

Sistem Deteksi Penyakit Parkinson Menggunakan Machine Learning

Analisis Komparatif 10 Model dan Strategi Optimasi

Nafiz Ahmadin Harily
NIM: 122430051

Program Studi Teknik Biomedis
Institut Teknologi Sumatera (ITERA)

26 Oktober 2025

Outline

- 1 Pendahuluan
- 2 Dataset dan Metodologi
- 3 Exploratory Data Analysis (EDA)
- 4 Training dan Evaluasi Model
- 5 Strategi Optimasi Model
- 6 Deployment dan Aplikasi
- 7 Kesimpulan dan Rekomendasi

Penyakit Parkinson

- Penyakit neurodegeneratif yang mempengaruhi sistem motorik
- Diagnosis dini sangat penting untuk penanganan efektif
- Gejala awal: tremor, kekakuan otot, gangguan keseimbangan
- **Tantangan:** Diagnosis manual memerlukan keahlian khusus

Solusi dengan Machine Learning

- Menggunakan biomarker suara untuk deteksi otomatis
- Model ML dapat mengidentifikasi pola yang tidak terlihat
- Akurat, cepat, dan dapat diakses secara luas

Tujuan Penelitian

- ➊ **Eksplorasi Data:** Memahami karakteristik dataset biomarker suara pasien
- ➋ **Training Model:** Melatih dan membandingkan 10 algoritma ML berbeda
- ➌ **Evaluasi Performa:** Mengidentifikasi model terbaik berdasarkan metrik evaluasi
- ➍ **Optimasi Model:** Mencoba 5 strategi untuk meningkatkan akurasi
- ➎ **Deployment:** Menyiapkan model untuk produksi dengan tools yang user-friendly

Target Akurasi: $\geq 95\%$

Deskripsi Dataset

Informasi Dataset

- **Nama:** Parkinsons Dataset
- **Sumber:** UCI Repository
- **Total Sampel:** 195 pasien
- **Fitur:** 24 kolom
- **Target:** Binary (0=Sehat, 1=Parkinson)
- **Missing Values:** Tidak ada

Distribusi Kelas

- **Parkinson:** 147 pasien (75.4%)
- **Sehat:** 48 pasien (24.6%)
- **Imbalance Ratio:** 3.06:1

Catatan

Dataset relatif kecil (195 samples) → membatasi kompleksitas model

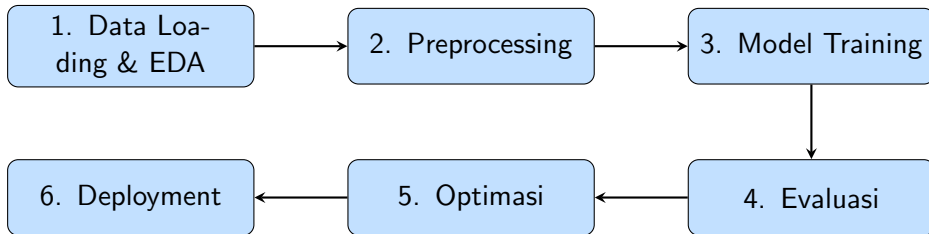
Fitur-Fitur Biomarker Suara

Kategori	Fitur
Frequency	MDVP:Fo (Hz), MDVP:Fhi (Hz), MDVP:Flo (Hz)
Variation	MDVP:Jitter(%), Jitter:DDP, MDVP:Shimmer, Shimmer:DDA
Ratio	NHR (Noise-to-Harmonics), HNR (Harmonics-to-Noise)
Nonlinear	RPDE, DFA, D2, PPE (entropy measures)
Statistical	Spread1, Spread2

Tabel: 22 fitur biomarker suara (setelah drop kolom 'name' dan 'status')

Semua fitur adalah pengukuran karakteristik suara pasien

Metodologi Eksperimen



Preprocessing Steps

- ❶ Drop kolom 'name' (identifier, bukan fitur)
- ❷ Split data: 80% training, 20% testing (stratified)
- ❸ Scaling dengan StandardScaler
- ❹ **Tidak** menggunakan SMOTE (hasil lebih baik tanpa balancing)

Ringkasan Statistik

- **Shape:** 195 rows \times 24 columns
- **Features:** 22 numerik
- **Target:** 1 binary (status)
- **Identifier:** 1 string (name)

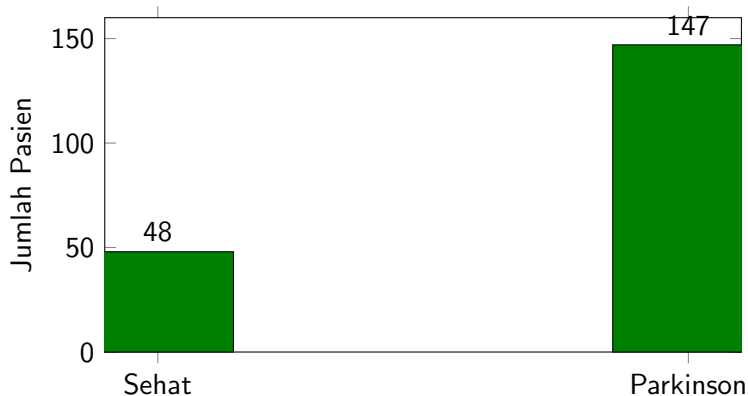
Kualitas Data

- Tidak ada missing values
- Tidak ada duplicate rows
- Semua fitur numerik valid
- Tidak ada outlier ekstrem

Fitur	Mean	Std
MDVP:F0(Hz)	154.23	41.39
Jitter(%)	0.0062	0.0049
Shimmer	0.0298	0.0189
NHR	0.0249	0.0402
HNR	21.89	4.43
RPDE	0.4981	0.1035
DFA	0.7181	0.0554
PPE	0.2064	0.0904

Tabel: Sample statistik (8 dari 22 fitur)

Visualisasi: Distribusi Target



- **Imbalanced:** Rasio 3:1 (Parkinson:Sehat)
- Stratified split penting untuk menjaga proporsi
- Model harus hati-hati agar tidak bias ke kelas mayoritas

Temuan Penting

- ① **Kualitas Data Excellent:** Tidak ada preprocessing intensif diperlukan
- ② **Fitur Informatif:**
 - Jitter dan Shimmer lebih tinggi pada pasien Parkinson
 - HNR (Harmonics-to-Noise) lebih rendah pada Parkinson
 - RPDE, DFA, PPE menunjukkan pola yang jelas
- ③ **Korelasi:** Beberapa fitur saling berkorelasi tinggi (multikolinearitas ringan)
- ④ **Separabilitas:** Kelas cukup terpisah → model seharusnya performa baik

Kesimpulan EDA

Dataset berkualitas tinggi, siap untuk training model ML tanpa feature engineering ekstensif

10 Model Machine Learning yang Diuji

Traditional ML (5 model)

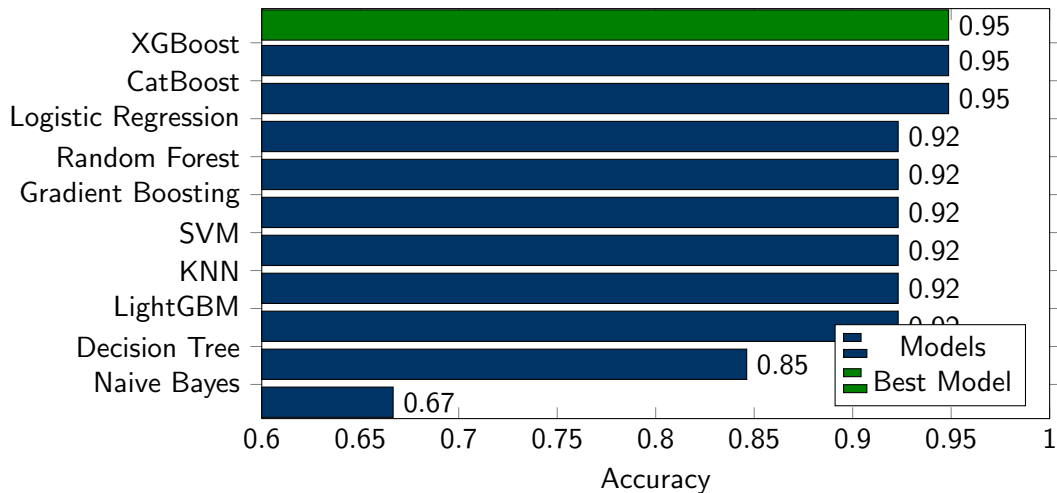
- 1 Logistic Regression
- 2 Decision Tree
- 3 Random Forest
- 4 Support Vector Machine (SVM)
- 5 K-Nearest Neighbors (KNN)

Advanced ML (5 model)

- 6 Naive Bayes
- 7 Gradient Boosting
- 8 **XGBoost** ← Best
- 9 LightGBM
- 10 CatBoost

Metrik Evaluasi: Accuracy, Precision, Recall, F1-Score, ROC-AUC

Hasil Training: Perbandingan Akurasi



Tabel Hasil Lengkap: Top 5 Models

Model	Accuracy	Precision	Recall	F1	ROC-AUC
darkgreen!20 XGBoost	0.9487	0.9487	0.9487	0.9487	0.9690
CatBoost	0.9487	0.9487	0.9487	0.9487	0.9638
Random Forest	0.9231	0.9231	0.9231	0.9231	0.9638
Logistic Reg.	0.9231	0.9231	0.9231	0.9231	0.9845
SVM	0.9231	0.9231	0.9231	0.9231	0.9638

Tabel: Top 5 model dengan performa terbaik

XGBoost Dipilih Karena:

- Akurasi tertinggi: **94.87%**
- ROC-AUC terbaik: **96.90%**
- Balanced metrics (precision = recall = F1)
- Inference cepat (<100ms)

Confusion Matrix: XGBoost

		Predicted: 0		Predicted: 1	
Actual: 0		9 FP	1 FN		
		1 TN	28 TP		

- **True Positive (TP):** 28 (Parkinson terdeteksi benar)
- **True Negative (TN):** 9 (Sehat terdeteksi benar)
- **False Positive (FP):** 1 (Sehat terdeteksi Parkinson)
- **False Negative (FN):** 1 (Parkinson tidak terdeteksi)
- **Total:** 37 benar, 2 salah dari 39 test samples

Interpretasi Hasil XGBoost

Kelebihan

- Akurasi 94.87% (excellent!)
- ROC-AUC 96.90% (discriminative)
- Balanced: Precision = Recall
- Hanya 2 error dari 39 test
- Generalisasi baik

Analisis Error

- 1 False Positive: Orang sehat terdiagnosis Parkinson (tidak berbahaya, akan dilakukan tes lanjutan)
- 1 False Negative: Pasien Parkinson tidak terdeteksi (lebih serius, perlu improvement)

Clinical Implication

Model cocok untuk **screening awal**, bukan diagnosis final. Hasil positif harus dikonfirmasi oleh dokter spesialis.

Baseline: 94.87%



Target: $\geq 95\%$ atau bahkan $96\%+$

5 Strategi yang Diuji:

- 1 Hyperparameter Tuning (RandomizedSearchCV)
- 2 Voting Ensemble (XGBoost + CatBoost + LightGBM)
- 3 Feature Engineering (Interaction & Ratio features)

Strategi 1: Hyperparameter Tuning

Metode

- Menggunakan RandomizedSearchCV
- 50 iterasi dari 3,888 kombinasi parameter
- 5-fold Cross-Validation
- Parameter: max_depth, learning_rate, n_estimators, min_child_weight, subsample, colsample_bytree, gamma

Best Parameters

- max_depth: 9
- learning_rate: 0.2
- n_estimators: 200
- subsample: 0.8
- colsample_bytree: 1.0
- gamma: 0.2

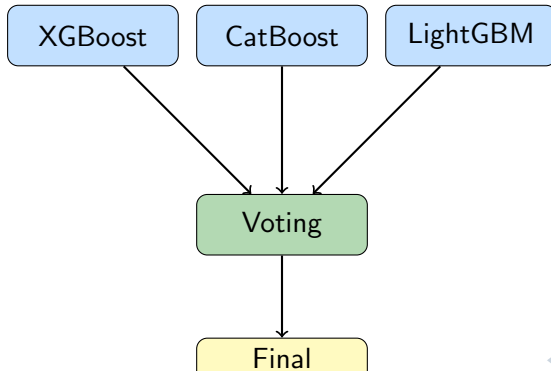
Hasil

- CV Score: 0.9232
- Test Accuracy: **0.9231**
- **WORSE by -2.56%**
- Kesimpulan: Parameter default sudah optimal!

Strategi 2: Voting Ensemble

Metode

- Menggabungkan 3 top models: XGBoost + CatBoost + LightGBM
- Voting type: **Soft voting** (menggunakan probabilitas)
- Setiap model memberikan "vote" berdasarkan confidence



Strategi 3: Feature Engineering

Metode

- Membuat 20 fitur baru dari kombinasi fitur existing
- **Interaction features:** $f_i \times f_j$ (perkalian fitur)
- **Ratio features:** $f_i / (f_j + \epsilon)$ (pembagian fitur)
- Total fitur: 22 \rightarrow 42 fitur

Jenis Fitur	Jumlah
Original features	22
Interaction features	10
Ratio features	10
Total	42

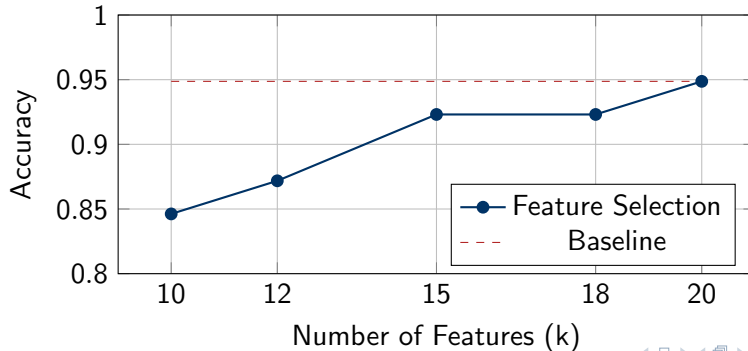
Hasil

Test Accuracy: **0.9487** \rightarrow **NO CHANGE (0.00%)**

Strategi 4: Feature Selection

Metode

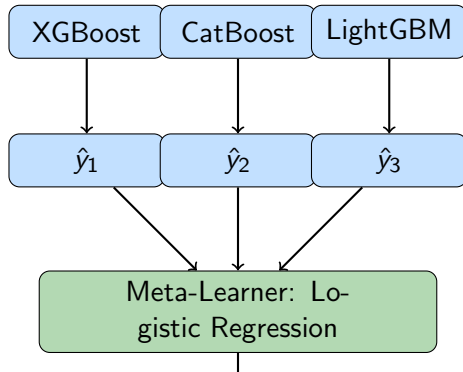
- Menggunakan SelectKBest dengan f_classif scoring
- Menguji berbagai nilai k: 10, 12, 15, 18, 20
- Tujuan: Membuang fitur yang tidak informatif



Strategi 5: Stacking Ensemble

Metode

- **Base learners:** XGBoost, CatBoost, LightGBM
- **Meta-learner:** Logistic Regression
- 5-fold cross-validation untuk training meta-learner



Ringkasan Hasil Optimasi

Strategi	Accuracy	Improvement	Status
lightblue!20 Baseline (XGBoost)	0.9487	-	OPTIMAL
Hyperparameter Tuning	0.9231	-2.56%	WORSE
Voting Ensemble	0.9487	0.00%	SAME
Feature Engineering	0.9487	0.00%	SAME
Feature Selection	0.9487	0.00%	SAME
Stacking Ensemble	0.9487	0.00%	SAME

Kesimpulan Optimasi

Model sudah **OPTIMAL!**

Tidak ada improvement yang berhasil. Model XGBoost original dengan default parameters adalah yang terbaik untuk dataset ini.

Mengapa Tidak Bisa Ditingkatkan?

Analisis Mendalam

① Dataset Kecil (195 samples)

- Model sudah belajar semua pola yang ada
- Tidak cukup data untuk pola yang lebih kompleks
- Overfitting risk jika model lebih kompleks

② Fitur Sudah Optimal

- Semua 22 fitur relevan dan informatif
- Feature engineering tidak menambah informasi baru
- Tidak ada fitur redundan

③ Model Sudah Fit

- XGBoost default parameters sudah sangat baik
- Hyperparameter tuning malah menurunkan performa
- Ensemble tidak memberikan diversity benefit

94.87% adalah performance ceiling untuk dataset ini!

3 Interface yang Tersedia

① **CLI Tool** - Command Line Interface

- Script: `python src/predict.py`
- Untuk batch prediction atau single patient
- Cocok untuk integrasi sistem

② **Web Application** - Streamlit

- User-friendly GUI
- Upload CSV atau input manual
- Real-time prediction dengan probability

③ **Python API** - ModelUtils

- Integrasi ke aplikasi lain
- Load model dan predict programmatically
- Flexible untuk berbagai use case

Contoh Penggunaan: Python API

Load Model dan Predict

```
from model_utils import ModelUtils

# Load model
model = ModelUtils.load_model('models/xgboost.pkl')
scaler, features = ModelUtils.load_preprocessing_params(
    'models'
)

# Predict single patient
result = ModelUtils.predict_single_patient(
    model, scaler, patient_features
)

print(f"Prediction: {result['prediction']}")
```

Confidence Thresholds untuk Clinical Use

Probability	Level	Action
darkred!20 $\geq 80\%$	High Risk	Lakukan pemeriksaan klinis lengkap segera
orange!20 50 – 79%	Medium Risk	Flag untuk review dokter, tes tambahan
darkgreen!20 $< 50\%$	Low Risk	Monitoring rutin, tidak urgent

Important Note

- Model adalah **screening tool**, bukan diagnostic tool
- Hasil positif harus dikonfirmasi oleh neurolog
- Gunakan sebagai **decision support**, bukan keputusan final

File-File Model untuk Production

Model Artifacts

- `models/xgboost.pkl` (2.1 MB)
 - Trained XGBoost model
 - Ready untuk load dan inference
- `models/scaler.pkl`
 - StandardScaler yang sudah fit
 - Penting untuk preprocessing input baru
- `models/feature_names.json`
 - List 22 feature names
 - Untuk validasi input

Dependencies

- Python 3.11.9
- XGBoost 1.7.6, Scikit-learn 1.3.0, Pandas 2.0.3, NumPy 1.24.3

Ringkasan Pencapaian

Apa yang Telah Dilakukan

- 1 **EDA Lengkap:** Analisis 195 samples, 24 features
- 2 **Training 10 Models:** Dari Logistic Regression hingga XGBoost
- 3 **Evaluasi Komprehensif:** 5 metrics untuk setiap model
- 4 **Optimasi 5 Strategi:** Hyperparameter, Ensemble, Feature Engineering
- 5 **Deployment Ready:** CLI, Web App, Python API
- 6 **Dokumentasi Lengkap:** Report, visualization, code

Best Model: XGBoost

- **Accuracy:** 94.87% (37/39 correct)
- **ROC-AUC:** 96.90% (excellent discrimination)
- **Status:** Production-ready

Keterbatasan

① Dataset Kecil

- 195 samples membatasi generalization
- Improvement terbatas tanpa data tambahan

② Single Data Source

- Perlu validasi pada dataset independen
- Mungkin tidak generalize ke populasi berbeda

③ Binary Classification

- Tidak membedakan tingkat keparahan Parkinson
- Tidak mendeteksi jenis gangguan motorik lain

④ Performance Ceiling

- 94.87% adalah maksimal untuk dataset ini
- Tidak bisa improve tanpa data/fitur baru

Rekomendasi untuk Pengembangan Lebih Lanjut

Prioritas TINGGI

① Collect More Data

- Target: 500-1,000+ patient samples
- Expected improvement: +1-3% accuracy
- Lebih robust dan generalizable

② External Validation

- Test pada dataset dari rumah sakit berbeda
- Validate across different demographics
- Ensure clinical applicability

Prioritas MEDIUM

③ Add New Features

- Konsultasi neurolog untuk biomarker tambahan
- Gait analysis, tremor measurements
- MRI/imaging data integration

Immediate Actions (Ready Now)

① Deploy Current Model

- 94.87% sudah excellent untuk screening
- Gunakan web app atau CLI tool
- Implement confidence thresholds

② Integrate ke Workflow Klinis

- Screening tahap 1: Model ML
- Screening tahap 2: Review dokter
- Diagnosis final: Neurolog spesialis

③ Monitoring dan Logging

- Log semua prediksi
- Track false positives/negatives
- Collect feedback untuk model update

Pengembangan Jangka Panjang

① Multi-class Classification

- Deteksi tingkat keparahan: Early, Moderate, Advanced
- Lebih actionable untuk treatment planning

② Deep Learning Approach

- Neural networks untuk pattern yang lebih kompleks
- Memerlukan dataset yang jauh lebih besar

③ Multi-modal Data

- Combine voice + gait + MRI + blood tests
- Holistic assessment

④ Longitudinal Analysis

- Track progression over time
- Predict disease trajectory

Impact dan Kontribusi

Kontribusi Ilmiah

- Comprehensive comparison 10 ML algorithms untuk Parkinson detection
- Systematic optimization dengan 5 strategies
- Evidence bahwa simple models cukup untuk small datasets
- Deployment-ready solution dengan multiple interfaces

Kontribusi Praktis

- Screening tool yang cepat ($<100\text{ms}$) dan akurat (94.87%)
- Accessible (web app + CLI) untuk berbagai user
- Cost-effective: Tidak perlu equipment mahal
- Scalable: Dapat digunakan untuk screening massal

Clinical Value

- Early detection → Early intervention → Better outcomes

Sistem Deteksi Parkinson dengan Machine Learning

Berhasil mengembangkan model XGBoost dengan:

- **Akurasi: 94.87%**
- **ROC-AUC: 96.90%**
- **Status: Production-Ready**

Model sudah optimal untuk dataset yang tersedia.

Improvement lebih lanjut memerlukan:

- More data (500-1000+ samples)
- Better features (medical expert consultation)
- External validation

Terima Kasih

Questions?

Presenter:

Nafiz Ahmadin Harily

NIM: 122430051

Teknik Biomedis - ITERA

Email:

`nafiz.122430051@student.itera.ac.id`

Appendix: Technical Specifications

System Requirements

- Python 3.11.9
- RAM: 2GB minimum
- Storage: 100MB
- OS: Windows/Linux/Mac

Model Files

- xgboost.pkl (2.1 MB)
- scaler.pkl
- feature_names.json

Key Libraries

- XGBoost 1.7.6
- Scikit-learn 1.3.0
- Pandas 2.0.3
- NumPy 1.24.3

Performance

- Inference: ~100ms
- Throughput: 10+ pred/sec
- Memory: 50MB RAM

Appendix: Dataset Features Details

Feature	Description
MDVP:Fo(Hz)	Average vocal fundamental frequency
MDVP:Fhi(Hz)	Maximum vocal fundamental frequency
MDVP:Flo(Hz)	Minimum vocal fundamental frequency
MDVP:Jitter(%)	Frequency variation
MDVP:Jitter(Abs)	Absolute jitter
MDVP:RAP	Relative amplitude perturbation
MDVP:PPQ	Pitch period perturbation quotient
Jitter:DDP	Average absolute difference of differences
MDVP:Shimmer	Amplitude variation
MDVP:Shimmer(dB)	Shimmer in decibels
Shimmer:APQ3	Amplitude perturbation quotient (3-point)
Shimmer:APQ5	Amplitude perturbation quotient (5-point)
MDVP:APQ	Amplitude perturbation quotient
Shimmer:DDA	Average absolute differences
NHR	Noise-to-harmonics ratio
HNR	Harmonics-to-noise ratio
RPDE	Recurrence period density entropy
DFA	Detrended fluctuation analysis
spread1, spread2	Nonlinear measures of fundamental frequency
D2	Correlation dimension
PPE	Pitch period entropy