

## Implementasi Algoritma K-Nearest Neighbor (KNN) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba

M. Syukri Mustafa<sup>1</sup>, I Wayan Simpen<sup>2</sup>  
<sup>1,2</sup> Teknik Informatika STMIK Dipanegara  
<sup>1</sup>[moh.syukri@gmail.com](mailto:moh.syukri@gmail.com), <sup>2</sup>[simpen.dp@gmail.com](mailto:simpen.dp@gmail.com)

### Abstrak

Penelitian ini dimaksudkan untuk melakukan pengujian terhadap kemungkinan seorang pasien baru pada puskesmas Manyampa kabupaten Bulukumba dapat terkena penyakit diabetes militus atau tidak dengan menggunakan analisis data mining untuk menggali tumpukan histori data menggunakan algoritma K-Nearest Neighbor (KNN). Aplikasi yang dihasilkan pada penelitian ini menggunakan beberapa atribut yang klasifikasikan data mining antara lain : usia, anak ke, jumlah saudara, jenis kelamin, status perkawinan, pekerjaan, keluarga terkena diabetes militus, sehingga dengan menerapkan analisis KNN dapat dilakukan suatu prediksi berdasarkan kedekatan histori data yang ada dengan data yang baru, apakah pasaien tersebut diprediksi terkena diabete atau tidak. Data training yang digunakan dalam penelitian ini sebanyak 200 data pasien yang telah melakukan pemeriksaan dalam 2 tahun terakhir. Dari hasil pengujian penulis menggunakan data testing/sampel sebanyak 104 data pasien yang diambil diluar data training, hal ini dimaksudkan agar hasil prediksi Status Diabetes Militus Pasien dari sistem yang dirancang dapat dibandingkan dengan hasil Status Diabetes Militus Pasien yang sesuai dengan data. Hasil akurasi yangdipeloleh dari pengujian tersebut sebesar 68,30%.

*Kata kunci: diabetes militus, algoritma KNN, puskesmas Manyampa*

### Abstract

*This study is intended to examine the possibility of a new patient at the Pusekesmas Manyampa in Bulukumba district who can get diabetes militus or not by using data mining analysis to explore the data stack using the K-Nearest Neighbor (KNN) algorithm. The application produced in this study uses several attributes that classify data mining include: age, child, number of siblings, gender, marital status, occupation, family affected by diabetes militus, so that by applying KNN analysis a prediction can be made based on the proximity of data history there is a new data, is it predicted to be hit or not. The training data used in this study were 200 data of patients who had carried out examinations in the last 2 years. From the results of testing the authors used testing / sample data as many as 104 patient data taken outside the training data, this is so that predictive results of Diabetes Status Patients from the system designed can be compared with the results of Diabetes Patient Status in accordance with the data. The accuracy results obtained from the test were 68.30%.*

*Keywords: diabetes militus, KNN algorithm, Puskesmas Manyampa*

### 1. Pendahuluan

Diabetes mellitus atau pada masyarakat umum dikenal dengan nama penyakit gula merupakan penyakit kronis yang terjadi karena kelainan sekresi insulin pada kenaikan glukosa yang tidak teratur. Diabetes mellitus akan meningkatkan gula darah dalam tubuh sehingga terjadi penyakit komplikasi yang dapat menyebabkan beberapa risiko seperti stroke, penyakit jantung, kebutaan, gagal ginjal dan kematian [7]. Sementara itu Menurut Ahmed et al [1] penyakit diabetes tergolong penyakit yang mematikan dan meningkatkan gula darah. Munculnya penyakit disebabkan oleh masyarakat yang memiliki kebiasaan aktivitas fisik dengan diet tinggi kalori dan lemak yang tidak cukup serat pada makanan. Identifikasi penyakit diabetes diperlukan sebagai pencegahan penyakit diabetes. Dengan pemanfaatan pendekatan Data mining dapat mengekstraksi informasi yang tidak diketahui sebelumnya.

Puskesmas Manyampa adalah salah satu puskesmas yang berada di kabupaten Bulukumba. Berdasarkan obeservasi awal yang dilakukan penulis, hingga saat ini sudah memiliki pasien yang terdaftar sebanyak 2000 lebih orang, dari jumlah tersebut, penulis menggunakan data tersebut sebagai acuan untuk memprediksi pasien yang baru apakah berpotensi terkena terkena penyakit diabetes mellitus atau tidak dengan menggunakan teknik data mining.

## 2. Metode Penelitian

### 2.1. Penelitian yang relevan

Ikmalia Fernanda, Silvia, dkk membuat sistem untuk mengidentifikasi penyakit diabetes mellitus menggunakan metode Modified K-Nearest Neighbor (MKNN). Modified K-Nearest Neighbor (MKNN), Terdapat 15 gejala dan 2 tipe penyakit yang digunakan sebagai parameter dalam pengembangan sistem. Keluaran yang dihasilkan sistem berupa diagnosis tipe penyakit serta cara pengendaliannya. Berdasarkan pengujian yang telah dilakukan memperoleh hasil akurasi terbaik 93,33% dengan error rate 6,67%. [3].

Mustafa S.M & I Wayan Simpen dalam penelitiannya melakukan prediksi terhadap kemungkinan mahasiswa baru dapat menyelesaikan studi tepat waktu dengan menggunakan algoritma K-Nearest Neighbor (KNN). Aplikasi yang dihasilkan pada penelitian ini menggunakan berbagai atribut antara lain nilai ujian nasional (UN), asal sekolah/ daerah, jenis kelamin, pekerjaan dan penghasilan orang tua, jumlah bersaudara, dan lain-lain yang menjadi dasar untuk melakukan prediksi berdasarkan kedekatan histori data yang ada dengan data yang baru, apakah mahasiswa tersebut berpeluang untuk menyelesaikan studi tepat waktu atau tidak. Dari hasil pengujian dengan menerapkan algoritma KNN dan menggunakan data sampel alumni tahun wisuda 2004 s.d. 2010 untuk kasus lama dan data alumni tahun wisuda 2011 untuk kasus baru diperoleh tingkat akurasi sebesar 83,36%. [5]

Salim F, Marko dalam penelitiannya dengan judul Analisis Rekam Medis Pasien Diabetes Mellitus Melalui Implementasi Teknik Data Mining di RSUP Dr. Sardjito Yogyakarta, Mengidentifikasi karakteristik pasien Diabetes mellitus, kecenderungan dan tipe Diabetes melitus melalui penerapan teknik data mining. Penelitian ini merupakan penelitian deskriptif observasional dengan rancangan cross sectional. Teknik pengumpulan data dilakukan secara retrospektif melalui observasi dan studi dokumentasi rekam medis elektronik di RSUP Dr. Sardjito Yogyakarta. Data yang terkumpul kemudian dilakukan analisis dengan menggunakan aplikasi Weka. Hasil: Pasien Diabetes mellitus di RSUP Dr. Sardjito tahun 2011-2016 berjumlah 1.554 orang dengan tren yang cenderung menurun. Pasien paling banyak berusia 56 - 63 tahun (27,86%). Kejadian Diabetes mellitus didominasi oleh Diabetes mellitus tipe 2 dengan komplikasi tertinggi adalah hipertensi, nefropati, dan neuropati. Dengan menggunakan teknik data mining dengan algoritma decision tree J48 (akurasi 88.42%) untuk analisis rekam medis pasien telah menghasilkan beberapa rule. [6]

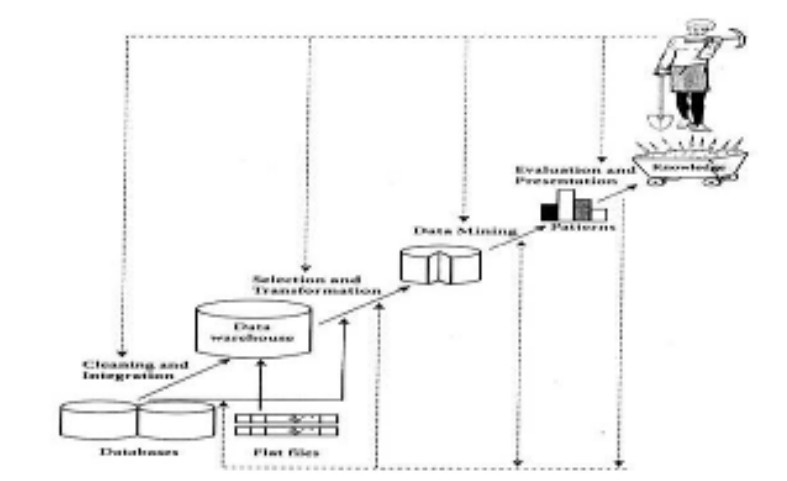
## 2.2. Landasan Teori

### 2.2.1. Pengertian data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini

Sedangkan menurut Han dan Kamber, “Data mining adalah proses menambang (*mining*) pengetahuan dari sekumpulan data yang sangat besar”. Data mining merupakan suatu langkah dalam *knowledge discovery in database* (KDD) [4].

Langkah-langkah dalam menemukan pengetahuan (*discovery knowledge*) pada Gambar 1 meliputi :



Gambar 1. Langkah-langkah dalam menemukan pengetahuan  
(Sumber : Kursini & Emha Taufiq Luthfi, (2009))

1. *Data cleaning*  
Menghilangkan noise dan data yang tidak konsisten.
2. *Data integration*  
Menggabungkan berbagai sumber data
3. *Data selection*  
Menerima data yang berhubungan dengan analisa dari database.
4. *Data transformation*  
Mengubah data ke bentuk yang sesuai untuk mining dengan melakukan agregasi atau summary.
5. *Data mining*  
Melakukan proses mining untuk mengekstrak data.
6. *Pattern evaluation*  
Mengidentifikasi pola yang menggambarkan pengetahuan (*knowledge*).
7. *Knowledge presentation*  
Menampilkan mined knowledge kepada pengguna.

### 2.2.2. Teknik data Mining

Kegunaan data mining adalah untuk mengklasifikasikan pola yang harus ditemukan dalam data mining. Secara umum, data mining dapat diklasifikasikan dalam dua kategori yaitu deskriptif dan prediktif. [2].

Adapun operasi-operasi dan teknik-teknik yang berhubungan :

1. Operasi Predictive modeling : (*classification, value prediction*)
2. Database segmentation : (*demographic clustering, neural clustering*)
3. Link Analysis : (*association discovery, sequential pattern discovery, similar timesequence discovery*)
4. Deviation detection : (*statistics, visualization*) [2].

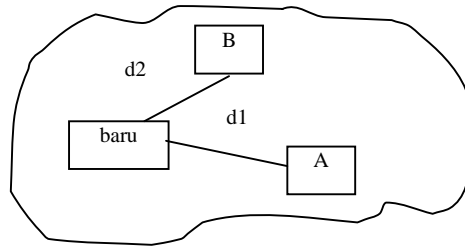
### 2.2.3. Algoritma K-Nearest Neighbor (KNN)

#### A. Pengertian Algoritma KNN

Algoritma *K-Nearest Neighbor* (KNN) adalah merupakan sebuah metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan (K) tetangga terdekatnya. KNN termasuk algoritma *supervised learning*, dimana hasil dari query instance yang baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi [4].

Nearest Neighbor adalah suatu pendekatan untuk menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada. Ilustrasi kedekatan kasus pada Gambar 2. memberikan gambaran tentang proses mencari solusi terhadap seorang pasien baru dengan menggunakan mengacu pada solusi dari pasien terdahulu. Untuk mencari kasus pasien mana yang akan digunakan, maka dihitung kedekatan anatara kasus pasien baru dengan semua

kasus pasien lama. Kasus pasien lama dengan kedekatan terbesar-lah yang akan diambil solusinya untuk digunakan pada kasus pasien baru [6].



Gambar 2. Ilustrasi kedekatan kasus  
(Sumber : Kursini & Emha Taufiq Luthfi, 2009)

Rumus untuk melakukan penghitungan kedekatan antara kedua dua kasus adalah sebagai berikut : [5]

$$\frac{\sum_{i=1}^n f(T_i, S_i) * w_i}{w_i}$$

Similarity(T,S) :  $w_i$  (1)

Dimana :

T : kasus baru

S : kasus dalam penyimpanan

n : jumlah atribut tiap kasus

i : atribut individu 1 s/d n

f : fungsi kesamaan atribut  $i$  antara kasus T dan S

w : bobot pada atribut yang ke  $i$

B. Urutan Algoritma KNN [4]

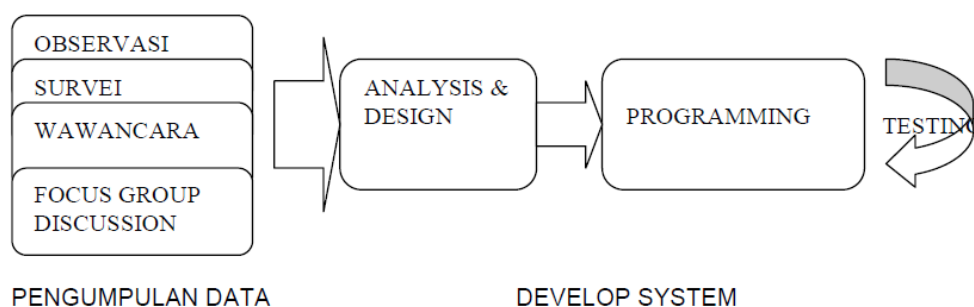
1. Menentukan parameter K (jumlah tetangga paling dekat)
2. Menghitung kuadrat jarak euclidian (*euclidean distance*) masing-masing obyek terhadap data sampel yang diberikan

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

3. Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil
4. Mengumpulkan kategori Y (klasifikasi *nearest neighbor*)
5. Dengan menggunakan kategori mayoritas, maka dapat diprediksikan nilai query instance yang telah dihitung

### 2.3. Metode Perancangan

Metodologi yang digunakan dalam penelitian ini adalah menggunakan pendekatan Objek (UML) yang diimplementasikan dengan menggunakan bahasa pemrograman PHP. Adapun tahapan pengembangan system dapat dilihat pada Gambar 3 berikut :



Gambar 3. Tahapan Pembangunan Sistem pendekatan Objek  
(Sumber : M. Shalahuddin, Rosa A.S. 2010)

Kegiatan dalam Pengumpulan Data adalah :

1. Observasi, yakni dengan melakukan pengamatan terhadap proses-proses yang sedang berlangsung/berjalan. Kegiatan ini didahului dengan identifikasi dan inventarisasi kondisi organisasi, teknologi informasi dan sumber daya manusia yang ada.
2. Survei, bertujuan untuk memperoleh gambaran kondisi kesiapan fungsional dan non fungsional akan kebutuhan sistem untuk tujuan migrasi ke sistem baru.
3. Analisa Dokumen, dilakukan dengan tujuan mempelajari arus dokumen yang menggambarkan sistem yang sedang berjalan.
4. Focus Group Discussion, yaitu diskusi ahli yang memiliki kompetensi dibidangnya guna mendapatkan masukan dan respon bagi penyempurnaan pada tahap pengumpulan data [5].

Kegiatan dalam Tahap Development Sistem adalah sebagai berikut :

1. Analisis Sistem, dilakukan berdasarkan dari hasil pada tahap pengumpulan data. Hal ini dilakukan dengan memperhatikan permasalahan yang terjadi, tujuan dibangunnya sistem informasi, identifikasi input dan output yang dibutuhkan pada laporan/tampilan informasi.
2. Desain Sistem, pada tahap ini akan dibuat rancangan sistem informasi yang detail berdasarkan kebutuhan yang diinginkan pada tahap analisis. Pada tahap ini juga dilakukan penetapan standar baku yang diutuhkan dalam perencanaan basis data.
3. Programming, yaitu proses implementasi dari tahap desain dengan menggunakan bahasa pemrograman komputer.
4. Testing, dilakukan untuk pengujian modul dan sistem keseluruhan (unit testing and sistem testing) [5].

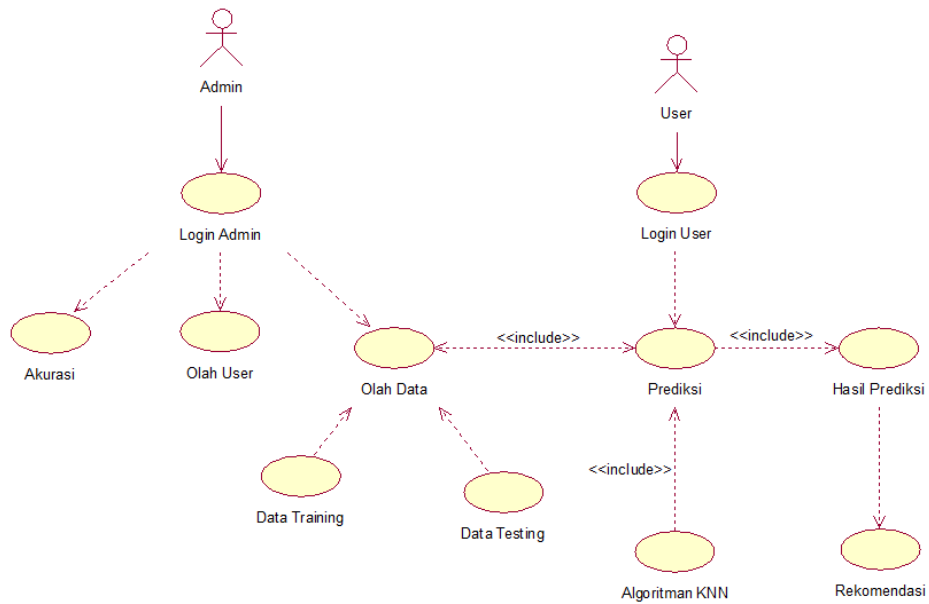
### 3. Hasil dan Pembahasan

#### 3.1. Desain Sistem Pendekatan Objek

Dalam penelitian ini, metode yang digunakan dalam perancangansistem adalah pendekatan objek UML (Univied Model Language) dengan menggunakan diagram Use Case, Activity diagram, sequence diagram dan class diagram.

##### 3.1.1. Use Case Diagram

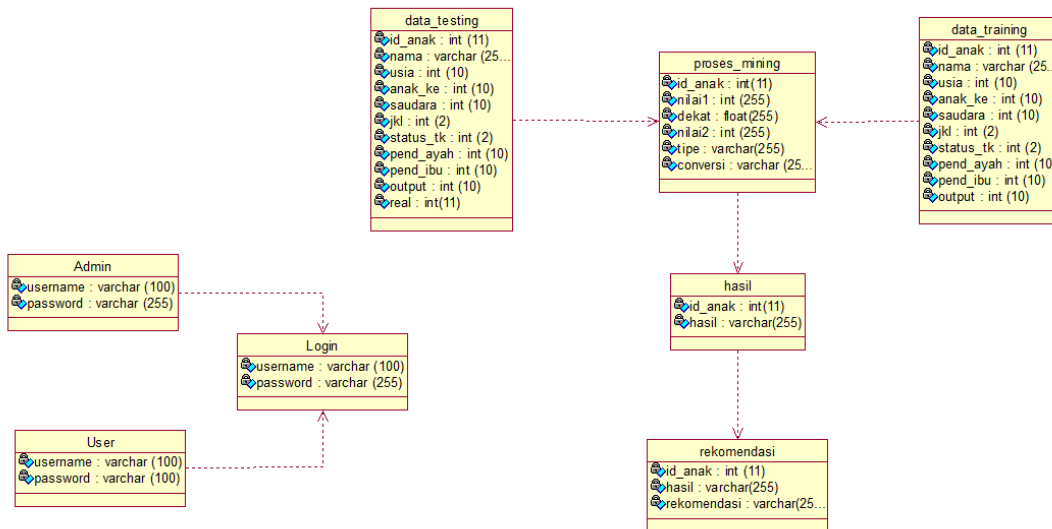
Use case diagram melibatkan 2 actor yaitu admin dan user, admin dan user harus melakukan proses login terlebih dahulu sebelum masuk dalam proses selanjutnya. Admin dapat melakukan penginputan data training dan data testing, sedang user dapat melihat hasil prediksi dengan memasukkan data pasien sebagai data kasus. Use case diagram dapat dilihat pada gambar 4.



Gambar 4. Use Case Diagram Sistem

3.1.2 Class Diagram

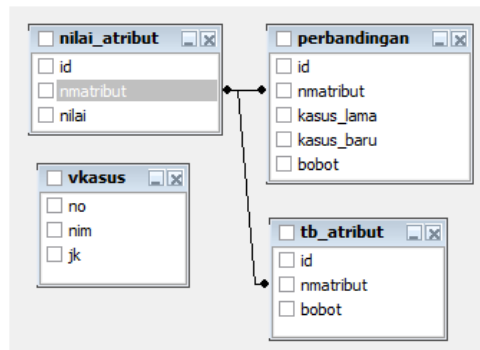
Class diagram merupakan gambaran objek yang mewakili sistem yang digunakan sebagai dasar membangun sistem dalam sisi pemrograman dan basis data yang terdiri dari beberapa class yakni login admin, login user, data training, data testing, prediksi keaddan pasien, hasil prediksi, dan rekomendasi. Class diagram dapat dilihat pada gambar 5.



Gambar 5. Class Diagram Aplikasi

**3.2. Desain Database**

Database yang digunakan dalam produk aplikasi datamining ini terdiri dari 4 tabel yaitu tabel tb\_atribut, tabel nil\_atribut, tabel perbandingan dan tabel kasus. Desain database ini dapat dilihat pada gambar 6

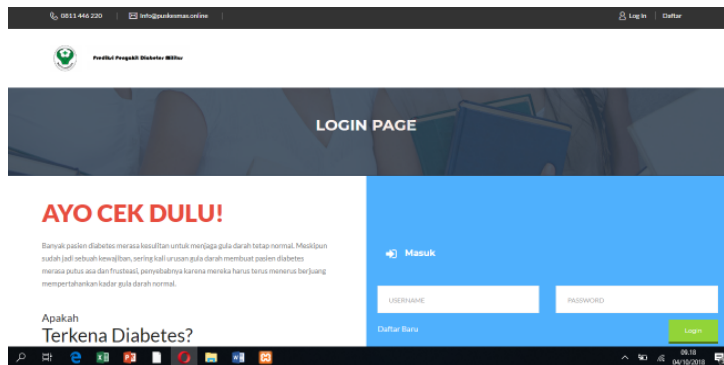


Gambar 6. Desain database

**3.3. Implementasi Sistem**

Setelah tahap desain sistem selesai dibuat, selanjutnya tahap pembuatan aplikasi dengan menggunakan bahasa pemrograman PHP. Adapun implementasi dari hasil aplikasi yang dibuat sebagai berikut :

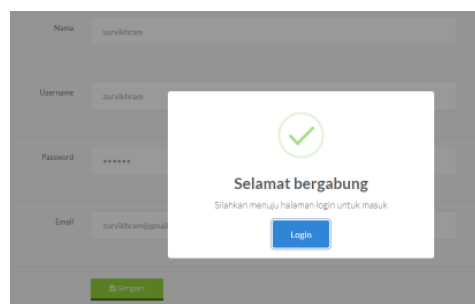
Sebelum program dijalankan, pengguna harus memasukkan login dan password terlebih dahulu. Pengguna dalam hal ini adalah Admin atau pengguna biasa. Tampilan login page dapat dilihat pada gambar 7



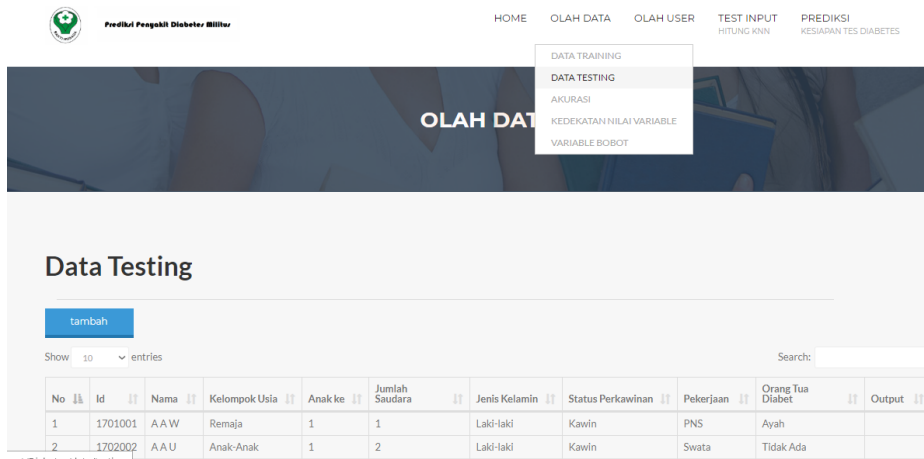
Gambar 7 Halaman Login

**3.3.1 Halaman Login Admin**

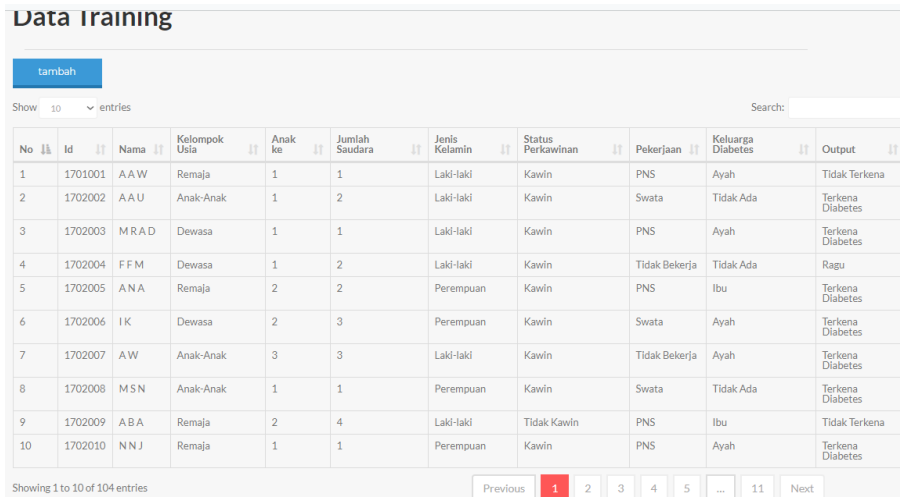
Proses yang dapat dilakukan oleh admin antara lain verifikasi pendaftaran user baru, olah data training, olah data testing, atur kedekatan dan bobot variable serta pengujian akurasi sistem, yang masing-masing dapat dilihat pada gambar 8, 9,10 dan ?



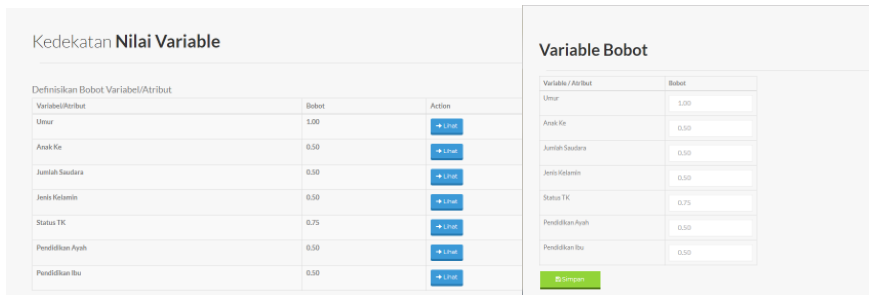
Gambar 8. verifikasi pendaftar user baru



Gambar 9. Olah Data Testing



Gambar 10. Olah Data Training



Gambar 11. Pengaturan bobot dan kedekatan variable



**OLAH DATA**

**Perhitungan Nilai Akurasi**

Jumlah Total Data Test				Jumlah Benar				Jumlah Salah				Akurasi			
104				71				33				68%			

No	Id Pasien	Nama	Kelompok Usia	Anak Ke	Jumlah Saudara	Jenis Kelamin	Status Perkawinan	Pekerjaan	Keluarga Diabetes	Prediksi	Kenyataan	Penilaian
1	1701001	A.A.W	Remaja	1	1	Laki-laki	Kawin	PNS	Ayah	Tidak Terkena	Terkena Diabetes	Salah
2	1702002	A.A.U	Anak-Anak	1	2	Laki-laki	Kawin	Swata	Tidak Ada	Terkena Diabetes	Ragu	Salah
3	1702003	M.R.A.D	Dewasa	1	1	Laki-laki	Kawin	PNS	Ayah	Terkena Diabetes	Terkena Diabetes	Benar

Gambar 12. Hasil perhitungan nilai akurasi KNN

**3.3.2 Halaman Login User/Pasien**

Proses yang dapat dilakukan oleh user/pasien adalah melakukan prediksi diabetes dengan terlebih dahulu mengisi data pasien yang terdiri dari : Nama, kelompok usia, anak ke, jumlah saudara, jenis kelamin, keluarga yang diabetes. Hasil prediksi yang diperoleh dapat dilihat pada gambar 13.

Prediksi Penyakit Diabetes Milliter
HOME PREDIKSI KESIAPAN TES DIABETES

**PREDIKSI**

**Data Pasien**

Daftar Baru

Nama	Anak ke	Jenis Kelamin	Output	Aksi
Syukri	2	3	<input checked="" type="checkbox"/> Terkena Diabetes	<span style="background-color: #f1c40f; padding: 5px 10px; border-radius: 3px;">History</span> <span style="background-color: #3498db; padding: 5px 10px; border-radius: 3px; margin-left: 5px;">Ulang</span>

Gambar 13. Hasil Prediksi pasien.

**3.4. Pengujian Sistem**

Dalam menguji akurasi dan ketepatan hasil pengklasifikasian sistem ini, digunakan 104 data arsip yang diambil. 104 data tersebut tidak terdapat di dalam data training, hal ini dimaksudkan agar hasil prediksi Status Diabetes Pasien dari sistem yang dirancang dapat dibandingkan dengan hasil Status Diabetes Pasien yang sesuai dengan data.

$$Akurasi = \frac{\text{jumlah data} - \text{data salah}}{\text{jumlah data}} \times 100\%$$

$$Akurasi = \frac{104 - 33}{104} \times 100\%$$

$$= 68.3 \%$$

Hasil pengujian menunjukkan besar 68.3 %. Detail perbandingannya dapat dilihat pada gambar 14 dibawah ini.

Perhitungan Nilai Akurasi												
Jumlah Total Data Test					Jumlah Benar			Jumlah Salah		Akurasi		
104					71			33		68%		
No	Id Pasien	Nama	Kelompok Usia	Anak Ke	Jumlah Saudara	Jenis Kelamin	Status Perkawinan	Pekerjaan	Keluarga Diabetes	Prediksi	Kenyataan	Penilaian
1	2	3	4	5	6	7	8	9	10	11	12	13
1	1701001	A A W	Remaja	1	1	Laki-laki	Kawin	PNS	Ayah	Tidak Terkena	Tidak Terkena	Benar
2	1702002	A A U	Anak-Anak	1	2	Laki-laki	Kawin	Swata	Tidak Ada	Terkena Diabetes	Ragu	Salah
3	1702003	M R A D	Dewasa	1	1	Laki-laki	Kawin	PNS	Ayah	Terkena Diabetes	Terkena Diabetes	Benar
4	1702004	F F M	Dewasa	1	2	Laki-laki	Kawin	Tidak Bekerja	Tidak Ada	Ragu	Terkena Diabetes	Salah
5	1702005	A N A	Remaja	2	2	Perempuan	Kawin	PNS	Ibu	Terkena Diabetes	Terkena Diabetes	Benar
6	1702006	I K	Dewasa	2	3	Perempuan	Kawin	Swata	Ayah	Terkena Diabetes	Terkena Diabetes	Benar
7	1702007	A W	Anak-Anak	3	3	Laki-laki	Kawin	Tidak Bekerja	Ayah	Terkena Diabetes	Terkena Diabetes	Benar
8	1702008	M S N	Anak-Anak	1	1	Perempuan	Kawin	Swata	Tidak Ada	Terkena Diabetes	Tidak Terkena	Salah
9	1702009	A B A	Remaja	2	4	Laki-laki	Tidak Kawin	PNS	Ibu	Tidak Terkena	Tidak Terkena	Benar
10	1702010	N N J	Remaja	1	1	Perempuan	Kawin	PNS	Ayah	Terkena Diabetes	Terkena Diabetes	Benar

Gambar 14. Hasil Prediksi pasien.

#### 4. Kesimpulan

Aplikasi data *mining* ini dapat memprediksikan Status Diabetes Pasien yang dibagi menjadi tiga kategori usia yaitu anak-anak, remaja dan dewasa, dengan hasil prediksi yaitu terkena diabetes, tidak terkena diabetes dan ragu-ragu dengan menggunakan algoritma *K-Nearest Neighbor*.

Hasil pengujian data dari sistem ini yang menggunakan 104 data pasien pada puskesmas Manyampa memperoleh hasil prediksi yang benar sebanyak 71 dan salah atau ragu-ragu sebesar 33 dengan tingkat akurasi sebesar 68.3%.

#### Referensi

- [1] Ahmed, K., Jesmin, T., & Fatima, U. 2012. "Intelligent and Effective Diabetes Risk Prediction System Using Data Mining. *Oriental Journal of Computer Science & Technology*", 5(2), 215–221.
- [2] Ian H. Witten, Frank Eibe, Mark A. Hall, 2011. "Data mining: Practical Machine Learning Tools and Techniques" 3rd Edition, Elsevier.
- [3] Ikmalia Fernanda, Silvia & Eka Ratnawati, Dian & Adikara, Putra, 2017. "Identifikasi Penyakit Diabetes Mellitus Menggunakan Metode Modified K-Nearest Neighbor (MKNN)". *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* Vol. 1, No. 6, Juni 2017, hlm. 507-513
- [4] Kursini & Emha Taufiq Luthfi, 2009. "Algoritma data Mining", ANDI, Yogyakarta
- [5] Mustafa S.M & I Wayan Simpen. 2014. "Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus: Data Akademik Mahasiswa STMIK Dipanegara Makassar)". Makassar: *Jurnal Citek*. Vol. 1, No. 4.
- [6] Salim F, Marko Sugeng, 2017. "Analisis Rekam Medis Pasien Diabetes Mellitus Melalui Implementasi Teknik Data Mining di RSUP Dr. Sardjito Yogyakarta", *Jurnal Kesehatan Vokasional* Vol 2, No 2, <https://doi.org/10.22146/jkesvo.30331>
- [7] V, V. V., & Ravikumar, A. 2014. "Study of Data Mining Algorithms for Prediction and Diagnosis of Diabetes Mellitus". *International Journal of Computer Application*, 95(17), 12–16.
- [8] M. Shalahuddin, Rosa A.S. 2010. "Pemrograman Berorientasi Objek"., Modula: Bandung.