

ABSTRAK

Diabetes merupakan salah satu penyebab utama kematian, kecacatan dan kerugian ekonomi di seluruh dunia. Tipe 2 diabetes lebih umum (90-95% di seluruh dunia) jenis diabetes. Prediksi dan diagnosis penyakit yang akurat akan menjadi masalah penelitian yang menantang bagi banyak peneliti. Oleh karena itu bertujuan untuk meningkatkan akurasi diagnosis yang kita miliki, mengusulkan pengklasifikasi K-nearest neighbor. Metode KNN dipilih karena kemampuannya yang sederhana namun efektif dalam melakukan klasifikasi berdasarkan atribut-atribut yang ada. Dalam penelitian ini akan dilakukan pengumpulan data penyakit diabetes mellitus yang terdiri dari atribut diabetes yang relevan, seperti age, pregnancies, glucose, blood pressure, skin thickness, insulin, BMI (Body Mass Index), DiabetesPedigreeFunction, yang digunakan. Kemudian, akan dilakukan tahap pra-pemrosesan data untuk membersihkan dan mengubah data menjadi bentuk yang sesuai dengan metode KNN. Dataset yang digunakan pada penelitian ini adalah dataset Diabetes Mellitus yang diambil dari website Kaggle Repository yang terdiri dari 767 data, 9 fitur dan binary class. Berdasarkan uji coba yang telah dilakukan dalam penelitian ini metode KNN dalam mengklasifikasikan penyakit diabetes mellitus dari pembagian data 9 : 1 mendapatkan nilai evaluasi terbaik dengan nilai presisi nilai presisi 0.728, recall 0.734, f1score 0.729, dan nilai K = 9.

Kata Kunci : *Data mining, K-Nearest Neighbor, Penyakit Diabetes Mellitus, Pearson Correlation*

DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
HALAMAN JUDUL.....	iii
HALAMAN PERNYATAAN ORISINALITAS.....	iv
KATA PENGANTAR	v
ABSTRAK	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	2
1.2.1Permasalahan	2
1.2.2Metode Usulan.....	3
1.2.3Pertanyaan Peneliti	3
1.3 Tujuan dan Manfaat.....	3
1.3.1Tujuan Peneliti.....	3
1.3.2 Manfaat Penelitian	3
1.4 Batasan – Batasan.....	3
1.5 Sistematika Proposal	4
BAB II KAJIAN PUSTAKA	5
2.1. Data Mining.....	5
2.2. Penyakit Diabetes	5
2.3. Normalisasi <i>Min-Max</i>	6
2.4. Pearson Correlation	6
2.5. KNN (Modified K-Nearest Neighbor)	7
2.6. Evaluasi	8

2.6.1Metode Evaluasi	8
2.6.2Matrix Evaluasi.....	9
2.7. Orange Data Mining.....	10
2.8. Penelitian Terkait	10
BAB III METODE USULAN.....	16
3.1 Metodologi Penelitian	16
3.2 Arsitektur Sistem.....	17
3.3 Flowchat Sistem	18
3.4 Dataset	19
3.5 Jadwal Penelitian.....	20
BAB IV HASIL DAN PEMBAHASAN.....	22
4.1 Lingkungan Uji Coba	22
4.2 Tampilan Sistem.....	22
4.3 Tahapan Sistem	23
4.3.1 Input Dataset	23
4.3.2 Data Diabetes.....	24
4.3.3 Fitur Statistik.....	24
4.3.4 Impute	25
4.3.5 Select Kolom.....	27
4.3.6 Tahapan <i>Preprocessing</i>	29
4.3.7 Data Min-Max.....	30
4.3.8 Seleksi Fitur Pearson Correlation	30
4.3.9 Scatter Plot.....	31
4.3.10 Pembagian Dataset.....	32
4.3.11 Tahapan Klasifikasi	35
4.3.12 Tahapan Evaluasi.....	36
4.4 Skenario Pengujian.....	37
4.5 Hasil Pengujian.....	39
4.5.1 Pembagian Data	39

4.5. Correlation	40
BAB V KESIMPULAN DAN SARAN.....	41
5.1 KESIMPULAN	41
5.2 SARAN	41
REFERENSI	42

DAFTAR GAMBAR

Gambar 2. 1 Metode evaluasi K-Fold Cross Validation	9
Gambar 3. 1 Arsitektur Sistem.....	17
Gambar 3. 2 <i>Flowchat</i> Sistem.....	18
Gambar 4. 1 Tampilan Sistem	20
Gambar 4. 2 Input Dataset	21
Gambar 4. 3 Proses Input	21
Gambar 4. 4 Dataset Diabetes	22
Gambar 4. 5 Tampilan dataset	22
Gambar 4. 6 Fitur Statistik	25
Gambar 4. 7 Tampilan Statistik Dataset	25
Gambar 4. 8 Impute	26
Gambar 4. 9 Proses Impute	27
Gambar 4. 10 Select kolom	28
Gambar 4. 11 Proses Target	28
Gambar 4. 12 Preprocessing	29
Gambar 4. 13 Normalisasi Fitur	29
Gambar 4. 14 min-max dataset	30
Gambar 4. 15 Outcome min-max	30
Gambar 4. 16 Pearson Correlation	31
Gambar 4. 17 Outcome Pearson Correlation	31
Gambar 4. 18 Scatter Plot	32
Gambar 4. 19 Hasil Scatter Plot	32
Gambar 4. 20 Data Sampler	33
Gambar 4. 21 Garis Testing	33
Gambar 4. 22 Data Testing	34
Gambar 4. 23 Garis Training	34
Gambar 4. 24 Data Training	35
Gambar 4. 25 KNN	35
Gambar 4. 26 Proses KNN	36
Gambar 4. 27 Test dan Score	36
Gambar 4. 28 Proses Evaluasi	37

Gambar 4. 29 Evaluasi Correlation	40
---	----

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	9
Tabel 2. 2 Penelitian terkait	12
Tabel 3. 1 Keterangan Dataset.....	19
Tabel 3. 2 Jadwal Penelitian	20
Tabel 4. 1 Spesifikasi Perangkat Keras dan Perangkat Lunak	22
Tabel 4. 2 Hasil Normalisasi Min-Max	37
Tabel 4. 3 Korelasi fitur Pearson Correlation.....	38
Tabel 4. 4 Pembagian Data.....	38
Tabel 4. 5 Hasil	39