

**PENERAPAN RANDOM FOREST UNTUK DETEKSI DINI PENYAKIT  
PARKINSON'S DENGAN DATA FREKUENSI SUARA**

**SKRIPSI**

Diajukan Untuk Memenuhi Salah Satu Syarat Guna  
Memperoleh Gelar Sarjana Komputer ( S.Kom. )  
Pada Program Studi Sistem Informasi



OLEH :

**MOHAMMAD ANNAN MAKRUH MUSTOFA**

NPM : 2113030041

FAKULTAS TEKNIK DAN ILMU KOMPUTER (FTIK)  
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA  
**UN PGRI KEDIRI**

2025

Skripsi oleh:

**MOHAMMAD ANNAN MAKRUF MUSTOFA**

NPM : 2113030041

Judul:

**PENERAPAN RANDOM FOREST UNTUK DETEKSI DINI PENYAKIT  
PARKINSON'S DENGAN DATA FREKUENSI SUARA**

Telah Disetujui untuk diajukan kepada

Panitia Ujian/Sidang Skripsi Program Studi Sistem Informasi

FTIK UN PGRI Kediri

Tanggal: 2 Juli 2025

Pembimbing I



Sucipto, S.Kom, M.Kom  
NIDN. 0721029101

Pembimbing II



Arie Nugroho, S.Kom, M.M, M.Kom  
NIDN. 0712108103

Skripsi oleh:

**MOHAMMAD ANNAN MAKRUH MUSTOFA**

NPM: 2113030041

Judul:

**PENERAPAN RANDOM FOREST UNTUK DETEKSI DINI PENYAKIT  
PARKINSON'S DENGAN DATA FREKUENSI SUARA**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi  
Program Studi Sistem Informasi FTIK UN PGRI Kediri

Tanggal: 11 Juli 2025

**Dan Dinyatakan telah Memenuhi Persyaratan**

Panitia Penguji :

1. Ketua Penguji : Sucipto, S.Kom, M.Kom.

.....

2. Penguji 1 : Aidina Ristyawan, S.Kom, M.Kom.

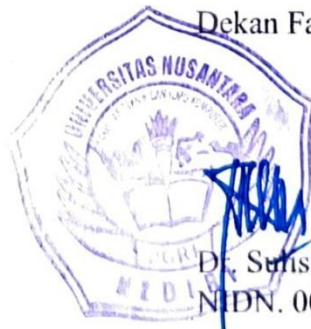
.....

3. Penguji 2 : Arie Nugroho, S.Kom, M.M, M.Kom.

.....

Mengetahui,

Dekan Fakultas Teknik dan Ilmu Komputer



Dr. Sunistiono, M.Si  
NIDN. 0007076801

### **PERNYATAAN**

Yang bertanda tangan di bawah ini saya,

Nama : Mohammad Annan Makruf Mustofa

Jenis Kelamin : Laki-laki

Tempat/ tgl. lahir : Nganjuk, 26 Oktober 2002

NPM : 2113030041

Fak/ Prodi. : FTIK/S1-Sistem Informasi

Menyatakan dengan sebenarnya, bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini disebutkan dalam Daftar Pustaka.

Kediri, 2 Juli 2025

Yang menyatakan

Mohammad Annan Makruf Mustofa

NPM: 2113030041

## **MOTTO**

"Know what you own, and know why you own it."

(Peter Lynch)

## ABSTRAK

**Mohammad Annan Makruf Mustofa:** Penerapan Random Rorest Untuk Deteksi Dini Penyakit Parkinson's Dengan Data Frekuensi Suara, Skripsi, Sistem Informasi, Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri, 2025.

**Kata Kunci:** Penyakit Parkinson, Deteksi Dini, Random Forest, SMOTE, Frekuensi Suara, Stratified 10-Fold Cross Validation.

Penyakit Parkinson merupakan gangguan neurodegeneratif progresif yang berdampak serius pada sistem saraf pusat dan sering kali sulit dideteksi pada tahap awal. Gejala awal yang menyerupai tanda-tanda penuaan normal menjadi tantangan dalam proses diagnosis, sehingga banyak pasien kehilangan kesempatan untuk mendapatkan intervensi medis secara dini. Deteksi dini sangat penting untuk memperlambat progresi penyakit dan meningkatkan kualitas hidup penderita. Seiring berkembangnya teknologi kecerdasan buatan, pemanfaatan analisis data frekuensi suara menjadi salah satu pendekatan non-invasif yang menjanjikan dalam mendeteksi penyakit ini.

Penelitian ini bertujuan membangun model klasifikasi berbasis algoritma Random Forest untuk mendeteksi dini penyakit Parkinson melalui data akustik suara. Penelitian menggunakan dataset dari platform Kaggle yang terdiri dari 195 sampel dengan 24 atribut numerik. Untuk mengatasi ketidakseimbangan kelas, diterapkan teknik Synthetic Minority Oversampling Technique (SMOTE). Model divalidasi menggunakan metode Stratified 10-Fold Cross Validation agar hasil evaluasi lebih stabil dan representatif. Seluruh proses penelitian mengikuti tahapan CRISP-DM, mulai dari pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, hingga deployment.

Hasil evaluasi menunjukkan bahwa model Random Forest mampu memberikan performa klasifikasi yang unggul dalam mendeteksi penyakit Parkinson. Model ini menghasilkan akurasi rata-rata sebesar 91,21%, dengan precision 92%, recall 91,2%, dan F1-score 91,3%. Penerapan teknik SMOTE terbukti efektif dalam meningkatkan sensitivitas terhadap kelas minoritas tanpa menurunkan kinerja keseluruhan. Validasi menggunakan Stratified 10-Fold Cross Validation memastikan bahwa hasil evaluasi model bersifat stabil, representatif, dan tidak bias terhadap pembagian data acak.

Kesimpulan dari penelitian ini menunjukkan bahwa kombinasi algoritma Random Forest, teknik SMOTE, dan validasi Stratified 10-Fold Cross Validation merupakan pendekatan yang efektif dalam mendeteksi dini penyakit Parkinson secara non-invasif. Namun, keterbatasan penelitian ini terletak pada ukuran dataset yang relatif kecil. Pengembangan lebih lanjut dapat dilakukan dengan memperluas jumlah data, menggunakan fitur suara tambahan, atau menerapkan algoritma hybrid untuk meningkatkan akurasi model pada data dunia nyata.

## **KATA PENGANTAR**

Puji Syukur Kami panjatkan Kehadirat Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenan-Nya penyusunan skripsi ini dapat diselesaikan. Skripsi dengan judul “Penerapan Random Rorest Untuk Deteksi Dini Penyakit Parkinson’s Dengan Data Frekuensi Suara” ini ditulis guna memenuhi sebagian syarat untuk memperoleh gelar Sarjana Komputer, pada Program Studi Sistem Informasi Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri. Pada kesempatan ini diucapkan terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. Dr. Zainal Afandi, M.Pd. selaku Rektor Universitas Nusantara PGRI Kediri.
2. Dr. Sulistiono, M.Si. selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri.
3. Sucipto, S.Kom, M.Kom. selaku Ketua Program Studi Sistem Informasi Universitas Nusantara PGRI Kediri.
4. Sucipto, S.Kom, M.Kom. dan Arie Nugroho, S.Kom, M.M, M.Kom. selaku dosen pembimbing yang telah memberikan arahan, bimbingan, dan motivasi selama penyusunan skripsi ini.
5. Orang tua tercinta yang senantiasa memberikan doa, dukungan moral, serta materiil yang tiada henti, sehingga saya dapat menyelesaikan tugas ini dengan baik.

Disadari bahwa skripsi ini masih banyak kekurangan, maka diharapkan tegur sapa, kritik dan saran-saran, dari berbagai pihak sangat diharapkan. Akhirnya, disertai harapan semoga skripsi ini ada manfaatnya bagi kita semua, khususnya bagi dunia pendidikan, meskipun hanya ibarat setitik air bagi samudra yang luas.

Kediri, 2 Juli 2025

Mohammad Annan Makruf Mustofa  
NPM : 2113030041

## DAFTAR ISI

|                                               |             |
|-----------------------------------------------|-------------|
| Halaman Sampul .....                          | i           |
| Halaman Persetujuan.....                      | ii          |
| Halaman Pengesahan .....                      | iii         |
| Halaman Pernyataan.....                       | iv          |
| Halaman Motto.....                            | v           |
| <b>ABSTRAK .....</b>                          | <b>vi</b>   |
| <b>KATA PENGANTAR.....</b>                    | <b>vii</b>  |
| <b>DAFTAR ISI.....</b>                        | <b>viii</b> |
| <b>DAFTAR GAMBAR.....</b>                     | <b>xi</b>   |
| <b>DAFTAR TABEL .....</b>                     | <b>xii</b>  |
| <b>DAFTAR LAMPIRAN .....</b>                  | <b>xiii</b> |
| <b>BAB I PENDAHULUAN.....</b>                 | <b>1</b>    |
| A. Latar Belakang Masalah.....                | 1           |
| B. Identifikasi Masalah.....                  | 3           |
| C. Batasan Masalah .....                      | 4           |
| D. Rumusan Masalah.....                       | 4           |
| E. Tujuan Penelitian .....                    | 5           |
| F. Manfaat Penelitian .....                   | 5           |
| <b>BAB II KAJIAN TEORI DAN HIPOTESIS.....</b> | <b>6</b>    |
| A. Kajian Teori .....                         | 6           |
| 1. Penyakit <i>Parkinson's</i> .....          | 6           |
| 2. <i>Data Mining</i> .....                   | 6           |
| 3. Klasifikasi .....                          | 7           |
| 4. <i>Random Forest</i> .....                 | 8           |
| 5. <i>Preprocessing</i> .....                 | 9           |
| 6. <i>SMOTE</i> .....                         | 10          |
| 7. <i>K-Fold Cross Validation</i> .....       | 12          |
| 8. <i>Confusion Matrix</i> .....              | 13          |
| 9. <i>Google Collaboratory</i> .....          | 15          |
| B. Kajian Hasil Penelitian Terdahulu .....    | 16          |
| C. Kerangka Berpikir.....                     | 18          |
| D. Hipotesis .....                            | 18          |
| <b>BAB III METODOLOGI PENELITIAN .....</b>    | <b>20</b>   |



|                                                                          |    |
|--------------------------------------------------------------------------|----|
| A. <i>Business Understanding</i> .....                                   | 20 |
| 1. <i>Determine Business Objectives</i> .....                            | 21 |
| 2. <i>Assess Situation</i> .....                                         | 21 |
| 3. <i>Determine Data Mining Goals</i> .....                              | 21 |
| 4. <i>Produce Project Plan</i> .....                                     | 22 |
| B. <i>Data Understanding</i> .....                                       | 22 |
| 1. <i>Pengumpulan Data Awal (Collect Initial Data)</i> .....             | 23 |
| 2. <i>Deskripsi Data (Describe Data)</i> .....                           | 23 |
| 3. <i>Eksplorasi Data (Explore Data)</i> .....                           | 23 |
| C. <i>Data Preparation</i> .....                                         | 23 |
| D. <i>Modeling</i> .....                                                 | 23 |
| E. <i>Evaluation</i> .....                                               | 24 |
| F. <i>Deployment</i> .....                                               | 25 |
| <b>BAB IV HASIL DAN PEMBAHASAN</b> .....                                 | 26 |
| A. <i>Business Understanding</i> .....                                   | 26 |
| B. <i>Data Understanding</i> .....                                       | 26 |
| 1. <i>Collect Initial Data</i> .....                                     | 26 |
| 2. <i>Describe Data</i> .....                                            | 27 |
| 3. <i>Explore Dataset</i> .....                                          | 51 |
| C. <i>Data Preparation</i> .....                                         | 53 |
| 1. <i>Penghapusan Fitur</i> .....                                        | 53 |
| 2. <i>Pemisahan Fitur dan Target</i> .....                               | 55 |
| D. <i>Modeling</i> .....                                                 | 55 |
| 1. <i>Validasi Model: Stratified K-Fold Cross Validation</i> .....       | 55 |
| 2. <i>Penanganan Ketidakseimbangan Data dengan SMOTE</i> .....           | 57 |
| 3. <i>Pelatihan Model</i> .....                                          | 58 |
| 4. <i>Perhitungan Manual Random Forest</i> .....                         | 59 |
| a) <i>Pemilihan Fitur dan Pembentukan Split</i> .....                    | 59 |
| b) <i>Perhitungan Gini Index</i> .....                                   | 60 |
| c) <i>Proses Prediksi</i> .....                                          | 60 |
| E. <i>Evaluation</i> .....                                               | 61 |
| 1. <i>Parameter Model Klasifikasi dan Evaluasi</i> .....                 | 61 |
| 2. <i>Perbandingan Hasil Evaluasi Model dengan dan tanpa SMOTE</i> ..... | 62 |
| F. <i>Deployment</i> .....                                               | 64 |
| 1. <i>Potensi Implementasi dan Skalabilitas</i> .....                    | 64 |
| 2. <i>Dokumentasi dan Distribusi</i> .....                               | 65 |

|                               |           |
|-------------------------------|-----------|
| <b>BAB V KESIMPULAN .....</b> | <b>66</b> |
| A. Kesimpulan .....           | 66        |
| B. Saran .....                | 67        |
| <b>DAFTAR PUSTAKA .....</b>   | <b>68</b> |
| <b>LAMPIRAN.....</b>          | <b>72</b> |

## DAFTAR GAMBAR

|                                                      |    |
|------------------------------------------------------|----|
| Gambar 2.1 Kerangka Berpikir .....                   | 18 |
| Gambar 3. 1 Alur <i>CRISP-DM</i> .....               | 20 |
| Gambar 4.1 Distribusi <i>MDVP:F0(Hz)</i> .....       | 28 |
| Gambar 4.2 Distribusi <i>MDVP:Fhi(Hz)</i> .....      | 29 |
| Gambar 4.3 Distribusi <i>MDVP:Flo(Hz)</i> .....      | 30 |
| Gambar 4.4 Distribusi <i>MDVP:Jitter(%)</i> .....    | 31 |
| Gambar 4.5 Distribusi <i>MDVP:Jitter(Abs)</i> .....  | 32 |
| Gambar 4.6 Distribusi <i>MDVP:RAP</i> .....          | 33 |
| Gambar 4.7 Distribusi <i>MDVP:PPQ</i> .....          | 34 |
| Gambar 4.8 Distribusi <i>Jitter:DPP</i> .....        | 35 |
| Gambar 4.9 Distribusi <i>MDVP:Shimmer</i> .....      | 36 |
| Gambar 4.10 Distribusi <i>MDVP:Shimmer(dB)</i> ..... | 37 |
| Gambar 4.11 Distribusi <i>Shimmer:APQ3</i> .....     | 38 |
| Gambar 4.12 Distribusi <i>Shimmer:APQ5</i> .....     | 39 |
| Gambar 4.13 Distribusi <i>MDVP:APQ</i> .....         | 40 |
| Gambar 4.14 Distribusi <i>Shimmer:DDA</i> .....      | 41 |
| Gambar 4.15 Distribusi <i>NHR</i> .....              | 42 |
| Gambar 4.16 Distribusi <i>HNR</i> .....              | 43 |
| Gambar 4.17 Distribusi <i>RPDE</i> .....             | 44 |
| Gambar 4.18 Distribusi <i>DFA</i> .....              | 45 |
| Gambar 4.19 Distribusi <i>Spread1</i> .....          | 46 |
| Gambar 4.20 Distribusi <i>Spread2</i> .....          | 47 |
| Gambar 4.21 Distribusi <i>D2</i> .....               | 48 |
| Gambar 4.22 Distribusi <i>PPE</i> .....              | 49 |
| Gambar 4.23 Distribusi <i>Status</i> .....           | 50 |
| Gambar 4.24 Sebelum <i>SMOTE</i> .....               | 57 |
| Gambar 4.25 Sesudah <i>SMOTE</i> .....               | 58 |

## DAFTAR TABEL

|                                                                                 |    |
|---------------------------------------------------------------------------------|----|
| Tabel 2.1 <i>Confusion Matrix</i> .....                                         | 14 |
| Tabel 4.1 <i>Dataset Parkinson's</i> .....                                      | 26 |
| Tabel 4.2 Eksplorasi <i>Dataset</i> .....                                       | 51 |
| Tabel 4.3 Fitur yang digunakan .....                                            | 53 |
| Tabel 4.4 Data Tiap <i>fold</i> .....                                           | 56 |
| Tabel 4.5 Sampel Data Perhitungan Manual <i>Random Forest</i> .....             | 59 |
| Tabel 4.6 Parameter Model Klasifikasi dan Evaluasi .....                        | 61 |
| Tabel 4.7 Perbandingan Hasil Evaluasi Model dengan dan tanpa <i>SMOTE</i> ..... | 62 |
| Tabel 4.8 Perbandingan Hasil Tiap <i>Fold</i> .....                             | 63 |

## DAFTAR LAMPIRAN

|                                                                                              |    |
|----------------------------------------------------------------------------------------------|----|
| Lampiran 1 Kartu Bimbingan .....                                                             | 72 |
| Lampiran 2 Surat Keterangan Bebas Similarity.....                                            | 73 |
| Lampiran 3 Bukti Halaman Awal Cek Similarity.....                                            | 74 |
| Lampiran 4 Bukti Screenshoot Submit artikel/ LOA Artikel/ Screenshot Artikel<br>Terbit ..... | 75 |
| Lampiran 5 Lembar Berita Acara Ujian.....                                                    | 76 |
| Lampiran 6 Lembar Revisi Ujian.....                                                          | 77 |
| Lampiran 7 Kode Program.....                                                                 | 78 |

## **BAB I**

### **PENDAHULUAN**

#### **A. Latar Belakang Masalah**

Kecerdasan buatan (*Artificial Intelligence*) telah menjadi salah satu teknologi yang berperan penting dalam transformasi berbagai sektor, termasuk bidang kesehatan. Kemampuannya untuk mengolah data dalam skala besar dan mengidentifikasi pola-pola yang kompleks telah memberikan manfaat signifikan dalam mendukung proses diagnosa penyakit secara cepat, akurat, dan efisien. Salah satu cabang *AI* yang paling sering diterapkan adalah *machine learning*, yang memungkinkan sistem belajar dari data historis dan menghasilkan prediksi tanpa memerlukan intervensi langsung dari manusia. Dengan kemampuan menyajikan insight prediktif secara otomatis, *machine learning* mendukung pengambilan keputusan berbasis data untuk memprediksi dan memahami pola penyakit yang kompleks seperti penyakit *Parkinson's* (Wardhana et al., 2023).

Penyakit *Parkinson's* merupakan gangguan neurologis yang ditandai dengan kerusakan progresif pada sel-sel saraf di area tertentu otak, khususnya di substansia nigra dan basal ganglia. Kerusakan ini mengakibatkan gangguan pada proses neurokimia yang berperan penting dalam mengatur koordinasi gerakan tubuh manusia. Akibatnya, penderita mengalami kesulitan dalam melakukan gerakan-gerakan yang biasa dan teratur (Iskandar, 2023). Salah satu tantangan terbesar dalam menangani *Parkinson* adalah diagnosa yang sering kali terlambat. Hal ini disebabkan oleh perkembangan gejala yang bertahap serta kesamaan gejala awalnya dengan tanda-tanda penuaan normal. Metode konvensional dalam mendeteksi penyakit ini umumnya bergantung pada pengamatan klinis dan laporan subyektif pasien, yang tidak cukup sensitif untuk mendeteksi penyakit pada tahap awal (Wulandari et al., 2024). Akibatnya, banyak pasien kehilangan kesempatan untuk mendapatkan intervensi dini yang dapat memperlambat perkembangan penyakit dan meningkatkan kualitas hidup mereka.

Seiring dengan perkembangan teknologi, pemanfaatan data frekuensi suara telah menjadi salah satu pendekatan inovatif dalam deteksi dini *Parkinson*. Perubahan pada karakteristik suara, seperti kestabilan frekuensi, kecepatan berbicara, dan intonasi, telah diidentifikasi sebagai indikator awal yang dapat

muncul sebelum gejala motorik yang lebih nyata. Analisis terhadap data suara ini membuka peluang untuk mengembangkan model prediktif berbasis *machine learning*. Beberapa penelitian sebelumnya telah menghasilkan model prediktif dengan akurasi tinggi menggunakan berbagai algoritma *machine learning*. Salah satunya adalah penelitian oleh Elshewey et al. (2023) yang berhasil mencapai akurasi 92,3% menggunakan metode *Bayesian Optimization Support Vector Machine* (BO-SVM). Namun, dalam penelitian tersebut belum diterapkan validasi silang (*cross-validation*), sehingga hasilnya rentan terhadap variasi acak dari pembagian data pelatihan dan pengujian.

Sebaliknya, penelitian ini menggunakan pendekatan yang lebih ketat secara statistik dengan menerapkan *Stratified K-Fold Cross Validation (10-fold)*. Selain itu, untuk mengatasi masalah ketidakseimbangan kelas yang umum ditemukan pada dataset medis seperti ini, digunakan teknik *Synthetic Minority Oversampling Technique* (SMOTE). Algoritma utama yang digunakan dalam penelitian ini adalah *Random Forest*, yang memiliki kemampuan kuat dalam menangani data kompleks yang baik. Algoritma ini bekerja dengan membangun sekumpulan pohon keputusan yang berkolaborasi untuk menghasilkan prediksi yang lebih stabil dan akurat. Algoritma ini sangat cocok untuk dataset yang bertipe numerik. Semakin banyak fitur yang digunakan, semakin banyak variasi pohon yang ada (Nugroho et al., 2021). Dalam konteks deteksi *Parkinson*, *Random Forest* dapat mengintegrasikan berbagai fitur suara untuk mengidentifikasi pola-pola yang mengindikasikan keberadaan penyakit.

Keunggulan *Random Forest* juga terletak pada kemampuannya menangani data yang hilang dan ketahanan terhadap *outlier*, menjadikannya algoritma yang efisien dalam berbagai situasi. Selain itu, algoritma ini memiliki proses seleksi fitur yang memungkinkan pemilihan fitur terbaik, sehingga meningkatkan performa klasifikasi secara signifikan (Devella & Novia Rahmawati, 2020). Dengan memanfaatkan algoritma ini, data frekuensi suara dapat dianalisis secara simultan untuk menghasilkan keputusan berdasarkan kolektivitas pohon-pohon keputusan. Hal ini memungkinkan deteksi dini *Parkinson* dilakukan dengan tingkat akurasi yang tinggi, memberikan peluang untuk intervensi medis yang lebih awal dan efektif. Lebih jauh, implementasi model ini tidak hanya

memberikan manfaat dalam konteks klinis, tetapi juga memiliki potensi untuk digunakan dalam aplikasi berbasis komunitas, seperti pemeriksaan skrining massal, yang dapat menjangkau lebih banyak populasi berisiko.

Dengan mempertimbangkan tantangan dalam deteksi dini *Parkinson* dan potensi besar yang dimiliki analisis data frekuensi suara, pengembangan model prediktif berbasis algoritma *Random Forest* menjadi solusi yang menjanjikan. Algoritma ini tidak hanya andal dalam menangani dataset kompleks dan numerik, tetapi juga mampu mengintegrasikan berbagai fitur suara untuk mengidentifikasi pola yang menunjukkan keberadaan *Parkinson* secara akurat. Selain itu, keunggulan *Random Forest* dalam menangani data hilang, mengelola *outlier*, dan melakukan seleksi fitur secara otomatis menjadikannya alat yang efisien untuk menghasilkan prediksi yang stabil. Pendekatan ini membuka peluang baru untuk diagnosa dini yang lebih cepat dan tepat, baik dalam konteks klinis maupun aplikasi komunitas seperti skrining massal. Dengan demikian, implementasi model ini diharapkan tidak hanya dapat mempercepat intervensi medis bagi pasien *Parkinson*, tetapi juga berkontribusi pada pengurangan beban penyakit secara global, meningkatkan pengelolaan penyakit, dan mendukung kualitas hidup yang lebih baik bagi pasien serta keluarga mereka.

Sejalan dengan kebutuhan tersebut, penelitian ini mengembangkan model klasifikasi berbasis algoritma *Random Forest* yang dikombinasikan dengan teknik *Synthetic Minority Oversampling Technique* (SMOTE) dan validasi *Stratified 10-Fold Cross Validation*. Tidak seperti sejumlah penelitian terdahulu yang belum secara optimal menangani ketidakseimbangan kelas maupun validasi performa secara menyeluruh, pendekatan yang diterapkan dalam penelitian ini menawarkan kombinasi yang komprehensif untuk meningkatkan akurasi, dan kestabilan. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang lebih kuat dalam pengembangan sistem deteksi dini penyakit *Parkinson* berbasis suara secara cerdas, adaptif, dan dapat diimplementasikan secara luas di berbagai konteks aplikasi.

## **B. Identifikasi Masalah**

Penyakit *Parkinson's* sering terlambat didiagnosis karena gejalanya berkembang secara perlahan dan sering disalahartikan sebagai tanda penuaan



normal, sehingga banyak pasien kehilangan peluang untuk mendapatkan intervensi medis dini yang dapat memperlambat perkembangan penyakit. Metode diagnosis konvensional, yang bergantung pada observasi klinis dan tes neurologis, sering kali kurang efektif dalam mendeteksi gejala non-motorik, seperti perubahan suara, yang sebenarnya dapat muncul pada tahap awal penyakit. Berbagai penelitian sebelumnya telah menunjukkan bahwa algoritma seperti BO-SVM mampu memberikan akurasi yang cukup tinggi dalam mendeteksi pola penyakit berdasarkan data suara. Namun, peningkatan performa model masih diperlukan dengan memanfaatkan algoritma yang lebih kompleks dan andal seperti *Random Forest*, yang memiliki kemampuan untuk mengolah data secara lebih mendalam, meningkatkan akurasi, dan memperbaiki generalisasi model terhadap variasi data suara yang lebih luas.

### C. Batasan Masalah

Penelitian ini akan fokus pada penerapan algoritma *Random Forest* untuk mendeteksi dini penyakit *Parkinson's* dengan menggunakan data frekuensi suara. Ruang lingkup penelitian ini dibatasi pada beberapa aspek berikut:

1. Dataset yang digunakan berasal dari sumber publik yang berisi parameter frekuensi suara pasien. Dataset dapat diakses melalui tautan berikut: <https://www.kaggle.com/datasets/jainaru/parkinson-disease-detection>.
2. Teknik *oversampling* yang digunakan adalah SMOTE untuk menyeimbangkan distribusi kelas.
3. Model yang digunakan adalah *Random Forest* dengan evaluasi menggunakan *Stratified K-Fold Cross Validation (10-fold)*.
4. Bahasa pemrograman yang digunakan adalah *Python* dengan library *scikit-learn*, *imbalanced-learn*, *pandas*, dan *matplotlib/seaborn*.
5. Penelitian ini hanya mencakup tahap deteksi dini, bukan diagnosis klinis atau prediksi progresi penyakit.

### D. Rumusan Masalah

1. Seberapa akurat algoritma *Random Forest* dalam mendeteksi penyakit *Parkinson* secara dini berdasarkan data frekuensi suara?
2. Apakah penerapan SMOTE dapat meningkatkan performa model, terutama pada kelas minoritas?

3. Bagaimana pengaruh penggunaan *Stratified K-Fold Cross Validation* terhadap stabilitas dan reliabilitas hasil prediksi?

#### **E. Tujuan Penelitian**

Penelitian ini memiliki tujuan sebagai berikut:

1. Membangun model klasifikasi untuk mendeteksi penyakit *Parkinson* secara dini berdasarkan data frekuensi suara menggunakan algoritma *Random Forest*.
2. Meningkatkan performa model pada kelas minoritas dengan menerapkan teknik SMOTE.
3. Mengevaluasi keandalan model menggunakan *Stratified K-Fold Cross Validation* agar hasil prediksi lebih stabil dan representatif.
4. Memberikan kontribusi dalam pengembangan sistem pendukung keputusan berbasis machine learning untuk aplikasi diagnostik medis.

#### **F. Manfaat Penelitian**

Penelitian ini diharapkan memberikan kontribusi signifikan dalam pengembangan metode deteksi dini penyakit *Parkinson* melalui analisis data frekuensi suara. Dengan memanfaatkan algoritma *Random Forest* dan teknik SMOTE untuk menangani ketidakseimbangan kelas, penelitian ini dapat meningkatkan akurasi serta stabilitas prediksi model secara efektif. Selain itu, penerapan *Stratified K-Fold Cross Validation* menjadikan evaluasi model lebih andal dan representatif dibandingkan metode pembagian data sederhana. Bagi dunia medis, hasil penelitian ini dapat menjadi dasar pengembangan alat bantu diagnosis yang cepat, efektif, dan non-invasif, sehingga memungkinkan intervensi medis lebih dini dan tepat sasaran. Secara teknologi, penelitian ini juga memperkaya literatur ilmiah terkait pemanfaatan machine learning dalam bidang kesehatan, khususnya pada penggunaan data suara untuk mendukung proses diagnostik.

## DAFTAR PUSTAKA

- Desiani, A., Ramayanti, I., Arhami, M., Studi Matematika, P., & Matematika dan Ilmu Pengetahuan Alam, F. (2023). DIAGNOSIS OF PARKINSON'S DISEASE USING K-NEAREST NEIGHTBOUR AND DECISION TREE C4.5 ALGORITHMS. *Jurnal SimanteC*, 12(1).  
<https://doi.org/10.21107/simantec.v12i1.21167>
- Devella, S., & Novia Rahmawati, F. (2020). Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(2).  
<https://doi.org/10.35957/jatisi.v7i2.289>
- Elshewey, A. M., Shams, M. Y., El-Rashidy, N., Elhady, A. M., Shohieb, S. M., & Tarek, Z. (2023). Bayesian Optimization with Support Vector Machine Model for Parkinson Disease Classification. *Sensors*, 23(4).  
<https://doi.org/10.3390/s23042085>
- Fahira, N. R., Lawi, A., & Aqsha, M. (2023). Early Detection Model of Parkinson's Disease Using Random Forest Method on Voice Frequency Data. *Journal of Natural Sciences and Mathematics Research J. Nat. Scien. & Math. Res*, 9(1), 30–38. <https://doi.org/10.21580/jnsmr.2023.9.1.13148>
- Farida, Y., Ulinnuha, N., Sari, S. K., & Desinaini, L. N. (2023). Comparing Support Vector Machine and Naïve Bayes Methods with A Selection of Fast Correlation Based Filter Features in Detecting Parkinson's Disease. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 14(2), 80.  
<https://doi.org/10.24843/lkjiti.2023.v14.i02.p02>
- Fatkhudin, A., Adi Artanto, F., Abiyu Safli, N., & Wibowo, D. (2024). Decision Tree Berbasis SMOTE dalam Analisis Sentimen Penggunaan Artificial Intelligence untuk Skripsi. *Remik: Riset Dan E-Jurnal Manajemen Informatika Komputer*, 8, 494–505.  
<https://doi.org/10.33395/remik.v8i2.13531>
- FUADAH, Y. N., UBAIDULLAH, I. D., IBRAHIM, N., TALININGSING, F. F., SY, N. K., & PRAMUDITHO, M. A. (2022). Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, 10(3), 728. <https://doi.org/10.26760/elkomika.v10i3.728>
- Gelar Guntara, R. (2023). Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(1), 55–60.  
<https://doi.org/10.47233/jteksis.v5i1.750>
- Haganta Depari, D., Widiastiwi, Y., & Mega Santoni, M. (2022). Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi

- Klasifikasi Penyakit Jantung. *JURNAL INFORMATIK*, 18, 2022.  
<https://doi.org/10.52958/iftk.v18i3.4694>
- Haryanti, M. F., Fauzi, A., Jelita, A. A., Setiyowati, A., Octarina, A., Putra Edina, E., Zahra Aulia, R., & Fitriana, S. (2024). Pengaruh Data Mining, Strategi Perusahaan Terhadap Laporan Kinerja Perusahaan. *PORTOFOLIO: Jurnal Manajemen Dan Bisnis*, 3(1). <https://doi.org/10.70704/jpjmb.v3i1.285>
- Iskandar, A. (2023). Sistem Pakar Dalam Mendiagnosa Penyakit Parkinson Menerapkan Metode Dempster-Shafer. *Journal of Information System Research (JOSH)*, 4(3), 847–854. <https://doi.org/10.47065/josh.v4i3.3320>
- Ismail, N. H., Du, M., Martinez, D., & He, Z. (2019). Multivariate multi-step deep learning time series approach in forecasting Parkinson's disease future severity progression. *ACM-BCB 2019 - Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 383–389. <https://doi.org/10.1145/3307339.3342185>
- Lubis, A. A., Wong, N. P., & Sinaga, F. M. (2020). Prediksi Akurasi Perusahaan Saham Menggunakan SVM dan K-Fold Cross Validation. *Jurnal SIFO Mikroskil*, 21(1). <https://doi.org/10.55601/jsm.v21i1.718>
- Mustofa, M. A. M., Wahiid, H. N., Islami, B. M., Ristyawan, A., & **Daniati, E.** (2024). Penggunaan Algoritma KNN dalam Deteksi Awal Kanker Paru-Paru Menggunakan Data Medis. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 8, 485–493. <https://doi.org/10.29407/inotek.v8i1.4979>
- Nugroho, A.**, Fanani, A. Z., & Shidik, G. F. (2021). Evaluation of Feature Selection Using Wrapper For Numeric Dataset With Random Forest Algorithm. *International Seminar on Application for Technology of Information and Communication (ISemantic)*. <https://doi.org/10.1109/iSemantic52711.2021.9573249>
- Nugroho, A., & **Harini, D.** (2024). Teknik Random Forest untuk Meningkatkan Akurasi Data Tidak Seimbang. *JSITIK*, 2(2). <https://doi.org/10.53624/jsitik.v2i2.XX>
- Pramayasa, K., Md, I., Maysanjaya, D., Ayu, G., & Diatri Indradewi, A. (2023). Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE. *SCIENCE AND INFORMATION TECHNOLOGY*, 6. <https://doi.org/10.31598>
- Putra, F., Tahiyat, H. F., Ihsan, R. M., Rahmaddeni, R., & Efrizoni, L. (2024). Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 273–281. <https://doi.org/10.57152/malcom.v4i1.1085>
- Sholihah, N. N., & Hermawan, A. (2023). IMPLEMENTATION OF RANDOM FOREST AND SMOTE METHODS FOR ECONOMIC STATUS

- CLASSIFICATION IN CIREBON CITY. *Jurnal Teknik Informatika (Jutif)*, 4(6), 1387–1397. <https://doi.org/10.52436/1.jutif.2023.4.6.1135>
- Sucipto**, Kusriani, & Taufiq, E. L. (2016). Classification Method of Multi-class on C4.5 Algorithm for Fish Diseases. *International Conference on Science in Information Technology*. <https://doi.org/10.1109/ICSITech.2016.7852598>
- Sucipto**, S., Dwi Prasetya, D., & Widiyaningtyas, T. (2024). Educational Data Mining: Multiple Choice Question Classification in Vocational School. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 23(2), 379–388. <https://doi.org/10.30812/matrik.v23i2.3499>
- Sunarya, U., & Haryanti, T. (2022). Perbandingan Kinerja Algoritma Optimasi pada Metode Random Forest untuk Deteksi Kegagalan Jantung. *Jurnal Rekayasa Elektrika*, 18(4). <https://doi.org/10.17529/jre.v18i4.26981>
- Supriyanto, Y., Ilhamsyah, M., & Enri, U. (2022). Prediksi Harga Minyak Kelapa Sawit Menggunakan Linear Regression Dan Random Forest. *Jurnal Ilmiah Wahana Pendidikan*, 8. <https://doi.org/10.5281/zenodo.6559603>
- Syahputri, C. N., & Hasibuan, M. S. (2024). OPTIMASI KLASIFIKASI DECISION TREE DENGAN TEKNIK PRUNING UNTUK MENGURANGI OVERFITTING. *JSiI (Jurnal Sistem Informasi)*, 11(2), 87–96. <https://doi.org/10.30656/jsii.v11i2.9161>
- Syifa, N. A. H., Nugroho, A., & **Firliana, R.** (2023). Perbandingan Algoritma Naïve Bayes Classifier Dan K-Nearest Neighbors Untuk Analisis Sentimen Covid-19 Di Twitter. *Jurnal Ilmiah Informatika*, 11 No. 01. <https://doi.org/10.33884/jif.v11i01.7069>
- Tchagna Kouanou, A., Mih Attia, T., Feudjio, C., Djeumo, A. F., Ngo Mouelas, A., Nzogang, M. P., Tchito Tchapg, C., & Tchiotsop, D. (2021). An Overview of Supervised Machine Learning Methods and Data Analysis for COVID-19 Detection. In *Journal of Healthcare Engineering* (Vol. 2021). Hindawi Limited. <https://doi.org/10.1155/2021/4733167>
- Wardhana, R. G., Wang, G., & Sibuea, F. (2023). PENERAPAN MACHINE LEARNING DALAM PREDIKSI TINGKAT KASUS PENYAKIT DI INDONESIA. In *Journal of Information System Management (JOISM) e-ISSN* (Vol. 5, Issue 1). <https://doi.org/10.24076/joism.2023v5i1.1136>
- Wulandari, K. A., Nugraha, A., Luthfiarta, A., & Nisa, L. R. (2024). Peningkatan Akurasi Deteksi Dini Penyakit Parkinson melalui Pendekatan Ensemble Learning dan Seleksi Fitur Optimal. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 575–584. <https://doi.org/10.29408/edumatic.v8i2.27788>
- Yudha, E. P., & Muhammad, N. F. (2023). Sistem Otomatis untuk Deteksi Penyakit Parkinson Menggunakan Fuzzy K-NN. *Jurnal Teknik Komputer*, 9(2), 96–102. <https://doi.org/10.31294/jtk.v9i2.15933>

Yulian Pamuji, F., Puspaning Ramadhan, V., & Artikel, R. (2021). Komparasi Algoritma Random Forest Dan Decision Tree Untuk Memprediksi Keberhasilan Immunotherapy. *Jurnal Teknologi Dan Manajemen Informatika*, 7, 46–50. <https://doi.org/10.26905/jtmi.v7i1.5982>