pengubahan data atau penggabungan data yang harus dilakukan sebelum proses data mining, karena didalam proses data mining membutuhkan format data yang khusus sebelum diaplikasikan.
5. *Process mining*: Proses ini merupakan proses utama yang dilakukan berdasarkan metode yang sudah ditentukan untuk menghasilkan suatu informasi yang berguna.
6. *Evaluation*: Proses ini mengidentifikasi model berdasarkan pengetahuan yang ada dimana hasil dari teknik data mining dievaluasi untuk melakukan penilaian apakah hipotesa telah tercapai, apabila tpemrosesan tidak tercapai maka dilakukan perbaikan pada proses data mining, dan dilakukan dengan metode yang lain.
7. *Knowledge presentation*: Teknik ini direpresentasikan untuk menyajikan pengetahuan. Tahap ini berfokus pada bagaimana visualisasi darai hasil yang telah diperoleh sehingga dapat mudah dipahami oleh pengguna.
   1. Klasifikasi

Klasifikasi merupakan suatu proses yang memiliki tujuan menemukan model yang dapat membedakan dan menggambarkan kelas pada suatu data (Setiawati et al., 2019). Model tersebut dapat terbentuk berdasarkan analisis data pelatihan. Metode klasifikasi adalah salah satu teknik dari data mining yang popular dan merupakan metode yang sederhana sehingga telah banyak digunakan masyarakat didunia. Metode klasifikasi memiliki dua tahap prosedur yaitu pengembangan model dan evaluasi model dengan menggunakan pengujian. Ada beberapa algoritma dan teknik analisis dari metode klasifikasi antara lain Analisis regresi, Model Bayesian, Analisis Korelasi, Algoritma berdasarkan jarak, Algoritma berdasarkan Jaringan Syaraf atau *Neural Network*, Algoritma berdasarkan Pohon Keputusan atau *Decision Tree* (Jalota & Agrawal, 2019).

## 2.6 Machine Learning

*Machine learning* atau pembelajaran mesin adalah cabang ilmu komputer yang meneliti bagaimana mesin memecahkan masalah tanpa diprogram secara eksplisit. *Machine learning* diciptakan oleh Arthur Samuel pada tahun 1952 dengan sebuah program permainan catur yang terdapat pada komputer IBM, komputer melakukan pembelajaran dari gerakan agar bisa memenangkan permainan tersbut dengan menyimpannya dimemori.

*Machine learning* memiliki beberapa algoritma, namun untuk secara luas *machine learning* memiliki dua teknik dasar belajar antara lain:

1. *Supervised*

*Supervised learning* adalah teknik pembelajaran mesin yang dapat menerima informasi yang sudah ada pada data dengan memberikan label tertentu, yang nanti dari data tersebut akan mempelajari hubungan dan ketergantungan antar data. Teknik ini memiliki cara kerja dengan memasukkan informasi sebagai input dan data berlabel sebagai output.

1. *Unsupervised*

*Unsupervised learning* atau bisa disebut pembelajaran mesin tanpa pengawasan. Teknik pembelajaran yang dilakukan yaitu pada dataset mentah atau data langsung tanpa memiliki label, sehingga teknik *machine learning* akan mengidentifikasi pola secara komputasi tanpa bantuan dari pengembang.

1. *Reinformation Learning*

*Reinformation learning* merupakan teknik *machine learning* yang memiliki kebalikan dari *supervised learning*, data tidak berlabel dan sistem tidak mengetahui output yang sesuai. Teknik machine learning ini memiliki tujuan untuk mengeksplorasi data dan menemukan pola didalamnya.

## 2.7 *Confusion Matrix*

*Confusion Matrix* merupakan sebuah tabel yang berisi baris data yang diuji atau diprediksi benar dan salah oleh suatu metode algoritma. Didalam data mining *confusion matrix* merupakan metode evaluasi yang menghitung akurasi nilai dari model klasifikasi. Nilai akurasi didalam *confusion matrix* terdapat dua kelas yaitu kelas positif dan kelas negative yang membentuk tabel matrix (Ramdhani, 2016). Kelas tersebut terdiri dari *True positive, True negative, False positive, False Negative.*

Tabel 2. 1 Model *Confusin Matrix*

|  |  |  |
| --- | --- | --- |
| Kelas aktual | Kelas prediksi | |
| + | - |
| + | *true positive* | *false negative* |
| - | *false positive* | *true negative* |

Sumber: (Bramer, 2020)

Didalam tahap evaluasi menggunakan *confusion matrix* akan mendapatkan nilai *accuracy, precision, recall*.

* Akurasi

Akurasi merupakan nilai evaluasi oleh *confusion matrix* keseluruhan jumlah data yang memiliki prediksi benar atau juga bisa disebut tingkat kedekatan nilai aktual dengan nilai prediksi.

*Accuracy =*

* Presisi

Presisi merupakan nilai evaluasi oleh *confusion matrix* yang memiliki rasio prediksi benar posiif yang berasal dari keseluruhan data prediksi positif.

*Precision* =

* Sensivisitas *(Recall)*

Sensivisitas merupakan nilai evaluasi oleh *confusion matrix* yang memiliki rasio prediksi benar positif yang membandingkan dengan keseluruhan data benar positif

*Recall* =

Keterangan:

TP = jumlah data positif yang telah diklasifikasikan sebagai positif.

FP = jumlah data negatif yang telah diklasifikasikan sebagai positif.

FN = jumlah data positif yang telah diklasifikasikan sebagai negatif.

TN = jumlah data negatif yang telah diklasifikasikan sebagai negatif.

## 2.8 *Naïve Bayes*

*Naïve Bayes* merupakan salah satu algoritma machine learning untuk masalah klasifikasi. Algoritma *Naïve Bayes* termasuk kedalam *Supervised Learning*, maksud dari supervise learning adalah dimana data yang akan dijadikan sumber memiliki label, kelas, atau target sebagai acuan. Algoritma ini diciptakan oleh seorang ahli statistic dan filsuf, Thomas Bayes. Naïve Bayes menggunakan teori probabilitas dalam memecahkan suatu kasus. Dalam menggunakan algoritma *Naïve Bayes* ada beberapa fase yang wajib dilakukan untuk membuat keputusan yang akurat. Ada lima fase didalam metode ini yaitu *Training Data*, *Algorithm*, *Model*, *Testing Data*, dan *Accuracy*. Pengertian learning didalam algoritma ini adalah fase pembelajaran yang menghitung probabilitas dalam suatu kasus menggunakan algoritma. Yang dimaksud testing yaitu pengujian dalam menggunakan model yang dibuat berdasarkan Training Data dan *Learning Algortihm* sebelum nya pada suatu data baru. *Accuracy* yaitu fase pengukuran performa untuk mendapatkan tingkat akurasi (Nasruddin, 2021). Tahapan dari proses perhitungan algoritma *Naïve Bayes* antara lain:

1. Mehitung nilai *prior probability* setiap kelas dan hitung mean setiap kelas
2. Menghitung nilai *likehood* untuk menghitung peluang setiap atribut pada kelas
3. Menghitung nilai *posterior*
4. Bandingkan hasil per kelas.

Dasar teori *Naïve Bayes* menggunakan rumus dalam menentukan nilai peluang dari suatu hipotesis yaitu seperti dibawah ini:

Keterangan:

A = Data dengan kelas yang masih belum diketahui.

B = Hipotesis data A merupakan kelas.

P(B|A) = Probabilitas akhir (*posterior* probability) suatu kelas B apabila kondisi kelas A terjadi.

P(B) = Probabilitas awal (*prior probability*) kelas B terjadi, tanpa adanya kondisi atribut apapun.

P(A|B) = Probabilitas awal attribut A terjadi, akan mempengaruhi kelas B (*likehood probability*).

P(A) = Probabilitas awal atribut A.

Ada beberapa kasus didalam perhitungan menggunakan *Naïve Bayes* yaitu ketika data yang digunakan merupakan bernilai kontinyu atau numerik, sedangkan algoritma *Naïve Bayes* bisa dilakukan klasifikasi jika data tersebut bernilai diskrit atau *polynominal*. Untuk menyelesaikan klasifikasi dari atribut yang benilai kontinyu atau numerik adalah dengan mencari nilai mean dan standar deviasi dari masing-masing parameter yang menunjukkan data angka. Formula yang digunakan dalam menghitung nilai rata-rata (*mean*) sebagai berikut:

Dimana:

= nilai rata-rata (*mean*).

= nilai sampel ke -i.

n = jumlah sampel.

Dan untuk persamaan menghitung standar deviasi sebagai berikut:

Dimana:

= standar deviasi.

= nilai x ke -i.

= nilai rata-rata (*mean*).

n = jumlah sampel.

Lalu setelah mengetahui nilai *mean* dan standar deviasi untuk atribut yang bernilai kontinyu atau numerik, selanjutnya menghitung nilai probabilitas pada setiap atribut disetiap kelas. Untuk menghitung nilai probabilitas untuk setiap atribut maka harus menghitung jumlah data sesuai dengan kategori yang sama lalu dibagi dengan jumlah data pada kategori tersebut.

Selanjutnya yaitu menghitung nilai probabilitas untuk fitur data uji yang memiliki atribut kontinyu atau numerik. Formula untuk persamaan dalam mencari nilai distribusi gaussian sebagai berikut:

Dimana:

P = peluang.

Xi = atribut ke -i.

Xj = atribut ke -j.

= menunjukkan rata-rata seluruh atribut.

= diviasi standar, varian dari seluruh atribut.

## 2.9 *Adaboost*

Algoritma *Adaptive Boosting* (*Adaboost*) merupakan algoritma yang diperlukan untuk mengambil suatu keputusan. Metode *Adaboost* merupakan teknik dari metode *Ensemble Methods/Boosting* yang digunakan untuk menyeimbangkan dan menggabungkan record-record dalam suatu class dengan memberikan koefisien bobot berdasarkan performansi hasil pelatihan sebelumnya untuk meningkatkan nilai akurasi dengan meminimalkan kesalahan fungsional dalam klasifikasi (Pareza Alam Jusia, 2018).

Dalam boosting, sampel yang tidak diprediksi dengan benar oleh pengklasifikasi di sirkuit akan dipilih lebih sering daripada sampel yang diprediksi dengan benar. Jadi *adaboost* mencoba membuat pengklasifikasi dasar baru mana yang lebih baik untuk memprediksi sampel dengan kinerja buruk dari pengklasifikasi dasar sebelumnya. *Adaboost* dapat dikombinasikan dengan algoritma klasifikasi lainnya untuk meningkatkan kinerja klasifikasi. Dalam meningkatkan kinerja maka bobot akan dikaitkan dengan setiap pelatihan diikuti dengan klasifikasi yang berbeda akan menggunakan bobot ini untuk pelatihan. Bobot akan diubah secara iterasi berdasarkan kinerja klasifikasi (C.Aggarwal, 2015). Metode *Adaboost* memiliki banyak varian didalam perkembangannya antara lain yaitu *Adaboost.M1, Adaboost.M1W, Kullback-Leibler Boosting* (*KLBoosting*) dan *Jensen-Shannon Boosting* (*JSBoost*). *Adaboost.M1* merupakan generalisasi secara langsung dari *Adaboost* untuk masalah multikelas. Untuk varian *KLBoosting* dan *JSBoosting* digunakaan untuk masalah dalam pendeteksian pola atau gambar (objek). Dalam penelitian ini menggunakan varian *Adaboost.M1*. Berikut ini merupakan perumusan dalam algoritma *Adaboost.M1*:

Input:

Dataset

*Weak Learn* (L) atau Algoritma Pembelajaran Dasar

Jumlah Iterasi (T)

Proses:

Inisiasi nilai bobot:

Keterangan :

t = Iterasi

T = Jumlah Iterasi

Pengujian *base learn* ht, terhadap distribusi Dt

keterangan:

ht = Model pengklasifikasi rendah

L = Algoritma pengklasifikasi rendah

Dt = Distribusi

Menghitung eror sampel pelatihan dari persamaan:

Keterangan:

€ = Kesalahan Bobot

Menghitung nilai bobot dari persamaan:

keterangan:

αt = hipotesis bobot.

Perbarui nilai bobot:

Keterangan:

-αt = klasifikasi benar (hasil)

αt = klasifikasi salah (hasil)

Zt = Perbaruhi distribusi, faktor normalisasi yang mengaktifkan Dt+1 menjadi distibusi.

Output:

end for

Keterangan:

H(x) = Hasil akhir hipotesis

## 3.0 *Python*

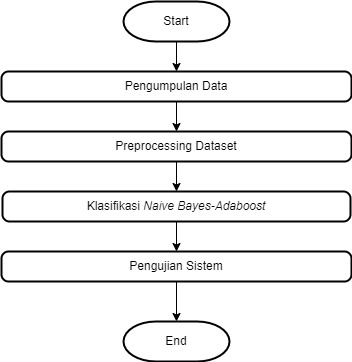
*Python* adalah bahasa pemrograman tingkat tinggi yang sekarang menjadi standar dalam dunia komputasi ilmiah. *Python* adalah bahasa pemrograman *open-source multi-platform* yang tersedia untuk berbagai sistem operasi (Windows, Linux, dan Sistem Apple). Selain itu, *Python* adalah bahasa pemrograman yang fleksibel dan mudah digunakan untuk pemula. Program yang ditulis dengan *Python* umumnya lebih mudah dibaca, dan banyak lagi Ringkas dibandingkan dengan menulis program dalam C atau Fortran. *Python* juga memiliki Modul standar yang menyediakan banyak fungsi dan algoritma untuk dipecahkan seperti mengurai data teks, memanipulasi dan menemukan file pada disk, membaca/tulis file terkompresi dan unduh data dari server web. Dengan menggunakan *Python*, pemrogram juga dapat dengan mudah menerapkan teknik komputasi tingkat lanjut seperti pemrograman berorientasi objek atau OOP (Herho, 2017).

# BAB III

# METODOLOGI PENELITIAN

## 3.1 Gambaran Umum

Penelitian ini dapat bekerja dengan baik apabila langkah-langkah yang akan dilakukan dapat konsisten dengan mengikuti alur yang telah dibuat. Penelitian ini dimulai dari tahap studi literatur hingga prediksi penyakit diabetes melitus. Kerangka penelitian ini berfungsi sebagai pedoman dalam melakukan penelitian secara terstruktur agar mencapai hasil sesuai yang diinginkan. Gambar 3.1 berikut ini adalah alur atau kerangka dari penelitian ini**.**



Gambar 3. 1 Gambaran Umum

## 3.2 Analisa Kebutuhan

Pada tahap Analisa kebutuhan ini meliputi kebutuhan *software, hardware,* data yang dibutuhkan untuk implementasi diagnosis penyakit diabetes melitus menggunakan metode *Naïve Bayes-Adaboost.*

### 3.2.1 Kebutuhan Software

1. *Jupyter NoteBook* (anaconda 3).
2. Sistem Operasi Windows 10 Pro 64bit.

### 3.2.2 Kebutuhan Hardware

1. Intel Core i5-7200U CPU 2.50GHz.
2. Memory 8 GB.
3. VGA NVIDIA GeForce 930MX 2GB.

### 3.3.3 Data yang dibutuhkan

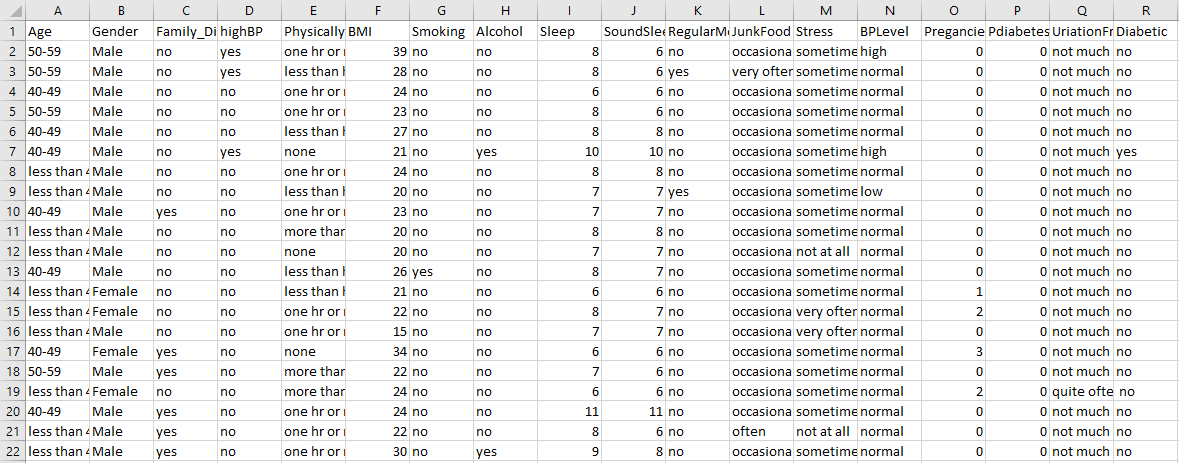
1. Data penyakit diabetes melitus.
2. Data yang berhubungan dengan faktor penyakit diabetes melitus.

## 3.3 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah *Diabetes Dataset* 2019 (Neha Prerna Tigga and Dr. Shruti Garg) dari *dataset public* yang bersumber dari *Kaggle*. Kumpulan data ini berjumlah 952 kasus dengan 17 variabel dan 1 target biner diabetes. Tabel 3.1 dibawah ini mendeskripsikan atribut dari *dataset diabetes* 2019.

|  |  |
| --- | --- |
| 1 | <https://www.kaggle.com/datasets/tigganeha4/diabetes-dataset-2019?resource=download> |

Gambar dataset yang diambil dari *website Kaggle* dapat dilihat pada gambar 3.2 berikut ini:



Gambar 3. 2 Dataset Diabetes Melitus

Tabel 3. 1 Variabel Dataset Diabetes

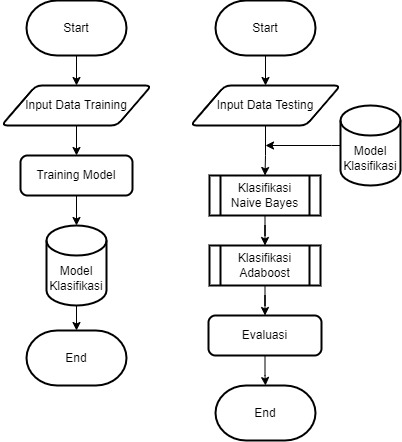
|  |  |  |  |
| --- | --- | --- | --- |
| No | Atribut | Deskripsi | Tipe Data |
| 1 | Age | Umur Pasien (18 or above) | Numerik |
| 2 | Gender | Laki-laki, Perempuan | Kategorikal |
| 3 | Family diabetes | Iya/Tidak | Kategorikal |
| 4 | High Blood Pleasure | Iya/Tidak | Kategorikal |
| 5 | Physically Active | -Tidak sama sekali  - Kurang dari setengan jam  - Lebih dari setengah jam  - Satu jam atau lebih | Kategorikal |
| 6 | BMI | Indeks Massa Tubuh | Numerik |
| 7 | Smoking | Iya/Tidak | Kategorikal |
| 8 | Alcohol | Iya/Tidak | Kategorikal |
| 9 | Sleep | Jam tidur | Numerik |
| 10 | Sound Sleep | Jam tidur nyenyak | Numerik |
| 11 | Regular Medicine | Iya/Tidak | Kategorikal |
| 12 | Junk Food | - Kadang-kadang  - Sering  - Sangat sering  - Selalu | Kategorikal |
| 13 | Stress | - Tidak pernah  - Kadang-kadang  - Sering  - Selalu | Kategorikal |
| 14 | Blood Pleasure Level | Tinggi/normal/rendah | Kategorikal |
| 15 | Pregnancies | Jumlah kehamilan | Numerik |
| 16 | Pdiabetes | Iya/Tidak | Kategorikal |
| 17 | Urination Frequency | - Tidak banyak  - Cukup banyak | Kategorikal |
| 18 | Diabetic | Iya/Tidak | Kategorikal |

## 3.4 Preprocessing Dataset

Pada tahap ketiga ini yaitu pengolahan data. Tujuan dari tahapan ini yaitu melakukan *preprocessing* data yang bertujuan untuk mengubah data mentah menjadi data yang mudah digunakan untuk proses data mining. Beberapa data terdapat *missing value* yaitu nilai kosong. Maka pada tahap ini beberapa data yang kosong akan dilakukan pengisian dengan menggunakan nilai rata-rata dari keseluruhan data dari variable yang terdapat nilai kosong tersebut. Pengisian data tersebut bertujuan untuk mengatasi kesalahan dalam proses klasifikasi.

## 3.5 Klasifikasi *Naïve Bayes - Adaboost*

Rancangan algoritma sistem merupakan gambaran tentang bagaimana menjelaskan proses klasifikasi menggunakan metode algoritma *Naïve Bayes – Adaboost* sehingga menghasilkan output berupa diagnosis penyakit diabetes melitus. Didalam sistem klasifikasi terdapat dua tahap yaitu *training* dan *testing*. Untuk tahap training dilakukan untuk melatih model agar klasifikasi dapat berjalan dengan baik, sementara untuk tahap testing dilakukan untuk memperoleh hasil prediksi klasifikasi menggunakan model yang sudah dilatih pada tahap *training*. Berikut *flowchart* proses klasifikasi tersebut dapat dilihat pada gambar 3.3.

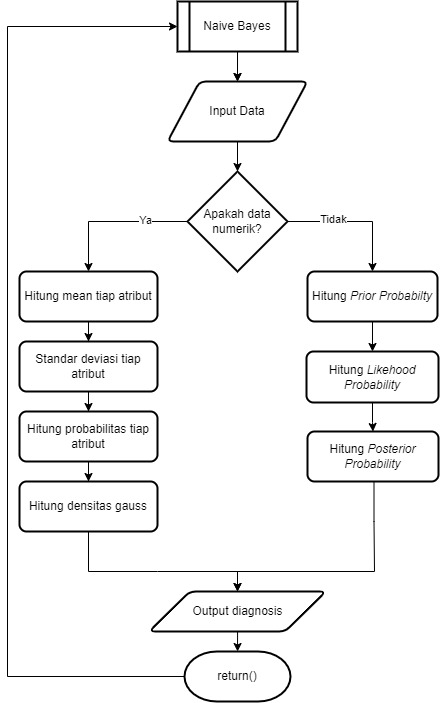


Gambar 3. 3 Flowchart Algoritma Sistem

Langkah-langkah *flowchart* akan dijelaskan dibawah ini:

1. Input data training dari dataset yang telah dibagi dengan data *testing*.
2. Data sebelumnya dilakukan preprocessing untuk menjadikan data siap diolah.
3. Lalu proses training model menggunakan algoritma yang sudah ditentukan.
4. Setelah mendapatkan model dari *training* yang telah dilakukan selanjutnya model tersebut disimpan untuk digunakan dalam proses *testing*.
5. Input data *testing* dari dataset yang sudah dibagi dengan data *training.*
6. Selanjutnya proses data *testing* untuk memperoleh hasil diagnosis dengan menggunakan model yang sudah dilatih pada data *training*.
7. Tahap evaluasi yaitu untuk mengukur tingkat keberhasilan sistem yang telah dilakukan.
8. Output yang akan didapatkan berupa diagnosis penyakit diabetes melitus positf atau negative dan mengetahui tingkat akurasi dari kinerja algoritma tersebut.

### 3.5.1 Flowchart Klasifikasi *Naïve Bayes*

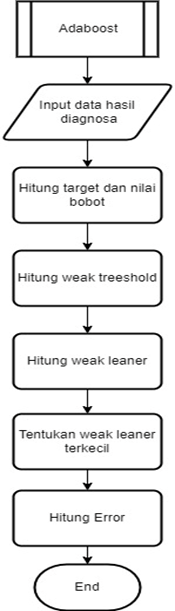


Gambar 3. 4 Flowchart Naïve Bayes

Keterangan:

1. Menginputkan data yang merupakan faktor penyakit *diabetes melitus*.
2. Jika dataset berupa kategori hitung nilai probabilitas priorpada dataset yang memiliki 2 kelas, yaitu kelas diabetes dan tidak diabetes.
3. Langkah selanjutnya yaitu menghitung nilai probabilitas *likehood* yang diperoleh dari seluruh jumlah faktor penyakit *diabetes melitus* dibagi dengan jumlah seluruh data tiap kelas. Perhitungan nilai likehood tiap gejala menggunakan persamaan berikut.
4. Lalu langkah selanjutnya setelah mendapatkan nilai *prior* dan *likehood*, maka untuk nilai probabilitas posterior didapatkan dengan mengalikan nilai *prior* dari tiap kelas dengan nilai *likehood* tiap faktor pada setiap kelas.
5. Apabila dataset berupa numerik maka dihitung menggunakan persamaan gaussian.
6. Hitung nilai *mean* pada setiap atribut.
7. Hitung nilai standar deviasi pada setiap atribut.
8. Lalu menghitung nilai probabilitas setiap atribut.
9. Lalu selanjutnya menghitung densitas gauss.
10. Setelah menghitung semua yang diperlukan selanjunya yaitu melakukan pengujian dengan melakukan perhitungan data uji, maka sistem akan mendapatkan hasil diagnosa penyakit *diabetes melitus*.

### Flowchart Klasifikasi *Adaboost*



Gambar 3. 5 Flowchart Adaboost

Keterangan:

1. Langkah pertama dalam klasifikasi *adaboost* yaitu input data dari hasil diagnosis
2. Langkah selanjutnya yaitu hitung target dan nilai bobot, untuk nilai target dapat diperoleh dari banyaknya nilai data *training*, untuk nilai bobot didapatkan dari persamaan D=1/m.
3. Selanjutnya menghitung nilai *weak treeshold* dengan mengambil nilai maksimum dan minimun menggunakan teknik regresi linear, lalu membandingkan nilai +1 dan -1 pada masing-masing x.
4. Setelah itu mencari *weak leaner* dengan memilah fitur-fitur lemah, lalu mencari error dengan membandingkan nilai threshold a atau threshold b dengan nilai target.
5. Setelah memilah fitur lemah selanjutnya yaitu menentukan *weak leaner* terkecil
6. Tahap terakhir hitung nilai *error* pada masing-masing nilai x. Nilai *error* diperoleh dari nilai eror tertinggi.

### 3.5.3 *Preprocessing* Data

Sebelum klasifikasi dilakukan dataset perlu dilakukan *preprocessing* data untuk menghilangkan data noise yang ditemukan seperti missing value. Berikut tabel 3.1 merupakan contoh dataset dengan variabel BMI yang terdapat missing value.

Tabel 3. 2 Variabel BMI Sebelum di *Preprocessing*

|  |
| --- |
| BMI |
| 28 |
| 30 |
| 30 |
|  |
| 27 |
| 22 |
|  |

Lalu dari variabel yang terdapat *missing value*, untuk memeperbaiki nilai yang kosong akan diisi dengan nilai mean dari seluruh nilai yang terdapat pada data variabel BMI. Mengisi nilai yang kosong lebih baik dari pada menghapus nilai yang kosong tersebut.

Mean =

=

=

= 27,4

Dari perhitungan tersebut telah mendapatkan nilai mean dari seluruh data, selanjutnya memasukkan nilai tersebut kedalam nilai yang kosong. Tabel 3.3 menunjukkan hasil data setelah dilakukan *preprocessing*.

Tabel 3. 3 Variabel BMI Setelah di *Preprocessing*

|  |
| --- |
| BMI |
| 28 |
| 30 |
| 30 |
| 27,4 |
| 27 |
| 22 |
| 27,4 |

### 3.5.4 Perhitungan Manual *Naïve Bayes – Adaboost*

Pada tahap awal perhitungan yaitu menggunakan klasifikasi *Naïve Bayes* dengan melakukan pengambilan data latih dari *dataset diabetes* 2019. Variabel yang digunakan dalam mengklasifikasikan data yaitu:

1. *Age*

Merupakan variabel usia pasien yang dikelompokkan menjadi empat kategori yaitu 40 tahun kebawah, 40-49 tahun, 50-59 tahun, dan 60 tahun keatas.

1. *Gender*

Merupakan variabel jenis kelamin pasien yang dikelompokkan menjadi dua yaitu laki-laki dan perempuan.

1. *Family diabetes*

Merupakan variabel riwayat keluarga pasien yang menderita diabetes, dikelompokkan menjadi dua yaitu iya atau tidak

1. *High Blood Pleasure*

Merupakan variabel tekanan darah dimana pasien memiliki tekanan darah atau tidak, dikelompokkan menjadi dua yaitu iya atau tidak.

1. *Physically Active*

Merupakan variabel aktifitas fisik yang dilakukan pasien, dikelompokkan menjadi empat yaitu tidak sama sekali**,** kurang dari setengan jam**, l**ebih dari setengah jam**, s**atu jam atau lebih.

1. *BMI*

Merupakan variabel indeks massa tubuh pasien yang dapat dikategorikan berdasarkan besar indeks massa tubuh pasien.

1. *Smoking*

Merupakan variabel dimana pasien memiliki riwayat perokok atau tidak, dikelompokkan menjadi dua yaitu iya atau tidak.

1. *Alcohol*

Merupakan variabel dimana pasien memiliki riwayat peminum alkohol atau tidak, dikelompokkan menjadi dua yaitu iya atau tidak.

1. *Sleep*

Merupakan variabel durasi jam tidur yang dilakukan pasien.

1. *Sound Sleep*

Merupakan variabel durasi jam tidur pulas pasien.

1. *Regular Medicine*

Merupakan variabel dimana pasien mengkonsumsi obat-obatan secara rutin dari

1. *Junkfood*

Merupakan variabel dimana pasien mengkonsumsi makanan cepat saji, dikelompkkan menjadi tiga yaitu kadang-kadang, sering, dan sangat sering, selalu.

1. *Stress*

Merupakan variabel dimana pasien mengalami stress atau tidak, dikelompokkan menjadi empat yaitu tidak pernah, kadang-kadang, sering, dan selalu.

1. *Blood Pleasure Level*

Merupakan variabel level tekanan darah pasien, dikelompokkan menjadi tiga yaitu rendah, normal, dan tinggi.

1. *Pregencies*

Merupakan variabel dimana riwayat pasien dalam jumlah melahirkan.

1. *Pdiabetes*

Merupakan diabetes yang berlangsung selama masa pasien mengandung sampai persalinan, dikelompokkan menjadi dua yaitu iya atau tidak.

1. *Urination Frequency*

Merupakan variabel bagaimana frekuensi urin pasien, dikelompokkan menjadi dua yaitu tidak banyak dan cukup banyak.

1. *Diabetic*

Merupakan variabel dimana pasien dinyatakan postif atau negatif terdiagnosa diabetes.

Dari dataset diatas terdapat 13 variabel bersifat kategori dan 4 variabel bersifat numerik. Klasifikasi ini menggunakan dataset dari web Kaggle yaitu *Diabetes Dataset 2019* sebanyak 20 data training dan 5 data testing, data diambil secara acak. Data testing dapat dilihat pada gambar 3.6 berikut ini.



Gambar 3. 6 Gambar Data Training

1. Menghitung Nilai *Prior*

Langkah pertama yaitu hitung nilai *prior* pada setiap kelas, setelah menentukan data training yang dapat dilihat pada gambar 3.6. Berikut perhitungan nilai *prior*:

P(K1) = Jumlah Diabetes / Seluruh Data

= 7 / 19

= 0,35

P(K2) = Jumlah Tidak Diabetes / Seluruh Data

= 12 / 19

= 0,63

Tabel 3. 4 Nilai Prior

|  |  |
| --- | --- |
| P (Kelas Diagnosis) | Nilai *Prior* |
| Diabetes | 0,35 |
| Tidak Diabetes | 0,63 |

1. Menghitung Nilai *Likehood*

Langkah kedua setelah mendapatkan nilai *prior probability* adalah mencari nilai *likehood probability*, dimana nilai tersebut didapatkan dari setiap jumlah faktor penyakit diabetes yang terjadi pada setiap kelas lalu dibagi dengan jumalh keseluruhan data pada faktor penyakit diabetes setiap kelas. Contoh perhitungan nilai *likehood* dapat dilihat pada tabel 3.5 dan 3.6 berikut:

Tabel 3. 5 Nilai *Likehood* Kelas Diabetes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Faktor | Jumlah  faktor terdiagnosis (diabetes) | Jumlah kelas terdiagnosis (diabetes) | Nilai Likehood  (diabetes) |
| 1 | Umur (dibawah 40 tahun) | 2 | 7 | 2/7 = 0,29 |
| 2 | Umur (40-49 tahun) | 1 | 7 | 1/7 = 1,14 |
| 3 | Umur (50-59 tahun) | 2 | 7 | 2/7 = 0,29 |
| 4 | Umur (diatas 60 tahun) | 2 | 7 | 2/7 = 0,29 |
| 5 | Jenis Kelamin (laki-laki) | 5 | 7 | 5/7 =0,71 |
| 6 | Jenis Kelamin (perempuan) | 2 | 7 | 2/7 = 0,29 |
| 7 | Riwayat keluarga (iya) | 5 | 7 | 5/7 = 0,71 |
| 8 | Riwayat keluarga (tidak) | 2 | 7 | 2/7 = 0,29 |
| 9 | Tekanan darah (iya) | 4 | 7 | 4/7 = 0,57 |
| 10 | Tekanan darah (tidak) | 3 | 7 | 3/7 = 0,43 |
| 11 | Aktifitas fisik (tidak sama sekali) | 3 | 7 | 3/7 = 0,43 |
| 12 | Aktifitas fisik (kurang dari setengah jam) | 1 | 7 | 1/7 = 1,14 |
| 13 | Aktifitas fisik (lebih dari setengah jam) | 1 | 7 | 1/7 = 1,14 |
| 14 | Aktifitas fisik (satu jam atau lebih) | 2 | 7 | 2/7 = 0,29 |
| 15 | Perokok (iya) | 1 | 7 | 1/7 = 1,14 |
| 16 | Perokok (tidak) | 6 | 7 | 6/7 = 0,86 |
| 17 | Alkohol (iya) | 4 | 7 | 4/7 = 0,57 |
| 18 | Alkohol (tidak) | 3 | 7 | 3/7 = 0,43 |
| 19 | Obat tanpa resep dokter (iya) | 5 | 7 | 5/7 = 0,71 |
| 20 | Obat tanpa resep dokter (tidak) | 2 | 7 | 2/7 = 0,29 |
| 21 | *Junkfood* (kadang-kadang) | 5 | 7 | 5/7 = 0,71 |
| 22 | *Junkfood* (sering) | 1 | 7 | 1/7 = 1,14 |
| 23 | *Junkfood* (sangat sering) | 1 | 7 | 1/7 = 1,14 |
| 24 | *Junkfood*(selalu) | 0 | 7 | 0/7 = 0 |
| 25 | Stress (tidak pernah) | 1 | 7 | 1/7 = 1,14 |
| 26 | Stress (kadang-kadang) | 4 | 7 | 4/7 = 0,57 |
| 27 | Stress (sering) | 1 | 7 | 1/7 = 1,14 |
| 28 | Stress (selalu) | 1 | 7 | 1/7 = 1,14 |
| 29 | Level tekanan darah (rendah) | 0 | 7 | 0/7 = 0 |
| 30 | Level tekanan darah (normal) | 2 | 7 | 2/7 = 0,29 |
| 31 | Level tekanan darah (tinggi) | 5 | 7 | 5/7 = 0,71 |
| 32 | Diabetes dimasa kehamilan (iya) | 1 | 7 | 1/7 = 1,14 |
| 33 | Diabetes dimasa kehamilan (tidak) | 6 | 7 | 6/7 = 0,86 |
| 34 | Frekuensi urin (tidak banyak) | 5 | 7 | 5/7 = 0,71 |
| 35 | Frekuensi urin (cukup banyak) | 2 | 7 | 2/7 = 0,29 |

Tabel 3. 6 Nilai Likehood Kelas Tidak Diabetes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Faktor | Jumlah  faktor terdiagnosis  (tidak diabetes) | Jumlah kelas terdiagnosis (tidak diabetes) | Nilai Likehood  (tidak diabetes) |
| 1 | Umur (dibawah 40 tahun) | 5 | 12 | 5/12 = 0,42 |
| 2 | Umur (40-49 tahun) | 3 | 12 | 3/12 = 0,25 |
| 3 | Umur (50-59 tahun) | 3 | 12 | 3/12 = 0,25 |
| 4 | Umur (diatas 60 tahun) | 1 | 12 | 1/12 = 0,08 |
| 5 | Jenis Kelamin (laki-laki) | 9 | 12 | 9/12 = 0,75 |
| 6 | Jenis Kelamin (perempuan) | 3 | 12 | 3/12 = 0,25 |
| 7 | Riwayat keluarga (iya) | 2 | 12 | 2/12 = 0,17 |
| 8 | Riwayat keluarga (tidak) | 10 | 12 | 10/12 = 0,83 |
| 9 | Tekanan darah (iya) | 2 | 12 | 2/12 = 0,17 |
| 10 | Tekanan darah (tidak) | 10 | 12 | 10/12 = 0,83 |
| 11 | Aktifitas fisik (tidak sama sekali) | 1 | 12 | 1/12 = 0,08 |
| 12 | Aktifitas fisik (kurang dari setengah jam) | 3 | 12 | 3/12 = 0,25 |
| 13 | Aktifitas fisik (lebih dari setengah jam) | 2 | 12 | 2/12 = 0,17 |
| 14 | Aktifitas fisik (satu jam atau lebih) | 6 | 12 | 6/12 = 0,5 |
| 15 | Perokok (iya) | 3 | 12 | 3/12 = 0,25 |
| 16 | Perokok (tidak) | 9 | 12 | 9/12 = 0,75 |
| 17 | Alkohol (iya) | 2 | 12 | 2/12 = 0,17 |
| 18 | Alkohol (tidak) | 10 | 12 | 10/12 = 0,83 |
| 19 | Obat tanpa resep dokter (iya) | 1 | 12 | 1/12 = 0,08 |
| 20 | Obat tanpa resep dokter (tidak) | 11 | 12 | 11/12 = 0,92 |
| 21 | *Junkfood* (kadang-kadang) | 9 | 12 | 9/12 = 0,75 |
| 22 | *Junkfood* (sering) | 2 | 12 | 2/12 = 0,17 |
| 23 | *Junkfood* (sangat sering) | 0 | 12 | 0/12 = 0 |
| 24 | *Junkfood* (selalu) | 1 | 12 | 1/12 = 0,08 |
| 25 | Stress (tidak pernah) | 1 | 12 | 1/12 = 0,08 |
| 26 | Stress (kadang-kadang) | 9 | 12 | 9/12 = 0,75 |
| 27 | Stress (sering) | 1 | 12 | 1/12 = 0,08 |
| 28 | Stress (selalu) | 1 | 12 | 1/12 = 0,08 |
| 29 | Level tekanan darah (rendah) | 1 | 12 | 1/12 = 0,08 |
| 30 | Level tekanan darah (normal) | 8 | 12 | 8/12 = 0,67 |
| 31 | Level tekanan darah (tinggi) | 3 | 12 | 3/12 = 0,25 |
| 32 | Diabetes dimasa kehamilan (iya) | 1 | 12 | 1/12 = 0,08 |
| 33 | Diabetes dimasa kehamilan (tidak) | 11 | 12 | 11/12 = 0,92 |
| 34 | Frekuensi urin (tidak banyak) | 9 | 12 | 9/12 = 0,75 |
| 35 | Frekuensi urin (cukup banyak) | 3 | 12 | 3/12 = 0,25 |

1. Menghitung Nilai Gaussian

Langkah ketiga yaitu hitung nilai gauss pada variabel yang bersifat numerik yaitu *BMI*, *Sleep, Sound Sleep, Pregancies*. Untuk mendapatkan nilai untuk variabel bersifat numerik yaitu hitung nilai rata-rata setiap variabel pada masing-masing kelas.

* Hitung nilai mean

Tahap pertama mengelompokkan jumlah diabetes dan tidak diabetes pada masing-masing kelas pada data training pada gambar 3.6.

Diabates : 7

Tidak diabetes : 12

Hitung nilai mean setiap variabel bersifat numerik pada masing masing kelas seperti berikut:

**BMI:**

µ(diabetes) =

=

=

µ(tidak diabetes) =

=

= 26,75

**Sleep:**

µ(diabetes) =

=

= 7

µ(tidak diabetes) =

=

= 7,25

**Sound Sleep:**

µ(diabetes) =

=

= 6

µ(tidak diabetes) =

=

= 5,75

**Pregancies:**

µ(diabetes) =

=

= 0,714

µ(tidak diabetes) =

=

=

* Hitung nilai standar deviasi

Tahap selanjutnya setelah mendapatkan nilai mean yaitu menghitung nilai standar deviasi.

**BMI:**

σ (diabetes) =

=

=

= 7,2997

σ (tidak diabetes) =

=

=

= 4,535

**Sleep:**

σ (diabetes) =

=

=

= 2,309

σ (tidak diabetes) =

=

=

= 1,215

**Sound Sleep:**

σ (diabetes) =

=

=

= 2,236

σ (tidak diabetes) =

=

=

= 1,603

**Pregancies:**

σ (diabetes) =

=

=

= 1,254

σ (tidak diabetes) =

=

=

= 0,651

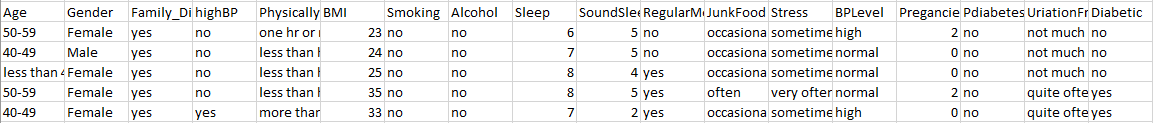
Hasil perhitungan dalam mencari nilai mean dan standar deviasi dapat dilihat pada tabel 3.7 probabilitas berikut:

Tabel 3. 7 Probabilitas Variabel Numerik

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Variabel | µ Mean | | σ Standar deviasi2 | | σ Standar deviasi | |
| D | TD | D | TD | D | TD |
| 1 | BMI | 30,571 | 26,75 | 53,286 | 20,568 | 7,2997 | 4,535 |
| 2 | Sleep | 7 | 7,25 | 5,333 | 1,447 | 2,309 | 1,215 |
| 3 | Sound Sleep | 6 | 5,75 | 5 | 2,568 | 2,236 | 1,603 |
| 4 | Pregancies | 0,714 | 0,333 | 1,571 | 0,424 | 1,254 | 0,651 |

* Hitung Nilai Gaussian

Tahap selanjutnya yaitu menghitung nilai gauss dengan data testing yang sudah ditetapkan. Data testing dapat dilihat pada gambar 3.7.



Gambar 3. 7 Data Testing

**BMI:**

P(BMI = 23 | Diabetes) =

P(BMI = 23 | Diabetes) =

P(BMI = 23 | Diabetes) =

P(BMI = 23 | Diabetes) =

P(BMI = 23 | Diabetes) = 0,086

P(BMI = 23 | Tidak Diabetes) =

P(BMI = 23 | Tidak Diabetes) =

P(BMI = 23 | Tidak Diabetes) =

P(BMI = 23 | Tidak Diabetes) =

P(BMI = 23 | Tidak Diabetes) = 0,133

***Sleep*:**

P(*Sleep* = 6 | Diabetes) =

P(*Sleep* = 6 | Diabetes) =

P(*Sleep* = 6 | Diabetes) =

P(*Sleep* = 6 | Diabetes) =

P(*Sleep* = 6 | Diabetes) = 0,239

P(*Sleep* = 6 | Tidak Diabetes) =

P(*Sleep* = 6 | Tidak Diabetes) =

P(*Sleep* = 6 | Tidak Diabetes) =

P(*Sleep* = 6 | Tidak Diabetes) =

P(*Sleep* = 6 | Tidak Diabetes) = 0,211

***Sound Sleep*:**

P(*SoundSleep* = 5 | Diabetes) =

P(*SoundSleep* = 5 | Diabetes) =

P(*SoundSleep* = 5 | Diabetes) =

P(*SoundSleep* = 5 | Diabetes) =

P(*SoundSleep* = 5 | Diabetes) = 0,241

P(*SoundSleep* = 5 | Tidak Diabetes) =

P(*SoundSleep* = 5 | Tidak Diabetes) =

P(*SoundSleep* = 5 | Tidak Diabetes) =

P(*SoundSleep* = 5 | Tidak Diabetes) =

P(*SoundSleep* = 5 | Tidak Diabetes) = 0,282

***Pregancies*:**

P(*Pregancies* = 2 | Diabetes) =

P(*Pregancies* = 2 | Diabetes) =

P(*Pregancies* = 2 | Diabetes) =

P(*Pregancies* = 2 | Diabetes) =

P(Pregancies = 2 | Diabetes) = 0,211

P(*Pregancies* = 2 | Tidak Diabetes) =

P(*Pregancies* = 2 | Tidak Diabetes) =

P(*Pregancies* = 2 | Tidak Diabetes) =

P(*Pregancies* = 2 | Tidak Diabetes) =

P(*Pregancies* = 2 | Tidak Diabetes) = 0,019

1. Menghitung Nilai Probabilitas Akhir Setiap Kelas

Nilai probabilitas yang dihitung diperoleh dengan mengalikan jumlah data yang sesuai untuk kategori yang sama dengan jumlah data untuk kategori tersebut.

P(X | Diabetes) = P( Age | Diabetes ) \* P( Gender | Diabetes ) \* P( FamilyDiabetes | Diabetes ) \* P( HighBP | Diabetes ) \* P( Physically | Diabetes ) \* P( BMI | Diabetes ) \* P( Smoking | Diabetes ) \* P( Alcohol | Diabetes ) \* P( Sleep | Diabetes ) \* P( SoundSleep | Diabetes ) \* P( RegularMadicine | Diabetes ) \* P( JunkFood | Diabetes ) \* P( Stress | Diabetes ) \* P( BPLevel | Diabetes ) \* P( Pregancies | Diabetes ) \* P( Pdiabetes | Diabetes ) \* P( UrinationFrequency | Diabetes )

P(X | Diabetes) = 0,29 \* 0,29 \* 0,71 \* 0,43 \* 0,29 \* 0,086 \* 0,86 \* 0,43 \* 0,239 \* 0,241 \* 0,29 \* 0,71 \* 0,57 \* 0,71 \* 0,211 \* 0,86 \* 0,71

P(X | Diabetes) = 0,000000146429947

P(X | Tidak Diabetes) = P( Age | Tidak Diabetes ) \* P( Gender | Tidak Diabetes ) \* P( FamilyDiabetes | Tidak Diabetes ) \* P( HighBP | Tidak Diabetes ) \* P( Physically | Tidak Diabetes ) \* P( BMI | Tidak Diabetes ) \* P( Smoking | Tidak Diabetes ) \* P( Alcohol | Tidak Diabetes ) \* P( Sleep | Tidak Diabetes ) \* P( SoundSleep | Tidak Diabetes ) \* P( RegularMadicine | Tidak Diabetes ) \* P( JunkFood | Tidak Diabetes ) \* P( Stress | Tidak Diabetes ) \* P( BPLevel | Tidak Diabetes ) \* P( Pregancies | Tidak Diabetes ) \* P( Pdiabetes | Tidak Diabetes ) \* P( UrinationFrequency | Tidak Diabetes )

P(X | Tidak Diabetes) = 0,25 \* 0,25 \* 0,17 \* 0,83 \* 0,5 \* 0,133 \* 0,75 \* 0,83 \* 0,21 \* 0,282 \* 0,92 \* 0,75 \* 0,75 \* 0,25 \* 0,019 \* 0,92 \* 0,75

P(X | Tidak Diabetes) = 0,000000036668191

Untuk contoh perhitungan menggunakan 1 sampel data testing, untuk sampel selanjutnya menggunakan langkah penyelesaian yang sama. Hasil probabilitas akhir setiap kelas dapat dilihat pada tabel 3.8 berikut.

Tabel 3. 8 Hasil Probabilitas Akhir Setiap Kelas

|  |  |  |
| --- | --- | --- |
| No | Diabetes | Tidak Diabetes |
| 1 | 0,0000001464 | 0,0000001464299 |
| 2 | 0,0000040746 | 0,0000066574605 |
| 3 | 0,0000007853 | 0,0000001878177 |
| 4 | 0,0000010747 | 0,0000000000132 |
| 5 | 0,0000017309 | 0,0000000001131 |

1. Menghitung Probabilitas Akhir

Setelah mendapatkan nilai probabilitas akhir setiap kelas, tahap berikutnya yaitu probabilitas akhir. Probabilitas akhir didapatkan dengan perhitungan menggunakan persamaan berikut.

P(Diabetes | X) = P(Diabetes) \* P(X)

= 0,35 \* 0,0000001464

= 0,00000005124

P(Tidak Diabetes | X) = P(Tidak Diabetes) \* P(X)

= 0,63 \* 0,0000001464299

= 0,00000009225

Perhitungan diatas menggunakan 1 sampel dari data testing, untuk sampel berikutnya juga dilakukan perhitungan probabilitas akhir hingga mendapatkan hasil dari semua sampel. Probabilitas akhir untuk semua sampel dapat dilihat pada tabel 3.9 berikut.

Tabel 3. 9 Tabel Probabilitas Akhir

|  |  |  |
| --- | --- | --- |
| No | Diabetes | Tidak Diabetes |
| 1 | 0,00000005124 | 0,00000009225 |
| 2 | 0,00000142611 | 0,0000041942 |
| 3 | 0,00000027485 | 0,00000011833 |
| 4 | 0,00000037615 | 0,000000000008316 |
| 5 | 0,00000060582 | 0,00000000007125 |

1. Menghitung Nilai *Confusion Matrix*

Tahap selanjutnya yaitu menghitung kinerja algoritma *naïve bayes* menggunakan *confusion matrix* sebagaimana tabel 3.10 berikut:

Tabel 3. 10 Confusion Matrix

|  |  |  |
| --- | --- | --- |
| Data = 5 | Aktual Positif | Aktual Negatif |
| Diagnosis Positif | TP = 2 | FP = 1 |
| Diagnosis Negatif | FN = 0 | TN = 2 |
|  | 2 | 3 |

Maka dari hasil *confusion matrix* pada tabel 3.10 dapat dihitung nilai *accuracy*:

*Accuracy* =

=

= \*100%

= 0,8 \* 100%

= 80%

1. Hitungan *Adaboost*

Pada tahap ini yaitu melakukan perhitungan dengan algoritma *adaboost* untuk meningkatkan performa kinerja dari metode *naïve bayes.* Perhitungan dapat dilihat seperti berikut.

* Langkah pertama tentukan nilai bobot awal dari total data training dengan persamaan berikut;
* Langkah kedua untuk mendapatkan nilai kesalahan dari data sampel dengan menggunakan persamaan berikut:
* Langkah ketiga yaitu menetapkan nilai bobot dengan menggunakan persamaan berikut:
* Langkah keempat update bobot sampel dengan menggunakan persamaan berikut:

Klasifikasi salah

Klasifikasi benar

Iterasi berhenti dilakukan pada perhitungan awal, dikarenakan data sampel yang digunakan tidak mendukung atau kurang banyak.

## 3.6 Pengujian Sistem

Tahap ini merupakan pengujian terhadap sistem yang telah diproses dengan cara membandingkan dari data aktual dengan data yang telah diproses klasifikasi. Setelah itu dilakukan pengujian akurasi menggunakan *confusion matrix* sesuai hasil dari klasifikasi.

# DAFTAR PUSTAKA

Adelita, M., Arto, K. S., & Deliana, M. (2020). Kontrol Metabolik pada Diabetes Melitus Tipe-1. *Fakultas Kedokteran Universitas Sumatera Utara/RS Pendidikan Universitas Sumatera Utara, Medan, Indonesia*, *47*(3), 227–232. http://www.cdkjournal.com/index.php/CDK/article/viewFile/377/175

Ahsana, R., Saedudin, R. R., & Widharta, V. P. (2021). *Perbandingan Akurasi Algoritma Adaboost Dan Algoritma Lightgbm Untuk Klasifikasi Penyakit Diabetes*. *8*(5), 9757–9764.

Algoritma, C., Carolina, A., Ade, K., & Kunci, K. (2020). Penerapan Data Mining Dengan Menggunakan Algoritma C4.5 Pada Klasifikasi Fasilitas Kesehatan Provinsi Di Indonesia. *Jurnal Ilmiah Komputasi*, *19*(1), 27–38. https://doi.org/10.32409/jikstik.19.1.153

Anggeria, E., & Siregar, P. S. (2019). Efektivitas perawatan ulkus diabetikum terhadap penerimaan diri pasien diabetes melitus tipe II. *Jurnal Jumantik*, *4*(2), 178–187.

Bramer, M. (2020). *Measuring the Performance of a Classifier*. 175–187. https://doi.org/10.1007/978-1-4471-7493-6\_12

Byna, A., & Basit, M. (2020). *Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes*. *09*(November), 407–411.

C.Aggarwal, C. (2015). Data mining The Textbook. In *Cancer Letters* (Vol. 14, Issue 3). https://doi.org/10.1016/0304-3835(81)90152-X

Febrinasari, R. P., Sholikah Agusti, T., Pakha Nasirochim, D., & Putra Erdana, S. (2020). Buku Saku Diabetes Melitus. *UNS Press*, *November*, 70.

Fitriyani, F. (2021). Prediksi Diabetes Menggunakan Algoritma Naive Bayes dan Greedy Forward Selection. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *7*(2), 61–69. https://doi.org/10.25077/teknosi.v7i2.2021.61-69

Herho, S. H. S. (2017). Tutorial Pemrograman Python 2 Untuk Pemula. *WCPL Press*, 1–140.

Imelda, S. I. (2019). Faktor-Faktor Yang Mempengaruhi Terjadinya diabetes Melitus di Puskesmas Harapan Raya Tahun 2018. *Scientia Journal*, *8*(1), 28–39. https://doi.org/10.35141/scj.v8i1.406

Jalota, C., & Agrawal, R. (2019). Analysis of Educational Data Mining using Classification. *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019*, 243–247. https://doi.org/10.1109/COMITCon.2019.8862214

Khairani. (2019). Hari Diabetes Sedunia Tahun 2018. *Pusat Data Dan Informasi Kementrian Kesehatan RI*, 1–8.

Kurniawan, L. B. (2016). Patofisiologi, Skrining, dan Diagnosis Laboratorium Diabetes Melitfile:///C:/Users/user/Downloads/Pelaksanaan Pengkajian Keperawatan Pada Pasien Diabetes Melitus di Rumah Sakit (1).pdfus Gestasional. *Cdk*, *43*(11), 811–813.

Lenhard, M. J., Maser, R. E., Patten, B. C., & DeCherney, G. S. (1998). The new diagnosis and classification of diabetes mellitus. *Delaware Medical Journal*, *70*(8), 355–359.

Nasruddin, A. (2021). *Prediksi Awal Penyakit Diabetes Mellitus Menggunakan Algoritma Naive Bayes*. *20*(2), 246–253.

Nuraisyah, F. (2018). Faktor Risiko Diabetes Mellitus Tipe 2. *Jurnal Kebidanan Dan Keperawatan Aisyiyah*, *13*(2), 120–127. https://doi.org/10.31101/jkk.395

Pareza Alam Jusia. (2018). Analisis komparasi pemodelan algoritma decision tree menggunakan metode particle swarm optimization dan metode adaboost untuk prediksi awal penyakit jantung. *Seminar Nasional Sistem Informasi 2018*, 1048–1056.

Ramdhani, L. S. (2016). Penerapan Particle Swarm Optimization ( Pso ) Untuk Seleksi Atribut Dalam Meningkatkan Akurasi Prediksi Diagnosis Penyakit Hepatitis Dengan Metode Algoritma C4 . 5. *Swabumi*, *IV*(1), 1–15.

Ririh, K. R., Laili, N., Wicaksono, A., & Tsurayya, S. (2020). Studi Komparasi dan Analisis Swot Pada Implementasi Kecerdasan Buatan (Artificial Intelligence) di Indonesia. *Jurnal Teknik Industri*, *15*(2), 122–133. https://ejournal.undip.ac.id/index.php/jgti/article/view/29183

Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. *None*, *13*(1), 76–84.

Setiawati, I., Wibowo, A. P., & Hermawan, A. (2019). Pendahuluan Tinjauan Pustaka Penelitian Sebelumnya Klasifikasi. *Jurnal of Information System Management*, *1*(1), 13–17.

Tigga, N. P., & Garg, S. (2020). Prediction of Type 2 Diabetes using Machine Learning Classification Methods. *Procedia Computer Science*, *167*(2019), 706–716. https://doi.org/10.1016/j.procs.2020.03.336

Yosmar, R., Almasdy, D., & Rahma, F. (2018). Jurnal Sains Farmasi Dan Klinis. *Survei Risiko Penyakit Diabetes Melitus Terhadap Kesehatan Masyarakat Kota Padang*, *5*(Agustus 2018), 134–141.