JUSTINDO
Jurnal Sistem dan Teknologi Informasi
http://jurnal.unmuhjember.ac.id/index.php/JUSTINDO
E-ISSN: 25541-5735, P-ISSN: 2502-5724
Volume 7, No. 2, Agustus 2022

Implementasi Metode *Adaboost* untuk Mengoptimasi Klasifikasi Penyakit Diabetes dengan Algoritma *Naïve Bayes*

Lidia Pebrianti[1], Fitrahuda Aulia[2], Halimatun Nisa[3], Kana Saputra S[4]

Program Studi Ilmu Komputer, FMIPA, Universitas Negeri Medan
Email: [1]lidiapebrianti503@gmail.com, [2]auliafitrahuda29@gmail.com, [3]halimatunnisa40@gmail.com, [4]kanasaputras@unimed.ac.id

(Naskah masuk: 3 Juli 2022, diterima: 4 agustus 2022, diterbitkan: 28 Agustus 2022)

ABSTRAK

Diabetes melitus adalah penyakit metabolik yang disebabkan oleh kegagalan tubuh menggunakan insulin atau tidak adanya insulin kimia, karena itu kadar gula dalam darah tidak dapat terkendali. Menurut *International Diabetes Federation* (IDF), saat ini dinilai untuk jumlah penderita diabetes di Indonesia bisa mencapai 28,57 juta pada tahun 2045. Jumlah ini 47% lebih besar daripada 19,47 juta dari tahun 2021. Penderita diabetes diketahui melonjak 167% dibandingkan dengan penderita diabetes pada tahun 2011 yang mencapai 7,29 juta. Secara umum, IDF mengukur jumlah penderita diabetes di dunia dapat mencapai 783,7 juta orang pada tahun 2045. Jumlah ini meningkat 46% dibandingkan dengan tahun 2021 yang mencapai 536,6 juta. Adaboost adalah Algoritma Boosting yang paling terkenal, dapat digunakan dengan tujuan untuk meningkatkan keakuratan kinerja pembelajaran Machine Learning *Naïve Bayes*, sehingga dapat mengurangi *noise* dalam kumpulan data yang berukuran besar dengan beberapa kelas atau multi kelas. .Dengan menggunakan split data 60/40 Algoritma *Naïve Bayes* menghasilkan akurasi sebesar 0.7608. Sedangkan untuk hasil *Naïve Bayes* yang di *boosting* dengan menggunakan algoritma *Adaboost* adalah sebesar 0,7694.

Kata kunci: *adaptive boosting, naïve bayes, diabetes, akurasi*

ABSTRACT

Mellitus diabetes is a metabolic disease caused by the body's failure to use insulin or lack of chemical insulin, which makes the blood's sugar level uncontrollable. According to the international diabetes federation (idf), the current rate for diabetes sufferers in Indonesia could be 28.57 million by 2045. This figure is 47% greater than 19.47 million from 2021. Diabetes sufferers are known to climb 167% compared with those who lived in 2011 to 7.29 million. In general, idf measures the world's number of diabetes sufferers may have been 783.7 million by 2045. The total increased by 46% compared with 2021, which reached 536.6 million. Adaboost is the most famous boosting algorithm, which can be used to improve the performance of the learning machine learning naive bayes, so that it can reduce the noise in large data collections with multiple classes or multi-classes.. By using a 60/40 data split naive baisa's file results in an accuracy of 0.7608, As for naive bayes's hit with an adaboost algorithm was. 7694.

Keywords: *adaptive boosting, naïve bayes, diabetes, accuracy*

1. PENDAHULUAN

Diabetes melitus adalah penyakit metabolik yang disebabkan oleh kegagalan tubuh menggunakan insulin atau tidak adanya insulin kimia, karena itu kadar gula dalam darah tidak dapat terkendali. (Marito Putry & Nurina Sari, 2022). Diabetes melitus (DM) ialah penyakit kronis yang terjadi saat pankreas tidak bisa memproduksi insulin yang cukup ataupun saat badan tidak bisa memanfaatkan insulin yang dihasilkan secara efisien. Insulin merupakan hormon yang mengendalikan gula darah. Hiperglikemia ataupun gula darah yang bertambah, merupakan dampak umum dari diabetes tidak terkendali yang menimbulkan kerusakan serius pada banyak sistem tubuh, khususnya saraf dan pembuluh darah. (Murtiningsih, Pandelaki & Sedli, 2021).

Menurut *International Diabetes Federation* (IDF), saat ini dinilai untuk jumlah penderita diabetes di Indonesia bisa mencapai 28,57 juta pada tahun 2045. Jumlah ini 47% lebih besar daripada 19,47 juta dari tahun 2021. Penderita diabetes diketahui melonjak 167% dibandingkan dengan penderita diabetes pada tahun 2011 yang mencapai 7,29 juta. Secara umum, IDF mengukur jumlah penderita diabetes di dunia dapat mencapai 783,7 juta orang pada tahun 2045. Jumlah ini meningkat 46% dibandingkan dengan tahun 2021 yang mencapai 536,6 juta. (Pahlevi, Reza, 2021).

Seiring berkembangnya waktu *Machine Learning* (ML) digunakan melalui pendekatan dalam kecerdasan buatan (AI) untuk menyelesaikan masalah ataupun melakukan optimisasi (Ula & Faridhatul Ulva, 2021). Salah satu Algoritma yang melakukan optimasi adalah algoritma *Adaboost. Adaboost* digunakan untuk memboosting hasil akurasi untuk menghasilkan nilai akurasi yang lebih baik. Keakuratan dengan memanfaatkan Algoritma *Adaboost* digunakan untuk mengevaluasi kinerja pada pembelajaran *machine learning* (Byna & Basit, 2020),(Zhang & Kong, 2020). Akurasi dapat diartikan sebagai tingkat korelasi antara nilai prediksi dengan nilai aktual (Argina, 2020).

Teori keputusan *bayes* adalah pendekatan statistik yang mendasar dalam pengenalan pola (*pattern recognition*). Pendekatan ini mengikuti pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas (Indra Borman, 2020). Klasifikasi adalah proses untuk menemukan model atau peranan yang memaparkan maupun membedakan konsep kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak dikenal (Lestari & Adhiva, 2022). Penerapan algoritma pengklasifikasi *Naïve Bayes* dapat mengurangi data *noise* pada *dataset* yang berukuran besar serta memiliki banyak kelas atau multi kelas sehingga dapat meningkatkan akurasi (Sutoyo, E. & Almaarif, A. (2020). *Naïve Bayes Classifier* merupakan metode klasifikasi mengikuti pada Teorema *Bayesian*. Metode ini dapat digunakan untuk memprediksi peluang di masa depan berdasarkan pengalaman yang terjadi di masa sebelumnya menggunakan perhitungan probabilitas dan statistika (Simanjuntak, Simatupang & Anita, 2022). Algoritma *Naïve Bayes* merupakan salah satu algoritma yang digunakan untuk klasifikasi tetapi tidak bisa mengatasi klasifikasi berjenis numerik pada kelasnya (Astuti *et al.*, 2022).

Pada penelitian ini, digunakan Algoritma *Adaboost* untuk mengoptimasi kinerja Algoritma *Naïve Bayes* dengan

p-ISSN : 2502-5724; e-ISSN : 2541-5735

cara meningkatkan nilai akurasi klasifikasi penyakit Diabetes melitus.

2. Metode Penelitian

Gambar 1. Tahapan Penelitian

Pada penelitian ini menggunakan metode pendekatan kuantitatif, dengan cara memberikan penekanan pengukuran dari data yang ada. Tahap penelitian dapat dilihat pada Gambar 1 (Abdurrahman, 2022).

2.1 *Collecting Data*

Hal pertama yang dilakukan sebelum mengumpulkan data yang akan digunakan dalam penelitian yaitu mencari data secara daring. Pada penelitian ini data diperoleh dari www.kaggle.com (Nur Rais, Rahmawati & Faizal Amir, 2021). *Dataset* berupa data tabular dari kondisi kesehatan pasien yang terindikasi diabetes maupun tidak.

2.2 *Preprocessing*

Preprocessing data adalah salah satu proses yang penting dilakukan untuk menghasilkan data dengan kualitas yang baik dengan cara di antaranya validasi, integrasi dan transformasi. Praproses data meliputi pemeriksaan dan pembuangan data yang inkonsisten, data ganda, data yang perlu diperbaiki dan penambahan data sesuai dengan yang dibutuhkan (Ramdan *et al.*, 2022). Adapun fitur yang diperbaiki yaitu *Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, Diabetes Pedigree Function*, *BMI* dan *Age*.

2.3 *Data Cleaning*

Pembersihan data adalah Teknik menghapuskan *noise* dan data yang tidak relevan (Sibarani, 2020). Pada bagian ini dilakukan penghapusan baris-baris data yang tidak lengkap.

2.4 *Split Data*

Selanjutnya adalah melakukan *split data* menjadi data *training* dan *testing*. Dalam penelitian ini, menggunakan persentase 60/40 (Byna & Basit, 2020). *Split* data sebanyak 768 baris berupa pembagian porsi 60% data *training*, dan 40% data *testing*.

2.5 Akurasi

Confussion matrix merupakan tabel matriks dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual yang dapat digunakan apabila kelas keputusan pada suatu *dataset* hanya terdiri dari dua kelas, yakni satu kelas dianggap positif dan kelas yang lain dianggap sebagai negatif (Defiyanti, 2015).

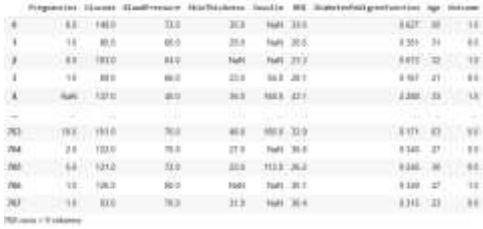
3. HASIL DAN PEMBAHASAN

3.1 *Preprocessing*

Sebelum pengolahan data dilakukan identifikasi terhadap adanya *missing values* pada data. Yaitu berupa nilai 0 *default* yang terdapat pada *dataset*. Pada Gambar 2 terlihat adanya *missing values* pada fitur.

Gambar 2. Identifikasi data

Untuk melihat jumlah data yang tidak lengkap terlebih dahulu nilai 0 pada setiap fitur perlu diubah menjadi nilai kosong (*NaN*) dengan menggunakan sintaks *replace(0, np.nan)* agar dapat dideteksi sebagai *missing values* seperti pada Gambar 3. Setelah itu akumulasi total dari data yang kosong pada Gambar 4 diperoleh dengan menggunakan sintaks *isnull().sum()*

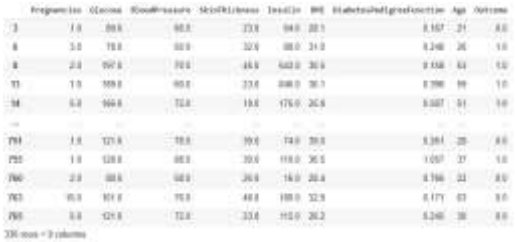


Gambar 3. *Missing values*

Gambar 4. Banyaknya *missing values*

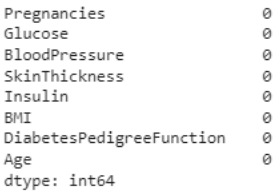
3.2 Data *Cleaning*

Dengan menggunakan sintaks *dropna()* untuk menghapus setiap baris yang memiliki *NaN* diperoleh data bersih berjumlah 336 baris disajikan pada Gambar 4.



Gambar 5. Proses data *cleaning*

Pada Gambar 5 berikut adalah identifkasi yang dilakukan setelah dilakukan *cleaning data*.



Gambar 6. *Missing values* setelah dilakukan *cleaning data*

3.3 Evaluasi

3.4.1 Pengujian Akurasi *Naïve Bayes*

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi yang mana menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang diberikan (Maulidah *et al.*, 2021).

Prediksi Hasil Akurasi *Naïve Bayes* menggunakan *python*, dengan *split* data 60%/40%:

Gambar 7. *Confusion matrix* dan hasil akurasi model *Naïve Bayes*

Pada Gambar 7 di atas hasil eksekusi diperoleh jumlah *True Positif* (TP) adalah 162, *True Negatif* (TN) adalah 62, *False Positif* (FP) adalah 37, dan untuk *False Negatif* (FN) adalah 47. Hasil akurasi *Naïve Bayes* dapat dilihat pada Gambar 7 yaitu 0.7608

3.4.2 Pengujian Akurasi *Naïve Bayes* dan *Adaboost*

Boosting dapat dikombinasikan dengan *classifier* algoritma yang lain untuk dapat meningkatkan performa klasifikasi. Tentunya secara intuitif, penggabungan

beberapa model dapat membantu jika model tersebut sangat berbeda satu sama lain (Gultom, 2020). Pada bagian ini *boosting Naïve Bayes* dilakukan dengan menjadikannya sebagai *base estimator* pada *Adaboost Classifier*.

```
Confusion matrix

[[171  28]
[ 47  62]]
```

Gambar 8. *Confusion matrix Naïve Bayes + Adaboost*

	Model	Accuracy	Recall	Precision	Specificity	F1_score
0	Naive Bayes + AdaBoost	0.769481	0.569	0.721	0.879	0.636

Gambar 9. Hasil akurasi *Naïve Bayes + Adaboost*

Pada Gambar 8 hasil eksekusi *Confusion matrix* menunjukkan jumlah *True Positif* (TP) adalah 171, *True Negatif* (TN) adalah 62, *False Positif* (FP) adalah 28, dan untuk *False Negatif* (FN) adalah 47. Hasil akurasi *Naïve Bayes + Adaboost* dapat dilihat pada Gambar 9 yaitu sebesar 0,7694.

3.4.3 Analisis Pembahasan

Dari hasil pengujian yang dipaparkan dengan dilakukannya evaluasi model tunggal maupun model kombinasi dapat dilihat hasil yang memiliki akurasi paling tinggi adalah optimasi *Adaboost* dengan *Naïve Bayes* dengan akurasi sebesar 0,7694. Berikut merupakan perbedaan hasil akurasi dapat dilihat pada Tabel 1 dibawah ini:

Tabel 1. Hasil akurasi beberapa model

Model	Akurasi
Naïve Bayes	0,7608
Adaboost + Naïve Bayes	0,7964
(Selisih)	0,008

Dari tabel di atas dapat dilihat *diagnose* kedua model adalah *Good Classification* yang mana keduanya menggunakan ukuran *split* data yang sama yaitu 60/40 sehingga akhirnya didapat selisih sebesar 0,008.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Hasil penelitian untuk nilai akurasi Algoritma *Naïve Bayes* memiliki nilai 0.7608 dengan *split* data 60/40, sedangkan untuk nilai akurasi optimasi *Adaboost* dan *Naïve Bayes* senilai 0.7694 dengan *split* data yang sama. Kedua model tersebut memiliki diagnosa *Good Classification*, dalam pengujian prediksi penyakit diabetes dengan menggunakan *dataset* berjumlah 336 data yang terdiri dari 9 variabel.

Kelebihan Algoritma *Adaboost* adalah mengoptimasi algoritma *Naïve Bayes* sebagai algoritma *estimator* atau *weak learner* sehingga menghasilkan akurasi yang lebih baik.

4.2 Saran

Untuk penelitian di masa depan dapat dilakukan dengan metode *boosting* lain seperti *XGBoost* yang memiliki kemampuan bawaan untuk menangani *missing values.* Dapat pula dikembangkan *smart system* untuk memprediksi penyakit Diabetes melitus berdasarkan klasifikasi yang sudah dilakukan.

DAFTAR PUSTAKA

Abdurrahman, G. (2022). Klasifikasi Penyakit Diabetes Melitus Menggunakan Adaboost Classifier. *Jurnal Sistem dan Teknologi Informasi*, 7(1), pp.59 - 65.

Argina, A.M. (2020). Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*, 1(2), pp.29–33.

Astuti, Y. *et al.* (2022). Naïve Bayes untuk Prediksi Tingkat Pemahaman Kuliah Online Terhadap Mata Kuliah Algoritma

Struktur Data. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 8(1), pp.28–32.

Byna, A. & Basit, M. (2020). Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 9(3), pp.407–411. https://doi.org/10.32736/sisfokom.v9i3.1023.

Defiyanti, S. (2015). Integrasi Metode Klasifikasi Dan Clustering dalam Data Mining. *KNIF 9 (Konferensi Nasional Informatika)*, pp.39–44.

Gultom, S.I. (2020). Implementasi Data Mining Menentukan Pola Hidup Sehat Bagi Pengguna KB Menggunakan Algoritma Adaboost (Studi Kasus :Dinas Serdang Bedagai). *Jurnal Infromasi dan Teknologi Ilmiah (INTI)*, 7(3), pp.298-304.

Indra Borman, R. (2020). Penerapan Data Maining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes. *Jurnal Ilmiah Ilmu Komputer*, 9(1), pp.25-34.

Lestari, E.T. a& Adhiva, J. (2022). Implementasi Algoritma Naive Bayes Classifier dan K-Nearest Neighbor Untuk Klasifiasi Status Gizi Obesitas Anak Disabilitas. *SENTIMAS: Seminar Masional Penelitian dan Penabdian Masyarakat*, pp.1–11.

Marito Putry, N. & Nurina Sari, B. (2022). Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Melitus. *Jurnal Sains dan Manajemen*, 10(1), pp.45-57.

Maulidah, N. *et al.* (2021). Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes. *Indonesian Journal on Software Engineering (IJSE)*, 7(1), pp.63–68.

Murtiningsih, M.K., Pandelaki, K. & Sedli, B.P. (2021). Gaya Hidup sebagai Faktor Risiko Diabetes Melitus Tipe 2. *e-Clinic (ECL)*, 9(2), pp.328-333. https://doi.org/10.35790/ecl.v9i2.32852.

Nur Rais, A., Rahmawati, E. & Faizal Amir, R. (2021). Prediksi Pima Indians Diabetes Database Dengan Ensemble Adaboost Dan Bagging. *Jurnal Sains dan Manajemen*, 9(2), pp.36-42.

Pahlevi, Reza. (2021). *Jumlah Penderita Diabetes di Indonesia Diproyeksikan Capai 28,57 Juta pada 2045*, https://databoks.katadata.co.id/datapublish/2021/11/24/jumlah-penderita-diabetes-di-indonesia-diproyeksikan-capai-2857-juta-pada-2045.

Ramdan, A. *et al.* (2022). Prediksi Jaringan TOR dan VPN menggunakan Algoritma K-Nearest Neighbour pada Trafik Darknet. *Jurnal Sistem Cerdas*, 05(01), pp.21–35.

Sibarani, A.J.P. (2020). Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 7(2), pp.262–276. https://doi.org/10.35957/jatisi.v7i2.195.

Simanjuntak, A.Y., Simatupang, I.S.S. & Anita (2022). Implementasi Data Mining Menggunakan Metode Naïve Bayes Classifier Untuk Data Kenaikan Pangkat Dinas Ketenagakerjaan Kota Medan. *Journal of Sience and Social Research*, 4307(1), pp.85–91.

Sutoyo, E. & Almaarif, A. (2020). Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 1(3), pp.95–101.

Ula, M. & Faridhatul Ulva, A. (2021). Implementasi Machine Learning Dengan Model Case Based Reasoning Dalam Mendagnosa Gizi Buruk Pada Anak. *JIK (Jurnal Informatika Kaputama)*, 5(2), pp.333–339.

Zhang, Q. & Kong, X. (2020). Design of Automatic Lung Nodule Detection System Based on Multi-Scene Deep Learning Framework. *IEEE Access*, p.8, 90380–90389. https://doi.org/10.1109/ACCESS.2020.2993872.