**IMPLEMENTASI CASE BASED REASONING**

**DAN K-NEAREST NEIGHBOR**

**UNTUK DIAGNOSIS PENYAKIT DIABETES**

I Gst Bgs Bayu Adi Pramana1 **,** I Wayan Supriana2

1,2Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,

Universitas Udayaana

Jl. Raya Kampus Universitas Udayana Bukit Jimbaran , Badung, Bali.

Kode Pos: 80364. Indonesia

[1gustibgsbayu@gmail.com](mailto:1gustibgsbayu@gmail.com), [2iwayansupriana@gmail.com](mailto:2iwayansupriana@gmail.com)

***Abstract***

Diabetes adalah suatu penyakit menahun yang ditandai dengan kadar glukosa darah (gula darah) melebihi nilai normal. Dari banyaknya faktor resiko yang menyebabkan seseorang menderita penyakit diabetes, menyebabkan lamanya proses diagnosis karena banyaknya faktor yang harus diuji oleh dokter. Dalam hal ini akan digunakan Case Based Reasoning dan K-Nearest Neighbor untuk Diagnosis Penyakit Diabetes. Case Based Reasioning (CBR) adalah sebuah metode pemecahan sebuah masalah atau kasus yang baru yang merujuk pada kasus-kasus sebelumnya. Dimana dalam proses mencari kedekatan nilai antar kasus baru dan kasus lama pada penelitian ini menggunakan metode KNN. KNN adalah sebuah klasifikasi terhadap terhadap objek berdasarkan jaraknya paling dekat dengan objek tersebut. Dalam melakukan proses indexing pada pelitian ini menggunakan aplikasi IBM SPSS untuk mencari korelasi dari tiap fitur dengan hasil dari diagnosis pada dataset. Pada pengujian yang dilakukan akan menginputkan semua data testing yang ada dengan k = 5. Berdasarkan hasil perhitungan diatas adapun data testing yang diprediksi benar adalah sejumlah 113 data dari 154 data testing yang mendapatkan hasil akurasi sebesat 73%.

***Keywords:*** *CBR, Diabetes, KNN*

1. **Pendahuluan**

Diabetes adalah suatu penyakit menahun yang ditandai dengan kadar glukosa darah (gula darah) melebihi nilai normal. Glukosa merupakan sumber energi utama bagi sel tubuh manusia. Glukosa yang menumpuk di dalam darah akibat tidak diserap sel tubuh dengan baik dapat menimbulkan berbagai gangguan organ tubuh. Jika diabetes tidak dikontrol dengan baik, dapat timbul berbagai komplikasi yang membahayakan nyawa penderita. Ada banyak gejala yang menjadi faktor dari penyakit diabetes ini diantaranya adalah sering merasa haus, sering buang air kecil, terutama di malam hari, sering merasa sangat lapar, turunnya berat badan tanpa sebab yang jelas, berkurangnya massa otot dan lain sebagainya. Seringkali seseorang akan enggan untuk memeriksakan dirinya secara rutin ke dokter untuk melakukan test apakah dirinya menderita penyakit diabetes atau tidak. Dari banyaknya faktor resiko yang menyebabkan seseorang menderita penyakit diabetes, menyebabkan lamanya proses diagnosis karena banyaknya faktor yang harus diuji oleh dokter. Dalam hal ini akan digunakan Case Based Reasoning dan K-Nearest Neighbor untuk Diagnosis Penyakit Diabetes. Case Based Reasoning digunakan karena CBR adalah metode penalaran yang menggunakan pengetahuan lama untuk mengatasi masalah baru [3]. Penggunaan dari K-Nearest Neighbor ini adalah pada pencarian kasus lama yang relevan dengan kasus baru yaitu pada tahap Retrieve yang akan menguji kedekatan kasus baru dan kasus lama dengan rumus euclidean.

Salah satu penelitian yang telah dilakukan dengan CBR-KNN adalah yang dilakukan oleh Kalam, Setiawan dan Ratnawati pada Tahun 2014 tentang Implementasi Metode Case Based Reasoning (CBR) dan K-Nearest Neighbor (K-NN) Dalam Sistem Pendukung Keputusan Untuk Menentukan Tingkatan Depresi. Pada penelitian yang dilakukan menggabungkan 2 metode yaitu Case Based Reasoning (CBR) dan K-Nearest Neighbor (K-NN) yang mendapatkan hasil akurasi sebesar 87% [2]. Berdasarkan penelitian tersebut pada penelitian ini akan dilakukan penelitian tentang Implementasi Case Based Reasoning dan K-Nearest Neighbor untuk Diagnosis Penyakit Diabetes. Data pada penelitian ini menggunakan dataset Diabetes Indian Pima yang diperoleh dari https://www.kaggle.com/uciml/pima-indians-diabetes-database sebanyak 768 record data. Pada penelitian ini juga akan menggunakan IBM SPSS untuk mengetahui korelasi tiap fitur data dengan output yang dihasilkan.

1. **Metodelogi Penelitian**
   1. **Pengumpulan Data**

Data yang digunakan adalah data sekunder yang didapatkan dari bank data https://www.kaggle.com/uciml/pima-indians-diabetes-database yaitu dataset Indian Pima Diabetes Patient. Pada dataset ini memiliki 9 fitur/atribut dimana terdapat 768 record data dengan 268 record untuk data pasien penyakit diabetes dan 500 record untuk data pasien bukan penyakit diabetes. Dataset Diabetes Indian Pima ditunjukkan pada Tabel 1, dengan tipe data pada dataset adalah numerik. Pada atribut output yang merupakan hasil klasifikasi dari dataset, pasien penyakit diabetes ditandai dengan nilai 1 dan untuk bukan pasien penyakit diabetes ditandai dengan nilai 0. Adapun deskripsi tiap fitur dapat dilihat pada tabel berikut :

**Tabel 1.** Informasi Tiap Fitur Dataset

|  |  |
| --- | --- |
| Fitur | Informasi |
| Pregnancies | Jumlah Kehamilan |
| Glucose | Kadar Glukosa |
| BloodPressure | Tekanan Darah |
| SkinThickness | Ketebalan Lipatan Kulit |
| Insulin | Insulin yang dihasilkan dalam 2 jam |
| BMI | Indeks Massa Badan |
| DiabetesPedigreeFunction | Riwayat hubungan genetik |
| Age | Umur |
| Outcome | Hasil Diagnosis 0 untuk tidak mengidap diabetes, dan 1 mengidap diabetes |

Berdasarkan fitur dataset yang telah dipaparkan, untuk mencari nilai korelasi tiap fitur dengan hasil diagnosis pada dataset akan menggunakan aplikasi IBM SPSS(Statistical Product and Service Solutions). Adapun hasil dari analisis korelasi menggunakan aplikasi IBM SPSS adalah sebagai berikut :

**Tabel 2.** Hasil Analisis Korelasi IBM SPSS

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Preg | Gluc | Blood | Skin | Insulin | BMI | DiabetesPF | Age |
| Pearson Correlation | -.024 | .067 | .031 | .110\*\* | .138\*\* | .100\*\* | .269\*\* | -.329\*\* |
| Sig. (2-tailed) | .504 | .062 | .391 | .002 | .000 | .006 | .000 | .000 |
| N | 768 | 768 | 768 | 768 | 768 | 768 | 768 | 768 |

Berdasarkan hasil analisis korelasi diatas berikut adalah perhitungan persentase tiap fitur dengan menggunakan jumlah korelasi sebesar = 1.068

|  |  |  |
| --- | --- | --- |
| Pregnancies | : | (0.024/1.068) \* 100% = 2% |
| Glucose | : | (0.067/1.068) \* 100% = 6% |
| BloodPressure | : | (0.031/1.068) \* 100% = 3% |
| SkinThickness | : | (0.110/1.068) \* 100% = 10% |
| Insulin | : | (0.138/1.068) \* 100% = 13% |
| BMI | : | (0.100/1.068) \* 100% = 9% |
| DiabetesPedigreeFunction | : | (0.269/1.068) \* 100% = 26% |
| Age | : | (0.329/1.068) \* 100% = 31% |

Pada hasil perhitungan korelasi diatas adapun fitur yang memiliki tingkat persentase korelasi yang tinggi ada pada fitur Age dan DiabetesPF. Berikut adalah domain tiap fitur dataset :

**Tabel 3.** Domain Tiap Fitur Dataset

|  |  |
| --- | --- |
| Fitur | Domain |
| Pregnancies | 0 - 17 |
| Glucose | 0 - 199 |
| BloodPressure | 0 - 122 |
| SkinThickness | 0 - 99 |
| Insulin | 0 - 846 |
| BMI | 0 – 67.1 |
| DiabetesPedigreeFunction | 0.08 – 2.42 |
| Age | 21 - 81 |
| Outcome | 0 for not patient diabetes and 1 for patient diabetes |

* 1. **Normalisasi Data**

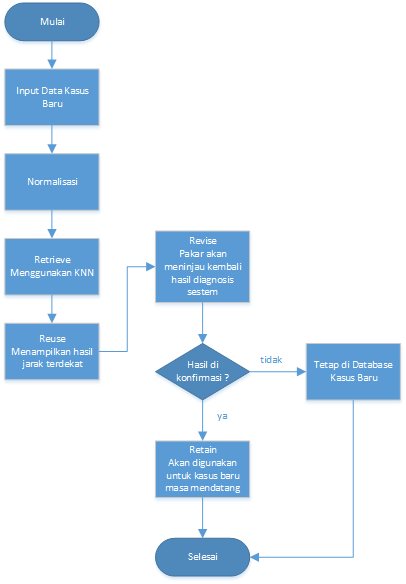
Normalisasi data digunakan untuk menyetarakan data tiap fitur agar tidak ada fitur yang memiliki nilai besar mendominasi hasil dari klasifikasi. Berikut adalah rumus untuk normalisasi min-max [1] :

(1)

Keterangan :

|  |  |  |
| --- | --- | --- |
|  | = | Nilai data sesudah normalisasi |
|  | = | Nilai data sebelum normalisasi |
|  | = | Nilai maksimum dari seluruh data asli |
|  | = | Nilai minimum dari seluruh data asli |

Dalam penelitian ini adapun alur penelitian dapat dilihat pada gambar dibawah ini, dimana pada tahap awal user akan memasukkan data kasus baru lalu data tersebut akan dinormalisasi oleh sistem. Selanjutnya akan dilakukan tahap retrieve CBR menggunakan metode KNN untuk mencari kedekatan antar kasus baru dan kasus lama. Setelah mendapat jarak terdekat akan menampilkan hasil diagnosis. Hasil diagnosis kasus baru ini akan disimpan pada sistem yang akan digunakan pada tahap revise, dimana pada tahap revise ini akan ditinjau kembali oleh pakar jika memenuhi kriteria akan dikonfirmasi ke basis kasus. Pada tahap retain akan menyimpan kasus baru yang telah dikonfirmasi ke tabel basis kasus untuk digunakan pada kasus baru mendatang dan jika tidak dikonfirmasi akan tetap berada di tabel kasus baru sampai dikonfirmasi oleh pakar. Berikut adalah alur penelitian sistem diagnosis :



**Gambar 1.** Alur Penelitian

* 1. **Case Based Reasioning (CBR)**

Case Based Reasioning (CBR) adalah sebuah metode pemecahan sebuah masalah atau kasus yang baru yang merujuk pada kasus-kasus sebelumnya. Case Based Reasioning menggunakan pendekatan kecerdasan buatan (Artificial Intelligent) atau komputasi yang menitikberatkan pemecahan masalah dengan didasarkan pada kasus-kasus sebelumnya [2]. Terdapat empat tahap dalam pemecahan kasus pada metode Case Based Reasioning (CBR) [2].

1. Retrieve

Pada proses ini sistem akan melakukan pengenalan parameter yang akan di jadikan acuan. Lalu mencari kasus lama yang memiliki kesamaan dengan kasus baru, selanjutnya sistem akan memilih kasus yang memiliki tingkat kecocokan tertinggi.

1. Reuse

Pada proses ini, sistem menggunakan informasi yang sudah didapat dari kasus sebelumnya atau melakukan adaptasi untuk memecahkan kasus baru tersebut.

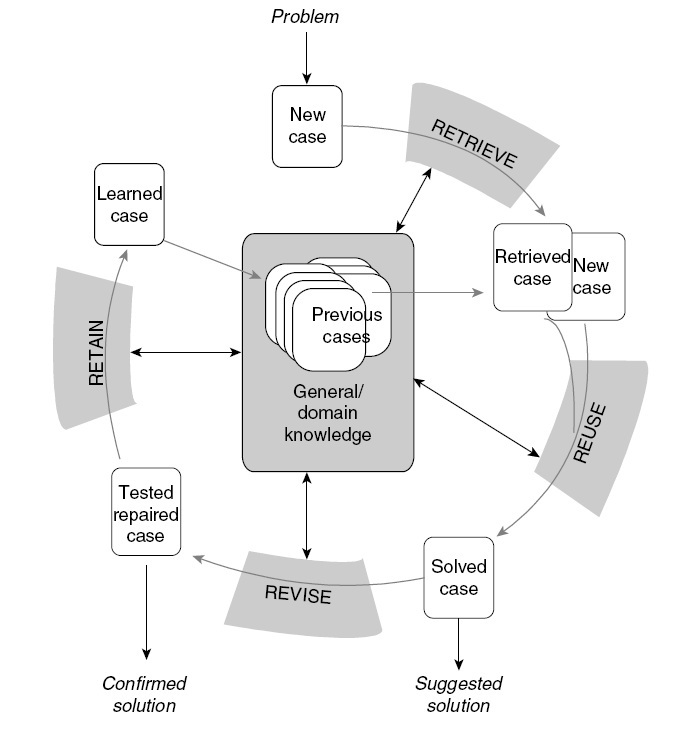
1. Revise

Pada proses ini sistem dan user akan meninjau kembali solusi yang telah didapatkan darikasus yang lama apakah solusi tersebut akan diterapkan pada kasus yang baru atau solusi tersebut perlu diperbaiki terlebih dahulu.

1. Retain

Pada proses ini apabila ternyata ditemukan solusi baru yang lebih baik dari solusi yang telah ada sebelumnya, maka solusi baru tersebut akan diberi indeks dan disimpan untuk kemudian digunakan kembali pada kasus serupa pada masa yang akan datang.

Berikut adalah alur proses Case Based Reasoning :



**Gambar 2.** Alur Proses Case Based Reasoning

* 1. **K-Nearest Neighbor(KNN)**

Algoritma KNN adalah sebuah klasifikasi terhadap terhadap objek berdasarkan jaraknya paling dekat dengan objek tersebut. Klasifikasi diperkirakan berdasarkan data jarak yang paling dekat (k = 1, nilai k ditentukan random) disebut dengan algoritma nearest neighbor [2]. Untuk mengitung nilai nearest neighbor mengunakan rumus :

Keterangan :

= jarak ke-i

= data uji ke-i

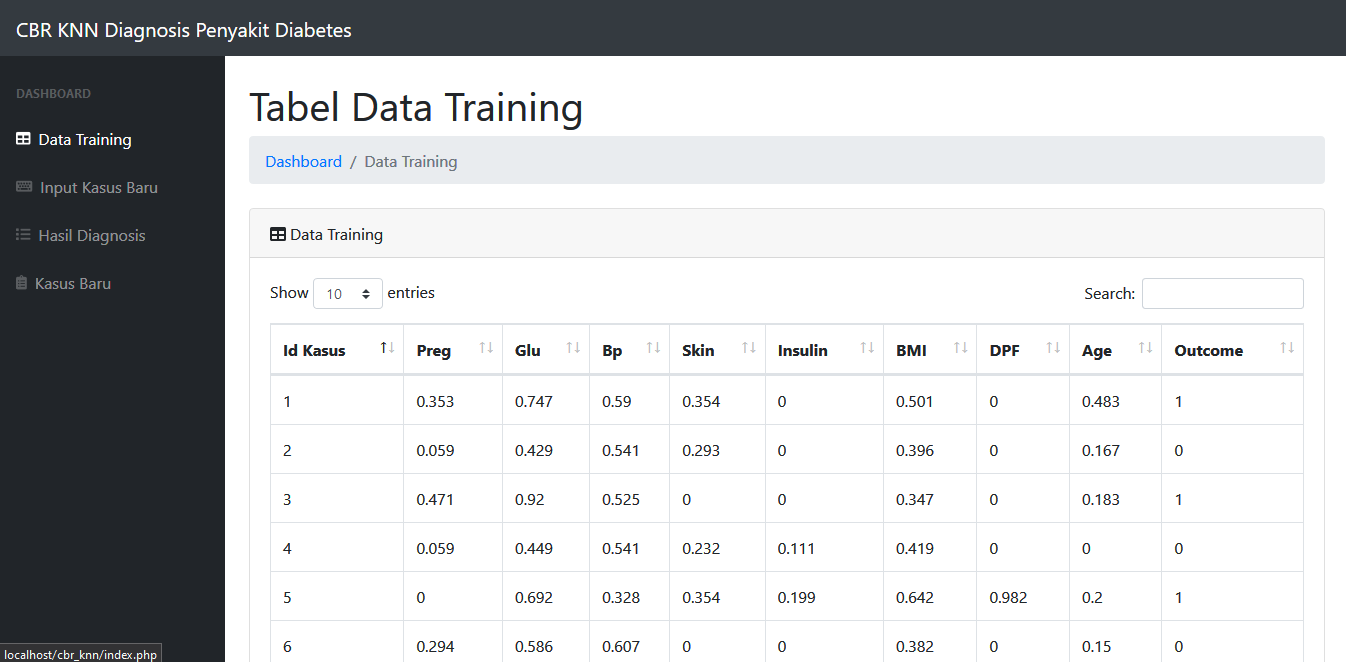
= basis kasus ke-i

*n* = maksimal iterasi

*i* = iterasi

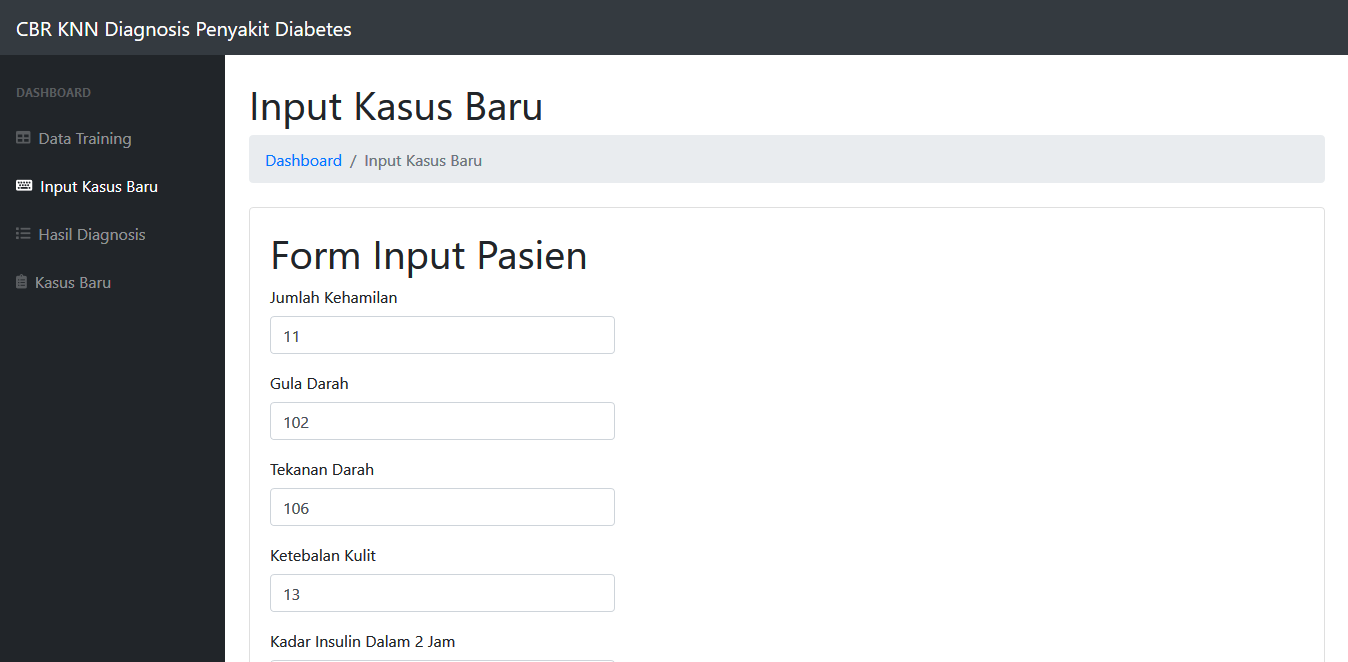
1. **Hasil dan Pembahasan**
   1. **Implementasi Sistem**

Pada implementasi sistem yang dilakukan menggunakan aplikasi berbasis web, dimana pada halaman dashboard terdapat 4 menu diantaranya adalah Data Training untuk melihat basis kasus dari sistem, Input Kasus Baru untuk menginput kasus baru, Hasil Diagnosis untuk melihat hasil dari diagnosis sistem berdasarkan kasus baru yang diinputkan, dan terakhir adalah Kasus Baru untuk melihat kasus baru yang diinputkan jika memenuhi kriteria jika dikonfirmasi oleh pakar akan disimpan ke basis kasus. Berikut adalah tampilan sistem diagnosis diabetes :



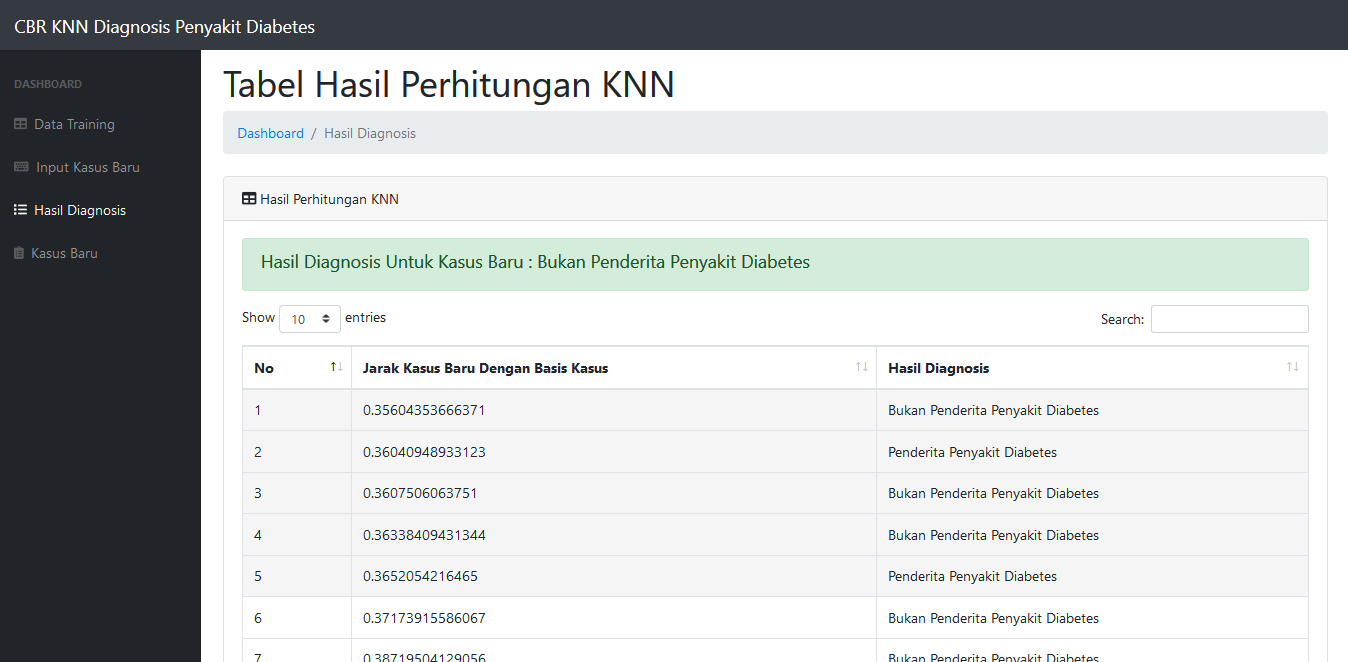
**Gambar 2.** Menu Data Training

Berikut adalah tampilan menu Input Kasus Baru, dimana user akan menginputkan nilai tiap fitur berdasarkan rentang yang sudah dideskripsikan :



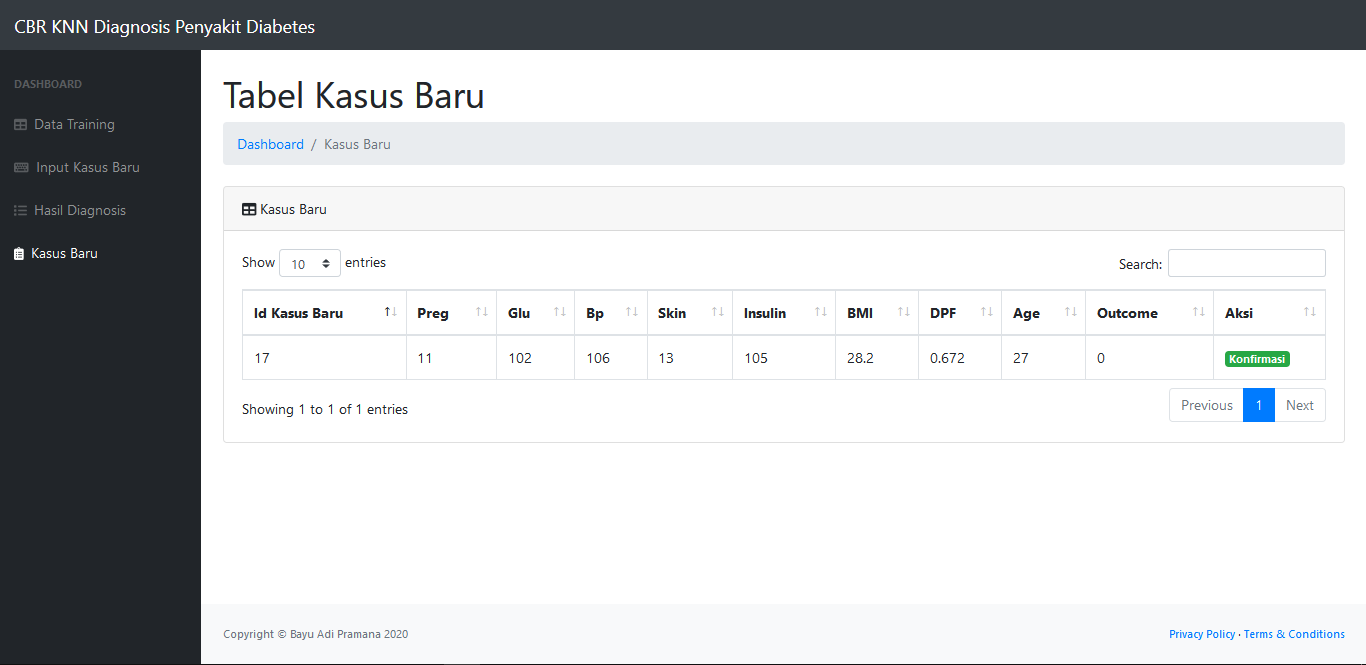
**Gambar 3.** Menu Input Kasus Baru

Berikut adalah tampilan menu Hasil Diagnosis, dimana pada menu ini akan menampilkan hasil perhitungan kasus baru yang dicari nilai similarity/kedekatan kasusnya dengan menggunakan metode KNN dan akan dipakai sejumlah K yang ditentukan untuk dijadikan acuan untuk mendiagnosis hasil dari inputan kasus baru :



**Gambar 4.** Menu Hasil Diagnosis

Berikut adalah tampilan dari menu Kasus Baru, dimana pada menu ini akan menampilkan kasus baru yang telah diinputkan user yang akan ditinjau oleh pakar jika sesuai kriteria maka pakar akan mengkonfirmasi kasus baru yang diinputkan user dan akan menyimpannya pada basis kasus yang akan digunakan untuk kasus baru mendatang.



**Gambar 5.** Menu Kasus Baru

# **Hasil dan Implementasi**

Pada tahap implementasi ini akan membagi dataset yang ada menjadi 2 bagian yaitu 80% data training dan 20% data testing. Data training berjumlah 614 record data dan data testing berjumlah 154 record data. Pada pengujian yang dilakukan akan menginputkan semua data testing yang ada dengan k = 5. Hasil prediksi yang benar akan digunakan untuk mencari nilai akurasi dari sistem. Adapun perhitungannya adalah sebagai berikut :

x 100%

Berdasarkan hasil perhitungan diatas adapun data testing yang diprediksi benar adalah sejumlah 113 data yang mendapatkan hasil akurasi sebesat 73%.

1. **Kesimpulan**

Case Based Reasioning (CBR) adalah sebuah metode pemecahan sebuah masalah atau kasus yang baru yang merujuk pada kasus-kasus sebelumnya. Dimana dalam proses mencari kedekatan nilai antar kasus baru dan kasus lama pada penelitian ini menggunakan metode KNN. KNN adalah sebuah klasifikasi terhadap terhadap objek berdasarkan jaraknya paling dekat dengan objek tersebut. Dalam melakukan proses indexing pada pelitian ini menggunakan aplikasi IBM SPSS untuk mencari korelasi dari tiap fitur dengan hasil dari diagnosis pada dataset. Pada tahap implementasi ini akan membagi dataset yang ada menjadi 2 bagian yaitu 80% data training dan 20% data testing. Data training berjumlah 614 record data dan data testing berjumlah 154 record data. Pada pengujian yang dilakukan akan menginputkan semua data testing yang ada dengan k = 5. Adapun data testing yang diprediksi benar adalah sejumlah 113 data dari 154 data testing yang mendapatkan hasil akurasi sebesat 73%.

**Referensi**

|  |  |
| --- | --- |
| [1] | M. S. A. Fauzi, B. Rahayudi and C. Dewi, "Perbandingan Jaringan Saraf Tiruan LVQ Dengan Backpropagation Dalam Deteksi Dini Penyakit Jantung Koroner," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* pp. 1952-1960, 2019. |
| [2] | M. E. Kalam, B. D. Setiawan and D. E. Ratnawati, "Implementasi Metode Case Based Reasoning (Cbr) Dan K-Nearest Neighbor (K-Nn) Dalam Sistem Pendukung Keputusan Untuk Menentukan Tingkatan Depresi," pp. 1-7, 2014. |
| [3] | N. W. Wiantari and I. W. Supriana, "Implementation of the CBR (Case-Based Reasoning) Method in Cases of Caesarean section," *Jurnal Elektronik Ilmu Komputer Udayana,* pp. 181-189, 2019. |