

IMPLEMENTASI DATA MINING DALAM MELAKUKAN PREDIKSI PENYAKIT DIABETES MENGGUNAKAN METODE RANDOM FOREST DAN XGBOOST

Oleh:

Muhammad Salsabil

Nuril Lutvi Azizah

Program Studi Informatika

Universitas Muhammadiyah Sidoarjo

Agustus, 2023

Pendahuluan

Diabetes mellitus adalah masalah kesehatan global yang signifikan yang ditandai oleh peningkatan kadar gula darah akibat masalah dalam produksi atau penggunaan hormon insulin di dalam tubuh. Untuk menghadapi tantangan ini, pengembangan metode klasifikasi yang akurat dan efisien dalam mendiagnosis diabetes menjadi sangat penting.

Dalam beberapa tahun terakhir, penggunaan metode pembelajaran mesin untuk klasifikasi penyakit telah menjadi fokus penelitian yang meningkat. Metode ini memungkinkan para peneliti untuk menganalisis dataset yang besar dan kompleks dengan lebih efisien, sehingga dapat menghasilkan model yang dapat memprediksi dan mengklasifikasikan penyakit dengan tingkat akurasi yang tinggi.

Pendahuluan

Tujuan penelitian ini adalah menganalisis kinerja metode Random Forest dan XGBoost dalam klasifikasi penyakit diabetes. Penelitian ini menggunakan dataset yang terdiri dari berbagai fitur klinis dan biokimia yang relevan dengan diabetes, dan melatih model menggunakan kedua metode tersebut. Selanjutnya, penelitian ini akan mengevaluasi kinerja model menggunakan metrik yang umum digunakan seperti akurasi, presisi, recall, dan F1-score.

Harapannya, penelitian ini dapat memberikan pemahaman yang lebih baik tentang efektivitas dan keunggulan masing-masing metode dalam klasifikasi penyakit diabetes, serta memberikan wawasan baru dalam penggunaan Random Forest dan XGBoost untuk tujuan ini. Selain itu, hasil penelitian ini dapat menjadi dasar untuk pengembangan metode klasifikasi yang lebih baik dan akurat dalam bidang medis.

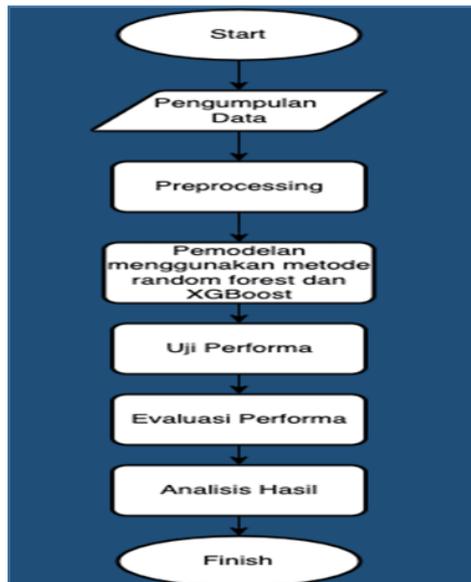
Pendahuluan

Pada penelitian Implementasi Data Mining Dalam Melakukan Prediksi Penyakit Diabetes ini, peneliti menggunakan database yang berasal dari kaggle, sebanyak 768 data dengan mencakup 9 indikator penyakit diabetes. Penelitian ini menggunakan 2 metode yaitu Random forest dan XGBoost untuk menganalisa prediksi dalam penyakit diabetes. Penelitian ini melalui beberapa tahapan-tahapan preprocessing untuk mengolah data-data awal sebelum di ujikan dengan kedua pemodelan topik antara random forest dan XGBoost, dilakukan pengujian menggunakan cross validation 5 untuk menentukan parameter terbaik. Menggunakan matrix evaluasi akurasi, presisi, recall dan f1 score.

Metode Penelitian

A. Desain Sistem

Perancangan perangkat lunak diawali dengan tahap preprocessing. Setelah tahap Preprocessing, maka dilakukan modelling menggunakan metode Random Forest dan XGBoost. Untuk mendapatkan parameter terbaik, peneliti menggunakan teknik Grid Search Cross Validation dengan cross validation 5 untuk melakukan tuning pada saat pemodelan. Berikut desain sistem alur program yang dibuat,



Gambar di samping, merupakan alur perancangan sistem pada penelitian ini, berjalan melalui beberapa proses mulai dari pengumpulan data, preprocessing data awal, processing dan pemodelan metode random forest dan Xgboost, hingga tahap analisis hasil supaya menciptakan hasil yang akurat.

Metode Penelitian

B. Data

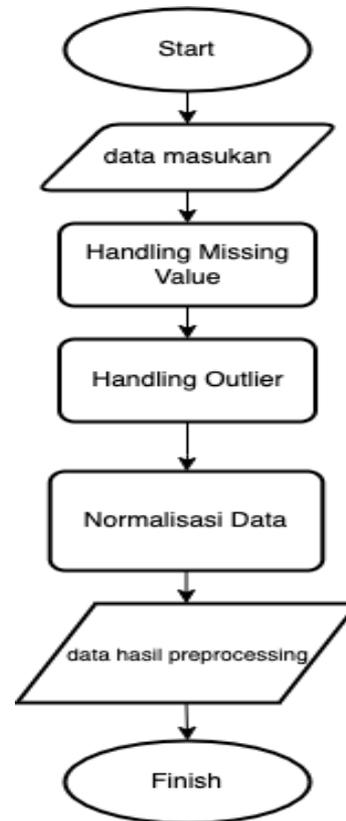
Data yang digunakan dalam penelitian ini berasal dari Kaggle dan terdiri dari 768 entri. Dataset ini mencakup 9 indikator yang terkait dengan penyakit diabetes.

Attribut	Keterangan
Pregnancies	Jumlah Kehamilan
Glucose	Konsentrasi glukosa plasma 2 jam setelah uji toleransi glukosa oral
Blood Pressure	Tekanan darah diastolik
Skin Thickness	Ketebalan lipatan kulit trisep
Insulin	Insulin serum 2 jam
BMI	Indikator untuk menentukan kategori berat badan
Diabetes Pedigree Function	Fungsi silsilah diabetes
Age	umur
Outcome	kelas

Metode Penelitian

B. Preprocessing

Flowchart preprocessing dalam penelitian ini sebagai berikut:



Metode Penelitian

Handling Missing Value

Pada tahapan awal preprocessing, data masukan dilakukan handling missing value terlebih dahulu. Missing value terjadi karena kesalahan penginputan data atau memang data tersebut tidak ada. Karena algoritma machine learning tidak dapat memproses data yang terdapat missing value, maka sebelum dilakukan modelling harus dilakukan handling missing value terlebih dahulu. Peneliti menggunakan Teknik imputasi mean. Sehingga data yang terdapat missing value diisi dengan nilai rata-rata dari kolom tersebut.

Handling Outlier

Selanjutnya, tahapan berikutnya adalah melakukan penanganan outlier menggunakan teknik Z-score. Z-score digunakan untuk membantu mengidentifikasi apakah suatu data termasuk dalam kategori outlier atau bukan. Data outlier merujuk pada data yang memiliki nilai yang sangat jauh dari rata-rata. Aturan umum yang digunakan adalah jika nilai Z-score kurang dari -3 atau lebih dari +3, maka data tersebut dianggap sebagai nilai ekstrem. Oleh karena itu, data yang melebihi batas bawah atau batas atas tersebut akan dihapus dari dataset.

Normalisasi Data

Normalisasi data adalah salah satu teknik penting dalam tahap preprocessing. Hal ini penting karena seringkali data memiliki rentang nilai yang berbeda antar variabelnya. Dalam penelitian ini, peneliti menggunakan metode min-max scaler untuk melakukan normalisasi data. Metode ini akan menyesuaikan nilai-nilai data ke dalam rentang yang ditentukan, biasanya 0 hingga 1, sehingga memungkinkan perbandingan yang lebih adil antar variabel.

Metode Penelitian

C. Processing

Klasifikasi dengan Random Forest dan XGBoost :

Langkah pertama dari tahap klasifikasi dengan Random Forest dan XGBoost adalah membuka data yang telah dilakukan ekstraksi fitur ke dalam jupyter notebook menggunakan library pandas. Ketika data sudah berhasil diload, maka dilakukan pembagian data antara data X dan data y dimana data X merupakan kolom fitur dan data y merupakan kolom target.

Setelah melakukan pembagian data X dan y, kemudian dilakukan pembagian data train dan data test pada data X menggunakan modul scikit-learn yaitu `train_test_split`. Besaran pembagian data yaitu 80% untuk data train dan 20% untuk data test.

Untuk mendapatkan parameter yang paling optimal pada kasus ini, peneliti menggunakan Teknik `GridSearchCV` yang mana Teknik ini dapat mencari parameter optimal dari algoritma yang digunakan untuk kasus yang sedang dianalisa. `GridSearchCV` adalah bagian dari modul scikit-learn yang bertujuan secara otomatis dan sistematis melakukan validasi beberapa model dan setiap hyperparameter. Ketika proses running `GridSearchCV` sudah selesai, maka akan didapatkan model beserta score test dan score train.

Metode Penelitian

D. Tahap Evaluasi

Tahapan ini digunakan untuk mengukur performa dari model machine learning yang telah dibuat. terdapat tiga metrik evaluasi yang dapat digunakan yaitu precision, recall dan confusion matrix. Seperti yang telah dijelaskan pada sub bab sebelumnya, data akan dibagi menjadi data train dan data test dengan perbandingan.

Agar mendapatkan hasil terbaik, beberapa perbandingan data train dan data test nantinya akan dilakukan percobaan pada masing-masing perbandingan. Pada penelitian ini, peneliti menggunakan Randomized Search Cross Validation dengan cross validation adalah 5. Sehingga dataset dibagi menjadi 5 bagian data sama banyak. Jika bagian 1 menjadi data test maka bagian 2 hingga 5 menjadi data train. Sedangkan jika bagian 2 menjadi data test maka bagian 1, 3, 4 dan 5 menjadi data train. Begitu seterusnya hingga bagian 5 menjadi data test.

Hasil dan Pembahasan

A. Data

Penelitian ini berfokus pada analisis data yang terdiri dari 768 entri. Dataset ini mencakup 9 indikator yang memiliki keterkaitan dengan penyakit diabetes. Dalam penelitian ini, 9 kategori yang mewakili indikator-indikator tersebut diidentifikasi. Fokus penggunaan 9 indikator ini adalah untuk membatasi ruang lingkup masalah dalam prediksi penyakit diabetes. Indikator-indikator tersebut meliputi:

1. Pregnancies
2. Glucose
3. Blood Pressure
4. Skin Thickness
5. Insulin
6. BMI
7. Diabetes Pedigree Function
8. Age
9. Outcome

Hasil dan Pembahasan

Tabel Indikator Data

Attribut	Keterangan
Pregnancies	Jumlah Kehamilan
Glucose	Konsentrasi glukosa plasma 2 jam setelah uji toleransi glukosa oral
Blood Pressure	Tekanan darah diastolik
Skin Thickness	Ketebalan lipatan kulit trisep
Insulin	Insulin serum 2 jam
BMI	Indikator untuk menentukan kategori berat badan
Diabetes Pedigree Function	Fungsi silsilah diabetes
Age	umur
Outcome	kelas

Dengan mempertimbangkan 9 indikator ini, penelitian ini bertujuan untuk lebih memahami dan merumuskan prediksi penyakit diabetes dengan kerangka kerja yang kokoh.

Hasil dan Pembahasan

B. Preprocessing

1. Handling Missing Value

Output yang dihasilkan dari langkah ini adalah data yang telah mengalami penyaringan dan perubahan.

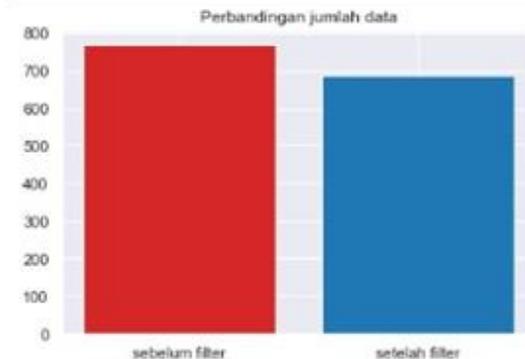


Dapat diamati dari gambar di atas, proses pengisian data dilakukan secara komprehensif sehingga tidak ada nilai yang terlewatkan. Presentasi visual dalam gambar mencerminkan keseragaman warna yang sama. Pengamatan ini mengindikasikan bahwa pengisian nilai telah berhasil dilakukan dengan sukses.

Hasil dan Pembahasan

2. Handling Outlier

Tahap berikutnya melibatkan penanganan outlier menggunakan metode Z-score. Z-score digunakan untuk membantu mengidentifikasi apakah suatu data termasuk dalam kategori outlier atau tidak. Data outlier merujuk pada data yang memiliki nilai yang signifikan dari rata-rata. Pedoman umum adalah jika nilai Z-score lebih kecil dari -3 atau lebih besar dari $+3$, data dianggap sebagai nilai ekstrem. Karena itu, data yang melampaui batas-batas ini akan dihilangkan dari dataset.



Setelah menerapkan metode Z-score untuk mengatasi outlier, gambar di atas menunjukkan bahwa dari jumlah data awal sebanyak 768, jumlah data yang tersisa setelah proses tersebut adalah 688.

Hasil dan Pembahasan

3. Normalisasi Data

Normalisasi data adalah langkah kunci dalam preprocessing, mengingat variasi nilai antar variabel. Dalam penelitian ini, metode yang diterapkan adalah skala min-max. Pendekatan ini mengatur nilai data dalam rentang 0 hingga 1 untuk perbandingan yang lebih adil.

```
Data awal:
Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI \
0           6.0    148.0         72.0           35.0        0.0   33.6
1           1.0    85.0          66.0           29.0        0.0   26.6
2           8.0    183.0         64.0           0.0         0.0   23.3
3           1.0    89.0          66.0           23.0        94.0  28.1
5           5.0    116.0         74.0           0.0         0.0   25.6
...
763        10.0   101.0         76.0           48.0       180.0  32.9
764         2.0   122.0         70.0           27.0        0.0   36.8
765         5.0   121.0         72.0           23.0       112.0  26.2
766         1.0   126.0         60.0           0.0         0.0   30.1
767         1.0   93.0          70.0           31.0        0.0   30.4

DiabetesPedigreeFunction  Age
0                0.627  50.0
1                0.351  31.0
2                0.672  32.0
3                0.167  21.0
5                0.201  30.0
...
763              0.171  63.0
764              0.340  27.0
765              0.245  30.0
766              0.349  47.0
767              0.315  23.0

[688 rows x 8 columns]
Data yang sudah dinormalisasi:
[[0.46153846 0.67096774 0.48979592 ... 0.41847826 0.39696312 0.61702128]
 [0.07692308 0.26451613 0.42857143 ... 0.22826087 0.19739696 0.21276596]
 [0.61538462 0.89677419 0.40816327 ... 0.13858696 0.42950108 0.23404255]
 ...
 [0.38461538 0.49677419 0.48979592 ... 0.2173913 0.12075199 0.19148936]
 [0.07692308 0.52983226 0.36734694 ... 0.32336957 0.19595083 0.55319149]
 [0.07692308 0.31612903 0.46938776 ... 0.33152174 0.17136659 0.04255319]]
```

Metode yang digunakan melibatkan penerapan skala min-max. Pendekatan ini mengelola nilai-nilai data agar berada dalam kisaran 0 hingga 1, untuk memastikan kesetaraan perbandingan.

Hasil dan Pembahasan

C. Processing

Cross-validation adalah pendekatan yang membagi dataset menjadi beberapa bagian untuk tujuan pelatihan dan pengujian model. Dalam konteks Random Forest, strategi ini membantu mengukur sejauh mana model mampu menggeneralisasi data yang belum pernah dilihatnya[20]. Model diperlakukan sebagai pelatihan dan diuji pada setiap bagian dataset, sambil menghitung metrik evaluasi yang relevan[21]. Hasil dari setiap tahapan diambil untuk memberikan gambaran yang lebih akurat tentang performa keseluruhan model Berikut merupakan parameter-parameter terbaik dalam random forest dan XGBoost:

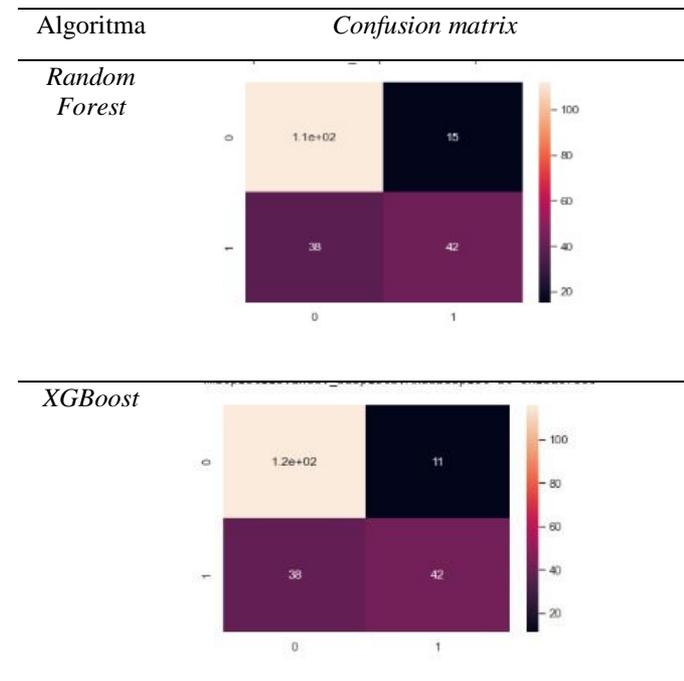
Algoritma	Parameter
<i>Random Forest</i>	<pre>Fitting 5 folds for each of 10 candidates, totalling 50 fits [Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers. [Parallel(n_jobs=-1)]: Done 34 tasks elapsed: 10.8s [Parallel(n_jobs=-1)]: Done 50 out of 50 elapsed: 12.7s finished 0.9646569646569647 0.713273195876289 0.7439613526570048]: model_rf.best_params_]: {'algo_max_depth': 29, 'algo_max_features': 0.9051404267682241, 'algo_min_samples_leaf': 3, 'algo_n_estimators': 110}</pre>
<i>XGBoost</i>	<pre>Fitting 5 folds for each of 10 candidates, totalling 50 fits [Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers. [Parallel(n_jobs=-1)]: Done 34 tasks elapsed: 49.5s [Parallel(n_jobs=-1)]: Done 50 out of 50 elapsed: 49.8s finished 0.7879417879417879 0.7754295532646047 0.7294685990338164 model_xgb.best_params_ {'algo_colsample_bytree': 0.293530703818476, 'algo_gamma': 10, 'algo_learning_rate': 0.35561914931255884, 'algo_max_depth': 2, 'algo_n_estimators': 160, 'algo_reg_alpha': 1.5611961512098822, 'algo_reg_lambda': 0.04563169215430241, 'algo_subsample': 0.32688076857370457}</pre>

Hasil dan Pembahasan

D. Evaluasi

Setelah melalui serangkaian langkah-langkah penting dalam preprocessing dan tahap klasifikasi data, langkah selanjutnya yang dijalankan adalah tahap evaluasi. Pada tahap ini, peneliti melanjutkan dengan menggunakan confusion matrix sebagai instrumen untuk melakukan evaluasi mendalam terhadap performa model yang telah dibangun.

Algoritma	Clasification Report				
<i>Random Forest</i>		precision	recall	f1-score	support
	0.0	0.75	0.88	0.81	127
	1.0	0.74	0.53	0.61	80
	accuracy			0.74	207
	macro avg	0.74	0.70	0.71	207
	weighted avg	0.74	0.74	0.73	207
	<i>XGBoost</i>		precision	recall	f1-score
0.0		0.75	0.91	0.83	127
1.0		0.79	0.53	0.63	80
accuracy				0.76	207
macro avg		0.77	0.72	0.73	207
weighted avg		0.77	0.76	0.75	207



Hasil dan Pembahasan

D. Evaluasi

Berdasarkan tabel di atas, berikut hasilnya:

Random Forest :

1. Akurasi keseluruhan: 74%
2. Rata-rata (macro avg): Presisi 0.74, Recall 0.70, F1-Score 0.71
3. Rata-rata berbobot (weighted avg): Presisi 0.74, Recall 0.74, F1-Score 0.73

XGBoost :

1. Akurasi keseluruhan: 76%
2. Rata-rata (macro avg): Presisi 0.77, Recall 0.72, F1-Score 0.73
3. Rata-rata berbobot (weighted avg): Presisi 0.77, Recall 0.76, F1-Score 0.75

Penutup

Dalam Penelitian Implementasi Data Mining Dalam Melakukan Prediksi Penyakit Diabetes Menggunakan Metode Random Forest Dan Xgboost, Penerapan teknik data mining melalui Random Forest dan XGBoost dalam meramalkan penyakit diabetes keduanya memberikan hasil prediksi yang akurat dan konsisten berdasarkan analisis dataset klinis dan biokimia yang didapat dari situs kaggle yang berjumlah 768 dan 9 indikator. Data diolah melalui tahap preprocessing diantaranya handling missing value, handling outlier dan normalisasi data, dan didapatkan data yang akan diolah sebesar 688. Setelah didapat data hasil preprocessing, data dilakukan tahapan pelatihan dan pengujian dengan Cross Validation dan dilakukan pengujian untuk mengetahui parameter-parameter terbaik yang akan digunakan, lalu dilakukan evaluasi kinerja model Random Forest dan XGBoost menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil evaluasi model menunjukkan performa yang baik dalam penelitian ini, didapatkan hasil akurasi keseluruhan dalam penggunaan random forest sebesar 74% dan penggunaan XGBoost sebesar 76%. Kesimpulan ini menyoroti peran penting data mining dalam mengatasi tantangan kesehatan global, penelitian ini berpotensi mendukung upaya deteksi dini dan pengelolaan penyakit khususnya diabetes.

Daftar Pustaka

- [1] T. Hidayat, S. S. Anelia, R. I. Pratiwi, N. Salsabila, and D. S. Prasvita, "Perbandingan Akurasi Klasifikasi Penyakit Diabetes Menggunakan Algoritma Adaboost- Random Forest Dan Adaboost- Decision Tree Dengan Imputasi Median Dan Knn," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, no. April, pp. 616–623, 2020.
- [2] G. Abdurrahman, H. Oktavianto, and M. Sintawati, "Optimasi Algoritma XGBoost Classifier Menggunakan Hyperparameter Gridsearch dan Random Search Pada Klasifikasi Penyakit Diabetes," *INFORMAL Informatics J.*, vol. 7, no. 3, p. 193, 2022, doi: 10.19184/isj.v7i3.35441.
- [3] A. Fauzi and A. H. Yunial, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Optimasi Algoritma Klasifikasi Naive Bayes, Decision Tree, K-Nearest Neighbor, dan Random Forest menggunakan Algoritma Particle Swarm Optimization pada Diabetes Dataset," *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 3, pp. 470–481, 2022.
- [4] F. ANISHA, Dodi Vionanda, Nonong amalita, and Zilrahmi, "Application of Random Forest for The Classification Diabetes Mellitus Disease in RSUP Dr. M. Jamil Padang," *UNP J. Stat. Data Sci.*, vol. 1, no. 2, pp. 45–52, 2023, doi: 10.24036/ujsds/vol1-iss2/30.
- [5] N. N. Pandika Pinata, I. M. Sukarsa, and N. K. Dwi Rusjyanthi, "Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 8, no. 3, p. 188, 2020, doi: 10.24843/jim.2020.v08.i03.p04.

Daftar Pustaka

- [6] Gde Agung Brahmana Suryanegara, Adiwijaya, and Mahendra Dwifabri Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 114–122, 2021, doi: 10.29207/resti.v5i1.2880.
- [7] S. D. M. I. Gunawan Ichsan Muhammad, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode Grid Search pada Algoritma Logistic Regression," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, pp. 280–284, 2020.
- [8] B. A. Candra Permana and I. K. Dewi Patwari, "Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 63–69, 2021, doi: 10.29408/jit.v4i1.2994.
- [9] H. S. W. Hovi, A. Id Hadiana, and F. Rakhmat Umbara, "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)," *Informatics Digit. Expert*, vol. 4, no. 1, pp. 40–45, 2022, doi: 10.36423/index.v4i1.895.
- [10] M. Syukri Mustafa and I. Wayan Simpen, "Implementation of the K-Nearest Neighbor (KNN) Algorithm to Predict Patients Affected by Diabetes at the Manyampa Health Center, Bulukumba Regency," *Pros. Semin. Ilm. Sist. Indormasi dan Teknol. Inf.*, vol. VIII, no. 1, pp. 1–10, 2019.

Daftar Pustaka

- [11] S. Ucha Putri, E. Irawan, F. Rizky, S. Tunas Bangsa, P. A. -Indonesia Jln Sudirman Blok No, and S. Utara, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5," Januari, vol. 2, no. 1, pp. 39–46, 2021.
- [12] E. C. P. Witjaksana, R. R. Saedudin, and V. P. Widartha, "Perbandingan Akurasi Algoritma Random Forest dan Algoritma Artificial Neural Network untuk Klasifikasi Penyakit Diabetes," e-Proceeding Eng., vol. 8, no. 5, pp. 9765–9772, 2021.
- [13] M. D. Purbolaksono, M. Irvan Tantowi, A. Imam Hidayat, and A. Adiwijaya, "Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 5, no. 2, pp. 393–399, 2021, doi: 10.29207/resti.v5i2.3008.
- [14] Z. Munawar, N. I. Putri, and D. Z. Musadad, "Meningkatkan Rekomendasi Menggunakan Algoritma Perbedaan Topik," J. Sist. Inf., vol. 01, no. 02, pp. 17–26, 2020.
- [15] A. E. Pramadhani and T. Setiadi, "Penerapan Data Mining untuk Klasifikasi Penyakit ISPA dengan Algoritma Decision Tree," J. Sarj. Tek. Inform. e-ISSN 2338-5197, vol. 2, no. 1, pp. 831–839, 2014.
- [16] N. Chamidah, W. Jurusan Teknik Informatika, U. Salamah, and K. Kunci Backpropagasi, "Pengaruh Normalisasi Data pada Jaringan Syaraf Tiruan Backpropagasi Gradient Descent Adaptive Gain (BPGDAG) untuk Klasifikasi," J. Itsmart, vol. 1, no. 1, pp. 28–33, 2012.

Daftar Pustaka

- [17] W. Aprilia, I. Kurniawan, M. Baydhowi, and T. Haryati, "SISTEMASI: Jurnal Sistem Informasi Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *J. Sist. Inf.*, vol. 10, no. 1, pp. 163–171, 2021, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>.
- [18] R. Bonetto and V. Latzko, "Machine learning," *Comput. Commun. Networks From Theory to Pract.*, pp. 135–167, 2020, doi: 10.1016/B978-0-12-820488-7.00021-9.
- [19] N. L. Rachmawati and M. Lentari, "Penerapan Metode Min-Max untuk Minimasi Stockout dan Overstock Persediaan Bahan Baku," *J. INTECH Tek. Ind. Univ. Serang Raya*, vol. 8, no. 2, pp. 143–148, 2022, doi: 10.30656/intech.v8i2.4735.
- [20] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [21] S. Kasus, K. Data, E. Gen, and S. Muscle, "Implementasi Metode Svm , Mlp Dan Xgboost," 2020.



Universitas
Muhammadiyah
Sidoarjo

Terima Kasih



www.umsida.ac.id



[umsida1912](https://www.instagram.com/umsida1912)



[umsida1912](https://twitter.com/umsida1912)



universitas
muhammadiyah
sidoarjo



[umsida1912](https://www.youtube.com/umsida1912)