Jurnal Smart Teknologi
Vol. 3, No. 4, Mei 2022, Halaman 366 – 373
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

Perbandingan Algoritma K-Nearest Neighbor Dan Gaussian Naïve Bayes Pada Klsifikai Penyakit Diabetes Melitus

Comparison Of K-Nears Neighbor And Gaussian Naïve Bayes Algorithm On The Classification Of Diabetes Mellitus

Puput Tri Rahayu[1], Daryanto [2*] , Qurrota A'yun[3]
Mahasiswa Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Email: puputtriktb11@gmail.com
[2]Dosen Fakultas Teknik, Universitas Muhammadiyah Jember *Koresponden Author
Email : daryanto@gmail.com
[3]Dosen Fakultas Teknik, Universitas Muhammadiyah Jember
Email : qurrotaayun@unmuhjember.ac.id

Abstrak

Data mining merupakan pemrosesan suatu data menggunakan cara statistik, matematik, dll untuk mengidentifikasi suatu informasi pengetahuan potensial dan berguna yang tersimpan dalam basis data besar. Klasifikasi adalah salah satu tugas dari data mining yang bertujuan untuk memprediksi label kategori benda yang tidak diketahui sebelumnya, dalam membedakan antara objek yang satu dengan yang lainnya berdasarkan atribut atau fitur Terdapat beberapa algoritma yang dapat digunakan untuk mengklasifikan seperti algoritma K-Nearest Neighbor yang memiliki keunggulan yaitu lebih efektif didata training yang besar, dapat menghasilkan data yang lebih akurat dan Gaussian Naïve Bayes dimana metode ini hanya membutuhkan jumlah data trainng yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Setelah dilakukan penelitian menggunakan kedua algoritma tersebut, penelitian menggunakan algoritma K-Nearest Neigboar menghasilkan accuracy sebesar 71.27%, sensitivity sebesar 76.86%, specifity sebesar 81.59%%, precision sebesar 76.15% dan error rate sebesar 35.05% sedangkan pada saat mengguakan algoritma Gaussian Naïve Bayes menghasilkan accuracy sebesar 73.50%, sensitivity sebesar 96.14%, specifity sebesar 86.45%, precision sebesar 78.98% dan error rate sebesar 29.47%. berdasarkan hasil penelitian tersebut menunjukkan untuk algoritma Gaussian Naïve Bayes mempunyai accuracy yang lebih akurat bila dibandingkan dengan algortima K-Nears Neighboar dalam mengklasifikasi data Diabetes Melitus.

Kata kunci : Diabetes Mellitus, Gaussian Naïve Bayes, K-Nearest Neighboar

Abstract

Data mining is a statement of data using statistical, mathematical, etc. methods to identify potential and useful knowledge information stored in large databases. Classification is one of the tasks of data mining that aims to predict the labels of previously unknown object categories, in distinguishing between objects from one another based on attributes or features there are several algorithms that can be used to classify such as the K-Nearest Neighbor algorithm which has advantages. which is more effective in large training data, can produce more accurate data and Gaussian Naïve Bayes where this method only requires a small amount of training data to determine the parameters needed in the classification process. After doing research using these two algorithms, research using the K-Nearest Neigboar algorithm produces an accuracy of 71.27%, sensitivity of 76.86%, specifity of 81.59%, precision of 76.15% and error rate of 35.05 % while using the Gaussian Naïve Bayes algorithm, the accuracy is 73.50%, the sensitivity is 96.14%, the specifity is 86.45%, the precision is 78.98% and the error rate is 29.47%. Based on the results of this study, the Gaussian Naïve Bayes algorithm has more accuracy than the K-Nears Neighbor algorithm in classifying Diabetes Mellitus data.

Jurnal Smart Teknologi
Vol. 3, No. 4, Mei 2022, Halaman 366 – 373
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

Keywords: *Diabetes, K-Nearest Neighbor, Gaussian Naïve Bayes.*

1. PENDAHULUAN

Data mining merupakan pemrosesan suatu data menggunakan cara statistik, matematik, dll untuk mengidentifikasi suatu informasi pengetahuan potensial dan berguna yang tersimpan dalam basis data besar (Riadi., 2017). Klasifikasi adalah salah satu tugas dari data mining yang bertujuan untuk memprediksi label kategori benda yang tidak diketahui sebelumnya, dalam membedakan antara objek yang satu dengan yang lainnya berdasarkan atribut atau fitur (Parvin dkk., 2008).

Algoritma K-Neraest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. KNN memiliki keunggulan yaitu lebih efektif didata training yang besar, dapat menghasilkan data yang lebih akurat, dll. Naïve Bayes menurut Rani Puspita dkk, (2020) merupakan metode untuk klasifikasi text dengan kecepatan pemrosesan yang tinggi jika dalam jumlah besar. Terdapat beberapa kelebihan dari Naïve Bayes seperti efisien dalam dalam pelatihan dan penggunaannya, akurasi yang dihasilkan relatif tinggi dan metode ini hanya membutuhkan jumlah data trainng yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian (Saleh, 2015). Penyakit diabetes dapat menyerang atau menjangkit siapa saja maupun usia berapapun, di Indonesia khususnya pada perempuan berdasarkan data dari Riskesdas pada tahun 2018 menyatakan bahwa pravelensi diabetes mellitus pada perempuan lebih tinggi dibandingkan laki-laki dengan perbandingan perempuan 1,78% dan laki-laki 1,4% serta untuk 5 tahun terakhir diabetes mellitus pada perempuan mengalami peningkatan sedangkan pada laki- laki mengalami penurunan. Diabetes mellitus merupakan penyebab kematian terbesar nomer 3 di Indonesia setelah stroke dan jantung coroner. (Pusat Data dan Informasi Kementrian Kesehatan RI).

Data set dalam penelitian ini termasuk berskala banyak maka peneliti memutuskan untuk menggunakan metode Gaussian Naïve Bayes dan metode K- Nearest Neighbor. Kinerja dari kedua metode tersebut akan dibandingkan, sehingga dapat diketahui metode yang paling efektif dalam melakukan klasifikasi. Berdasarkan pada latar belakang diatas maka penulis ingin melakukan penelitian dengan judul "Perbandingan Algoritma K-Nearest Neighbor dan Gaussian Naïve Bayes pada klasifikasi penyakit Diabetes Melitus".

2. TINJAUAN PUSTAKA

A. PENELITIAN TERDAHULU

Pada penelitian yang dilakukan oleh Januar Adi Putra, dkk, (2016) Dengan judul "Klasifikasi Pengidap Diabtes Pada Perempuan Menggunakan Penggabungan Metode Support Vector Machine dan K-Nearest Neighbor". Dengan berdasarkan 768 data dan 8 atribut, hasil menunjukkan bahwa Algoritma Support Vector Mechine mempunyai akurasi prediksi maksimum untuk diabetes sebesar 77.60%, dan K-Nearest Neighbor sebesar 91.00%. Sedangkan pada penelitian yang dilakukan oleh Fuad Nurhasan, dkk, (2018) yang berjudul "Perbandingan Algoritma C4.5, KNN, dan Naïve Bayes Untuk Penentuan Model Klasifikasi Penanggung Jawab BSI Enterpenture". Dengan 300 record data dan 12 atribut, menyimpulkan hasil dengan menggunakan metode C4.5 mempunyai nilai akurasi 73,33%, metode KNN mempunyai nilai akurasi 70% dan metode Naïve Bayes sebesar 80%.

B. DABETES MELITUS

Diabetes dikenal sebagai kencing manis, adalah penyakit yang berhubungan dengan meningkatnya kadar gula darah dalam tubuh, terutama setelah makan. Diabetes adalah penyakit yang mengancam jiwa yang disebabkan oleh kurangnya hormon insulin. Salah satu ciri-ciri diabetes adalah meningkatnya kadar gula darah yang digambarkan di atas normal atau hipertensi (120 mg / dl atau 120 mg% atau lebih).

Jurnal Smart Teknologi
Vol. 3, No. 4, Mei 2022, Halaman 366 – 373
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

Jadi diabetes mellitus yang dimaksud dalam penelitian ini adalah berkurangnya kemampuan metabolisme tubuh untuk memproduksi hormon insulin yang dapat mengakibatkan kenaikan pada kadar gula darah dalam dutubuh manusia.

C. DATA MINING

Menurut Kaylani M, (2012) Data mining adalah proses ekstraksi informasi dan pola yang bermanfaat yang diambil dari data yang banyak. Penerapan suatu metode terhadap sejumlah data yang besar untuk menggali atau mencari pola, informasi, maupun pengetahuan yang berguna yang tersimpan dalam data tersebut atau biasa di sebut dengan data mining.

D. KLASIFIKASI

Teknik klasifikasi merupakan teknik untuk memprediksi variabel target berdasarkan variabel input. Prediksi dari klasifikasi didasarkan pada model yang dibangun dari kumpulan data yang sebelumnya telah dikenal. Klasifikasi memprediksi variabel output dengan tipe data kategorikal ataupun polinominal (misalnya, prediksi keputusan ya atau tidak untuk menyetujui pinjaman).

E. K-NEARESR NEIGHBOR

Algoritma *K-Nearest Neighbor* menggunakan klasifikasi ketetanggaan sebagai nilai prediksi untuk *query* data yang baru atau data *testing*. Untuk menghitung jauh aatau dekatnya tetangga dapat dihitung menggunakan rumus *Euclidean Distance* dengan persamaan (3), sebagai berikut (Fitri Yunita, (2016):

$$d(x_i, x_j) = \sqrt{\sum_{n=1}^{p} (x_{1i} - x_{2i})^2}$$

Keterangan:
$d(x_i, x_j)$ = Jarak Euclidean
x_{1i} = Data Training
x_{2i} = Data testing

F. NAIVE BAYES

Menurut Bustami, (2014) *naïve bayes* nerupakan pengklasifikasian untuk memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya. Ketika dihadapkan dengan data kontinu maka akan menggunakan rumus *Densitas gauss* (Bustami, 2014):
Ketika dihadapkan dengan data kontinu maka akan menggunakan rumus *Densitas gauss* (Bustami, 2014):

$$P(X_i = x_i|Y = y_{j=} \frac{1}{\sqrt{2\pi\sigma ij}} e^{-\frac{(x_i-\mu_{ij})^2}{2\sigma^2 ij}}$$

Keterangan :
P: Peluang
X_i: Atribut ke i
x_i: Nilai atribut ke i
Y: Kelas yang dicari
y_j: Sub kelas Y yang dicari
μ: Mean, menyatakan rata rata dari seluruh atribut
σ: Deviasi standar, menyatakan varian dari seluruh atribut

G. CONFUSION MATRIX

Matriks konfusi merupakan alat ukur yang memiliki fungsi mengukur saat menganalisis pengklasifikasi. Apakah pengklasifikasi pandai mengenali tupel dari kategori yang berbeda? Ketika pengklasifikasi melakukan klasifikasi dan memiliki data yang benar, nilai True-Positive dan True-Negative akan berperan dalam memberikan informasi ini. Pada saat yang sama, jika pengklasifikasi membuat kesalahan dalam mengklasifikasikan data, nilai False-Positive dan False-Negative akan memberikan informasi ini (Han, Kamber, dan Pei 2011).

Table 1 Confusion Matrix

Prediksi	**Kelas Aktual**	
	Positif (+)	Negative (-)
Positif (+)	TP	FP
Negative (-)	TN	TN

Keterangan:

1. TP (True Positive) adalah banyaknya data dengan nilai benar positif dan nilai prediksi positif.

2. FP (False Positive), adalah banyaknya data yang memiliki nilai benar negatif dan nilai prediksi positif.

3. FN (False Negative), adalah banyaknya data yang memiliki nilai benar positif dan nilai prediksi negatif.

4. TN (True Negative), adalah banyaknya data dengan nilai benar negatif dan nilai prediksi negatif.

H. CROSS VALIDATION

Menurut Rohani, Abbas., et al. (dalam Jiang, Ping., 2017) K-Fold Cross Validation adalah salah satu dari jenis pengujian cross validation yang berfungsi untuk menilai kinerja proses sebuah metode algoritma dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai K k-fold. Kemudian salah satu kelompok k-fold tersebut akan dijadikan sebagai data uji sedangkan sisa kelompok yang lain akan dijadikan sebagai data latih. Nilai k atau jumlah fold yang digunakan dalam penelitian ini 2 sampai 10. Angka 2 sampai 10 digunakan untuk melakukan eksperimen dengan komposisi data latih dan data uji yang berbeda. Angka 10 digunakan sebagai batas akhir karena metode 10-fold validation merupakan metode yang paling umum digunakan dan memiliki estimasi performa yang akurat (Refaeilzadeh, dkk, 2019).

I. PYTHON

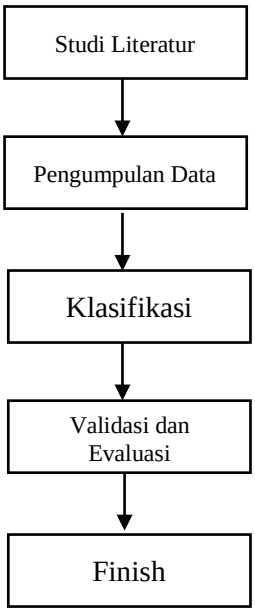
Bahasa pemrograman ini merupakan Bahasa pemrograman yang dapat dkembangkan oleh siapa saja karena bersifat open source yaitu Bahasa pemrograman ini dapat digunakan tanpa lisensi, dan dapat dikembangkan semaksimal mungkin. Ada beberapa tools yang ada dalam Bahasa pemrograman python tersebut, diantaranya Rstudio, Orange 3, Jupyter Lab, Jupyter notebook dll. Keunggulan dari python adalah mudah dipelajari, python memiliki struktur yang sederhana serta kata kunci yang sedikit, mudah diaplikasikan karena penulisan sintaksnya lebih sederhana(kode di python lebih singkat) dibandingkan dengan pemrograman lain untuk masalah yang sama. Selain itu python memiliki banyak library, python memiliki library powerful untuk proyek matematika dan statistika., untuk beberapa tujuan khususnya dibidang data, seperti library, numpy, pandas, statsmodel.

J. JUPYTER NOTEBOOK

Jupyter Notebook adalah sebauh aplikasi open source yang fungsinya untuk membuat dan berbagi dokumen yang berisi persamaan matematika, visualisasi data, tautan dan lain sebagainya. Pada Jupyter notebook input dan output berada pada satu line kode sehingga dapat dengan mudah dilakukan visualisai dari setiap kode tersebut. Jupyter Notebook dikenal sebelumnya sebagai IPython Notebook dan dalam waktu dekat akan berevolusi untuk mendukung ilmu data interaktif dan komputasi ilmiah di semua Bahasa pemrograman. (Avila Dkk dalam situsnya http://jupyter.org/).

3. METODOLOGI PENELITIAN

A. KERANGKA PENELITIAN



Gambar 1 Kerangka Kerja Penelitian
Sumber:hasil gambar menurut alur penelitian sendiri

Jurnal Smart Teknologi
Vol. 3, No. 4, Mei 2022, Halaman 366 – 373
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

B. STUDI LITERATUR

Tujuan dilakukan studi literatur adalah untuk mendapatkan landasan-landasan yang akan digunakan untuk penelitian ini, studi literartur ini bisa didapatkan melalui buku, jurnal, dll.

C. METODE PENGUMPULAN DATA

Berdasarkan data yang digunakan dari *Pima Indians* yang diambil dari *website kaggle*. Keseluruhan data ini mempumyai 8 atribut yaitu, kehamilan, glukosa, tekanan darah, ketebalan kulit, insulin BMI, keturunan dan umur.

Tabel 2 Atribut Dataset Beserta Deskripsinya

ATRIBUT	DESKRIPSI	SATUAN	TIPE DATA
Kehamilan	Berapa kali hamil	-	Numerik
Glukosa	Kadar Glukosa 2 jam	Mg/dL	Numerik
Tekanan Darah	Tekanan darah	Mm Hg	Numerik
Ketebalan Kulit	Ketebalan kulit lipatan trisep, diukur menggunakan skinfold	Mm	Numerik
Insulin	Hormon yang membantu mngendalikan gula	Mu U/ml	Numerik
BMI	Berat badan	Kg/m2	Numerik
Keturunan	Riwayat keturunan	-	Numerik
Umur	Umur	Years	Numerik
Outcome	Positif diabetes (1) dan	-	Numerik

Sumber: Data Diabetes Kaggle 2019

B. KLASIFIKASI

Proses Klasifikasi yang digunakan pada penelitian ini menggunakan metode K-Neraets Neighbor dan Gaussian Naïve Bayes.

C. VALIDASI DAN EVALUASI

Penentuan ini fold K yang digunakan pada penelitian ini yaitu 2-10. Pemilihan nilai fold K tersebut untuk memilih *kfold* mana yang akan menghasilkan nilai rata-rata yang paling tinggi. Proses evaluasi dilakukan dengan menghitung nilai TP, FP, TN, FN sehingga meghasilkan nilai *accuracy*, *sensitivity*, *specifity*, *precision* dan *error rate.*

4. HASIL DAN PEMBAHASAN

Berikut hasil terbaik yang didapatkan dari hasil klasifikasi algorima K-Nearest Neighbor.

Tabel 3 Hasil Klasifikasi KNN pada K2

K Fold	K-Nearest Neighbor-K2				
	Accuracy	Sensitivity	Specivity	Precission	Errorrate
2	65.29%	51.34%	51.20%	73.08%	34.70%
3	65.48%	47.93%	48.12%	74.61%	32.91%
4	65.67%	**51.53%**	**53.55%**	73.92%	34.32%
5	65.12%	49.73%	50.57%	73.12%	34.87%
6	**66.81%**	49.50%	52.31%	**76.15%**	33.18%
7	65.12%	47.88%	50.41%	73.40%	34.87%
8	65.29%	49.88%	53.34%	73.08%	34.70%
9	65.89%	48.44%	52.15%	74.38%	34.10%
10	64.94%	49.95%	53.33%	71.32%	**35.05%**

Sumber: Hasil penelitian sendiri

Tabel 4 Hasil Klasifikasi KNN pada K3

K Fold	K-Nearest Neighbor-K3				
	Accuracy	Sensitivity	Specivity	Precission	Errorrate
2	69.40%	**74.96%**	75.37%	69.41%	30.59%
3	**71.08%**	74.21%	75.63%	**69.88%**	28.91%
4	67.72%	71.04%	76.58%	67.09%	32.27%
5	68.28%	71.67%	75.09%	67.40%	31.71%
6	70.16%	73.60%	78.50%	68.60%	29.83%
7	68.28%	69.25%	75.80%	67.31%	31.71%
8	68.09%	69.75%	77.67%	66.94%	**31.90%**
9	69.22%	71.54%	79.11%	67.44%	30.775

Jurnal Smart Teknologi
Vol. 3, No. 4, Mei 2022, Halaman 366 – 373
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

10	69.22%	72.16%	**81.04%**	66.88%	30.77%

Sumber: Hasil penelitian sendiri

Tabel 5 Hasil Klasifikasi KNN pada K4

K Fold	K-Nearest Neighbor-K4				
	Accuracy	**Sensitivity**	**Specivity**	**Precission**	**Errorrate**
2	67.72%	**63.26%**	63.95%	71.39%	**32.27%**
3	68.65%	57.52%	59.04%	73.86%	31.34%
4	66.04%	57.86%	62.28%	69.98%	33.95%
5	**70.34%**	61.90%	64.95%	**75.03%**	29.65%
6	69.78%	61.08%	65.20%	73.94%	30.21%
7	68.85%	58.65%	64.63%	73.14%	31.14%
8	68.47%	58.33%	**72.45%**	72.45%	31.52%
9	69.43%	61.27%	66.63%	73.22%	30.56%
10	70.17%	60.98%	68.49%	73.65%	29.82%

Sumber: Hasil penelitian sendiri

Tabel 6 Hasil Klasifikasi KNN pada K5

K Fold	K-Nearest Neighbor-K5				
	Accuracy	**Sensitivity**	**Specivity**	**Precission**	**Errorrate**
2	69.02%	**76.86%**	78.10%	68.61%	30.97%
3	**71.27%**	73.37%	75.78%	70.02%	28.72%
4	68.84%	72.34%	77.68%	68.04%	**31.15%**
5	70.33%	74.12%	78.08%	69.17%	29.66%
6	72.20%	74.73%	80.225	**70.71%**	27.79%
7	71.09%	71.42%	77.67%	70.20%	28.90%
8	70.70%	73.19%	80.68%	69.16%	29.29%
9	71.66%	72.66%	80.21%	70.58%	28.33%
10	71.11%	73.40%	**81.59%**	69.46%	28.88%

Sumber: Hasil penelitian sendiri

Berikut hasil terbaik yang didapatkan dari hasil klasifikasi algoritma Gaussian Naïve Bayes

Tabel 7 Hasil Klasifikasi GNB

K Fold	Gaussian Naïve Bayes				
	Accuracy	**Sensitivity**	**Specivity**	**Precission**	**Errorrate**
2	**73.50%**	72.11%	85.39%	**78.98%**	26.49%
3	71.84%	69.84%	84.78%	76.87%	28.15%
4	70.52%	**96.14%**	79.75%	75.10%	**29.47%**
5	72.20%	70.89%	82.26%	76.37%	27.79%
6	72.01%	69.92%	84.13%	76.47%	27.98%
7	72.59%	70.36%	85.73%	77.14%	27.40%
8	71.82%	69.49%	83.47%	75.14%	28.17%
9	72.41%	70.60%	82.71%	75.73%	27.58%
10	72.61%	70.54%	**86.45%**	76.05%	27.38%

Sumber: Hasil penelitian sendiri

Setelah dilakukan beberapa kali percobaan dan menghasilkan nilai rata-rata berikut nilai tertinggi dari confusion matrix yang meliputi *accuracy*, *sensitivity*, *specifity*, *precission* dan *error rate* dari K-Nearest Neighbor dan Gaussian Naïve Bayes.

Tabel 48 Hasil tertinggi dari KNN dan GNB

	Accuracy	***Sensitivity***	***Specifity***	***Precission***	***Error Rate***
KNN	71.27%	76.86%	81.59%	76.15%	35.05%
GNB	73.50%	96.14%	86.45%	78.98%	29.47%

Sumber: Hasil penelitian sendiri

Berdasarkan tabel hasil perbandingan diatas untuk dataset Diabetes Melitus dengan Algoritma Gaussian Naïve Bayes lebih tepat dari algoritma K-Nearest Neighbor karena pada Gaussian Naïve Bayes memiliki nilai accuracy sebesar 73.50% untuk ketepatan keseluruhan data, sensitivity sebesar 96.14% untuk memprediksi berapa banyak yang benar pada penderita diabetes, specifity sebesar 86.45% untuk memprediksi berapa banyak yang benar pada yang tidak menderita diabetes , precission sebesar 78.98% untuk memprediksi seberapa banyak yang benar-benar menderita penyakit diabetes dan pada K-Nearest Neighbor memiliki nilai error rate sebesar 35.05% yang berarti memiliki akurasi yang rendah.

5. KESIMPULAN DAN SARAN

A. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan menggunakan algoritma K- Nearest Neighbor menghasilkan accuracy sebesar 71.27%, sensitivity sebesar 76.86%, specifity sebesar 81.59%, precision sebesar 76.15% dan error rate sebesar 35.05% sedangkan pada saat mengguakan algoritma Gaussian Naïve Bayes menghasilkan accuracy sebesar 73.50%, sensitivity sebesar 96.14%, specifity sebesar 86.45%, precision sebesar 78.98% dan error rate sebesar 29.47%. Dari kesimpulan diatas maka Algoritma Gaussian Naïve Bayes yang lebih tepat dalam klasifikasi Diabetes Melitus dibandingkan dengan K-Nearest Neighbor. Semakin sedikit data training yang digunakan maka hasil klasifikasinya menunjukan ketidakakuratan yang tinggi, sebaliknya jika data yang digunakan semakin banyak hasilnya akan semakin akurat.

B. SARAN

Pada penelitian selanjutnya dapat menggunakan data yang berbeda atau data training yang lebih banyak. Pada penelitian selanjutnya dapat menggunakan metode lain selain algoritma K-Nearest Neighbor untuk perbandingan.

6. REFERENSI

Bustami, B. (2014). Penerapan Algoritma Naïve bayes Untuk Mengklasifikasi Data Nasabah Asuransi. Jurnal Informatika, 8(1)

Bustan, (2015). Manajemen pengendalian penyakit tidak menular. Jakarta : Rineka Cipta.

Han, J, Kamber, M, & Pei, J. 2012. Data Mining: Concept and Techniques, Third Edition. Waltham: Morgan Kaufmann Publishers.

Hasdianah.(2012). Mengenal Diabetes Melitus pada Orang Dewasa dan Anak-

Anak dengan Solusi Herbal. Yogyakarta: Nuha Medika https://www.kaggle.com/salihacur/diabetes . https://pusdatin.kemkes.go.id/article/view/20111800001/diabetes-melitus.html https://www.who.int/news-room/fact-sheets/detail/diabetes

Januar Adi Putra, Dkk. 2016. Klasifikasi Pengidap Diabetes Pada Perempuan Menggunakan Penggabungan Metode Support Vector Machine danK- Nearest Neighbour. Fakultas Teknologi Informasi Institut Teknologi Sepuluh Nopember (ITS).

Kotu, V., & Deshpande, B. (2014). Predictive analytics and data mining: concepts and practice with rapidminer. Morgan Kaufmann. ISBN10: 0128014601

Lakshita, N. (2012). Anak Aktif, Bebas Diabetes. Jogjakarta: Javalitera.

Mahendra, Gede surya; Subawa, I. G. bendesa. (2019). Perancangan Metode AHP-WASPAS Pada Sistem Pendukung Keputusan Penempatan ATM. Seminar Nasional Pendidikan Teknik Informatika (SENAPATI), 122– 128. Singaraja: Pendidikan Teknik Informatika.

Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. Scientific Journal of Informatics, 3(1), 19–26. https://doi.org/10.15294/sji.v3i1.4610

Permana, A.Y., Ismasari, & Effendi, M.Makmun. (2018). Perbandingan Stemming Porter KBBI Dengan Tala Untuk Mencari Akurasi Klasifikasi Topik Soal UN Bhs. Indonesia Menggunakan Algoritma Naïve bayes. Prosiding SNTI VI 2018 Universitas Trisakti, 274-281.

Riadi Muchlisin, 2017. Pengertian, Fungsi, Proses dan Tahapan Data Mining, https://www.kajianpustaka.com/2017/09/data-mining.html. Diakes tanggal 20

Maret 2020.

Rosadi, R., Akmal, A., Hidayat, A., & Kharismawan, B. (2018, January). Aplikasi K-Means Clustering Untuk Mengelompokan Data Kinerja Akademik Mahasiswa. In Prosiding-Seminar Nasional Teknik Elektro UIN Sunan Gunung Djati Bandung (pp. 92-96).

Supartini, Ida Ayu, I Komang Gede Sukarsa, & I Gusti Ayu Made Srinadi. (Agustus 2017) " Analisis Diskriminan Pada Klasifikasi Desa Di Kabupaten Tabanan Menggunakan Metode K-Fold Cross Validation." E-Jurnal Matematika [Online].

Wibowo, A. (2017, November). Klasifikasi – MTI. Binus University. Yunita, F. (2018). Data Mining. Penerapan Data Mining Menggunakan

Algoritma K-Means Clustering (Stusi Kasus : Univeristas Islam Indragiri)