

Aplikasi Sistem Pakar Dengan Metode Naive Bayes untuk Mendeteksi Penyakit Diabetes

Expert System Application Using the Naive Bayes Method to Detect Diabetes

Ahmad Fauzi

Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

*E-mail: dosen0261@unpam.ac.id

Abstrak

Data IDF menyebutkan pengidap diabetes di Indonesia sebanyak 19,47 juta tahun 2021. Salah satu penyebabnya adalah kurangnya kesadaran konsumsi makanan yang sehat yang berdampak terhadap peningkatan berat badan dari tingginya gula darah. Dalam mengatasi kejadian tersebut perlu adanya penerapan aplikasi dalam mendeteksi penyakit diabetes, penggunaan sistem pakar memberikan banyak kemudahan bagi seseorang dalam pemeriksaan kesehatan dan juga petugas kesehatan ataupun dokter dalam mendiagnosa suatu penyakit. Dalam mendiagnosa, sistem pakar membutuhkan sebuah metode, salah satunya metode klasifikasi Naive Bayes. Pada penelitian ini data pengidap penyakit diabetes diambil dari situs kaggle sebagai data training, dengan total data 768 dari sembilan atribut dan salah satu atribut sebagai label. Untuk mempermudah dalam pengoperasian aplikasi atribut yang dijadikan sampel pengujian hanya empat berdasarkan penghitungan nilai korelasi tertinggi, yaitu pregnancies, Glucose, BMI, dan age. Selanjutnya opini pakar kesehatan, yaitu dokter spesialis penyakit dalam dan penghitungan nilai akurasi data diabetes menggunakan algoritma naïve bayes. Hasil nilai akurasi data 79%. Implementasi aplikasi sistem pakar berbasis web dengan bahasa pemrograman php dan database Mysql. Aplikasi sistem pakar ini bertujuan untuk deteksi penyakit diabetes berdasarkan dari hasil pengecekan kesehatan untuk diprediksi apakah positif diabetes atau negatif diabetes.

Kata kunci: Diabetes, Klasifikasi, Akurasi, Naive Bayes, Sistem Pakar

Abstract

Data IDF states that there will be 19.47 million people with diabetes in Indonesia in 2021. One of the causes is a lack of awareness of healthy food consumption which has an impact on increasing body weight due to high blood sugar. In overcoming this incident, it is necessary to implement an application to detect diabetes. The use of an expert system provides many conveniences for someone in health examinations and also for health workers or doctors in diagnosing a disease. In diagnosing, an expert system requires a method, one of which is the Naive Bayes classification method. In this study, data on diabetes sufferers was taken from the Kaggle site as training data, with a total of 768 data from nine attributes and one of the attributes as a label. To make it easier to operate the application, only four attributes were used as test samples based on calculating the highest correlation value, namely pregnancy, glucose, BMI and age. Next, the opinion of a health expert, namely a specialist in internal medicine, and calculating the accuracy value of diabetes data using the Naive Bayes algorithm. The resulting data accuracy value is 79%. Implementation of a web-based expert system application using the PHP programming language and MySQL database. This expert system application aims to detect diabetes based on the results of a health check to predict whether it is positive for diabetes or negative for diabetes.

Keywords: Diabetes, Classification, Accuracy, Naive Bayes, Expert system

Naskah diterima 24 Feb. 2024; direvisi 15 Jul. 2024; dipublikasikan 01 Apr. 2025.

JAMIKA is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.



I. PENDAHULUAN

Penyebab kematian terbanyak dialami oleh manusia salah satunya adalah penyakit diabetes termasuk Indonesia. Berdasarkan data *Internatonal Diabetes Federatiaon* (IDF), wilayah Asia Pasifik menjadi wilayah yang memiliki tingkat penderita diabetes terbanyak. 9,116.03 penderita yang ada di indonesia, berdasarkan perbandingan diabetes dunia 138 juta kasus (8.5%)[1]. Diperkirakan tahun 2035 diperkirakan usia penderita diabetes millitus 40-59 tahun menjadi 205 juta kasus. Sementara pada wilayah asia Tenggara jumlah terbanyak penderita diabetes adalah indonesia. Peningkatan jumlah diabetes karena keterlambatan diagnosa dan konsumsi makanan yang kurang sehat. Data tersebut diperkirakan akan terus bertambah atau meningkat, *word Health Organization* atau WHO memperkirakan bahwa pengidap penyakit diabetes di Indonesia pada tahun 2030

menembus angka pengidap sebanyak 21,3 juta jiwa [2]. Dengan tingginya perkiraan pengidap penyakit diabetes tersebut, sudah harus atau wajib menjadi perhatian yang sangat penting

Salah satu penyebab terjadinya penyakit diabetes adalah dengan konsumsi makanan yang kurang sehat sehingga berdampak terhadap terjadinya peningkatan berat badan, ditambah lagi kurangnya kesadaran oleh masyarakat dan belum maksimalnya alat atau aplikasi pendeteksi penyakit diabetes [3]. Hal tersebut yang membuat banyak masyarakat yang seolah mengabaikan efek atau bahaya dari penyakit diabetes tersebut. Sosialisasi yang dilakukan saat ini dari berbagai macam media belum dapat menurunkan tingkat kesadaran yang masyarakat akan bahaya penyakit diabetes [4].

Dengan semakin berkembangnya teknologi dengan digitalisasi diseluruh bidang kehidupan salah satunya yang sedang meningkat penggunaannya adalah dalam bidang kesehatan. Hal yang menjadi perhatian adalah dengan adanya teknologi sistem informasi yang dapat mempermudah pendeteksian sebuah penyakit, ini yang menjadikan sebuah proses perubahan dalam bidang kesehatan [5]. Penggunaan sistem pakar memudahkan seseorang dalam melakukan pengecekan Kesehatan dan juga membantu deteksi penyakit oleh petugas kesehatan [6]. Sistem Pakar bagian dari ilmu kecerdasan buatan dimana dapat melakukan atau memberikan sebuah pengetahuan berdasarkan seorang pakar [7].

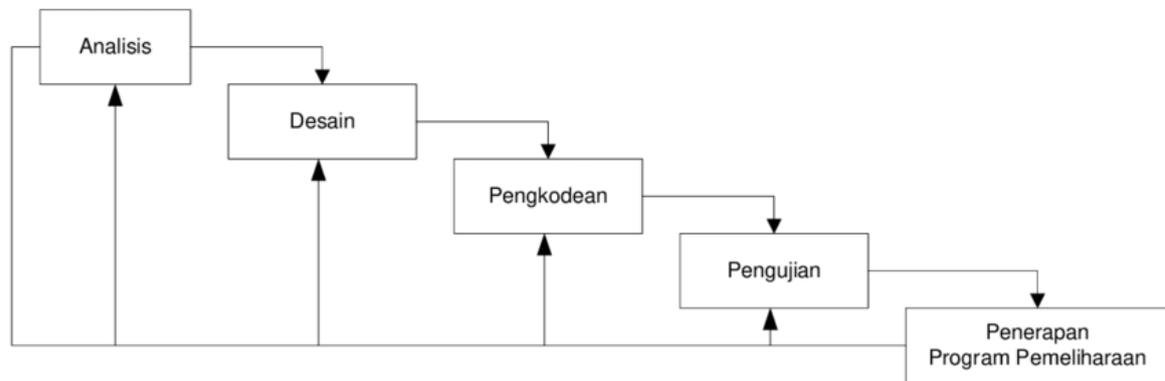
Pada penelitian yang sudah dilakukan sebelumnya, yaitu Sistem Pakar dalam melakukan Deteksi Penyakit Diabetes melalui Pendekatan Naive Bayesian Berbasis Web, dalam penelitian tersebut seorang pasien dalam melakukan pengecekan penyakit diabetes terlebih dahulu melakukan pemeriksaan melalui laboratorium, selanjutnya dilakukan penginputan data pada aplikasi untuk diprediksi dalam diagnosa penyakit diabetes. Hasil pada penelitian ini memberikan manfaat bagi puskesmas dalam mendeteksi penyakit diabetes sejak dini [3].

Selanjutnya dalam penelitian mengklasifikasikan kepuasan oleh para mahasiswa mengenai pembelajaran online dengan algoritma naive bayes, menghasilkan akurasi nilai 76.92%, dengan 100% class precision, dan 57.14% class recall, serta memiliki AUC 0.881 atau 88.1%, hasil yang diperoleh dari penghitungan menggunakan metode naive bayes tergolong cukup baik [8]. Selanjutnya penelitian tentang pembuatan sistem pakar dalam mendiagnosis diabetes menggunakan naive bayes, penelitian menghasilkan sebuah prediksi penyakit yang cukup baik, karena dari pengujian sebanyak 30 pengulangan hasil yang diperoleh sama dengan diagnosa oleh seorang pakar [1].

Berdasar dari penelitian di atas maka pemanfaatan teknologi dapat dioptimalkan dalam pendeteksian penyakit diabetes menggunakan sebuah aplikasi berbasis web, namun dalam penelitian diatas proses pengolahan data belum menerapkan penghitungan analisa korelasi variabel atau atribut terhadap label dalam dataset diabetes. Analisa korelasi penting diterapkan untuk melihat variabel yang memiliki pengaruh besar terhadap sebuah label, selain itu dengan analisa korelasi maka variabel atau atribut yang digunakan menjadi lebih sedikit sehingga memberikan efisiensi dalam variabel input pada aplikasi. Dari hal tersebut maka penelitian ini menggunakan analisa korelasi sebagai variabel input. Sehingga diharapkan dapat memberikan kemudahan kepada masyarakat dalam mengetahui secara langsung apakah positif diabetes atau negatif diabetes.

II. METODE PENELITIAN

Metode penelitian yang digunakan, yaitu *waterfall*, Metode Waterfall, bagaikan air terjun yang mengalir bertahap, menawarkan pendekatan terstruktur untuk pengembangan perangkat lunak dan alat penelitian. Alur yang jelas dan linier ini dimulai dengan identifikasi kebutuhan pengguna dan pemangku kepentingan, kemudian berlanjut ke desain sistem yang kokoh, implementasi kode yang terstruktur, pengujian menyeluruh untuk memastikan fungsionalitas, dan diakhiri dengan pemeliharaan berkelanjutan untuk menjamin kelancaran operasi. Metode ini unggul dalam kesederhanaannya, memungkinkan untuk memahami dan menjalankan proyek dengan mudah. Dokumentasi yang dihasilkan pada setiap tahap, bagaikan peta jalan yang terperinci, memudahkan pelacakan kemajuan dan pemeliharaan sistem di masa depan. Metode ini dipilih karena memiliki lima Langkah yang setiap langkahnya terdapat proses-proses tersendiri [9]. Selain itu metode *waterfall* juga banyak digunakan dalam melakukan pembuatan aplikasi yang memiliki tahapan-tahapan khusus atau terpisah salah satunya pembuatan aplikasi sistem pakar [10]. Berikut Langkah-langkah dalam metode:



Gambar 1. Langkah Metode *waterfall* (Air terjun)

Adapun Langkah-langkah pada metode *waterfall* dalam melakukan penelitian dapat diurutkan dengan penjelasan sebagai berikut:

Analisa Kebutuhan

Analisis kebutuhan merupakan tahap awal yang krusial dalam pengembangan sistem pakar diagnosa diabetes. Pada tahap ini, dilakukan pengumpulan data secara menyeluruh untuk memahami kebutuhan pengguna dan spesifikasi sistem yang diperlukan. Berbagai metode dan literatur dikaji untuk memperkaya pemahaman, termasuk metode sistem pakar, metode naïve bayes, analisis korelasi, dan wawancara dengan dokter spesialis penyakit dalam. Metode sistem pakar memberikan kerangka kerja untuk membangun sistem yang dapat meniru pemikiran pakar dalam mendiagnosis penyakit diabetes. Metode naïve bayes, algoritma klasifikasi probabilistik, handal dalam memprediksi kemungkinan diabetes berdasarkan gejala pasien. Analisis korelasi membantu mengidentifikasi hubungan antara variabel-variabel yang terkait dengan diabetes, seperti usia, riwayat keluarga, dan gaya hidup. Wawancara dengan dokter spesialis penyakit dalam memberikan wawasan berharga dari para ahli medis untuk memastikan variabel kunci dalam mendiagnosis diabetes dan memvalidasi sistem yang dikembangkan. Analisis kebutuhan yang komprehensif ini menjadi fondasi penting dalam pengembangan sistem pakar diagnosa diabetes yang akurat, andal, dan bermanfaat. Sistem yang akurat membantu dokter dalam pengambilan keputusan, sistem yang andal memberikan pelayanan kesehatan yang optimal, dan sistem yang bermanfaat memberikan informasi dan panduan yang tepat bagi pengguna. Analisis kebutuhan merupakan langkah awal yang esensial untuk membangun sistem pakar diagnosa diabetes yang handal dan bermanfaat, serta meningkatkan kualitas layanan kesehatan diabetes.. Berikut penjelasan dari metode tersebut:

Sistem Pakar

Sistem pakar merupakan sistem berbasis komputer yang dirancang untuk meniru keahlian manusia dalam menyelesaikan masalah dan mengambil keputusan di bidang tertentu. Sistem ini menggabungkan pengetahuan dan pengalaman pakar dengan teknik kecerdasan buatan (AI) untuk memberikan solusi yang akurat dan efisien. Pengembangan sistem pakar didasari oleh dua komponen utama: basis pengetahuan dan mesin inferensi. Basis pengetahuan berisi fakta, aturan, dan hubungan yang terkait dengan domain permasalahan. Mesin inferensi berperan sebagai otak sistem pakar, yang memproses informasi dari basis pengetahuan dan data pengguna untuk menghasilkan kesimpulan dan rekomendasi. Sistem pakar menawarkan beberapa keunggulan dibandingkan dengan metode tradisional dalam menyelesaikan masalah. Sistem ini dapat memberikan solusi yang konsisten dan terstandarisasi, terlepas dari keahlian atau pengalaman pengguna. Sistem pakar juga dapat bekerja dengan informasi yang tidak lengkap atau tidak pasti, dan mampu menjelaskan proses penalarannya, sehingga meningkatkan transparansi dan kepercayaan pengguna.. Sistem pakar menjadi salah satu ilmu dimana pada prosesnya menggunakan kepakaran seorang manusia dan menjadi bagian dari kecerdasan buatan [11]. Sistem pakar dapat diartikan sebagai implementasi dari fakta-fakta yang diperoleh seorang pakar berdasarkan kejadian-kejadian yang dialami [12]. Dalam proses memecahkan sebuah masalah sistem pakar melakukan penggabungan antara pengetahuan dari pakar dengan database yang ada [13]. Sistem pakar sama halnya dengan manusia yang memiliki kepakaran namun dibuat kedalam sebuah sistem [14]. Tujuan dari hal tersebut sebagai alat bantu sehingga banyak digunakan karena kemudahannya.

Metode Naïve Bayes

Metode Naïve Bayes merupakan salah satu algoritma klasifikasi probabilistik yang populer dalam berbagai aplikasi pembelajaran mesin, termasuk klasifikasi teks, spam filtering, dan deteksi intrusi. Algoritma ini terkenal dengan kesederhanaannya dan kemudahan implementasinya, namun memiliki performa yang mumpuni dalam menangani berbagai permasalahan klasifikasi. Metode Naïve Bayes didasarkan pada teorema Bayes dan asumsi kemerdekaan fitur, yang menyatakan bahwa kemunculan satu fitur tidak mempengaruhi kemunculan fitur lain. Asumsi ini memungkinkan perhitungan probabilitas kelas secara efisien dan akurat, bahkan dengan kumpulan data yang besar dan kompleks.

Metode Naïve Bayes menonjol di antara algoritma klasifikasi lainnya karena kemudahan interpretasinya. Model yang dihasilkan Naïve Bayes memiliki struktur yang sederhana dan mudah dipahami, memungkinkan pengguna untuk menelusuri dasar-dasar klasifikasi dan meningkatkan kepercayaan terhadap hasil. Hal ini menjadikannya pilihan ideal bagi pengguna yang ingin memahami alur penalaran di balik prediksi model. Keunggulan Naïve Bayes tidak berhenti di situ. Algoritma ini telah diterapkan dengan sukses di berbagai bidang, termasuk klasifikasi teks, yaitu mengkategorikan dokumen, email, dan pesan media sosial berdasarkan topik, genre, atau sentimen, klasifikasi email spam, yaitu memisahkan email spam dari email penting, deteksi intrusi jaringan, yaitu mengidentifikasi aktivitas berbahaya dalam jaringan komputer dan, analisis sentimen, yaitu menganalisis opini dan perasaan dalam teks, seperti ulasan produk atau komentar media sosial.

Kemampuan Naïve Bayes untuk menangani berbagai jenis data, termasuk teks, gambar, dan data numerik, semakin memperluas cakupan penerapannya. Algoritma ini efisien dalam komputasi, membuatnya ideal untuk aplikasi real-time dan perangkat dengan sumber daya terbatas.

Penghitungan nilai probabilitas dengan mengacu kepada teorema Bayesian disebut dengan metode naïve bayes [15],[16]. Proses penghitungannya adalah dengan memperhitungkan sebuah nilai dengan melihat dampak nilai attribute yang lainnya [17]. Selain itu metode naïve bayes algoritma klasifikasi statistik untuk memprediksi sebuah class dalam keanggotaannya berdasarkan nilai probabilitas [18]. Secara umum rumus metode naïve bayes adalah pada persamaan (1):

$$P(H|X) = \frac{p(X|Y)p(H)}{p(X)} \quad (1)$$

X adalah sebuah class yang tidak diketahui, sementara H adalah hipotesis dari X dimana X class spesifik. Maka $P(H|X)$ merupakan hipotesis H dari sebuah kondisi X, $P(H)$ probabilitas hipotesis H, $P(X|H)$ probabilitas X dengan kondisi tertentu, kemudian $P(X)$ merupakan probabilitas dari X.

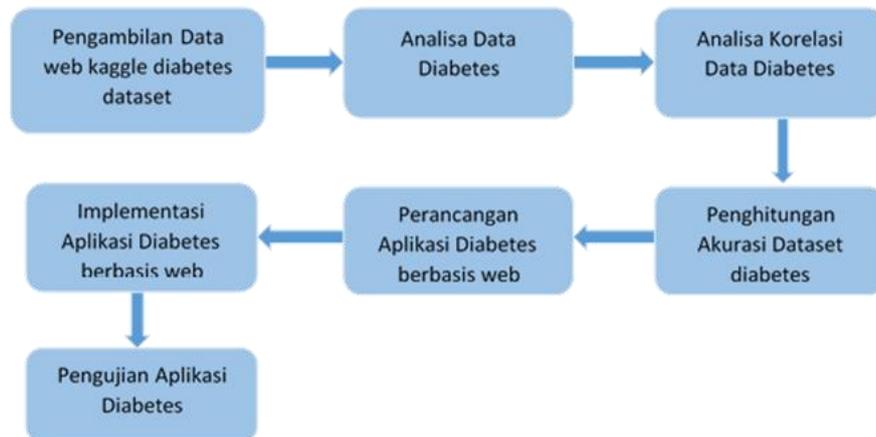
Korelasi

Korelasi bagaikan benang merah yang menghubungkan variabel-variabel dalam penelitian. Teknik statistik ini lazim digunakan untuk mengukur kekuatan dan arah hubungan antara dua atau lebih variabel, Sebagai alat ukur, korelasi membantu dalam memahami bagaimana perubahan pada satu variabel dapat memengaruhi variabel lain. Korelasi merupakan salah satu teknik statistik yang umum digunakan dalam penelitian untuk mengukur kekuatan dan arah hubungan antara dua atau lebih variabel. Hubungan ini dapat bersifat positif, negatif, atau tidak ada sama sekali dalam hal ini untuk mengetahui hubungan antar atribut atau variable dengan derajat hubungan linear, sehingga diketahui nilai hubungan terdekat antar attribute atau variable disebut korelasi [19]. Adapun tujuan analisa korelasi sebagai informasi tentang hubungan antara atribut bebas (X) dan atribut terikat (Y). berikut rumus analisa korelasi pada persamaan (2):

$$\frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (2)$$

Tahapan Penelitian

Dalam penelitian ini dataset diabetes terlebih dahulu dilakukan analisa data untuk memastikan data yang diambil tidak ada missing atau data kosong, selanjutnya dilakukan analisa korelasi untuk melihat hubungan antar attribute, kemudian dilakukan penghitungan akurasi untuk mengetahui tingkat akurasi dari klasifikasi data pengidap diabetes yang dijadikan prediksi. Tahapan akhir adalah membuat perancangan perangkat lunak sistem pakar berbasis web, menggunakan PHP dan MySQL. Berikut gambar alur dari penelitian ini:



Gambar 2. Tahapan penelitian

Pengumpulan data

Pengambilan dataset dengan data sekunder melalui pengambilan data pada situs dataset *Kaggle*. Dataset diabetes dikumpulkan oleh Akshay Dattatray Khare (*National Institute of Diabetes and Digestive and Kidney Diseases*). Dataset ini diperbarui pada tahun 2022, total dataset, yaitu sebanyak 768 data. Data terdiri dari delapan variabel atau atribut dan satu variabel atau atribut label. Berikut ini dataset diabetes.

No	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
1	6	148	72	35	0	33.6	0.627	50	1
2	1	85	66	29	0	26.6	0.351	31	0
3	8	183	64	0	0	23.3	0.672	32	1
4	1	89	66	23	94	28.1	0.167	21	0
5	0	137	40	35	168	43.1	2.288	33	1
6	5	116	74	0	0	25.6	0.201	30	0
7	3	78	50	32	88	31	0.248	26	1
8	10	115	0	0	0	35.3	0.134	29	0
9	2	197	70	45	543	30.5	0.158	53	1
10	8	125	96	0	0	0	0.232	54	1
11	4	110	92	0	0	37.6	0.191	30	0
12	10	168	74	0	0	38	0.537	34	1
13	10	139	80	0	0	27.1	1.441	57	0
14	1	189	60	23	846	30.1	0.398	59	1
-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-
764	10	101	76	48	180	32.9	0.171	63	0
765	2	122	70	27	0	36.8	0.34	27	0
766	5	121	72	23	112	26.2	0.245	30	0
767	1	126	60	0	0	30.1	0.349	47	1
768	1	93	70	31	0	30.4	0.315	23	0

Gambar 3. Dataset diabetes

Keterangan variabel atau atribut dataset diabetes diatas adalah sebagai berikut:

- a. *Pregnancies* : Informasi berapa kali hamil
- b. *Glucose* : Informasi tingkat gula darah
- c. *BloodPressure*: Informasi tekanan darah
- d. *SkinThickness* : Informasi ketebalan kulit
- e. *Insulin* : Informasi kadar Insulin dalam darah
- f. *BMI* : informasi indeks massa tubuh
- g. *DiabetesPedigreeFunction*: Informasi persentase Diabetes
- h. *Age* : Informasi usia
- i. *Outcome* : Informasi penyakit diabetes positif (1) dan negatif (0).

Setelah proses analisa kebutuhan selesai dilanjutkan dengan langkah metode *waterfall* kedua, yaitu proses desain atau perancangan pada aplikasi.berikut merupakan penjelasan dari perancangan pada penelitian ini:

Perancangan

Perancangan merupakan tahap krusial dalam proses penulisan jurnal ilmiah, di mana pada tahap ini dijabarkan secara detail metodologi dan strategi yang akan digunakan untuk mencapai tujuan penelitian. Tahap ini ibarat membangun fondasi kokoh bagi penelitian, memastikan arah yang jelas dan langkah-langkah yang terstruktur untuk mencapai kesimpulan yang valid dan reliabel dan dalam hal ini perancangan merupakan alur dalam pembuatan sebuah aplikasi atau perangkat lunak melalui hasil analisis sistem [2]. Dalam perancangan terdapat beberapa metode diantara *flow chart* yang berguna untuk mengetahui alur dari sebuah aplikasi, kemudian UML (*Unified Modelling Language*) yang bertujuan untuk memvisualisasikan perancangan aplikasi [5].

Implementasi atau pengkodean

Langkah dalam metode *waterfall* ke tiga adalah proses implementasi yang merupakan Langkah atau tahapan pembuatan kode program dari hasil perancangan. Pada sistem pakar deteksi penyakit diabetes bahasa pemrograman yang dipilih adalah PHP yang dilakukan pada server XAMPP. Adapun alat text editor menggunakan visual studio code dengan database Mysql sebagai web server dalam penyimpanan data [20].

Setelah dilakukan implementasi pengkodean pada aplikasi sistem pakar deteksi penyakit diabetes, Langkah atau tahapan selanjutnya adalah dengan melakukan pengujian aplikasi berikut penjelasan dari pengujian aplikasi sistem pakar deteksi penyakit diabetes:

Pengujian Aplikasi

Dalam melakukan pembuatan aplikasi atau perangkat lunak Langkah atau tahapan yang penting yang harus dilakukan dalam menjaga ataupun menghasilkan aplikasi yang berkualitas adalah dengan melakukan pengujian. Pengujian aplikasi dilakukan untuk memastikan fitur atau fungsi dapat berjalan baik dan tidak ditemukannya kesalahan baik dari alur aplikasi, logika, dan lain sebagainya [20]. Adapun dalam melakukan pengujian aplikasi terdapat metode atau cara salah satunya dengan *blackbox testing*. *Blackbox testing* merupakan pengujian dengan melakukan pengoperasian langsung terhadap fungsi-fungsi atau fitur yang terdapat pada aplikasi tanpa harus mengetahui kode program pada aplikasi.

III. HASIL DAN PEMBAHASAN

Dalam pemanfaatan teknologi dalam bidang kecedasan buatan prediksi penyakit diabetes dengan sistem pakar menggunakan algoritma naïve bayes terdapat beberapa langkah-langkah dalam prosesnya. Langkah-langkah tersebut merupakan penerapan dari metode *waterfall*, sehingga diharapkan pembuatan aplikasi memberikan kemudahan kepada Masyarakat dalam melakukan pengecekan Kesehatan untuk mencegah penyakit diabetes. berikut hasilnya dari proses penelitian ini:

Hasil Analisa Data dan Korelasi data

Penelitian ini merupakan penelitian dengan data sekunder dengan pengambilan data pada situs dataset Kaggle. Dataset diabetes dikumpulkan oleh Akshay Dattatray Khare (*National Institute of Diabetes and Digestive and Kidney Diseases*). Dataset ini diperbarui 2022, total dataset, yaitu sebanyak 768 data. Data terdiri dari delapan atribut dan satu atribut yang dijadikan prediksi atau label. Analisa korelasi dilakukan untuk melihat sebuah nilai terdekat atau hubungan antar atribut, dimana dalam dataset diabetes delapan variabel atau atribut akan dilihat kedekatan nilai tertinggi dengan atribut label, pada penelitian ini penghitungan nilai korelasi dilakukan dengan menggunakan bahasa pemrograman python.3. Gambar 4 menunjukkan sourcode dalam analisa korelasi menggunakan python.3:

```
correlation = df.corr()
# constructing a heatmap to understand the correlation
plt.figure(figsize=(12,12))
sns.heatmap(correlation, cbar=True, square=True, fmt='.1f', annot=True, annot_kws={'size':8}, cmap='Blues')
```

Gambar 4. Analisa korelasi dataset diabetes

Hasil dari analisa korelasi dapat diketahui kedekatan antar variabel atau atribut terhadap label atau variabel prediksi dalam hal ini pasien yang terkena penyakit diabetes. Berikut gambar hasil analisa korelasi dengan visualisasi nilai korelasi dataset diabetes:



Gambar 5. Hasil Analisa korelasi dataset diabetes

Dari gambar 5 dapat diketahui nilai korelasi setiap variabel atau atribut terhadap label pada dataset diabetes, dimana variabel atau atribut yang memiliki nilai korelasi tertinggi adalah *Pregnancies* dengan nilai korelasi 0.22, *glucose* 0.47, *BMI* 0.29, dan *age* 0.24. Sehingga berdasarkan hasil tersebut maka yang menjadi

variabel input pada aplikasi sistem pakar deteksi penyakit diabetes pada penelitian ini, yaitu *Pregnancies*, *Glucose*, BMI, dan age. Berikut tabel 1 dibawah ini hasil penghitungan korelasi dataset diabetes:

Hasil Penghitungan Nilai Akurasi Data

Penelitian ini menggunakan data publik sehingga diperlukan penghitungan nilai akurasi atau keakuratan data pada dataset diabetes, metode algoritma naïve bayes dapat melakukan penghitungan nilai akurasi data. Untuk penghitungan dataset diabetes tidak dilakukan penghitungan secara manual dikarenakan jumlah data yang banyak sehingga penghitungan dilakukan dengan bahasa pemrograman python.3, penghitungan dilakukan berdasarkan hasil pembagian data, yaitu data training 80% dengan data testing 20%. Gambar 5 adalah proses penghitungan nilai akurasi pada dataset diabetes dengan metode naïve bayes :

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
from sklearn.model_selection import cross_val_score

#Memanggil data diabetes dengan format csv
df = pd.read_csv('/content/drive/MyDrive/Project/predict-diabetes/diabetes.csv')
df.shape

(768, 9)
```

Gambar 6. Pemanggilan Dataset Diabetes

Gambar 6 merupakan tahap awal penghitungan dengan terlebih dahulu import library python.3 yang dibutuhkan, selanjutnya proses pemanggilan dataset diabetes dari penyimpanan google drive.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
...
763	10	101	76	48	180	32.9	0.171	63	0
764	2	122	70	27	0	36.8	0.340	27	0
765	5	121	72	23	112	26.2	0.245	30	0
766	1	126	60	0	0	30.1	0.349	47	1
767	1	93	70	31	0	30.4	0.315	23	0

768 rows x 9 columns

Gambar 7. Tampilan Dataset Diabetes

Gambar 7 merupakan isi dari dataset diabetes yang pada awal dilakukan pemanggilan dalam google drive. Terdapat 768 data dengan 9 kolom atau atribut.

```
#Milhat data missing atau hilang pada dataset  
print("Missing Value in Training Dataset")  
df.isnull().sum()
```

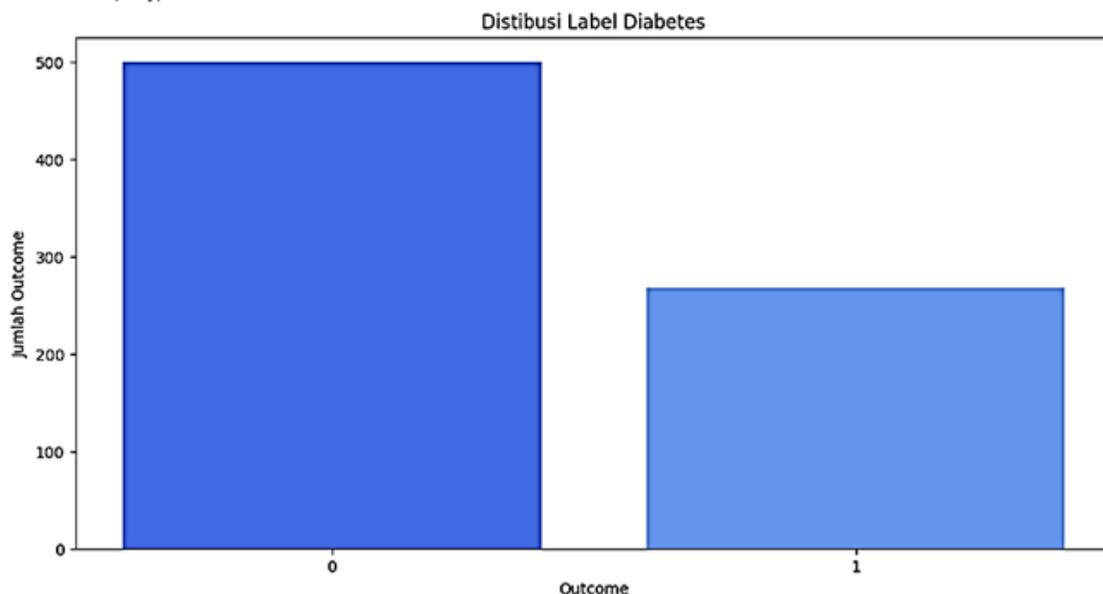
```
Missing Value in Training Dataset  
Pregnancies      0  
Glucose          0  
BloodPressure    0  
SkinThickness    0  
Insulin          0  
BMI              0  
DiabetesPedigreeFunction  0  
Age              0  
Outcome          0  
dtype: int64
```

Gambar 8. Hasil Missing Value Dataset Diabetes

Dari hasil gambar 8, dataset diabetes tidak memiliki data yang missing atau kosong, sehingga proses penghitungan dapat dilanjutkan dengan terlebih dahulu melihat perbandingan label dengan 0 merupakan pasien yang negatif diabetes dan 1 merupakan pasien yang positif diabetes.

```
#Visualisasi Label dengan grafik  
counter = df.Outcome.value_counts()  
index = [1,2]  
plt.figure(1,figsize=(12,6))  
plt.bar(index,counter,color=['royalblue','cornflowerblue'])  
plt.xticks(index,['0','1'],rotation=0)  
plt.xlabel('Outcome')  
plt.ylabel('Jumlah Outcome')  
plt.title('Distibusi Label Diabetes')  
df['Outcome'].value_counts()
```

```
0    500  
1    268  
Name: Outcome, dtype: int64
```



Gambar 9. Hasil Visualisasi Label Outcome Dataset Diabetes

Dari gambar 9 dapat dilihat perbandingan pengidap penyakit diabetes dan tidak, dimana untuk 1 merupakan positif diabetes dengan jumlah 268 dan 0 negatif diabetes dengan jumlah 500. Selanjutnya dilakukan split atau pemisahan dataset untuk menguji data dalam penghitungan nilai akurasi dengan 0.2 (20%) data uji dan 0.8 (80%) data latih, ditunjukkan pada gambar 10.

```
#Membagi dataset dengan data training 80% dan data testing 20%
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
print(len(x_train), len(y_train))
print(len(x_test), len(y_test))

614 614
154 154
```

Gambar 10. Pemisahan Data Training dan Testing

Berdasarkan hasil penghitungan dataset diabetes menggunakan metode naïve bayes memiliki nilai akurasi 0.79 atau 79%, hasil ini cukup baik untuk sebuah akurasi data dalam memprediksi penyakit diabetes. Gambar 11 menunjukkan hasil penghitungan akurasi dataset diabetes.

```
#Menghitung akurasi dataset dengan metode Naive Bayes
gnb = GaussianNB()
y_pred = gnb.fit(x_train, y_train).predict(x_test)

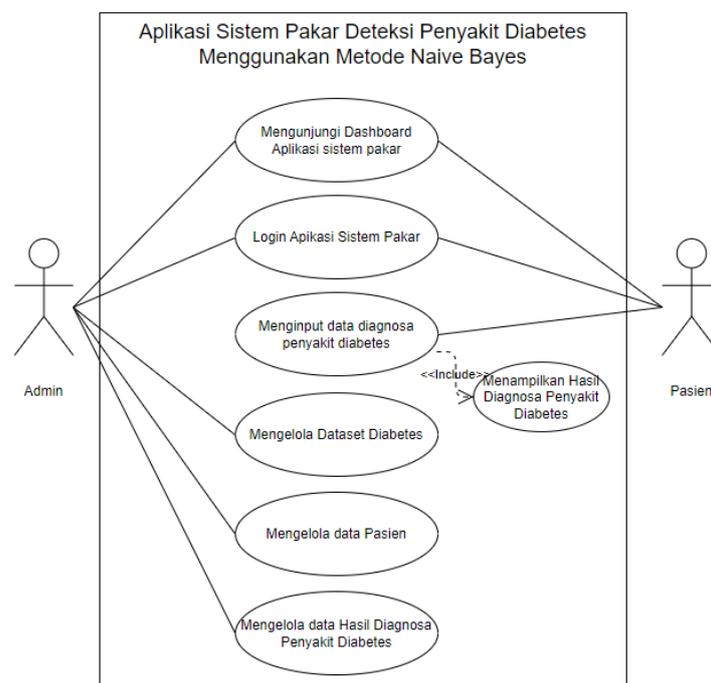
accuracy_GNB = accuracy_score(y_test, y_pred)
print (f"Gaussian Naive Bayes Accuracy: {accuracy_GNB:.2f}")

Gaussian Naive Bayes Accuracy: 0.79
```

Gambar 11. Hasil Penghitungan Akurasi Dataset Diabetes

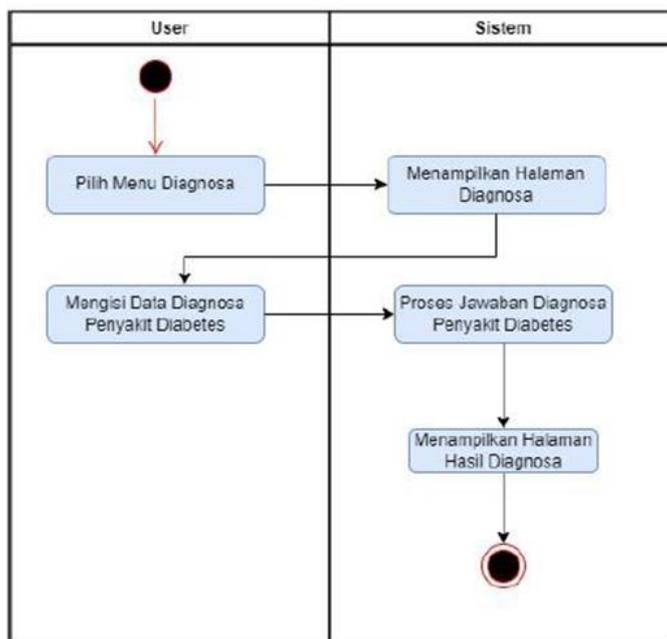
Hasil Perancangan Aplikasi

Dalam perancangan aplikasi deteksi penyakit diabetes langkah awal adalah dengan membuat sebuah use case diagram. Tujuan dari use case diagram adalah untuk mengetahui secara umum proses aplikasi. Gambar 12 merupakan gambar use case pada aplikasi sistem pakar deteksi penyakit diabetes.



Gambar 12. Use Case Diagram Aplikasi

Kemudian melakukan pembuatan activity diagram sebagai visualisasi alur dari Aplikasi Sistem Pakar deteksi Penyakit Diabetes. Gambar 13 menunjukkan activity diagram aplikasi.



Gambar 13. Activity Diagram Aplikasi

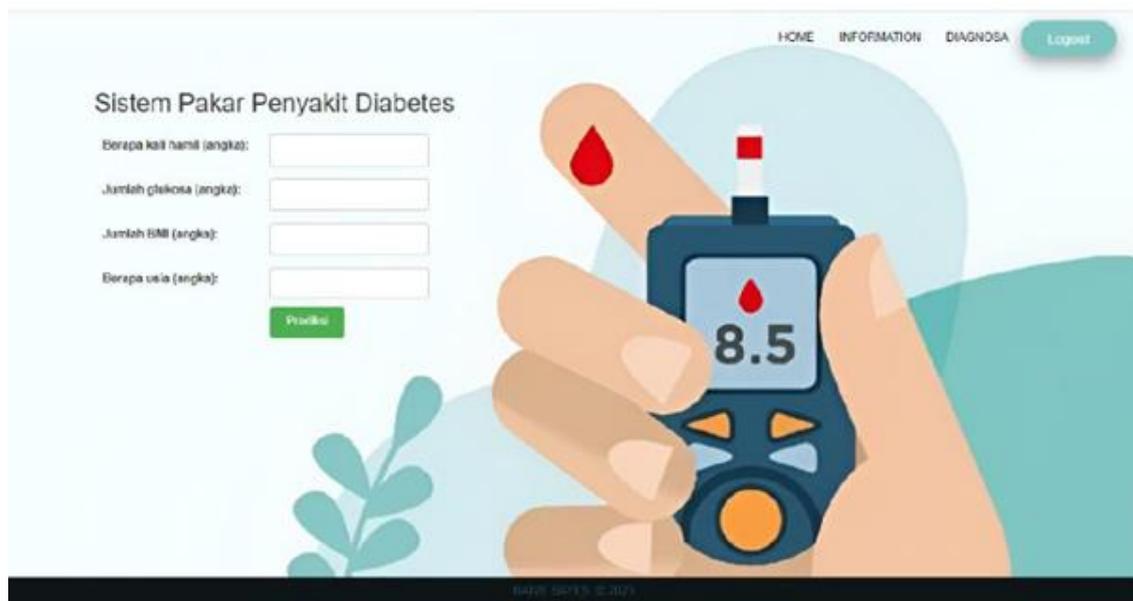
Implementasi Aplikasi

Dalam penerapan hasil dari analisa dilakukan sebuah perancangan, yaitu proses coding dan pembuatan database dengan MS SQL untuk menyimpan data aplikasi sistem pakar deteksi penyakit diabetes berbasis web dalam memprediksi atau mendeteksi penyakit diabetes. Berikut hasil penerapan aplikasi sistem pakar deteksi penyakit diabetes:



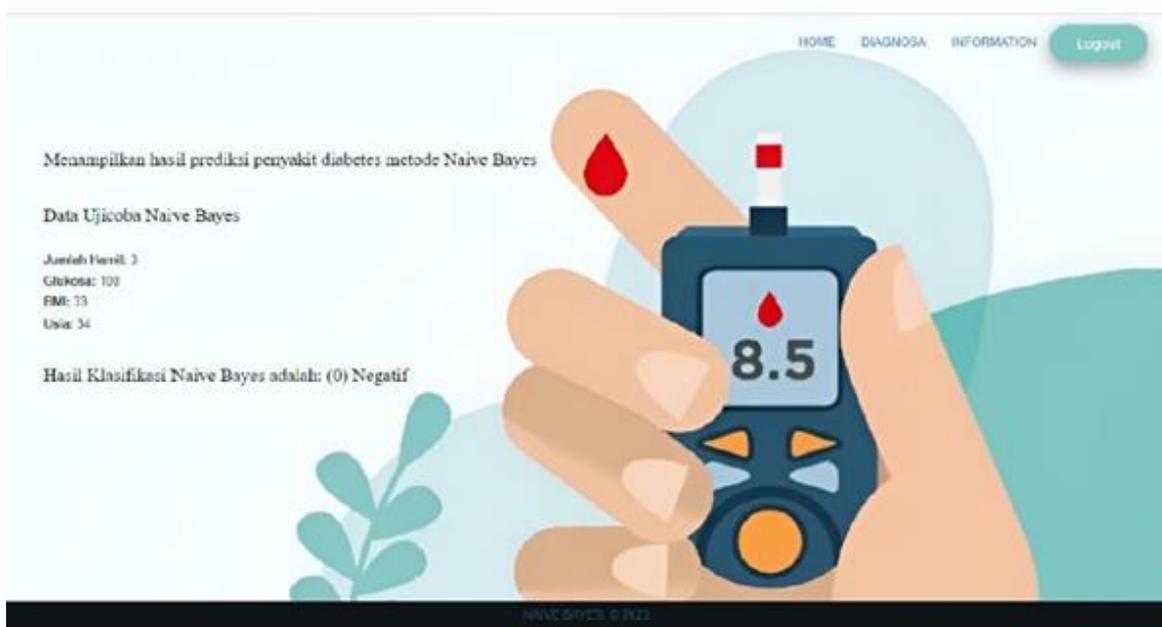
Gambar 14. Halaman Utama Aplikasi

Gambar 14 diatas merupakan tampilan dari user interface halaman utama pada aplikasi dengan bahasa pemogram php yang merupakan hasil dari implementasi berdasarkan hasil perancangan aplikasi.



Gambar 15. Halaman Diagnosa Penyakit

Gambar 15 merupakan gambar dari halaman Diagnosa pada aplikasi sistem pakar deteksi penyakit diabetes, untuk diagnosa penyakit, inputan yang dimasukkan dalam aplikasi, yaitu Berapa kali hamil (Pregnancies), Jumlah Glukosa (Glucose), Jumlah BMI dan Berapa Usia (Age).



Gambar 16. Halaman Hasil Diagnosa Penyakit

Gambar 16 merupakan tampilan hasil diagnosa pada aplikasi sistem pakar deteksi penyakit diabetes, setelah memasukan empat variabel input pada aplikais. Selajutnya aplikasi akan memproses berdasarkan data yang telah dimasukan pada aplikais dilakukan prediksi penyakit diabetes berdasarkan masukan atau inputan seorang user atau pasien dengan (0) Negatif dan (1) Positif.

Pengujian Aplikasi

Untuk memastikan kualitas pada aplikasi sistem pakar diagnose penyakit diabetes, dilakukan pengujian menggunakan salah satu metode pengujian, yaitu blackbox testing. Berikut merupakan tabel 1 dari hasil pengujian blackbox:

TABEL 1
 HASIL BLACK-BOX TESTING PADA APLIKASI

Test	Skenario <i>Test</i>	Hasil yang diharapkan	Hasil pengujian	Keterangan
Login User belum terdaftar	1. Mengunjungi halaman Login 2. Email: "test@gmail.com" 3. Input password: "test" 4. Klik login	Gagal login dan akan tampil pop up email atau password salah	Gagal login dan tampil pop up email atau password salah	Sukses
Login User sudah terdaftar	1. Mengunjungi halaman Login 2. Email: "admin@gmail.com" 3. Input password: "admin" 4. Klik login	Berhasil login dan tampil halaman utama	Berhasil login dan masuk halaman utama	Sukses
Menekan tombol prediksi tanpa mengisi data gejala terlebih dahulu	1. Masuk ke menu diagnosa 2. Tidak mengisi data diagnosa penyakit diabetes. Berapa kali hamil : "..." Jumlah glukosa : "..." Jumlah BMI : "..." Berapa usia : "..." 3. Klik prediksi	Sistem akan menolak atau tidak menampilkan	Sistem menolak atau tidak menampilkan	Sukses
Menekan tombol prediksi dengan mengisi data gejala terlebih dahulu	1. Masuk ke menu diagnosa 2. Mengisi data diagnose penyakit diabetes. Berapa kali hami : " 3 " Jumlah glukosa : " 100 " Jumlah BMI : " 33 " Berapa usia : " 34 " 3. Klik prediksi	Sistem akan menampilkan hasil diagnosa penyakit diabetes.	Sistem menampilkan hasil diagnosa penyakit diabetes.	Sukses

Dari hasil pengujian fungsi atau menu aplikasi sistem pakar dengan melakukan serangkaian pengujian mulai dari login sampai pada pendeteksian penyakit diabetes aplikasi dapat beroperasi sesuai dengan tujuan pembuatan aplikasi. Berdasarkan hasil tersebut aplikasi sitem pakar pendeteksian penyakit diabetes dapat berperan dan digunakan untuk melakukan identifikasi penyakit diabetes pada petugas kesehatan dan masyarakat.

IV. KESIMPULAN

Dengan langkah atau tahapan yang telah dilakukan diatas dapat diambil kesimpulan bahwa dataset yang diambil dari data publik Kaggle dapat digunakan untuk untuk mendeteksi penyakit diabetes dengan perancangan aplikasi menggunakan naïve bayes sebagai algoritma dalam melakukan penghitungan nilai akurasi serta prediksi pada penyakit diabetes. Sehingga dari hasil tersebut diharapkan masyarakat mendapatkan kemudahan dan pengetahuan serta informasi lebih cepat apakah terkena penyakit (positif diabetes) atau tidak (negatif diabetes) dengan melakukan penginputan data hasil pemeriksaan kesehatan kedalam aplikasi, serta akan berdampak pada kesadaran masyarakat terhadap penyakit diabetes dalam mengidentifikasi penyakit diabetes sejak dini. Penelitian ini menunjukkan bahwa dataset publik Kaggle dapat digunakan untuk mendeteksi diabetes menggunakan algoritma Naive Bayes. Aplikasi yang dikembangkan berdasarkan penelitian ini memungkinkan masyarakat untuk dengan mudah dan cepat mengetahui status diabetes mereka dengan memasukkan data hasil pemeriksaan kesehatan. Penggunaan aplikasi ini diharapkan dapat meningkatkan kesadaran masyarakat terhadap penyakit diabetes dan mendorong deteksi dini, yang pada akhirnya dapat meningkatkan kualitas hidup penderita diabetes. Di masa depan, pengembangan aplikasi ini

perlu dilanjutkan dengan menambahkan fitur-fitur baru dan mengintegrasikannya dengan sistem kesehatan. Diperlukan juga penelitian lebih lanjut untuk meningkatkan akurasi dan cakupan aplikasi. Diharapkan penelitian ini dapat memotivasi pihak-pihak terkait untuk mendukung implementasi aplikasi ini dalam upaya deteksi dini diabetes dan meningkatkan kesehatan masyarakat.

UCAPAN TERIMA KASIH

Dengan penuh rasa syukur dan bahagia, kepada Allah SWT, Tuhan Yang Maha Esa, atas segala Rahmat, karunia, dan hidayah-Nya yang tiada terkira, serta limpahan ilmu dan kesehatan yang telah diberikan, sehingga saya dapat menyelesaikan penelitian dengan judul “Aplikasi Sistem Pakar Dengan Metode Naive Bayes untuk Mendeteksi Penyakit Diabetes”. Peneliti juga ingin mengucapkan terima kasih kepada Universitas Pamulang atas dukungan yang diberikan dalam menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] H. Sitorus, V. Yasin, and A. B. Yulianto, “Perancangan sistem pakar diagnosis penyakit diabetes berbasis web menggunakan algoritma naive bayes,” *Jurnal Sains dan Teknologi Widyaloka*, vol. 1, no. 1, pp. 135–144, Jan. 2022.
- [2] Y. B. Widodo, S. A. Anggraeni, and T. Sutabri, “Perancangan Sistem Pakar Diagnosis Penyakit Diabetes Berbasis Web Menggunakan Algoritma Naive Bayes,” *Jurnal Teknologi Informatika dan Komputer*, vol. 7, no. 1, pp. 112–123, Mar. 2021, doi: 10.37012/jtik.v7i1.507.
- [3] R. Antoni, “SISTEM PAKAR DETEKSI PENYAKIT DIABETES MELLITUS DENGAN MENGGUNAKAN PENDEKATAN NAÏVE BAYESIAN BERBASIS WEB (Studi Kasus : PUSKESMAS Kelurahan Grogol 3),” vol. 2, no. 1, 2023, [Online]. Available: <https://journal.mediapublikasi.id/index.php/oktal>
- [4] R. Ariyanto, D. Puspitasari, Y. Alfani, J. T. Informasi, N. Malang, and A. A. Id, “PENERAPAN ENTROPY BASED DISCRETIZATION PADA METODE NAIVE BAYES DALAM KLASIFIKASI PENYAKIT DIABETES MELLITUS,” 2020.
- [5] E. Fathoni Alamsyah and Niki Ratama, “APLIKASIS SISTEM PAKAR DIAGNOSA PENYAKIT BALITA BERBASIS WEB MENGGUNAKAN METODE NAIVE BAYES (STUDI KASUS: PUSKESMAS SETU),” *JORAPI : Journal of Research and Publication Innovation*, vol. 1, no. 2, 2023.
- [6] P. Septiani, I. Pratiwi, M. Ghofar Rohman, and M. Sholihin, “Sistem Pakar Penyakit Telinga Menggunakan Metode Naive Bayes,” 2023.
- [7] F. Karim, G. W. Nurcahyo, and S. Sumijan, “Sistem Pakar dalam Mengidentifikasi Gejala Stroke Menggunakan Metode Naive Bayes,” *Jurnal Sistim Informasi dan Teknologi*, pp. 221–226, Aug. 2021, doi: 10.37034/jsisfotek.v3i4.69.
- [8] A. Natuzzuhriyyah, N. Nafisah, and R. Mayasari, “Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara Daring Menggunakan Algoritma Naive Bayes,” 2021.
- [9] M. Ruza, I. S. Wijaya, and E. Suratno, “Sistem Pendukung Keputusan untuk Menentukan Karyawan Terbaik dengan Metode TOPSIS pada PT. Sumbertama Nusa Pertiwi,” *Jurnal Manajemen Informatika*, vol. 13, no. 2, pp. 121–134, Sep. 2023, doi: 10.34010/jamika.v13i2.9901.
- [10] D. Pramaishella and M. Maryam, “Aplikasi Manajemen Kinerja Pegawai Dengan Metode Sosiometri pada PT. Siba Prima Utama Feed Mill,” *Jurnal Manajemen Informatika*, vol. 13, no. 2, pp. 94–107, Oct. 2023, doi: 10.34010/jamika.v13i2.9557.
- [11] T. Firdaus and F. Yanti, “IMPLEMENTASI METODE NAÏVE BAYES PADA SISTEM PAKAR DIAGNOSA PENYAKIT GIGI BERBASIS WEB,” *Dinamika Informatika*, vol. 14, no. 2, 2022.
- [12] S. Syaputra *et al.*, “Sistem Pakar Diagnosa Dampak Penggunaan Eyelash Extension Menggunakan Metode Naive Bayes Sistem Pakar Diagnosa Dampak Penggunaan Eyelash Extension Menggunakan Metode Naive Bayes,” *MEANS (Media Informasi Analisa dan Sistem)*, 2023, [Online]. Available: http://ejournal.ust.ac.id/index.php/Jurnal_Means/
- [13] F. Ramadhana, Winarsih, and Fauziah, “APLIKASI SISTEM PAKAR UNTUK MENDIAGNOSA PENYAKIT ISPA MENGGUNAKAN METODE NAIVE BAYES BERBASIS WEBSITE,” 2020.
- [14] M. Ridho Handoko, “SISTEM PAKAR DIAGNOSA PENYAKIT SELAMA KEHAMILAN MENGGUNAKAN METODE NAIVE BAYES BERBASIS WEB,” *Jurnal Teknologi dan Sistem Informasi*, vol. 2, no. 1, pp. 50–58, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTTSI>
- [15] N. Kadek, W. Patrianingsih, I. Kadek, and A. Sugianta, “Penerapan Naive Bayes pada Potensi Akademik Siswa SD Negeri 5 Singakerta,” *MEI*, 2023.

-
- [16] Y. Yuliyana and A. S. R. M. Sinaga, "Sistem Pakar Diagnosa Penyakit Gigi Menggunakan Metode Naive Bayes," *Fountain of Informatics Journal*, vol. 4, no. 1, p. 19, May 2019, doi: 10.21111/fij.v4i1.3019.
- [17] F. A. Nugroho, A. F. Solikin, M. D. Anggraini, and K. Kusriani, "Sistem Pakar Diagnosa Virus Corona Dengan Metode Naive Bayes," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, p. 81, Apr. 2021, doi: 10.30646/tikomsin.v9i1.553.
- [18] Supiyandi, A. Hussein, I. Gunawan, and W. L. Rahman Harjo, *Analisis Klasifikasi Broken Home pada Anak Menggunakan Metode Naive Bayes Classifier*, vol. 8. Hogarth Press, 2023.
- [19] I. N. Fitriani and O. S. Purwanti, "Lama Sakit Diabetes Berhubungan dengan Fungsi Kognitif pada Pasien Diabetes Mellitus," *Jurnal Keperawatan Silampari*, vol. 6, no. 2, pp. 1236–1243, Feb. 2023, doi: 10.31539/jks.v6i2.5488.
- [20] K. Nurwijayanti, "KLASIFIKASI DIAGNOSA PENYAKIT DIABETES DENGAN METODE NAIVE BAYES BERBASIS WEB," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 2, no. 3, pp. 115–121, 2023.