

SISTEM APLIKASI PREDIKSI PENYAKIT DIABETES MENGGUNAKAN *FUTURE SELECTION* KORELASI PEARSON DAN KLASIFIKASI NAÏVE BAYES

¹Mohammad Burhan Hanif, ²Khoirudin

^{1,2} Program Studi Teknik Informatika, Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang,
¹hanifburhan@usm.ac.id, ²khoirudin@usm.ac.id

Abstrak

Diabetes penyakit serius yang terkenal didunia dengan sebutan *silent killer* yang tercatat semakin meningkat dari tahun 1980 sampai 2014 sebanyak 422 juta jiwa. Hal ini perlu di perhatikan secara serius karena berimbas pada timbulnya beban kerja sumberdaya medis yang berlebihan serta tentunya beban keuangan yang akan timbul karena hal tersebut. Pada era teknologi maju saat ini data mining pada bidang kesehatan hadir untuk memberikaran analisa data khususnya data penyakit diabetes dengan tepat dan akurat. Klasifikasi data mining dipadukan dengan metode *research and development* dapat digunakan sebagai sebuah sistem aplikasi untuk memprediksi penyakit diabetes. Oleh sebab itu peneliti membuat sebuah sistem deteksi penyakit diabetes menggunakan *future selection* korelasi pearson dan klasifikasi naïve bayes yang diharapkan dapat membantu ahli medis dalam deteksi penyakit diabetes secara lebih cepat dan mengurangi beban biaya yang timbul akibat masalah ini. Sistem prediksi berbasis web akan menangani data diabetes berjumlah 768 baris data dengan menampilkan 10 *feature* yaitu Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age dan Outcome yang diperoleh pada Pima Indians Diabetes kaggle dataset. Penggunaan algoritma korelasi pearson dibutuhkan untuk meningkatkan performa dari algoritma naïve bayes dengan nilai akurasi dari 68,2% menjadi 79,13% pada data penyakit diabetes.

Keywords: Data mining, Diabetes, Naïve bayes, future selection

1. Pendahuluan

Diabetes penyakit yg tidak menampakan gejala yang membuat penderita terlena karena langsung merusak bagian vital manusia seperti ginjal, mata dan saraf manusia (Kumar & Abedin, 2017).

Di Indonesia penderita terus tumbuh dengan sangat subur di lihat dari data tahun 2007 hingga tahun 2013 terlihat peningkatan sebesar 2.1% dan terus meningkat hingga saat ini (Kesehatan, 2014). Tak hanya di Indonesia di luar negeri pun terjadi peningkatan yang serupa, lembaga kesehatan dunia mencatat dari tahun 1980 penderita hanya 108 jiwa namun di tahun 2014 meledak menjadi 422 jiwa (Chan, 2013).

Penyakit diabetes ini merupakan sumber utama penyakit jantung, serta menjadi salah satu momok kematian ibu saat melahirkan. Selain itu juga menjadi faktor penularan penyakit bagi bayi yang telah dilahirkannya (Colagiuri et al., 2011).

Hal ini sungguh sangat memprihatinkan jika penyakit ini tidak di tangani secara serius. Karena melihat dari akibat yang di timbulkan menjadi penyakit komplikasi parah seperti kebutaan, gagal ginjal, pecah pembuluh darah dan gangguan saraf (Mansourypoor & Asadi, 2017).

Selain akibat yang timbul pada diri manusia akibat lain juga akan terjadi seperti tingkat kenaikan

pelayanan kesehatan dan tentunya menjadi beban tambahan tenaga medis yang saat ini perkembangannya tidak sebanding dengan banyaknya masalah yang terjadi (Rigla et al., 2017). Efek karambol terjadi juga pada beban anggaran dana yang tentunya akan ikut meningkat pula (Kumar & Abedin, 2017).

Oleh sebab beban yang ditimbulkan cukup banyak maka penyakit diabetes penting untuk diteliti. Penyakit diabetes dalam data mining telah banyak di teliti, berikut beberapa penelitian relevan dengan topik ini diantaranya adalah

Penelitian yang dilakukan oleh Dewan Farid, Li Zhang, Chowdhury Mofizur, M A Hossain dan Rebecca Strachan. Penelitian ini menceritakan tentang masalah naïve bayes ketika menghadapi kondisi kelas independence sehingga mengalami gangguan ketika melakukan komputasi. Untuk mengatasinya digunakan algoritma decision tree pada tahap preprosesing. Hasil yang didapatkan adalah peningkatan akurasi dari 76.30% menjadi 79.55% (Farid et al., 2014).

Peneliti selanjutnya dilakukan oleh Yue Huang, Paul Mccullagh, Norman Black dan Roy Harper. Penelitian ini mengurai tentang masalah faktor – faktor utama penyakit yang masih secara subjektif di kemukakan oleh dokter sehingga belum

ada susunan baku tentang faktor utama penyakit diabetes.

Untuk mengatasinya maka digunakan teknik data mining dengan jalan menentukan faktor – faktor penting penyebab penyakit diabetes. Teknik data mining yang digunakan berupa metode feature selection via supervised model construction untuk mengurai feature penting dalam penyakit diabetes. Setelahnya di kombinasikan dengan klasifikasi naïve bayes sehingga didapati akurasi terbaik mencapai 95% (Huang et al., 2007).

Peneliti selanjutnya yaitu Tao Zheng, Wei Xie, Liling Xu, Xiaoying He, Ya Zhang, Mingrong You, Gong Yang, You Chen. Penelitian ini menerangkan tentang penggunaan metode machine learning yang memiliki gangguan low recall rate karena masih menggunakan data dengan filter biasa yang konservatif. Untuk mengatasinya digunakan metode feature selection yang cocok dikombinasikan dengan metode naïve bayes sehingga terjadi peningkatan nilai AUC dari 0.71 menjadi 0.98 (Zheng et al., 2016).

Peneliti selanjutnya ditulis oleh Nongyao Nai-arun, Rungruttikarn Moungrmai. Penilitan ini mengulas tentang penanganan penyakit diabetes menggunakan metode naïve bayes. Akan tetapi metode naïve bayes ini perlu ditingkatkan untuk mendapatkan nilai performa yang lebih baik. Dalam penelitian ini peningkatan performa dilakukan menggunakan metode boosting. Hasil dari peningkatan naïve bayes yang di optimasi menggunakan metode boosting didapati nilai akurasi yang meningkat dari 81.010% menjadi 81.019% (Nai-arun & Moungrmai, 2015).

Peneliti lain tentang penyakit diabetes ditulis oleh Liangxiao Jiang, Zhihua Cai, Harry Zhang, dan Dianhong Wang. Penelitian ini membahas tentang metode naïve bayes yang memiliki dasar metode bayesian network. Oleh sebab itu kinerjanya akan menurun ketika menghadapi dependensi feature yang kompleks. Untuk mengatasi hal itu digunakan pendekatan feature selection secara acak dengan nama Randomly Selected Naïve Bayes. Hasil yang didapat berupa peningkatan nilai akurasi dari 75.8094 menjadi 76.4604 (Jiang et al., 2012).

Penyakit diabetes sangat mempengaruhi manusia tak hanya di Indonesia namun secara

mendunia (Cheruku et al., 2017). oleh karena itu prediksi penyakit ini perlu dilakukan secara akurat untuk mencegah terjadinya komplikasi parah (Mansourypoor & Asadi, 2017). Serta bisa membantu tenaga medis dalam prediksi awal penyakit (Rigla et al., 2017).

Prediksi penyakit diabetes bisa dilihat dari data rekam medis pasien, maka dari itu penulis berkeinginan untuk membuat sistem prediksi penyakit diabetes secara mandiri melalui sebuah sistem aplikasi. sistem aplikasi ini akan di padukan dengan pengamatan data yaitu dengan ilmu data mining menggunakan algoritma naïve bayes dan algoritma korelasi pearson sebagai optimasi dalam tahap feature selection.

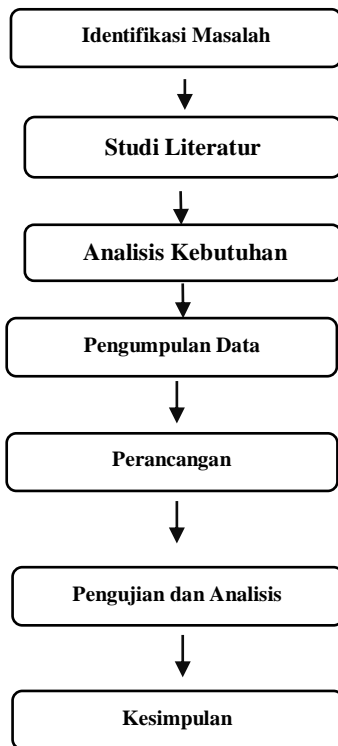
2. Metode Penelitian

Metode yang digunakan dalam penelitian ini menggunakan beberapa pendekatan yang mencakup model Research and development dengan menggabungkan model Knowledge Discovery in Database (KDD) dan Prototype Model..

Knowledge Discovery in Database (KDD) digunakan untuk mengurai suatu pola dalam data dengan bantuan perhitungan algoritma secara spesifik. Sedangkan untuk prototype model digunakan sebagai jalan yang menghubungkan antara perancangan kerja sistem perangkat lunak dan penggunaan kerja sistem secara menyeluruh.

Dalam sistem informasi perangkat lunak pelanggan seringkali mendefinisikan sebuah sistem itu secara umum tanpa menjelaskan kebutuhan input, proses, output. Oleh sebab itu dalam penelitian ini metode yang digunakan dalam perancangan perangkat lunak hanya menyediakan kebutuhan umum perangkat lunak tanpa melihat detail input, proses, output untuk menciptakan efisiensi kerja algoritma yang akan digunakan pada sistem informasi ini.

Untuk melakukan metode tersebut maka penelitian ini akan menggunakan kerangka pemikiran sebagai berikut :



Gambar 1 Kerangka Pemikiran

Identifikasi Masalah menjelaskan bahwa analisa mengenai permasalahan yang timbul tentang penyakit diabetes dan kenapa perlunya dilakukan prediksi penyakit diabetes menggunakan suatu sistem perangkat lunak.

Studi Literatur menerangkan tentang penelitian yang sudah dilakukan sebelum penelitian ini dilaksanakan. Tujuannya adalah sebagai dasar teori referensi yang terkait secara relevan dengan permasalahan yang terjadi.

Analisis kebutuhan yang didalamnya merupakan tahapan untuk melakukan identifikasi kebutuhan apa saja yang diperlukan untuk membuat sistem perangkat lunak ini.

Pengumpulan data diambil dari data public Pima Indians Diabetes Dataset kaggle. Format data berbentuk excel terdiri dari berjumlah 768 baris data dengan menampilkan 10 fitur yaitu Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age dan Outcome.

Perancangan diawali dengan tahapan preresesing data baik secara manual menggunakan Microsoft excel atau menggunakan tools rapidminer. Preprocessing data dilakukan dengan jalan transformasi data, cleansing data dan fitur selection data. Tahap selanjutnya membagi data menjadi dua yaitu data training dan data testing.

Pengujian serta analisis dikerjakan melalui hasil implementasi yang sudah dilaksanakan dilihat

dari nilai confusion matrix serta nilai AUC melalui proses klasifikasi yang dianalisa dari kelas true false positive dan negative. Sedangkan untuk pengujian sistem perangkat lunak digunakan metode blackbox validation sebagai dasar pengujian yang merupakan syarat fungsional sistem sebagai kualitas dashboard sistem perangkat lunak yang dihasilkan.

Kesimpulan serta saran penelitian selanjutnya akan ditulis pada tahap ini supaya bisa diteruskan dan dikembangkan jauh menjadi lebih baik lagi.

3. Analisis dan Hasil

Pembahasan dan analisis hasil penelitian yang dilakukan melalui beberapa tahapan sesuai dengan kerangka pemikiran yang telah diuraikan. Dimulai dari penjelasan perancangan diikuti dengan pengolahan serta hasil implementasi model dengan di ikuti pengujian dan analisis dari pengujian dan model yang telah dikerjakan.

Pengumpulan Data

Data yang digunakan pada penelitian kali ini merupakan data untuk memprediksi penyakit diabetes. Data diambil dari data Pima Indians Diabetes Dataset kaggle. Format data berbentuk excel terdiri dari berjumlah 768 baris data dengan menampilkan 10 fitur yaitu Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age dan Outcome seperti data table dibawah ini.

Tabel 1 Atribut Master Data Diabetes

NO	Nama Atribut	Keterangan
1	Pregnancies	berapakah hamil
2	Glucose	Konsentrasi plasma glukosa dalam 2 jam tes toleransi oral glukosa
3	BloodPressure	Tekanan darah diastolik (mm Hg)
4	SkinThickness	Ketebalan lipatan kulit trisep (mm)
5	Insulin	Insulin serum 2 Jam (mu U/ml)
6	BMI	Indeks massa tubuh (berat dalam kg/(tinggi dalam m) ²)
7	DiabetesPedigreeFunction	Fungsi silsilah diabetes
8	Age	Umur satuan tahun
9	Outcome	Variabel kelas (0 atau 1) 268 dari 768 adalah 1 penderita, yang lain adalah 0

Preprocessing Data

Data yang diperoleh adalah data penyakit diabetes sebanyak 768 baris data dengan 9 atribut pendukungnya. Untuk mendapat data yang

berkualitas sehingga dapat mendukung penghitungan algoritma yang berpengaruh pada akurasi ketepatan sistem perangkat lunak maka peneliti melakukan tahapan cleaning, selection, reduce dan transformation.

Data cleaning

Data cleaning pada penelitian ini merupakan sebuah proses untuk membersihkan data dari missing value dan menghilangkan data – data yang tidak lengkap. Pada penelitian ini missing value dilakukan secara manual yaitu dengan replace missing value menggunakan bantuan Microsoft Excel yaitu mengubah data record blank menjadi nilai 0 .

Selection Data dan reduce data

Selection Data pada penelitian ini merupakan sebuah proses reduce data yaitu menghilangkan attribute data yang tidak penting. Menghilangkan attribute data ini menggunakan perhitungan algoritma fitur selection Korelasi Pearson. Pada fase ini akan dihitung attribute mana yang penting dan bisa digunakan.. sedangkan yang menurut perhitungan tidak penting akan di reduce atau dihilangkan.

Data Training dan Testing

Data yang tersaji sebanyak 768 data akan dibagi menjadi data latih dan data uji. Pembagian data latih dan data uji akan menggunakan bantuan Ms. Excel yang di pilih secara random yang menghasilkan perbandingan 80:20 seperti table dibawah ini.

Tabel 2 Data Training dan Testing

Pembagian	Presentase	Total
Data Latih	80%	614
Data Uji	20%	154
Total	100%	768

Dari table 2 didapati presentase 2 kelompok yaitu kelompok data latih sebanyak 80% serta data uji sebanyak 20%. Dari keseluruhan data terdapat 268 data sebagai penderita diabetes sedangkan 500 data dinyatakan normal. Setelah melakukan perhitungan pasien penderita dan pasien normal secara keseluruhan maka selanjutnya dilakukan penghitungan pada data latih sebanyak 80% pasien penderita sebanyak 213 data sedangkan pasien normal sebanyak 401 data. Sedangkan untuk data uji 20% didapati data pasien penderita diabetes sebanyak 55 data dan pasien normal sebanyak 99

data. Untuk lebih jelasnya tersaji dalam tabel berikut ini

Tabel 3 Data Training dan Testing

	Data Uji	Data Latih	Total
Penderita	55	213	268
Normal	99	401	500
Total	154	614	768

Setelah melakukan pembagian data latih dan uji maka data sudah siap untuk dilakukan penghitungan algoritma naïve bayes.

Hasil Pembuatan Model

Pada tahap ini akan dijelaskan mengenai prediksi penyakit diabetes menggunakan algoritma naïve bayes dan algoritma naïve bayes + korelasi pearson.

a. Penyusunan Model Naïve Bayes

Perhitungan Model manual Naïve Bayes menggunakan sample dataset adapun dengan langkah-langkah algoritma sebagai berikut

1. Siapkan data
2. Hitung jumlah probabilitas
 - a. Hitung nilai mean dan standar deviasi dari setiap parameter
 - b. Hitung nilai probabilitistik
3. Hasil nilai mean, standar deviasi dan probabilitas.

b. Penyusunan Model Naïve Bayes + Korelasi Pearson

Perhitungan Model manual Naïve Bayes + Korelasi Pearson menggunakan sample dataset adapun dengan langkah-langkah algoritma sebagai berikut

1. Siapkan data
2. Seleksi data menggunakan Korelasi pearson
3. Penggunaan algoritma naïve bayes dengan menghitung jumlah probabilitas
 - a. Hitung nilai mean dan standar deviasi dari setiap parameter
 - b. Hitung nilai probabilitistik
4. Hasil nilai mean, standar deviasi dan probabilitas.

Hasil Komparasi

Untuk mengukur kinerja dari metode yang di pakai yaitu metode naïve bayes dengan naïve bayes + korelasi pearson maka digunakan pengukuran confusion matrix yang akan menunjukkan seberapa besar tingkat akurasi yang

telah di capai pada ke dua metode algoritma tersebut. Untuk melihat hasil dari confusion matrix metode naïve bayes akan di sajikan melalui tabel dibawah ini.

Tabel 4 confusion matrix naïve bayes

		Actual Class	
		+ (positif)	- (negative)
Prediction class	+ (positif)	41	27
	- (negative)	46	116

Sedangkan hasil dari confusion matrix metode naïve bayes + Korelasi Pearson akan di sajikan melalui tabel dibawah ini

Tabel 5 confusion matrix naïve bayes + korelasi pearson

		Actual Class	
		+ (positif)	- (negative)
Prediction class	+ (positif)	51	20
	- (negative)	28	131

Accuracy

Accuracy menjelaskan capain nilai dari model yang bisa mengklasifikasi dengan benar. Accuracy merupakan rasio prediksi benar dengan keseluruhan data. Berikut ini nilai akurasi pada kedua metode diatas

a. Accuracy metode naïve bayes

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \\
 &= \frac{41 + 116}{41 + 116 + 27 + 46} * 100 \\
 &= 68.26
 \end{aligned}$$

b. Accuracy metode naïve bayes + Korelasi Pearson

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \\
 &= \frac{51 + 131}{51 + 131 + 20 + 28} * 100 \\
 &= 79.13
 \end{aligned}$$

Precision

Precision menjelaskan tingkat keakuratan data yang diminta dengan data prediksi yang dihasilkan oleh model. Diukur dari perbandingan rasio benar positive dengan prediksi positive. Berikut nilai precision kedua metode

a. Precision metode naïve bayes

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP} \\
 &= \frac{41}{41 + 27} \\
 &= 0.6029
 \end{aligned}$$

b. Precision metode naïve bayes + Korelasi Pearson

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP} \\
 &= \frac{51}{51 + 20} \\
 &= 0.7183
 \end{aligned}$$

Recall

Recall menjelaskan tingkat keberhasilan perhitungan model untuk menemukan sebuah informasi kembali. Berikut nilai precision kedua metode

a. Recall metode naïve bayes

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} \\
 &= \frac{41}{41 + 46} \\
 &= 0.4712
 \end{aligned}$$

b. Recall metode naïve bayes + Korelasi Pearson

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} \\
 &= \frac{51}{51 + 28} \\
 &= 0.6455
 \end{aligned}$$

AUC

AUC (Area Under Curve) adalah ukuran numerik untuk membedakan kinerja model. Dimana mempermudah perbandingan antara model klasifikasi satu dengan model klasifikasi yang lain ketika tidak ada curve ROC yang mendominasi. Berikut nilai AUC kedua metode

a. AUC metode naïve bayes

$$\begin{aligned}
 \text{AUC} &= \frac{1 + TPrate - FPrate}{2} \\
 &= \frac{1 + 0.4712 - 0.1888}{2} \\
 &= 0.6412
 \end{aligned}$$

b. AUC metode naïve bayes + Korelasi Pearson

$$\begin{aligned}
 AUC &= \frac{1 + TPrate - FPrate}{2} \\
 &= \frac{1 + 0.6455 - 0.1324}{2} \\
 &= 0.75655
 \end{aligned}$$

Dari hasil pengujian confision matrix maka dapat dilihat bahwa metode algoritma naïve bayes dan korelasi pearson lebih baik jika dibandingkan dengan metode algoritma naïve bayes saja. Hal ini dapat dilihat dari nilai accuracy, precission, recall dan AUC pada tabel berikut ini.

Tabel 6 Hasil Perbandingan Performa Algoritma Naïve Bayes Dengan Naïve Bayes + Korelasi Pearson

	Accuracy	Precission	Recall	AUC
Naïve Bayes	68.26	0.60	0.47	0.64
Naïve Bayes + Korelasi Pearson	79.13	0.71	0.64	0.75

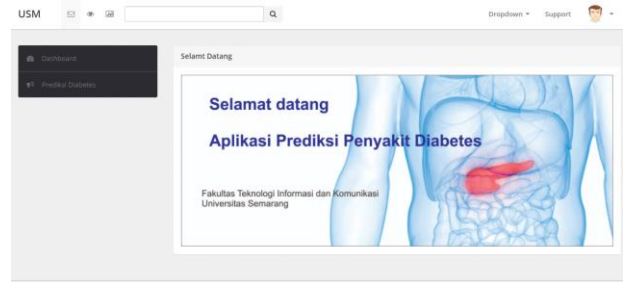
Penerapan Aplikasi Data Mining

Untuk menghasilkan aplikasi prototype yang bisa memprediksi penyakit diabetes agar dapat membantu ahli medis dalam penanganan awal penyakit. Maka alur yang dihasilkan adalah pembuatan aplikasi prototype berbasis website dengan menggunakan metode algoritma naïve bayes dan algoritma korelasi pearson.

Untuk prototype berbasis website akan menggunakan bahasa pemrograman PHP dan database menggunakan mysql. Adapun tampilan homepage hingga screen prediksi dapat di lihat melalui gambar berikut :

a. Tampilan Home Page

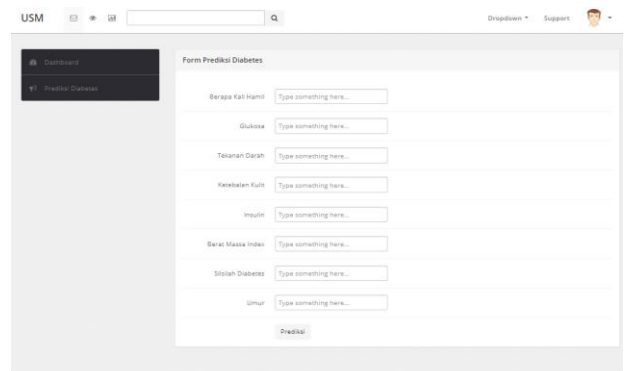
Pada Halaman Home Page user akan diperlihatkan sebuah tampilan berupa sambutan selamat datang aplikasi yang disertai logo sedang kan gambar di samping akan di perlihatkan menu dari aplikasi yaitu menu prediksi penyakit diabetes



Gambar 2 Home Page sistem

b. Tampilan Prediksi

Pada Menu Halaman ini akan di tampilkan menu untuk prediksi penyakit diabetes menggunakan data baru. Sehingga user akan menginputkan sebuah data yang nantinya akan keluar hasil prediknya terkena penyakit diabetes atau tidak.



Gambar 3 Menu Prediksi Penyakit diabetes

Pengujian Protoype Perangkat Lunak

Pengujian Prototype perangkat lunak pada penelitian ini akan menggunakan metode blackbox. Pengujian ini digunakan untuk mengetahui apakah prototype aplikasi yang dibangun telah sesuai dengan fungsional yang diharapkan. Pengujian dilakukan dengan memberikan beberapa jenis inputan yang nantinya akan di bandingkan dengan hasil yang sudah diharapkan.

Tabel 7 Pengujian Menu Prediksi

Input	Yang diharapkan	Kesimpulan
Input data pada kolom sistem yang sudah ditentukan	Sitem dapat menerima inputan yang dimasukkan	Diteima
Hasil dari Input data	Hasil akan menampilkan olahan dari sistem sesuai perhitungan algoritma	Diterima

Setelah dilakukan pengujian blackbox didapatkan hasil bahwa aplikasi yang dibangun telah

berjalan dengan baik sesuai dengan apa yang telah diharapkan.

4. Kesimpulan

Berdasarkan penelitian yang sudah dilakukan guna memprediksi penyakit diabetes menggunakan sistem yang berbasis pada algoritma datamining yaitu metode naïve bayes dan korelasi pearson. Maka kesimpulan yang bisa ditarik antara lain adalah sebagai berikut

- a. Penggunaan algoritma Naïve bayes dengan korelasi pearson lebih baik dari pada penggunaan algoritma naïve bayes saja.
- b. Hasil Pengujian dua metode naïve bayes dengan fitur selection korelasi pearson di dapat akurasi sebesar 79.13%. Sedangkan penggunaan metode independen naïve bayes yang tidak diikuti metode fitur selection didapati akurasi sebesar 68.26%. di sini terlihat bahwa ada peningkatan akurasi sebanyak 10.87%.

Saran

Penelitian ini bisa dikembangkan lagi dengan fitur atau studi kasus yang lain atau berbeda dengan penelitian ini. Sehingga bisa diperoleh hasil yang lebih baik yang dapat dilihat dari akurasi serta performa prediksi yang meningkat.

5. Daftar Pustaka

Chan, M. (2013). Global Report On Diabetes. *World Health Organization*.

Cheruku, R., Edla, D. R., Kuppili, V., & Dharavath, R. (2017). A Fuzzy Rule Miner Integrating Rough Set Feature Selection and Bat Optimization for Detection of Diabetes Disease Author: *Applied Soft Computing Journal*.
<https://doi.org/10.1016/j.asoc.2017.06.032>

Colagiuri, R., Brown, J., & Dian, K. (2011). Global Diabetes Plan At A Glance. *International Diabetes Federation*.

Farid, D., Zhang, L., Mofizur, C., Hossain, M. A., & Strachan, R. (2014). Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Systems With Applications*, 41(4), 1937–1946.
<https://doi.org/10.1016/j.eswa.2013.08.089>

Huang, Y., Mccullagh, P., Black, N., & Harper, R. (2007). *Feature selection and classification model construction on type 2 diabetic patients' data*.
<https://doi.org/10.1016/j.artmed.2007.07.002>

Jiang, L., Cai, Z., Zhang, H., & Wang, D. (2012).

Not so greedy : Randomly Selected Naive Bayes. *Expert Systems With Applications*, 39(12), 11022–11028.
<https://doi.org/10.1016/j.eswa.2012.03.022>

Kesehatan, K. (2014). *Situasi dan Analisis Diabetes*. Kesehatan, Infodatin pusat data dan informasi kementerian RI.

Kumar, N., & Abedin, M. (2017). Comparative Approaches for Classification of Diabetes Mellitus Data: Machine Learning Paradigm. *Computer Methods and Programs in Biomedicine*.
<https://doi.org/10.1016/j.cmpb.2017.09.004>

Mansourypoor, F., & Asadi, S. (2017). Development of a Reinforcement Learning-based Evolutionary Fuzzy Rule-Based System for Diabetes Diagnosis Fatemeh. *Computers in Biology and Medicine*.
<https://doi.org/10.1016/j.compbimed.2017.10.024>

Nai-arun, N., & Mounghmai, R. (2015). Comparison of Classifiers for the Risk of Diabetes Prediction. *Procedia - Procedia Computer Science*, 69, 132–142.
<https://doi.org/10.1016/j.procs.2015.10.014>

Rigla, M., Pons, B., & Elena, M. (2017). A web-based clinical decision support system for gestational diabetes : Automatic diet prescription and detection of insulin needs. *International Journal of Medical Informatics*.
<https://doi.org/10.1016/j.ijmedinf.2017.02.014>

Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., Yang, G., & Chen, Y. (2016). A Machine Learning-based Framework to Identify Type 2 Diabetes through Electronic Health Records. *International Journal of Medical Informatics*.
<https://doi.org/10.1016/j.ijmedinf.2016.09.014>