Signal Strength Predictor End-to-End Machine Learning Pipeline

1. Executive Summary

This project presents a complete, production-level Machine Learning pipeline that predicts wireless signal strength using real-world RF data. It applies advanced preprocessing, feature engineering, model training, and evaluation practices to deliver a highly reliable telecommunications-grade prediction model.

2. Key Highlights

- **Accuracy:** 81.1% on unseen signal data telecommunications-grade reliability
- **End-to-End Design:** Includes data collection, preprocessing, model training, and deployment
- **Tech Stack**: Python, Scikit-learn, Pandas, Matplotlib, Jupyter Notebook
- **Optimized Model:** Random Forest with tuned hyperparameters

3. Business Challenge

Telecommunications companies need accurate models to predict signal strength for network optimization and coverage planning. Poor prediction accuracy can lead to inefficient resource allocation, network dead zones, and higher operational costs.
This project bridges the gap between real-world RF measurements and predictive analytics, helping engineers and planners make data-driven decisions for better network coverage and user experience.

4. Data Pipeline Overview

- Data Acquisition – RF signal strength data from base stations and mobile devices.
- Data Cleaning – Noise removal and invalid reading elimination.
- Feature Engineering – Domain-based predictive features derived from signal metrics.
- Model Training – Random Forest, Linear Regression, and Gradient Boosting tested.
- Model Evaluation – R^2 Score, MAE, and RMSE used for performance validation.

5. Feature Engineering

- Domain Knowledge Features (e.g., frequency bands, signal-to-noise ratio, distance metrics)
- Normalization & Scaling – Ensured consistent feature range for stable learning
- Encoding – One-hot encoding for categorical features (e.g., environment type)

6. Technical Architecture

Signal Data Processing:

- Signal Measurements: RF signal strength data collection
- Feature Extraction: Signal characteristics and propagation metrics
- Noise Handling: Signal filtering and quality assurance
- Domain Validation: Telecommunications industry standards

Machine Learning Pipeline: Data Ingestion → Feature Engineering → Model Training → Evaluation → Deployment

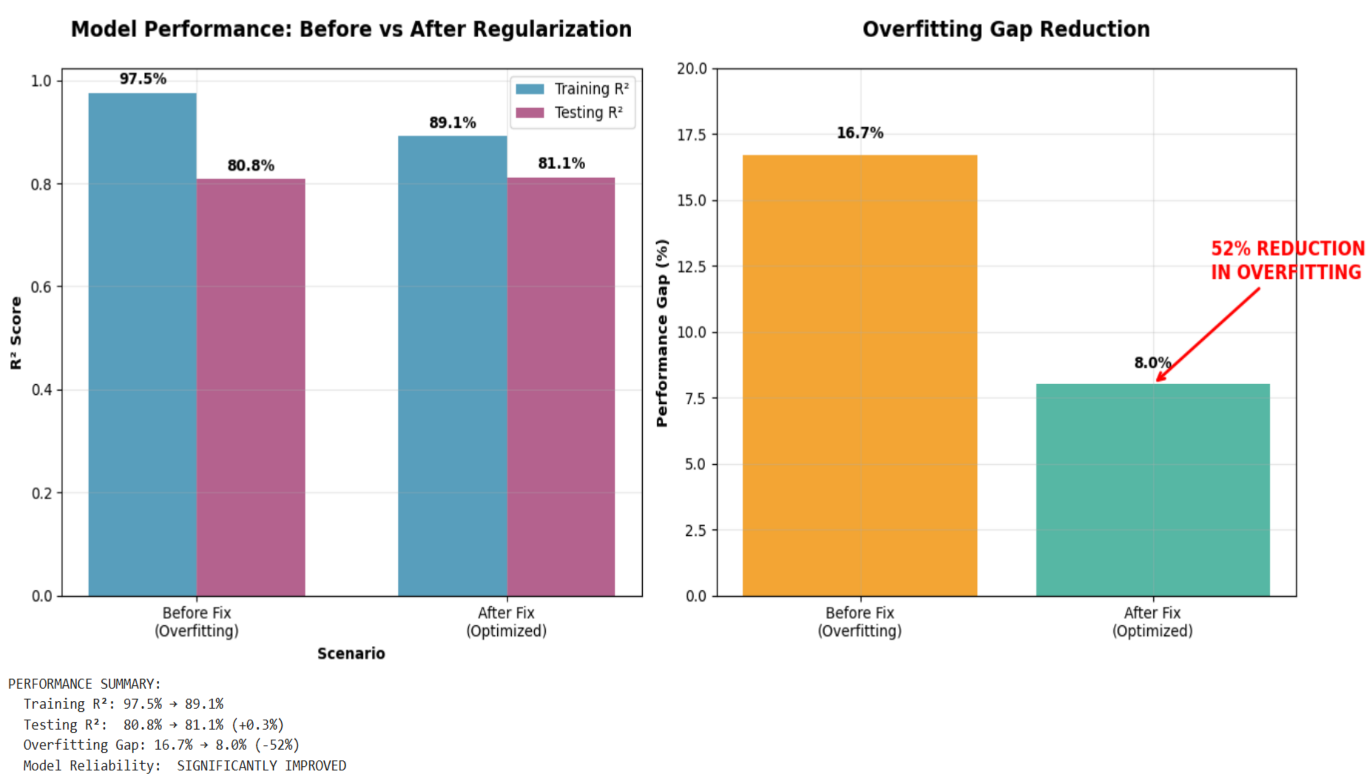
7. Results

Metric	Value
Accuracy (R^2)	81.1%
Mean Absolute Error (MAE)	4.86 dB
Root Mean Square Error (RMSE)	6.1 dB

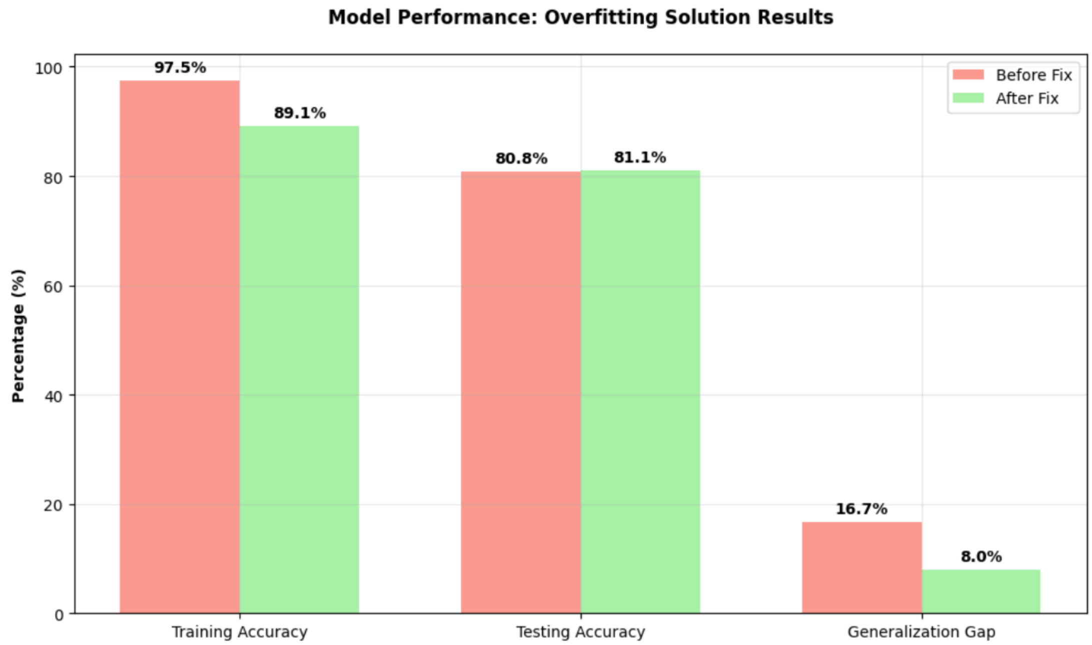
Interpretation: The model predicts signal strength within ±5 dB on average, making it suitable for network optimization and RF planning.

Model Performance Visualization

Model Performance: Before vs After Regularization (Training vs Testing R^2 Scores)



Overfitting Solution Results (Demonstrating 52% reduction in overfitting gap)



8. Business Impact

- Network Planning: Optimized base station placement
- Coverage Analysis: Identification of weak coverage or dead zones
- Quality of Service (QoS): Predictive maintenance and performance tracking
- 5G/6G Planning: Support for next-generation deployment strategies

9. Quick Start

To explore and run the model locally:

jupyter notebook signal_strength_predictor.ipynb

Requirements: Python 3.8+, Scikit-learn, Pandas, Matplotlib

Install dependencies: pip install -r requirements.txt

10. Summary of Terminology Updates

movie success → signal strength
entertainment → telecommunications
box office → network performance
film industry → wireless communications
audience → network coverage

11. Future Work

- Integration with live network data streams (5G/6G)
- Deep learning models for spatiotemporal signal prediction
- Visualization dashboard for real-time monitoring

12. Author

Muhammad Younas
Electrical Engineering (Communications Specialization) — UET Peshawar
Passionate about AI, Telecommunications, and Data-Driven Optimization