



DIABETES PREDICTION SYSTEM “DIAPRES” BERBASIS OPTIMASI PARAMETER PADA METODE *DECISION TREE*

Elly Muningsih¹⁾, Sutrisno²⁾, Fitriyani Rizki³⁾, Khurotul Dwi Asiffa⁴⁾

^{1,3,4} *Sistem Informasi, Universitas Bina Sarana Informatika*

² *Teknologi Komputer, Universitas Bina Sarana Informatika*

^{1,2,3,4} *Jl. Kramat Raya No.98, Kwitang, Kec. Senen, Kota Jakarta Pusat*

Email: ¹elly.emh@bsi.ac.id, ²sutrisno.stz@bsi.ac.id, ³rizkifitriyani02@gmail.com, ⁴khrtldwi@gmail.com

Abstrak

Diabetes merupakan penyakit menahun (kronis) dimana komplikasi yang sering terjadi pada penderita dan mematikan adalah serangan jantung dan stroke. Menurut Riskesdaa 2018, baru sekitar 25% penderita diabetes mengetahui bahwa dirinya menderita diabetes. Untuk menekan jumlah penderita penyakit diabetes yang semakin bertambah, salah satu yang bisa dilakukan adalah deteksi dini dengan mengembangkan aplikasi prediksi penyakit diabetes memanfaatkan metode klasifikasi Data Mining. Penelitian ini akan mengoptimasi metode klasifikasi *Decision Tree* dengan Optimasi Parameter untuk menghasilkan akurasi tertinggi. Parameter yang dimaksud adalah *Number of folds* pada *Cross Validation* dan pada metode *Decision Tree* adalah parameter *Criterion*, *maximal depth*, *apply pruning* dan *apply pre pruning*. Data yang digunakan diambil dari data publik Kaggle dengan jumlah 2000 record dan 9 atribut (1 atribut khusus). Evaluasi dilakukan dengan membandingkan hasil nilai akurasi sebelum dan sesudah optimasi. Dari eksperimen yang dilakukan diketahui sebelum optimasi nilai akurasi berkisar 76% - 80%, dan setelah optimasi nilai akurasi mencapai 98,00%. Kemudian hasil nilai akurasi tertinggi yang ditampilkan dalam pohon keputusan metode *Decision Tree* yang dikembangkan menjadi sebuah aplikasi *Diabetes Prediction System ‘DIAPRES’* berbasis web. Metode pengembangan perangkat lunak menggunakan metode *Waterfall* dengan PHP-MySQL sebagai bahasa pemrograman yang digunakan dan Pengujian Aplikasi *Black Box Testing*.

Kata Kunci: *diabetes, decision tree, optimasi parameter, diabetes prediction system*

Abstract

Diabetes is a chronic disease where complications that often occur in patients and deadly are heart attacks and strokes. According to Riskesdaa 2018, only about 25% of diabetics know that they have diabetes. To suppress the increasing number of diabetics, one thing that can be done is early detection by developing a diabetes prediction application utilizing the Data Mining classification method. This research will optimize the Decision Tree classification method with Parameter Optimization to produce the highest accuracy. The parameters in question are Number of folds in Cross Validation and in the Decision Tree method are the parameters of Criterion, maximal depth, apply pruning and apply pre pruning. The data used is taken from Kaggle's public data with a total of 2000 records and 9 attributes (1 special attribute). Evaluation is carried out by comparing the results of accuracy values before and after optimization. From the experiments conducted, it is known that before optimization the accuracy value ranged from 76% - 80%, and after optimization the accuracy value reached 98.00%. Then the results of the highest accuracy value displayed in the decision tree Decision Tree method developed into a web-based Diabetes Prediction System 'DIAPRES' application. The software development method uses the Waterfall method with PHP-MySQL as the programming language used and Black Box Testing Application Testing.

Kata Kunci: *diabetes, decision tree, parameter optimization, diabetes prediction system*

1. PENDAHULUAN

Menurut Pusat Data dan Informasi Kesehatan Republik Indonesia, diabetes adalah suatu penyakit menahun (kronis) berupa gangguan metabolik yang ditandai dengan kadar gula darah yang melebihi batas normal. Komplikasi yang sering terjadi pada penderita diabetes dan mematikan adalah serangan jantung dan stroke dimana ebagian besar kematian terjadi karena kenaikan kadar glukosa secara terus menerus sehingga mengakibatkan rusaknya pembuluh darah, saraf dan struktur internal lainnya [1]. Organisasi International Diabetes Federation (IDF) memperkirakan bahwa sedifitnya 463 juta orang pada usia 20-79 tahun di dunia menderita diabetes pada tahun 2019 atau sama dengan angka prevalessni sebesar



9,3% dari total penduduk pada usia yang sama [1]. Riset Kesehatan Dasar (Riskesdas) pada tahun 2018 telah melakukan pengumpulan data penderita diabetes mellitus pada penduduk yang berusia ≥ 15 tahun dimana kriteria penderita yaitu gejala sering lapar, sering haus, sering buang air kecil dan dalam jumlah banyak, dan juga berat badan turun.

Dari hasil Riskesdas 2018 juga diketahui bahwa prevalensi diabetes mellitus di Indonesia berdasarkan diagnosis dokter pada umur ≥ 15 tahun sebesar 2%. Prevalensi diabetes mellitus menurut hasil pemeriksaan gula darah juga meningkat 6,9% pada 2013 menjadi 8,5% pada tahun 2018. Menurut Riskesdas 2018, angka ini menunjukkan bahwa baru sekitar 25% penderita diabetes mengetahui bahwa dirinya menderita diabetes dan lebih dari sisa jumlah tersebut tidak mengetahui bahwa dirinya menderita diabetes. Untuk menekan jumlah penderita penyakit diabetes yang semakin bertambah, bisa dilakukan deteksi dini oleh tenaga ahli dan deteksi dini penyakit diabetes juga bisa dilakukana dengan mengembangkan suatu sistem prediksi penyakit diabetes dengan memanfaatkan berbagai metode, salah satunya adalah metode Data Mining [2].

Data Mining adalah proses untuk mengekstraksi informasi yang berguna dari database besar termasuk salah satunya data medis di rumah sakit yaitu data penderita diabetes. Data Mining merupakan bidang ilmu komputer multidisiplin yang melibatkan proses komputasi, machine learning. Teknik statistik, klasifikasi, pengelompokan dan penemuan pola [3]. Decision Tree adalah model prediksi yang paling umum digunakan untuk klasifikasi dimana setiap cabang mewakili hasil pengujian dan daun pohon mewakili kelas atau kelas distribusi dengan tujuan utama menghasilkan model yang memprediksi nilai variabel yang dibutuhkan berdasarkan banyaknya variabel input yang digunakan dimana aturan model klasifikasi ini adalah berbasis prediksi [4][4]. Decision Tree merupakan algoritma supervised machine learning yang digunakan untuk memecahkan masalah prediksi dan klasifikasi dimana dalam setiap tahapan pohon keputusan memilih setiap node dengan mengevaluasi perolehan informasi tertinggi di antara semua atribut [5]

Beberapa penelitian terkait prediksi diabetes yang pernah dilakukan sebelumnya oleh beberapa peneliti antara lain :

Prediksi penyakit diabetes dilakukan juga oleh [1] yang melakukan penelitian prediksi penyakit diabetes mellitus dengan class negative dan positive yang dilakukan dengan metode decision tree algoritma C4.5. Pada penelitian ini menggunakan dataset publik, dataset Early stage diabetes risk prediction dataset dari UCI Repository dan pengujian tingkat akurasi dilakukan dengan menggunakan confusion matrix. Penerapan metode algoritma C4.5 pada prediksi penyakit diabetes ini menghasilkan nilai akurasi sebesar 91,82%. Penelitian [5] melakukan komparasi metode klasifikasi Decision Tree, SVM dan Naïve Bayes untuk menghasilkan model berdasarkan akurasi tertinggi. Eksperimen dilakukan pada Pima Indians Diabetes Database (PIDD) yang bersumber dari Repository UCI. Kinerja ketiga algoritma dievaluasi dengan pengukuran Precision, Accuracy, F-Measure, and Recall. Hasil yang diperoleh menunjukkan Naive Bayes mengungguli dengan akurasi tertinggi yaitu 76,30% dibandingkan algoritma lainnya. Lain lagi penelitian yang dilakukan [6] dimana peneliti menggunakan metode Decision Tree untuk prediksi penyakit diabetes dengan perbandingan data training dan data testing untuk mencari akurasi tertinggi. Data yang digunakan adalah data pasien yang berjumlah 768 record dan untuk pemodelan digunakan bahasa pemrograman Python. Dari eksperimen yang dilakukan, diketahui dengan perbandingan data training 50% dan data testing 50% menghasilkan nilai akurasi tertinggi yaitu 71%. Sementara penelitian yang dilakukan [7] melakukan perbandingan metode klasifikasi Naïve Bayes, SVM, Linear Regression, Adaboost, Random Forest, K-Nearest Neighbor, Decision Tree dan Neural Network. Dataset yang digunakan diambil dari Repositori Machine Learning UCI yaitu Pima Indian Diabetes (PID) yang berjumlah 768 pasien dengan 9 atribut. Eksperimen dilakukan dengan tool Weka dan dari eksperimen diketahui akurasi tertinggi adalah algoritma Logistic Regression sebesar 78,85% dan kedua adalah algoritma Naïve Bayes sebesar 78,28%

Kemudian ada penelitian [8] yang membuat model system diagnosa diabetes dengan menerapkan metode clustering dan klasifikasi. Metode clustering yang diterapkan adalah metode K-Means dan metode klasifikasi yang digunakan adalah metode Naïve Bayes, Logistic Regression dan Artificial Neural Networks (ANN). Atribut yang digunakan adalah Age, Gender, Diet, Height, Weight, Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Random Blood Sugar (RBS) dan Class Variable (class). Pada penelitian ini belum dijelaskan evaluasi yang dilakukan menghasilkan akurasi berapa. Penelitian [9] melakukan dua tahapan dalam proses pengolahan data yang dilakukan. Tahapan pertama adalah mengumpulkan dataset yang digunakan yaitu Pima Indian Diabetes Dataset dari data publik UCI Repository dan dataset privat dari Bombay Medical Hall, India. Tahapan kedua adalah melakukan perbandingan performance dengan dua pendekatan. Pendekatan pertama adalah melakukan perbandingan metode klasifikasi yaitu Logistic Regression, K-Nearest Neighbor, ID3 DT, C4.5 DT dan Naïve Bayes. Pendekatan kedua dengan optimasi reduksi fitur menggunakan Principal Dimensionality Reduction (PCA) dan Particle Swarm Optimization (PSO) untuk metode klasifikasi yang digunakan. Hasil eksperimen menunjukkan penerapan PCA dan PSO menghasilkan akurasi yang lebih tinggi. Penelitian [10] melakukan komparasi metode Machine Learning Random Forest (RF), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), Gradient Boost (GB), Decision Tree (DT), and Logistic Regression (LR) untuk mendapatkan nilai akurasi tertinggi. Dataset yang digunakan diambil dari klinik kesehatan di distrik Bandipora, India dari April 2021–Feb2022. Urutan akurasi yang dihasilkan adalah MLP (90.99%), SVM (92%), GBC (97%), DT (96%) dan LR (69%),



Penelitian ini bertujuan untuk membangun sebuah sistem yaitu aplikasi ‘DIAPRES’ berbasis website yang berguna untuk identifikasi awal individu yang memiliki faktor resiko penyakit diabetes. Tujuan khusus dari penelitian adalah memudahkan individu melakukan deteksi dini penyakit diabetes agar resiko penyakit bisa diminimalisir. Urgensi penelitian ini adalah dengan adanya sistem yang dibangun maka bila ditemukan individu dengan masalah diabetes dapat dilakukan edukasi, intervensi dan atau dirujuk ke Fasilitas Pelayanan Kesehatan Tingkat Pertama (FKTP). Deteksi awal penyakit diabetes dilakukan agar dapat mengurangi faktor resiko jatuh sakit diabetes dan orang yang sudah menderita sakit diabetes dapat mengendalikan penyakitnya agar tidak terjadi komplikasi dan kematian dini. Aplikasi ini dikembangkan berbasis pada metode Data Mining yaitu metode Klasifikasi Decision Tree yang menghasilkan pohon keputusan yang digunakan untuk pedoman dalam pembuatan aplikasi yang dilakukan dengan Optimasi Parameter sehingga dihasilkan akurasi tertinggi. Penelitian ini akan menggunakan data penderita penyakit diabetes dari data publik Kaggle yang kemudian dibangun Aplikasi DIAPRES berbasis web menggunakan PHP dan MySQL. Aplikasi DIAPRES diharapkan dapat memudahkan setiap orang yang ingin mendapatkan akses informasi deteksi dini penyakit diabetes.

2. METODE PENELITIAN

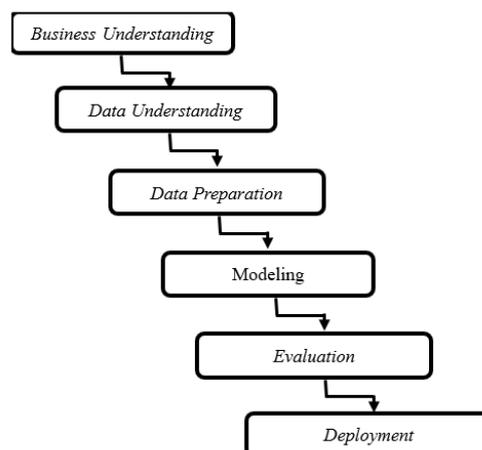
2.1 Desain Penelitian

Dataset diperoleh dari Public Dataset Kaggle yang memiliki 2000 record dan 9 atribut (8 atribut umum dan 1 atribut khusus). Atribut yang dimaksud adalah *Pregnancies*, *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, *BMI*, *DiabetesPedigreeFunction*, *Age* dan *Outcome*. *Outcome* adalah atribut yang digunakan untuk pelabelan dengan nilai 1 dan 0. Nilai 1 berarti positif (penderita diabetes) dan nilai 0 berarti negative (bukan penderita diabetes).

Metode atau model yang diusulkan dalam penelitian ini adalah metode klasifikasi Decision Tree dengan Optimasi Parameter. Parameter yang dimaksud yaitu *Number of folds* pada *Cross Validation* dan parameter *Criterion*, *maximal depth*, *apply pruning* dan *apply pre pruning* pada metode *Decision Tree*. Sedangkan metode penelitian yang diusulkan menggunakan metode *CRISP-DM* yang bertujuan mendapatkan pola atau pengetahuan baru dari data yang digunakan dalam memecahkan suatu masalah juga digunakan untuk pengambilan keputusan yang tepat. Metode *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) adalah metode yang memiliki beberapa tahapan utama yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation* dan *Deployment*, di mana dalam penelitian ini yang dilakukan antara lain pengumpulan dan *preprocessing* data, pemilihan atribut label, membangun model klasifikasi Decision Tree dengan Optimasi Parameter, evaluasi dan pengembangan aplikasi prediksi diabetes [11] [12].

2.2 Langkah-langkah Penelitian

Langkah-langkah penelitian sesuai metode *CRISP-DM* dijelaskan pada Gambar 3 berikut ini :



Gambar 1. Langkah-langkah Penelitian

Berikut penjelasan dari langkah-langkah Penelitian:



1. Business Understanding

Tahapan pertama yang dilakukan adalah menganalisa dan memahami masalah yang ada. Penelitian ini dilatar belakangi bahwa permasalahan akurasi pada prediksi penyakit diabetes yang belum tinggi atau belum akurat. Penelitian ini bertujuan untuk melakukan Optimasi Parameter pada metode klasifikasi *Decision Tree* agar dihasilkan pohon keputusan yang lebih akurat untuk prediksi penyakit diabetes.

2. Data Understanding

Dataset adalah data penderita diabetes yang diambil dari data publik Kaggle. Dataset berisi 2000 record dan 9 atribut (satu atribut khusus). Atribut yang dimaksud adalah *Pregnancies*, *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, *BMI*, *DiabetesPedigreeFunction*, *Age* dan *Outcome*. Dataset yang dimaksud ditampilkan dalam Tabel 1 berikut :

Tabel 1. Dataset

No	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigree Function	Age	Outcome
1.	0	145	0	0	0	44,200	0,630	31	1
2.	8	194	80	0	0	26,100	0,551	67	0
3.	3	80	0	0	0	0	0,174	22	0
4.	6	166	74	0	0	26,600	0,304	66	0
5.	5	110	68	0	0	26	0,292	30	0
6.	10	75	82	0	0	33,300	0,263	38	0
7.	8	120	78	0	0	25	0,409	64	0
...

3. Data Preparation

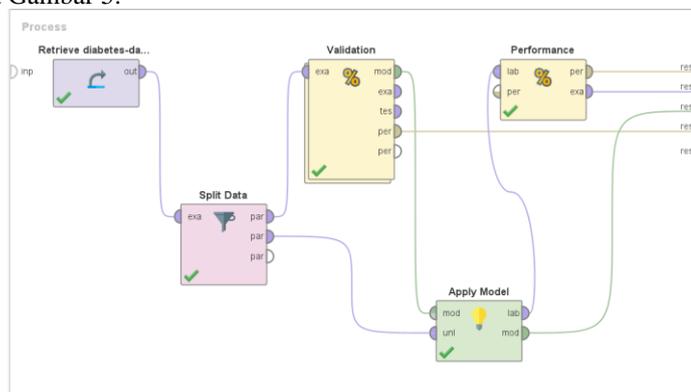
Sebelum data processing dilakukan maka dilakukan terlebih dahulu preprocessing, salah satunya adalah menetapkan atribut target atau label. Dan label yang digunakan pada dataset ini adalah atribut *Outcome*. *Outcome* merupakan atribut yang digunakan untuk pelabelan dengan nilai 1 dan 0. Nilai 1 berarti positif (penderita diabetes) dan nilai 0 berarti negatif (bukan penderita diabetes).

4. Modeling

Metode atau model yang diusulkan dalam penelitian ini adalah metode *Decision Tree* dengan Optimasi Parameter. Parameter yang dimaksud adalah *Number of folds* pada *Cross Validation* dan pada metode *Decision Tree* adalah parameter *criterion*, *maximal depth*, *apply pruning* dan *apply pre pruning*.

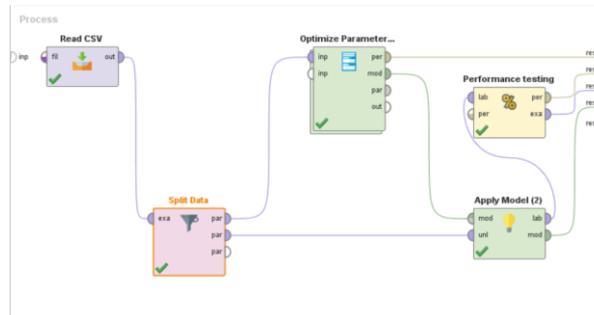
Dalam pemrosesan dan pengujian model, software yang digunakan adalah RapidMiner versi 10.

Modeling pertama dengan tools RapidMiner ditampilkan pada Gambar 2. yaitu menginputkan data, operator *Split Data* (membagi data training dan testing dengan perbandingan 80:20), *Validation* (melakukan sebuah *Cross Validation* untuk prediksi nilai kinerja model latih), *Apply Model* (mengaplikasikan sebuah model) dan *Performance* (evaluasi kinerja). Pada operator *Validation* jika diklik maka akan tampil penerapan metode klasifikasi didalamnya seperti ditampilkan pada Gambar 5.

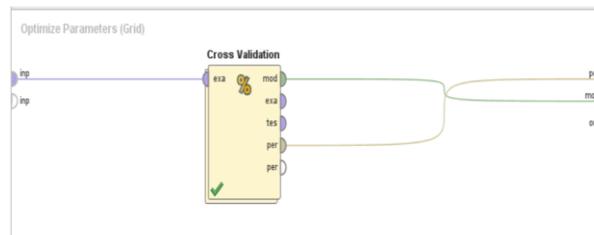


Gambar 2. Modeling Tanpa Optimasi

Modeling berikutnya ditampilkan pada Gambar 3. berikut ini dimana modelling sudah menambahkan parameter *Optimasi Parameters (Grid)* untuk melakukan optimasi parameter. Jika operator ini diklik, maka didalamnya akan tampil operator *Cross Validation* seperti ditampilkan pada Gambar 4.

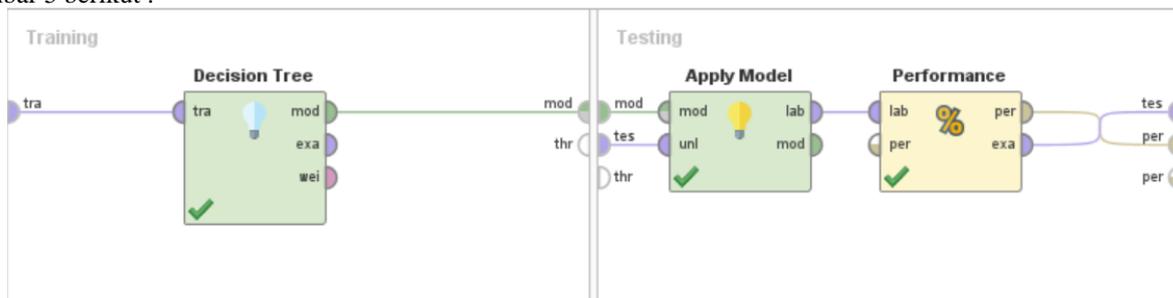


Gambar 3. Modeling dengan Optimasi Parameter



Gambar 4. Operator Cross Validation

Jika operator Cross Validation diklik maka akan ditampilkan modeling dari metode klasifikasi yang digunakan seperti Gambar 5 berikut :



Gambar 5. Modeling Metode Decision Tree

5. Evaluation

Proses evaluasi pada penelitian ini adalah membandingkan hasil akurasi yang belum dan sudah di optimasi parameternya. Diketahui bahwa nilai akurasi setelah di optimasi, hasilnya lebih tinggi.

6. Deployment

Setelah diperoleh model dengan akurasi yang tertinggi, maka tahap selanjutnya adalah merancang Diabetes Prediction System “Diapres” menggunakan pohon keputusan hasil Optimasi Parameter pada metode *Decision Tree*.

2.3 Metode Pengembangan Perangkat Lunak

Hasil tahapan terakhir dalam penelitian ini adalah pengembangan aplikasi system prediksi diabetes. Dalam pengembangan aplikasi metode pengembangan perangkat lunak yang digunakan adalah metode Waterfall. Tahapan dalam metode Waterfall yang diimplementasikan dalam penelitian ini yaitu [13] :

1. Requirement

Pada tahapan ini dilakukan analisa akan kebutuhan sistem. Deteksi awal penyakit diabetes dilakukan agar dapat mengurangi faktor resiko jatuh sakit diabetes dan orang yang sudah menderita sakit diabetes dapat mengendalikan penyakitnya agar tidak terjadi komplikasi dan kematian dini. *Diabetes Prediction System* dibutuhkan karena apabila ditemukan individu dengan masalah diabetes dapat dilakukan edukasi, intervensi dan atau dirujuk ke Fasilitas Pelayanan Kesehatan Tingkat Pertama (FKTP).



2. *Design System*
Desain sistem yang baik dan *user friendly* diterapkan dalam tahapan ini. Kebutuhan akan hak akses sistem dimana *user* siapa saja yang menggunakan aplikasi dan nterface dengan inputan yang mudah dimengerti juga dipahami menjadi poin penting disini.
3. *Coding and Testing*
Aplikasi yang akan dikembangkan adalah aplikasi berbasis website. Dan pada tahapan ini dilakukan pembuatan aplikasinya dengan *coding* bahasa pemrograman PHP-MySQL. Adapun dasar dari pembuatan aplikasi ini adalah hasil pohon keputusan berdasarkan eksperimen pengolahan data metode *Decision Tree* dengan Optimasi Parameter. Selain itu juga dilakukan testing untuk menguji aplikasi yang sudah dibuat apakah sudah sesuai dengan yang dibutuhkan atau belum.
4. *Program Implementation*
Tahapan ini adalah implementasi dari aplikasi yang dibuat. Implementasi dilakukan oleh pihak terkait yang berkepentingan.
5. *Operation and Maintenance*
Pada tahapan ini dilakukan pengoperasian dan *maintenance* pada aplikasi. *Maintenance* yang dilakukan antara lain melakukan *back-up* data secara berkala.

2.4 Metode Pengujian Perangkat Lunak

Pengujian Perangkat Lunak yang digunakan dalam penelitian ini adalah *Black Box Testing*. Dijelaskan bahwa pengujian *Black Box* adalah pengujian yang berfokus pada antarmuka atau interface dimana pengujian fungsional yang dilakukan langsung pada aplikasi, selain itu juga adanya kesesuaian pada alur fungsi yang dibutuhkan oleh pengguna (*user*) [14].

3. HASIL DAN PEMBAHASAN

3.1 Perbandingan Nilai Akurasi

Pada pengolahan data yang dilakukan dengan perbandingan data training dan data testing 80:20, peneliti membuat dua model. Yang pertama adalah model tanpa optimasi parameter dan yang kedua adalah model dengan optimasi parameter. Dari dua model tersebut didapatkan hasil dari operator Cross Validation yang menghasilkan Tabel Confusion Matrix sebagai berikut :

3.1.1 Nilai Akurasi Tanpa Optimasi

Pada Tabel 2. ditampilkan nilai akurasi yang dihasilkan sebesar 76% untuk data training dan nilai akurasi sebesar 80% untuk data testing. Hal ini menunjukkan bahwa nilai akurasi pada data testing ada peningkatan dibandingkan data training yaitu dari 76% menjadi 80%.

Tabel 2. Akurasi Data Training

Accuracy : 76.00 %			
	<i>True 1</i>	<i>True 0</i>	<i>Class Precision</i>
<i>Pred. 1</i>	224	61	78.60%
<i>Pred. 0</i>	323	992	75.44%
<i>Class Recall</i>	40.95%	94.21%	

Pada Tabel Data Testing diketahui bahwa yang diprediksi Diabetes dan prediksinya benar ada 60 orang dan prediksinya salah (tidak Diabetes) ada 3 orang. Sementara yang diprediksi tidak Diabetes tapi ternyata prediksinya adalah Diabetes ada 77 orang dan yang sesuai prediksi yaitu tidak Diabetes ada 260 orang.

Tabel 3. Akurasi Data Testing

Accuracy : 80.00 %			
	<i>True 1</i>	<i>True 0</i>	<i>Class Precision</i>
<i>Pred. 1</i>	60	3	95.24%
<i>Pred. 0</i>	77	260	77.15%
<i>Class Recall</i>	43.80	98.86%	



3.1.2 Sesudah di Optimasi

Parameter yang dioptimasi dalam pengolahan data penelitian ini adalah *Number of folds* pada *Cross Validation* dan pada metode *Decision Tree* adalah parameter *Criterion*, *maximal depth*, *apply pruning* dan *apply pre pruning*. Dan nilai pada parameter untuk menghasilkan akurasi tertinggi yaitu 98,00% dihasilkan dari dua kombinasi parameter seperti Tabel 4. Dibawah ini :

Tabel 4. Optimasi Parameter dan Hasil Akurasi

<i>Validation.number of folds</i>	<i>DT.criterion</i>	<i>DT.maximal_depth</i>	<i>DT.apply_pruning</i>	<i>DT.apply_prepruning</i>	<i>accuracy</i>
51	<i>Information_gain</i>	90	<i>False</i>	<i>False</i>	98.00%
31	<i>Information_gain</i>	60	<i>False</i>	<i>False</i>	98.00%

Dari tabel diatas diketahui bahwa optimasi parameter menghasilkan nilai dengan akurasi tertinggi yaitu 98% ada 2 skema yang bisa dilakukan dimana ada beberapa parameter yang menghasilkan nilai sama yaitu parameter *DT.criterion* menghasilkan *Information_gain*, *DT.apply_pruning* dan *DT.apply_prepruning* sama-sama menghasilkan *False*. Kemudian untuk parameter *Validation.number_of_folds* dan *DT.maximal_depth* menghasilkan nilai 51 dan 90, sedangkan yang kedua menghasilkan nilai 31 dan 60.

3.2 Deskripsi Pohon Keputusan

Dari pengolahan data metode *Decision Tree* dengan *Optimasi Parameter* diketahui deskripsi pohon keputusan. Deskripsi pohon keputusan adalah dasar dalam pengembangan aplikasi *Diabetes Prediction System* “Diapres”. Adapun deskripsi pohon keputusan ditampilkan sebagai berikut :

```

Glucose > 127.500
| Glucose > 154.500
| | Age > 62.500
| | | BloodPressure > 85: 1 {1=2, 0=0}
| | | BloodPressure ≤ 85: 0 {1=0, 0=9}
| | | Age ≤ 62.500
| | | | BMI > 46.100
| | | | | Insulin > 120: 0 {1=0, 0=12}
| | | | | Insulin ≤ 120: 1 {1=10, 0=0}
| | | | | BMI ≤ 46.100
| | | | | | BMI > 40.850: 1 {1=31, 0=0}
| | | | | | BMI ≤ 40.850
| | | | | | | BMI > 40.350: 0 {1=0, 0=4}
| | | | | | | BMI ≤ 40.350
| | | | | | | | Insulin > 542.500: 0 {1=0, 0=3}
| | | | | | | | Insulin ≤ 542.500
| | | | | | | | | DiabetesPedigreeFunction > 0.681: 1 {1=42, 0=0}
| | | | | | | | | DiabetesPedigreeFunction ≤ 0.681
| | | | | | | | | | DiabetesPedigreeFunction > 0.623
| | | | | | | | | | | Glucose > 163.500: 0 {1=0, 0=7}
| | | | | | | | | | | Glucose ≤ 163.500: 1 {1=2, 0=0}
| | | | | | | | | | | DiabetesPedigreeFunction ≤ 0.623
| | | | | | | | | | | DiabetesPedigreeFunction > 0.343: 1 {1=50, 0=0}
| | | | | | | | | | | DiabetesPedigreeFunction ≤ 0.343
| | | | | | | | | | | | Age > 30.500
| | | | | | | | | | | | DiabetesPedigreeFunction > 0.336: 0 {1=0, 0=2}
| | | | | | | | | | | | DiabetesPedigreeFunction ≤ 0.336
| | | | | | | | | | | | Age > 59.500: 0 {1=0, 0=2}
| | | | | | | | | | | | Age ≤ 59.500
| | | | | | | | | | | | | Pregnancies > 9.500
| | | | | | | | | | | | | | BloodPressure > 69: 0 {1=0, 0=4}
| | | | | | | | | | | | | | BloodPressure ≤ 69: 1 {1=5, 0=0}
| | | | | | | | | | | | | | Pregnancies ≤ 9.500
| | | | | | | | | | | | | | | BloodPressure > 69: 1 {1=37, 0=0}
| | | | | | | | | | | | | | | BloodPressure ≤ 69
| | | | | | | | | | | | | | | | Glucose > 160.500
| | | | | | | | | | | | | | | | | Glucose > 189: 0 {1=0, 0=1}
    
```



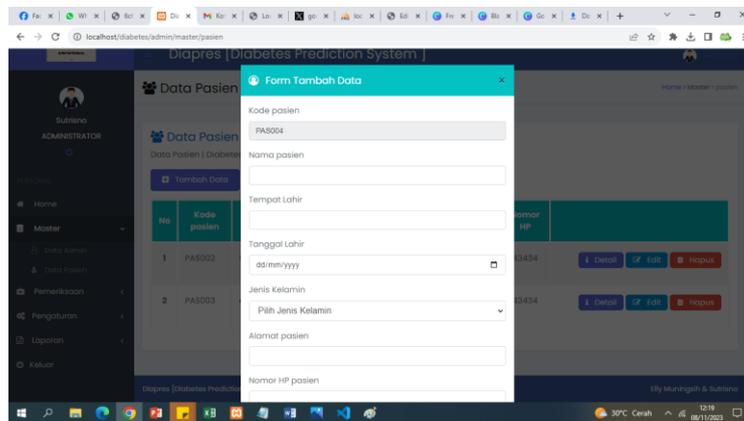

home, master, pemeriksaan, pengaturan, laporan dan keluar dari aplikasi



Gambar 6. Halaman Dashboard

3.3.2 Tambah Data Pasien

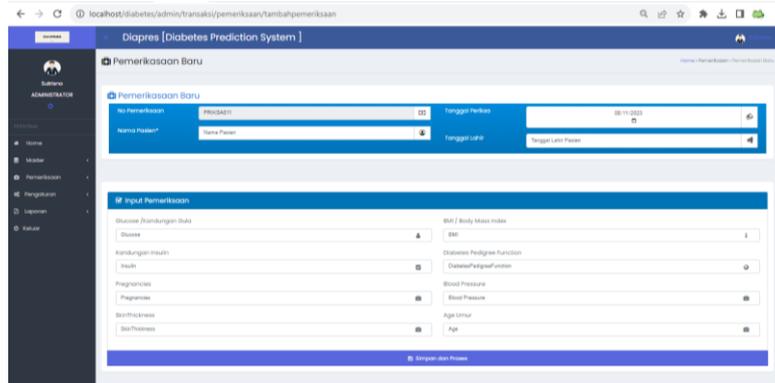
Halaman ini menampilkan Formulir Tambah Data Pasien. Pada aplikasi ini Data Pasien disimpan dalam database untuk efisiensi dan kemudahan dalam pengecekan prediksi penyakit Diabetes. Adapun atribut yang diinputkan dalam Form Tambah Data Pasien adalah Kode Pasien (didapatkan otomatis dari sistem), Nama pasien, Tempat Lahir, Tanggal Lahir, Jenis Kelamin, Alamat pasien dan Nomor HP pasien.



Gambar 7. Halaman Tambah Data Pasien

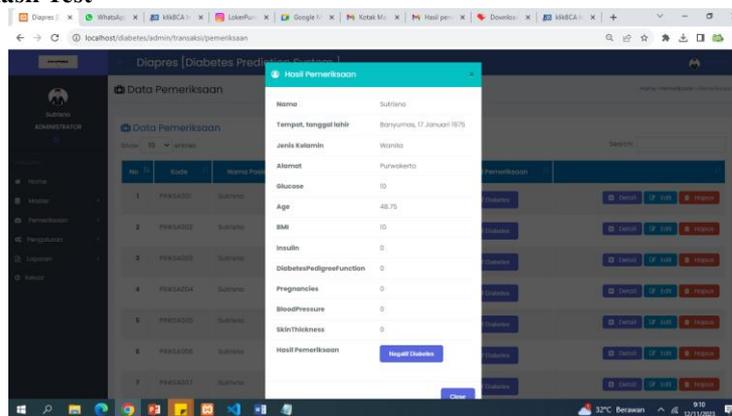
3.3.3 Halaman Pemeriksaan Baru

Untuk halaman tambah pemeriksaan/pemeriksaan baru berfungsi untuk menambahkan pemeriksaan pasien terdeteksi diabetes atau tidak. Langkahnya adalah user mengklik pasien, kemudian muncul pilihan pasien yang akan diperiksa, kemudian setelah itu Kandungan Insulin, *Diabetes Pedigree Function*, *Pregnancies*, *Blood Pressure*, *SkinThickness*, *Age Umur* [otomatis terisi berdasar tanggal lahir], setelah semua terisi klik tombol simpan dan proses, maka akan masuk ke menu pemeriksaan, dan hasil akan bisa dilihat di menu pemeriksaan



Gambar 8. Halaman Pemeriksaan Baru

3.3.4 Halaman Detail Hasil Test



Gambar 9. Halaman Detail Hasil Test

4. KESIMPULAN

Diabetes adalah suatu penyakit menahun (kronis) berupa gangguan metabolik yang ditandai dengan kadar gula darah yang melebihi batas normal. Komplikasi yang sering terjadi pada penderita dan mematikan adalah serangan jantung dan stroke. Deteksi dini penyakit diabetes perlu dilakukan karena baru sekitar 25% penderita diabetes mengetahui bahwa dirinya menderita diabetes dan lebih dari sisa jumlah tersebut tidak mengetahui bahwa dirinya menderita diabetes. Salah satu cara yang dapat dilakukan adalah dengan pengembangan aplikasi berbasis metode *Data Mining*. Penelitian yang dilakukan kali ini mengembangkan aplikasi berbasis metode klasifikasi Decision Tree dengan optimasi parameter. Parameter yang dimaksud adalah *Number of folds* pada *Cross Validation* dan pada metode *Decision Tree* adalah parameter *criterion*, *maximal depth*, *apply pruning* dan *apply pre pruning*. Dari hasil eksperimen terbukti efektif bahwa optimasi parameter dapat meningkatkan nilai akurasi, dimana diketahui bahwa tanpa optimasi parameter, akurasi yang dihasilkan berkisar 76.00-80.00%. Dan setelah dilakukan optimasi parameter, terbukti akurasi meningkat menjadi 98.00% (nilai tertinggi). Terdapat 2 skema yang bisa dilakukan untuk menghasilkan nilai akurasi tertinggi, dimana ada beberapa parameter yang menghasilkan nilai sama yaitu parameter *DT.criterion* menghasilkan *Information gain*, *DT.apply pruning* dan *DT.apply pre pruning* sama-sama menghasilkan *False*. Kemudian untuk parameter *Validation.number_of_folds* dan *DT.maximal_depth* menghasilkan nilai 51 dan 90, sedangkan yang kedua menghasilkan nilai 31 dan 60. Dari hasil pohon keputusan metode Decision Tree dikembangkan sebuah aplikasi prediksi penyakit diabetes yang diberi nama *Diabetes Prediction System* atau "Diapres". Aplikasi Diapres yang dikembangkan dengan metode *Waterfall* dan pengujian *Black Box* diharapkan bisa membantu deteksi dini penyakit diabetes agar seseorang yang terdeteksi diabetes bisa segera melakukan pemeriksaan, dan yang tidak terdeteksi bisa lebih menjaga diri dari penyakit diabetes.



UCAPAN TERIMAKASIH

Terima kasih tak terhingga penulis ucapkan kepada seluruh pimpinan dan civitas akademik Universitas Bina Sarana Informatika atas support dan dukungan sehingga penelitian dalam skema Penelitian Dosen Yayasan (PDY) ini dapat terselesaikan dengan baik.

DAFTAR PUSTAKA

- [1] R. A. Siallagan and Fitriyani, "Prediksi Penyakit Diabetes Mellitus Menggunakan Algoritma C4.5," *J. Responsif Ris. Sains dan Inform.*, vol. 3, no. 1, pp. 44–52, 2021, doi: 10.51977/jti.v3i1.407.
- [2] M. S. Efendi and H. A. Wibawa, "Prediksi Penyakit Diabetes Menggunakan Algoritma ID3 dengan Pemilihan Atribut Terbaik (Diabetes Prediction using ID3 Algorithm with Best Attribute Selection)," *Juita*, vol. VI, no. 1, pp. 29–35, 2018.
- [3] M. Komi, J. Li, Y. Zhai, and Z. Xianguo, "Application of data mining methods in diabetes prediction," *2017 2nd Int. Conf. Image, Vis. Comput. ICIVC 2017*, no. S IX, pp. 1006–1010, 2017, doi: 10.1109/ICIVC.2017.7984706.
- [4] E. Muningsih, C. Kesuma, Sunanto, Suripah, and A. Widayanto, "Combination of K-Means method with Davies Bouldin index and decision tree method with parameter optimization for best performance," *AIP Conf. Proc.*, vol. 2714, 2023, doi: 10.1063/5.0129119.
- [5] D. Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1578–1585, 2018, doi: 10.1016/j.procs.2018.05.122.
- [6] T. Dudkina, I. Menailov, K. Bazilevych, S. Krivtsov, and A. Tkachenko, "Classification and prediction of diabetes disease using decision tree method," *CEUR Workshop Proc.*, vol. 2824, pp. 163–172, 2021.
- [7] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, 2021, doi: 10.1016/j.ict.2021.02.004.
- [8] V. Prakash, R. Bhavani, and A. Anupriya, "Efficient diagnostic system for smart diabetes," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 3789–3792, 2019, doi: 10.35940/ijeat.F9392.088619.
- [9] D. K. Choubey, P. Kumar, S. Tripathi, and S. Kumar, "Performance evaluation of classification methods with PCA and PSO for diabetes," *Netw. Model. Anal. Heal. Informatics Bioinforma.*, vol. 9, no. 1, 2020, doi: 10.1007/s13721-019-0210-8.
- [10] S. S. Bhat, V. Selvam, G. A. Ansari, M. D. Ansari, and M. H. Rahman, "Prevalence and Early Prediction of Diabetes Using Machine Learning in North Kashmir: A Case Study of District Bandipora," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/2789760.
- [11] S. H. Sastry and P. M. S. P. Babu, "Implementation of CRISP Methodology for ERP Systems," vol. 2, no. 05, pp. 203–217, 2013, [Online]. Available: <http://arxiv.org/abs/1312.2065>
- [12] E. Muningsih, H. M. Nur, F. F. Dwi Imaniawan, Saifudin, V. R. Handayani, and F. Endiarto, "Comparative Analysis on Dimension Reduction Algorithm of Principal Component Analysis and Singular Value Decomposition for Clustering," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012101.
- [13] T. Tjahjanto, A. Arista, and E. Ermatita, "Application of the Waterfall Method in Information System for State-owned Inventories Management Development," *Sinkron*, vol. 7, no. 4, pp. 2182–2192, 2022.
- [14] B. A. Priyaungga, D. B. Aji, M. Syahroni, N. T. S. Aji, and A. Saifudin, "Pengujian Black Box pada Aplikasi Perpustakaan Menggunakan Teknik Equivalence Partitions," *J. Teknol. Sist. Inf. dan Apl.*, vol. 3, no. 3, p. 150, 2020, doi: 10.32493/jtsi.v3i3.5343.