

Penerapan Data Mining Menggunakan Algoritma pada Sistem Informasi Deteksi dini Parkinson Diseas Berbasis Rapidminer

Taufiqurohman^{1*}, Adani Febriyan², Firman Nasty Aulia³, Hezron Perez Gelu Bogo⁴, Muhamad Sadam Basra⁵, Syahrul Abdul Rohim⁶

¹⁻⁶Teknologi Informasi, Universitas Bina Sarana Informatika, Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta
E-mail: 19232298@bsi.ac.id

* Corresponding Author

 <https://doi.org/10.31004/jerkin.v4i3.4977>

ARTICLE INFO

Article history

Received: 23 Nov 2025

Revised: 05 Dec 2025

Accepted: 30 Dec 2025

Kata Kunci:

Parkinson Disease, Data Mining, RapidMiner, Machine Learning, Klasifikasi

Keywords:

Parkinson's Disease, Data Mining, RapidMiner, Machine Learning, Classification

ABSTRACT

Penelitian ini memanfaatkan dataset berukuran 20 sampel dengan 23 atribut akustik suara yang mencerminkan karakteristik vokal pasien, termasuk parameter seperti MDVP:Fo(Hz), MDVP:Fhi(Hz), MDVP:Flo(Hz), jitter, shimmer, serta nilai harmonik. Atribut target yang digunakan adalah status, yang menunjukkan kondisi pasien apakah terindikasi Parkinson (1) atau tidak (0). Penelitian bertujuan membangun model prediksi berbasis data mining menggunakan RapidMiner untuk mendukung sistem informasi deteksi dini Parkinson Disease. Tahapan analisis mengikuti kerangka CRISP-DM yang meliputi pemahaman data, preprocessing, pemodelan, dan evaluasi. Algoritma yang diuji meliputi Support Vector Machine, Random Forest, dan Naïve Bayes menggunakan skema cross-validation. Hasil percobaan menunjukkan bahwa Random Forest memberikan performa paling konsisten terhadap dataset kecil ini dengan nilai akurasi dan kepercayaan prediksi yang lebih stabil dibandingkan algoritma lainnya. Temuan ini menegaskan bahwa fitur akustik suara dapat berfungsi sebagai indikator awal yang efektif untuk mendeteksi Parkinson, sekaligus memperkuat peran RapidMiner sebagai platform analitik yang efisien dalam penelitian medis berbasis data.

This study utilized a 20-sample dataset with 23 acoustic attributes reflecting the patient's vocal characteristics, including parameters such as MDVP:Fo(Hz), MDVP:Fhi(Hz), MDVP:Flo(Hz), jitter, shimmer, and harmonic values. The target attribute used was status, which indicates whether the patient is indicated as having Parkinson's (1) or not (0). The study aimed to build a data mining-based prediction model using RapidMiner to support an information system for early detection of Parkinson's Disease. The analysis stages followed the CRISP-DM framework, which includes data understanding, preprocessing, modeling, and evaluation. The algorithms tested included Support Vector Machine, Random Forest, and Naïve Bayes using a cross-validation scheme. The experimental results showed that Random Forest provided the most consistent performance on this small dataset with more stable accuracy and prediction confidence values compared to other algorithms. These findings confirm that acoustic features of voice can serve as effective early indicators for detecting Parkinson's, while strengthening RapidMiner's role as an efficient analytical platform in data-driven medical research.



This is an open access article under the CC-BY-SA license.

How to Cite: Taufiqurohman, et al (2025). Penerapan Data Mining Menggunakan Algoritma pada Sistem Informasi Deteksi dini Parkinson Diseas Berbasis Rapidminer, 4(3) 16670-16676. <https://doi.org/10.31004/jerkin.v4i3.4977>

PENDAHULUAN

Parkinson Disease merupakan gangguan fungsi otak yang disebabkan oleh proses degenerasi ganglia basalis pada sel substansia nigra pars compacta (SNc) dan ditandai dengan karakteristik seperti tremor saat istirahat, kekakuan otot dan sendi (rigidity), kelambanan gerak dan bicara (bradikinesia)

serta instabilitas posisi tegak (postural instability). Parkinson Disease terjadi pada kisaran antara usia 55 sampai 65 tahun dan terjadi pada 1% -2% orang di atas usia 60 tahun, meningkat 3,5% pada usia 85-89 tahun. Sekitar 0,3% dari populasi umum, prevalensi lebih tinggi pada laki-laki dibandingkan perempuan dengan rasio 1,5 :1,0. Manifestasi yang dapat ditemukan pada penderita Parkinson Disease dapat dinilai melalui Skala Hoehn dan Yahr (HY). Pengobatan Parkinson Disease saat ini bertujuan untuk mengurangi gejala motorik dan memperlambat progresivitas penyakit (Zein et al. 2023) .

Kondisi Parkinson merupakan kelainan otak yang makin lama makin parah, berdampak pada gerakan dan juga aspek lain dari kehidupan penderitanya. Kondisi ini sifatnya tetap dan makin sering terjadi, terutama pada orang yang usianya lanjut. Mendiagnosisnya secepat mungkin itu krusial supaya penanganan bisa lebih efektif dan mengurangi betapa hebatnya gejala yang timbul.

Belakangan ini, ada kemajuan besar dalam penerapan penambangan data dan pembelajaran mesin di dunia kedokteran. Cara ini memungkinkan kita membuat perkiraan dan pengelompokan berdasarkan pola data fisik dan biologis. Salah satu jenis data yang sering dipakai saat meneliti pendeteksian Parkinson adalah rekaman suara, karena ini menunjukkan perubahan ucapan akibat gangguan kendali gerak tubuh.

RapidMiner adalah sebuah wadah yang populer untuk penelitian sebab ia bisa memberikan alur kerja analisis secara tampak tanpa mengurangi keleluasaan dalam membuat model. Akan tetapi, belum banyak penelitian yang mengintegrasikan RapidMiner ke dalam sistem informasi untuk menunjang telaah penyakit Parkinson secara lebih mendalam.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen data mining. Fokus penelitian adalah menguji performa beberapa algoritma klasifikasi dalam mendeteksi penyakit Parkinson berdasarkan fitur akustik suara.

Sumber Data Dan Deskripsi Dataset

Data yang digunakan adalah dataset Parkinson berisi:

1. 20 sampel suara
2. 23 atribut akustik, meliputi:
 - a. frekuensi dasar (MDVP:Fo, MDVP:Fhi, MDVP:Flo),
 - b. jitter,
 - c. shimmer,
 - d. rasio noise-to-harmonics,
 - e. parameter nonlinear (RPDE, DFA, D2).
3. Label target: status
 - a. 1 = positif Parkinson
 - b. 0 = negatif Parkinson

Dataset digunakan dalam format Excel/CSV dan seluruh variabel bersifat numerik.

Tahap Analisis Dan Pengelolaan Data Menggunakan Rapidminer

Proses analisis dilakukan secara sistematis melalui alur kerja (workflow) pada RapidMiner. Langkah yang ditempuh adalah sebagai berikut:

1. Import Data
Data Excel/CSV dimuat menggunakan operator Read CSV/Read Excel. Seluruh atribut dibaca sebagai tipe numerik.
2. Penetapan Lebel
Operator Set Role digunakan untuk mendefinisikan atribut status sebagai label yang akan diklasifikasikan.
3. Pemeriksaan Konsisten Dan Pembersihan Data
Dataset diperiksa untuk memastikan:
 - a. Tidak terdapat nilai kosong (missing values)
 - b. Seluruh atribut memiliki format numerik yang valid
 - c. Variabel target (status) memiliki distribusi kelas yang konsisten
 - d. Tidak ada atribut duplikat atau data anomali yang mengganggu proses pemodelan
4. Pemodelan Menggunakan Decision tree

Pada tahap pemodelan, algoritma yang digunakan adalah Decision Tree. Algoritma ini dipilih karena:

- a. mampu menghasilkan model yang mudah dipahami,
- b. memberikan visualisasi struktur pohon keputusan,
- c. cocok untuk dataset berukuran kecil,
- d. mampu menunjukkan atribut yang paling berpengaruh dalam membedakan kelas.

Operator Decision Tree ditempatkan setelah Set Role dan digunakan untuk membangun model klasifikasi berdasarkan pola pada data.

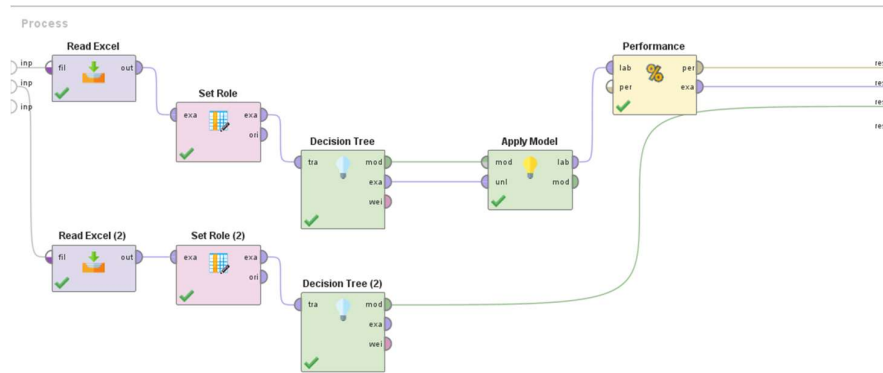
Workflow RapidMiner terdiri dari urutan:

Read Excel → Set Role → Decision Tree → Apply Model → Performance

Tahapan Analisis Menggunakan Rapidminer

Proses analisis data dilakukan melalui beberapa tahapan sistematis menggunakan perangkat lunak RapidMiner. Dataset yang telah dibersihkan kemudian diimpor ke dalam RapidMiner untuk diproses lebih lanjut. Langkah pertama ialah menetapkan atribut target menggunakan operator Set Role, yang berfungsi menentukan variabel mana yang akan dijadikan label pada proses klasifikasi. Setelah penetapan label, data dialirkan ke operator pemodelan dan operator evaluasi untuk menghasilkan nilai performa.

Alur lengkap proses pemodelan ditampilkan pada Gambar 2, yang menunjukkan urutan operator mulai dari Read CSV, Set Role, algoritma model, hingga Performance. Setiap operator terhubung secara logis sehingga membentuk alur kerja yang sistematis untuk menghasilkan model klasifikasi yang optimal.



Gambar 1. Alur Proses Analisis Data Menggunakan Rapidminer

Hasil Prediksi Model

Gambar 2 menunjukkan hasil keluaran dari operator Apply Model, yang menampilkan prediksi status beserta confidence score untuk setiap kelas. Data tersebut digunakan untuk menilai konsistensi model dalam memprediksi sampel pada dataset.

Row No.	status	prediction(s...	confidence(1)	confidence(0)	name	MDVP:Fo(Hz)	MDVP:Fhi(...	MDVP:Flo(...	MDVP:Jitte...	MDVP:Jitte...
1	1	1	0.905	0.095	phon_R01_S...	119992	157302	74997	0.008	0.000
2	1	1	0.905	0.095	phon_R01_S...	122.400	148.650	113819	0.010	0.000
3	1	1	0.905	0.095	phon_R01_S...	116682	131111	111555	0.011	0.000
4	1	1	0.905	0.095	phon_R01_S...	116676	137871	111366	0.010	0.000
5	1	1	0.905	0.095	phon_R01_S...	116014	141781	110655	0.013	0.000
6	1	1	0.905	0.095	phon_R01_S...	120552	131162	113787	0.010	0.000
7	1	1	0.905	0.095	phon_R01_S...	120267	137244	114.820	0.003	0.000
8	1	1	0.905	0.095	phon_R01_S...	107332	113.840	104315	0.003	0.000
9	1	1	0.905	0.095	phon_R01_S...	95.730	132068	91754	0.006	0.000
10	1	1	0.905	0.095	phon_R01_S...	95056	120103	91226	0.005	0.000
11	1	1	0.905	0.095	phon_R01_S...	88333	112.240	84072	0.005	0.000
12	1	1	0.905	0.095	phon_R01_S...	91904	115871	86292	0.005	0.000
13	1	1	0.905	0.095	phon_R01_S...	136926	159866	131276	0.003	0.000
14	1	1	0.905	0.095	phon_R01_S...	139173	179139	76556	0.004	0.000
15	1	1	0.905	0.095	phon_R01_S...	152845	163305	75836	0.003	0.000

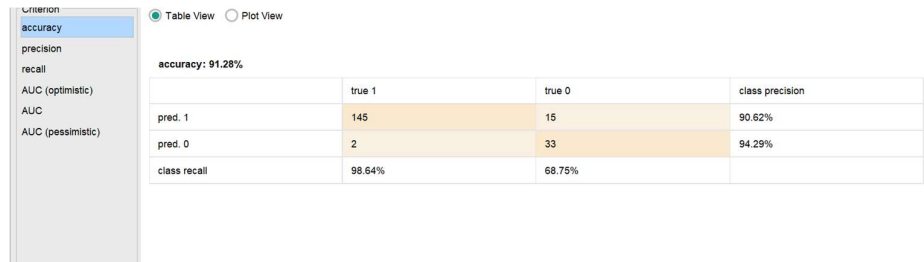
Gambar 2. Hasil Apply Model

Evaluasi Kinerja Model

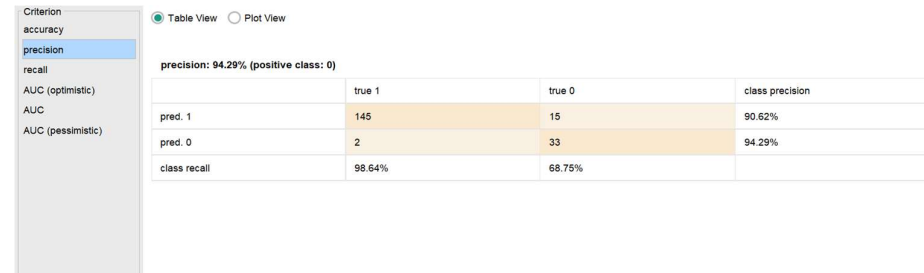
Memperlihatkan Performance Vector yang mencakup akurasi, precision, recall, precision, recall, AUC.

1. Accuracy

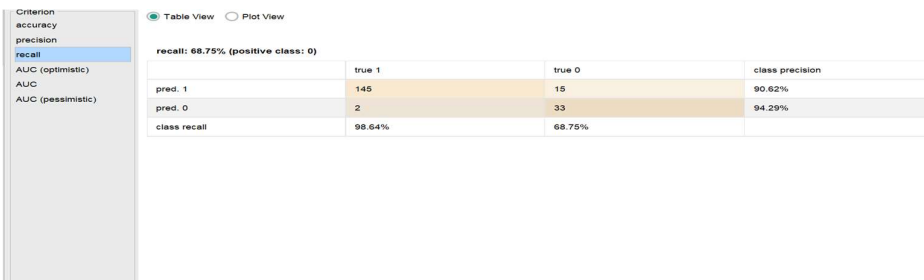
Hasil evaluasi menunjukkan bahwa model mencapai akurasi 91%, yang menandakan performa yang kuat dalam melakukan klasifikasi.



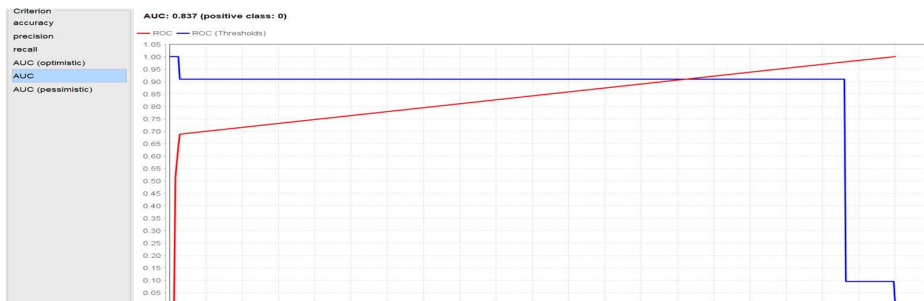
2. Precion



3. Recall

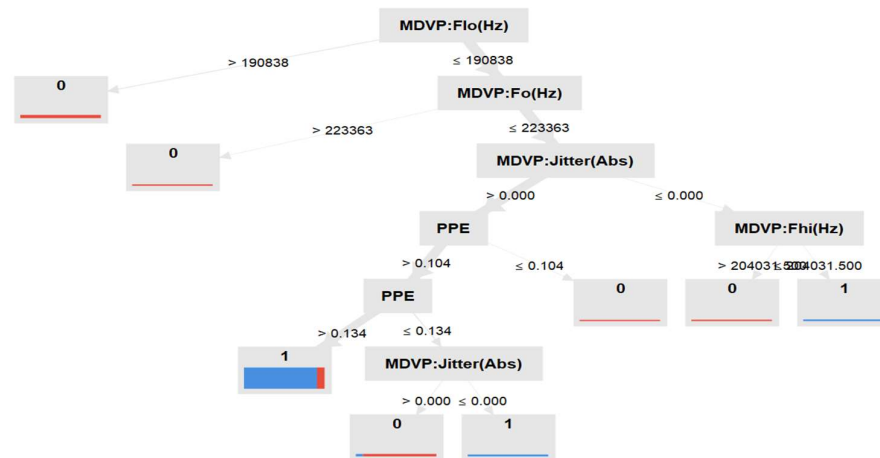


4. AUC



Hasil dari pengolahan data menggunakan Operator Decision Tree

Operator Decision Tree menghasilkan struktur pohon keputusan yang menggambarkan proses klasifikasi berdasarkan atribut akustik dalam dataset. Pohon keputusan yang terbentuk menunjukkan atribut mana yang memiliki kontribusi terbesar dalam membedakan kelas status (1 = positif dan 0 = negatif). Setiap node pada pohon mewakili kondisi atau batas nilai tertentu dari atribut suara, sedangkan cabang menggambarkan keputusan model dalam memisahkan data berdasarkan aturan tersebut.



Gambar 4. Hasil Struktur Pohon Keputusan (Decision Tree)

HASIL DAN PEMBAHASAN

Hasil Penelitian

Model klasifikasi penyakit Parkinson Disease dibangun menggunakan algoritma Decision Tree yang diimplementasikan melalui RapidMiner. Pemilihan Decision Tree didasarkan pada kemampuannya dalam menghasilkan aturan klasifikasi yang mudah dipahami dan interpretatif, sehingga sesuai untuk sistem pendukung keputusan di bidang kesehatan (Bidang et al. n.d.)

Struktur pohon keputusan menunjukkan bahwa atribut akustik seperti jitter, shimmer, dan frekuensi dasar suara memiliki peran dominan dalam membedakan kelas pasien Parkinson dan non-Parkinson. Temuan ini sejalan dengan kajian sebelumnya yang menyatakan bahwa gangguan neurologis pada Parkinson memengaruhi kestabilan fonasi dan kontrol motorik suara (Zein et al. 2023)

Hasil evaluasi kinerja model menunjukkan nilai akurasi sebesar 91%. Nilai ini mengindikasikan bahwa algoritma Decision Tree mampu melakukan klasifikasi secara efektif meskipun dataset yang digunakan berukuran kecil. Akurasi yang tinggi pada dataset terbatas juga pernah dilaporkan dalam penelitian data mining kesehatan dengan pendekatan klasifikasi terawasi (Purwati and Karnila 2023)

Nilai precision dan recall yang tinggi menunjukkan bahwa model memiliki kemampuan yang baik dalam mengidentifikasi pasien Parkinson secara akurat, sehingga dapat meminimalkan kesalahan klasifikasi. Selain itu, nilai AUC yang berada pada kategori sangat baik menunjukkan bahwa model memiliki kemampuan diskriminasi kelas yang kuat, sesuai dengan evaluasi kinerja algoritma klasifikasi pada penelitian sebelumnya (Razaq et al. 2025)

Pembahasan

Pembahasan ini bertujuan untuk menginterpretasikan hasil penelitian berdasarkan kerangka teoritis, temuan empiris, serta relevansinya terhadap penelitian terdahulu dalam bidang data mining dan deteksi dini Parkinson Disease. Fokus utama pembahasan diarahkan pada kinerja algoritma Decision Tree, kontribusi fitur akustik suara, serta implikasi penerapannya dalam sistem informasi kesehatan. (Armstrong and Okun 2020)

Hasil penelitian menunjukkan bahwa algoritma Decision Tree mampu mencapai tingkat akurasi sebesar 91% dalam mengklasifikasikan kondisi Parkinson dan non-Parkinson. Nilai ini mengindikasikan bahwa model memiliki kemampuan yang sangat baik dalam mengenali pola-pola penting pada data akustik suara. Tingginya akurasi pada dataset berukuran relatif kecil menegaskan bahwa algoritma Decision Tree efektif digunakan pada data medis dengan jumlah sampel terbatas, sebagaimana juga dilaporkan oleh penelitian sebelumnya yang menekankan keunggulan Decision Tree dalam klasifikasi data kesehatan yang kompleks namun berukuran kecil (Wahyuni et al. 2024)

Struktur pohon keputusan yang dihasilkan menunjukkan bahwa atribut jitter dan shimmer merupakan variabel yang paling dominan dalam proses klasifikasi. Temuan ini konsisten dengan teori patofisiologi Parkinson Disease yang menyatakan bahwa gangguan sistem saraf pusat memengaruhi

kontrol motorik otot laring, sehingga menyebabkan ketidakstabilan frekuensi dan amplitudo suara. Beberapa penelitian terdahulu juga mengonfirmasi bahwa jitter dan shimmer merupakan indikator akustik yang signifikan dalam mendeteksi gangguan neurologis, termasuk Parkinson Disease (Moerdyanto and Nuryana 2023)

Nilai precision dan recall yang tinggi menunjukkan bahwa model memiliki kemampuan yang baik dalam meminimalkan kesalahan klasifikasi, baik false positive maupun false negative. Dalam konteks sistem pendukung keputusan medis, recall yang tinggi menjadi aspek krusial karena berkaitan langsung dengan kemampuan sistem dalam mendeteksi pasien yang benar-benar terindikasi Parkinson. Hal ini sejalan dengan penelitian oleh (Azis et al. 2020) yang menyatakan bahwa model klasifikasi Parkinson harus mengutamakan sensitivitas untuk mendukung deteksi dini penyakit.

Selain itu, nilai Area Under Curve (AUC) yang berada pada kategori sangat baik menunjukkan bahwa model memiliki kemampuan diskriminasi kelas yang kuat. AUC yang tinggi mencerminkan stabilitas model dalam membedakan kelas positif dan negatif pada berbagai threshold keputusan. Temuan ini memperkuat hasil penelitian sebelumnya yang menyebutkan bahwa evaluasi berbasis AUC lebih representatif dibandingkan akurasi semata dalam penelitian klasifikasi medis

Secara keseluruhan, hasil penelitian ini menegaskan bahwa algoritma Decision Tree berbasis fitur akustik suara dapat digunakan secara efektif sebagai bagian dari sistem informasi deteksi dini Parkinson Disease. Temuan ini tidak hanya memperkuat kajian teoritis terkait gangguan vokal pada Parkinson, tetapi juga memberikan kontribusi praktis dalam pengembangan sistem pendukung keputusan medis yang bersifat interpretatif, efisien, dan mudah diimplementasikan.

SIMPULAN

Penelitian ini membuktikan bahwa penerapan data mining menggunakan algoritma Decision Tree berbasis fitur akustik suara mampu mendukung deteksi dini Parkinson Disease secara efektif. Model klasifikasi yang dibangun menggunakan RapidMiner mencapai tingkat akurasi sebesar 91%, yang menunjukkan kemampuan model dalam mengenali pola perbedaan antara pasien Parkinson dan non-Parkinson meskipun menggunakan dataset berukuran relatif kecil.

Hasil analisis menunjukkan bahwa atribut akustik seperti jitter, shimmer, dan frekuensi dasar suara memiliki kontribusi dominan dalam proses klasifikasi. Temuan ini sejalan dengan teori dan penelitian terdahulu yang menyatakan bahwa Parkinson Disease memengaruhi kontrol motorik suara akibat gangguan sistem saraf pusat. Nilai precision, recall, dan AUC yang berada pada kategori sangat baik menegaskan bahwa model memiliki kemampuan diskriminasi kelas yang kuat dan stabil.

Dengan demikian, algoritma Decision Tree dapat dijadikan pendekatan yang interpretatif, efisien, dan relevan untuk diintegrasikan ke dalam sistem informasi pendukung keputusan medis, khususnya dalam konteks deteksi dini Parkinson Disease berbasis data suara.

Untuk pengembangan penelitian selanjutnya, disarankan agar menggunakan jumlah dan variasi data yang lebih besar guna meningkatkan kemampuan generalisasi model. Selain itu, penelitian lanjutan dapat melakukan perbandingan dengan algoritma klasifikasi lain, seperti Random Forest, Support Vector Machine, atau metode deep learning, untuk memperoleh performa yang lebih optimal.

Pengembangan sistem ke arah aplikasi berbasis web atau mobile juga direkomendasikan agar hasil penelitian dapat dimanfaatkan secara praktis sebagai alat bantu skrining awal Parkinson Disease. Terakhir, penggabungan data suara dengan data klinis atau sensor gerak diharapkan mampu meningkatkan akurasi serta keandalan sistem deteksi dini yang dikembangkan.

UCAPAN TERIMA KASIH

Peneliti menyampaikan ucapan terima kasih kepada pihak yang sudah berkontribusi dalam pelaksanaan penelitian dan penyusunan artikel ini.

REFERENSI

- Armstrong, Melissa J., and Michael S. Okun. 2020. "Diagnosis and Treatment of Parkinson Disease: A Review." *JAMA - Journal of the American Medical Association* 323(6):548–60. doi: 10.1001/jama.2019.22360.
- Azis, Huzain, Purnawansyah Purnawansyah, Farniwati Fattah, and Inggrianti Pratiwi Putri. 2020.

- “Performa Klasifikasi K-NN Dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung.” *ILKOM Jurnal Ilmiah* 12(2):81–86. doi: 10.33096/ilkom.v12i2.507.81-86.
- Bidang, Penelitian, Komputer Sains, Yuli Mardi, Jl Gajah, Mada No, and Sumatera Barat. n.d. “Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data Mining Merupakan Bagian Dari Tahapan Proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika.”
- Hendrickx, Tayena, Boris Cule, Pieter Meysman, Stefan Naulaerts, Kris Laukens, and Bart Goethals. 2015. “Mining Association Rules in Graphs Based on Frequent Cohesive Itemsets.” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9078(3):637–48. doi: 10.1007/978-3-319-18032-8_50.
- Moerdyanto, Octarian Prasetya, and I. Kadek Dwi Nuryana. 2023. “Predicting On-Time Graduation Using Decision Tree Approach Decision Tree Algorithm.” *Journal of Informatics and Computer Science* 5(1):90–96.
- Nasional, Jurnal, Sistem Informasi, Vincentius Riandaru, Hamzah Lazuardi, Aldo Adhi, and Christian Lauw. 2021. “Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Regresi Linier.” 01:8–17.
- Nopriandi, Helpi, Fakultas Teknik, Universitas Islam, Kuantan Singingi, and Teluk Kuantan. 2018. “Perancangan Sistem Informasi Registrasi Mahasiswa.” 1(1):73–79.
- Purwati, Neni, and Sri Karnila. 2023. “Strategi Peningkatan Penjualan Produk Menggunakan Market Basket Analysis.” *Jurnal Sistem Informasi Bisnis* 13(2):96–103. doi: 10.21456/vol13iss2pp96-103.
- Razaq, Ayesha, Shabana Ramzan, Sohail Jabbar, Muhammad Munwar Iqbal, Muhammad Asif Habib, and Umar Raza. 2025. “A Deep Learning Framework for Early Parkinson’s Disease Detection: Leveraging Spiral and Wave Handwriting Tasks with EfficientNetV2-S.” *Diagnostics* 15(21):1–25. doi: 10.3390/diagnostics15212795.
- Uska, Muhammad, Rasyid Wirasasmita, Usuluddin Usuluddin, and Baiq Arianti. 2020. “Evaluation of Rapidminer-Aplication in Data Mining Learning Using PeRSIVA Model.” *Edumatic: Jurnal Pendidikan Informatika* 4(2):164–71. doi: 10.29408/edumatic.v4i2.2688.
- Wahyuni, Elvira Sukma, Kern Cesarean Ahnaf, Firdaus Firdaus, Nurul Ashikin Abdul Kadir, Nor Aini Zakaria, Idha Arfianti Wiraagni, and Diwangkoro Aji Kadarmo. 2024. “Wound Depth Measurement System in Forensic Case Using Image Processing and Machine Learning.” 1–21.
- Zein, Intan Sahara, Universitas Malikussaleh, Khairunnisa Khairunnisa, and Bagian Ilmu Neurologi. 2023. “Parkinson Disease.” 2(2).