



PERAMALAN PENDERITA DIABETES MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

A.Muhammad Syafar¹, Muhammad Syawal Idil Fitrah Baharuddin²,
Muhammad Zainal³, Nurul Fuadi⁴

^{1,2} Teknik Informatika, Fakultas Sains dan Teknologi, UIN Alauddin Makassar

³ Teknik Elektro, Fakultas Teknik, Universitas Muhammadiyah Pare-Pare

⁴ Fisika, Fakultas Sains dan Teknologi, UIN Alauddin Makassar

andi.syafar@uin-alauddin.ac.id ; 60200118004@uin-alauddin.ac.id ; enal_mza72@gmail.com ; nurul.fuadi@uin-alauddin.ac.id

Informasi Artikel

Riwayat Artikel:

Dikirim Author : 5-04-2021

Diterima Redaksi : 06-04-2021

Revisi Reviewer: 15-04-2021

Diterbitkan online: 05-05-2021

Keywords:

diabetes prediction ; knn ; cross validation

Kata kunci:

prediksi diabetes ; knn ; cross validation

ABSTRACT

Diabetes is a metabolic disorder caused by the pancreas not producing enough insulin. Pima Indian Dataset is to predict diagnostically whether a patient is potential or unpotential diabetes. All data taken in this dataset are women at least 21 years of age from the Pima Indians. For the parameters, we took several parameters such as blood pressure, triceps skin thickness, history of diabetes, and others. The existing data is then processed using Rapidminer tools using cross validation parameters 2,3, and 5. After testing, we get the results that the higher the cross validation value, the lower the accuracy. Even if we look at the K-1 sample, the accuracy value has no effect on cross validation.

ABSTRAK

Diabetes merupakan penyakit gangguan metabolik akibat pankreas tidak memproduksi cukup insulin. Pima Indian Dataset adalah untuk memprediksi secara diagnostik apakah seorang pasien potensial atau unpotensial diabetes.

Semua data yang diambil pada dataset ini adalah wanita minimal 21 tahun dari orang Indian Pima. Untuk parameternya, kami mengambil beberapa parameter seperti tekanan darah, ketebalan kulit trisep, riwayat diabet, dan lain lain. Data yang telah ada kemudian diolah menggunakan tools Rapidminer dengan menggunakan parameter cross validation 2,3, dan 5. Setelah melakukan pengujian, kami mendapatkan hasil bahwa semakin tinggi nilai cross validation akan menghasilkan akurasi yang semakin rendah. Walau apabila kita melihat dari sampel K-1 maka nilai akurasi tidak berpengaruh pada cross validation.

Penulis Korespondensi:

A.Muhammad Syafar
Teknik Informatika, Fak. Sains dan
Teknologi UIN Alauddin Makassar
andi.syafar@uin-alauddin.ac.id

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



I. PENDAHULUAN

Penyakit Tidak Menular (PTM), termasuk Diabetes, saat ini telah menjadi ancaman serius kesehatan global. Dikutip dari data WHO 2016, 70% dari total kematian di dunia dan lebih dari setengah beban penyakit. 90-95% dari kasus Diabetes adalah Diabetes Tipe 2 yang sebagian besar dapat dicegah karena disebabkan oleh gaya hidup yang tidak sehat [1].

Diabetes melitus (DM) didefinisikan sebagai suatu penyakit atau gangguan metabolisme kronis dengan multi etiologi yang ditandai dengan tingginya kadar gula darah disertai dengan gangguan metabolisme karbohidrat, lipid dan protein sebagai akibat insufisiensi fungsi insulin. Insufisiensi insulin dapat disebabkan oleh gangguan atau defisiensi produksi insulin oleh sel-sel beta Langerhans kelenjar pankreas, atau disebabkan oleh kurang responsifnya sel-sel tubuh terhadap insulin[2].

Kegiatan dalam melakukan prediksi terhadap berbagai penyakit telah banyak dilakukan dalam berbagai bidang keilmuan salah satunya bidang ilmu *computer science*.

Banyak percobaan yang dilakukan peneliti dengan menggunakan teknik data mining untuk memprediksi penyakit menggunakan berbagai algoritma seperti KNN (*k-nearest neighbor*) [3][4].

Data training yang digunakan adalah dataset PIMA Indian diabetes yang diperoleh dari UCI untuk menemukan faktor yang paling mempengaruhi diabetes dan pola klasifikasi pohon keputusan faktor yang mempengaruhi tersebut.

Pada penelitian ini kami berusaha melakukan prediksi pada data yang tersedia seberapa jauh tingkat akurasi seseorang terkena diabetes atau tidak berdasarkan tingkat glukosa, BMI, level insulin, usia, tekanan darah diabetes *Diabetes Pedigree*

Function dengan menggunakan algoritma KNN (*k-nearest neighbor*)

II. METODOLOGI PENELITIAN

Pada penelitian kali ini, kami mengambil referensi dataset dari kaggle. Kumpulan data yang kami ambil merupakan data yang diambil langsung dari *National Institute of Diabetes and Digestive and Kidney Diseases*. Dimana, tujuan dari pengambilan data ini adalah untuk memprediksi secara diagnostik apakah seorang pasien menderita diabetes atau tidak berdasarkan pengukuran diagnostik tertentu. Semua data yang diambil pada dataset ini adalah wanita minimal 21 tahun dari orang Indian Pima. Ada atribut yang kami tidak gunakan yaitu *pregnancies* (angka melahirkan)

Deskripsi tiap atribut :

- Glucose : Konsentrasi plasma glukosa
- Blood Pressure : Tekanan darah diastol (mmHg)
- Skin Thickness : Ketebalan kulit trisep (mm)
- Insulin : 2 jam serum insulin
- BMI : Index massa badan (kg untuk berat dan m untuk tinggi)
- Diabetes Pedigree Function : riwayat diabetes melitus
- Outcome : 1 untuk diabetes dan 0 untuk tidak diabetes menggunakan parameter Validation 2,3,dan 5

Data

Jumlah row : 769 Link source data :

<https://www.kaggle.com/uciml/pima-indians-diabetes-database>

IV. Hasil Dan Analisis

A. Skenario Pengujian

1. Pengolahan Data

Dataset Pima Indian diproses menggunakan rapidminer. Sebelum diolah, data tadi dilakukan normalisasi yaitu mengubah parameter label yang pada awalnya numerik (1 dan 0) menjadi string agar dapat dibaca oleh rapid miner. Karena pada Data tersebut awalnya berupa csv kemudian diubah ke xlsx an dipanggil dengan *operator read, x-validation, k-nn, apply model, dan performance classification*.

2. Hasil Pengolahan Data

Dalam pengujian, kami menetapkan nilai k yaitu 1 dan 3 dengan uji cross validation = 2,3,dan 5.

Dari pengujian tersebut kami mendapatkan hasil sebagai berikut :

- $k=1$

cross validation = 2

accuracy: 69.01% +/- 3.89% (piksi: 69.01%)			
pred. potensial	true potensial	true unpotensial	class precision
151	121	55.51%	75.41%
pred. unpotensial	117	379	
class recall	56.34%	75.60%	

cross validation = 3

accuracy: 64.58% +/- 1.21% (piksi: 64.58%)			
pred. potensial	true potensial	true unpotensial	class precision
139	143	49.29%	73.48%
pred. unpotensial	129	357	
class recall	51.87%	71.40%	

cross validation = 5

accuracy: 65.22% +/- 3.62% (piksi: 65.22%)			
pred. potensial	true potensial	true unpotensial	class precision
147	125	54.44%	75.70%
pred. unpotensial	121	377	
class recall	54.80%	75.40%	

- $k=3$

cross validation = 2

accuracy: 74.00% +/- 1.11% (piksi: 74.00%)			
pred. potensial	true potensial	true unpotensial	class precision
148	79	65.20%	77.82%
pred. unpotensial	120	421	
class recall	65.22%	84.20%	

cross validation = 3

accuracy: 75.00% +/- 3.38% (piksi: 75.00%)			
pred. potensial	true potensial	true unpotensial	class precision
142	85	68.80%	76.23%
pred. unpotensial	138	404	
class recall	52.90%	80.69%	

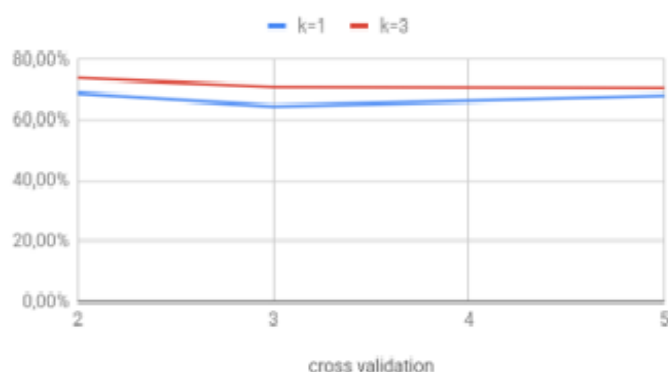
cross validation = 5

accuracy: 76.89% +/- 3.54% (piksi: 76.70%)			
pred. potensial	true potensial	true unpotensial	class precision
148	105	58.50%	76.70%
pred. unpotensial	120	395	
class recall	55.22%	79.30%	

Berikut merupakan hasil tabel dan grafik dari data diatas :

cross validation	k=1	k=3
2	69,01%	74,09%
3	64,58%	71,09%
5	68,23%	70,70%

k=1 dan k=3



V. KESIMPULAN

Dari hasil pengamatan yang telah dilakukan dapat disimpulkan bahwa semakin tinggi dari nilai cross validation pada algoritma KNN menghasilkan jumlah akurasi yang semakin rendah.

REFERENSI

[1] Kemenkes, (2018). CEGAH, CEGAH, dan CEGAH: Suara Dunia Perangi Diabetes. online, diakses pada 5 januari 2021. <https://www.kemkes.go.id/article/view/18121200001/prevent-prevent-and-prevent-the-voice-of-the-world-fight-diabetes.html>

[2] World Health Organization (WHO). (2000) Pencegahan Diabetes Mellitus. Jakarta : Hipokrates;

[3] Wu H, Yang S, Huang Z, He J, Wang X. Type 2 diabetes mellitus prediction model based on data mining. Informatics Med Unlocked [Internet]. 2018;10(August 2017):100–7. Available from: <https://doi.org/10.1016/j.imu.2017.12.00>

[4] M. Masnur, S. Alam, and M. Fikri Nasir, "RANCANG BANGUN SISTEM KEAMANAN MOTOR DENGAN PENGENALAN SIDIK JARI BERBASIS ARDUINO UNO," *J. Sintaks Log.*, vol. 1, no. 1, pp. 2412– 2775, Jan. 2021, [Online]. Available: <https://jurnal.umpar.ac.id/index.php/sylog>.

[5] Perveen S, Shahbaz M, Guergachi A, Keshavjee

K. Performance Analysis of Data Mining Classification Techniques to Predict Diabetes. *Procedia Comput Sci* [Internet]. 2016;82(March):115–21. Available from: <http://dx.doi.org/10.1016/j.procs.2016.04.016>

[6] Fransisca, K. (2012). *Awas Pankreas Rusak Penyebab Diabetes*. Jakarta: Penerbit Cerdas Sehat;

[7] Mulyahayati, L.I, (2018). *KNN Pima Indian Diabetes with python*. Yogyakarta : Universitas Islam Indonesia. online, diakses pada 5 januari 2021. <https://medium.com/@indiraluthfianam/knn-pima-indians-diabetes-database-dengan-menggunakan-python-part-1-2-8812372c0460>

[8] Hermawati, Fajar Astuti. 2013. *Data Mining*. Edisi ke-1. Yogyakarta: ANDI.

[9] Ponniah, P. 2001. *Data Warehouse Fundamentals: A Comprehensive Guide for IT Professionals*. John Wiley & Sons, Inc.

[10] Avelita, B. (2013). *Klasifikasi K-Nearest Neighbor*. Dipetik 06 2016, 22, dari www.academia.edu: https://www.academia.edu/9131959/A_Klasifikasi_K_Nearest_Neighbor online diakses pada 5 januari 2021

[11] Mulyahayati, L.I, (2018). *KNN Pima Indian Diabetes with python*. Yogyakarta : Universitas Islam Indonesia. online, diakses pada 5 januari 2021. <https://medium.com/@indiraluthfianam/knn-pima-indians-diabetes-database-dengan-menggunakan-python-1a0dcff25293>