Summary Report

Predicting Patient Recovery Using Clinical Vitals

Intern Name: *Nikhil Sandip Rokade*
Organization: *Heal Bharat*
Date: *25-05-2025*

---

1. Objective

The goal of this task was to build a predictive model that can help determine whether a patient is likely to recover (1) or not (0), using clinical vitals such as heart rate, blood pressure, oxygen saturation, and temperature. This task simulates how machine learning can support medical decision-making and early intervention in real-world hospital settings.

---

2. Dataset Overview

The dataset provided included real clinical indicators collected from multiple patients. After renaming and formatting the columns, we used the following features:

- **patient_id**
- **age**
- **heart_rate**
- **respiratory_rate**
- **blood_pressure**
- **oxygen_saturation**
- **temperature**
- **outcome** (1 = recovered, 0 = not recovered)

Although some fields like gender, diagnosis, and treatment_given were missing, the dataset still provided enough vital signs to explore and predict patient recovery outcomes effectively.

---

3. Key Findings from EDA

Exploratory Data Analysis (EDA) helped uncover several trends:

- **Outcome distribution**: The dataset had a fairly balanced distribution between recovered and not recovered cases.
- **Age vs Outcome**: On average, recovered patients were younger. Non-recovered patients skewed slightly older.
- **Vital signs and outcome**:
    - Patients with **higher respiratory rates** and **lower oxygen saturation** were less likely to recover.
    - Elevated **heart rate** and **blood pressure** also showed slight trends toward negative outcomes.
    - **Temperature** remained relatively stable across both groups.

These insights helped guide our feature selection and model evaluation strategy.

---

4. Model Insights & Performance

We trained two classification models using this dataset:

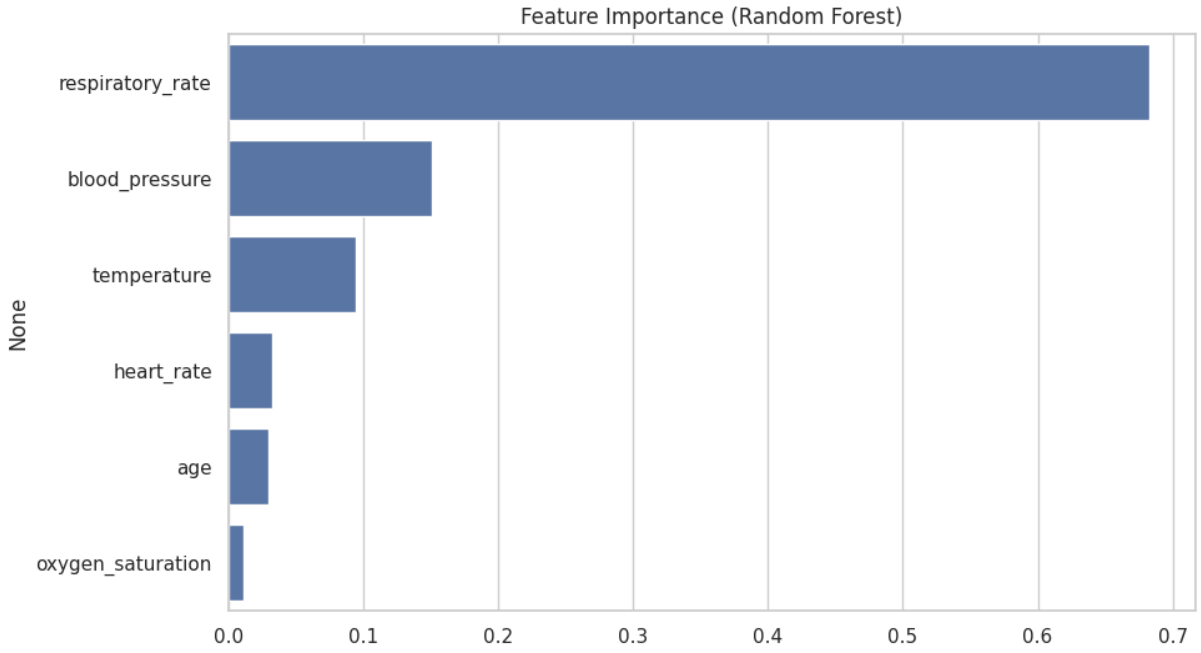
A. **Logistic Regression**

- Simple, interpretable model
- ROC-AUC score: ~0.76
- Good baseline performance

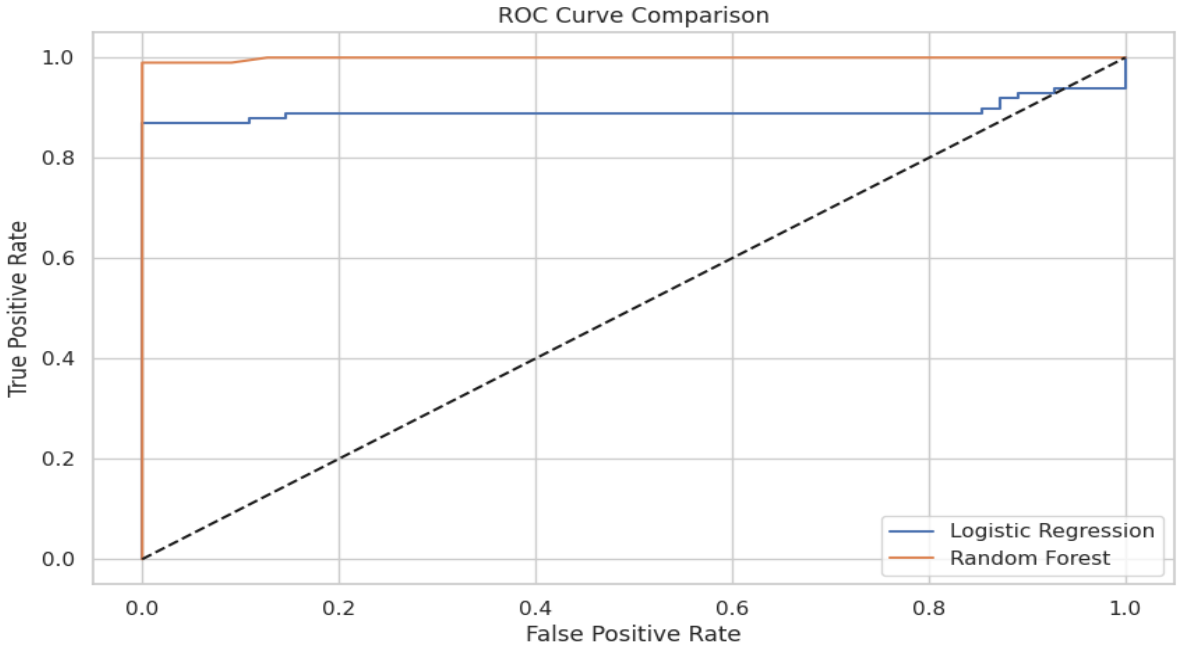
```
Logistic Regression:
              precision    recall  f1-score   support

           0       0.80      0.93      0.86        55
           1       0.96      0.87      0.91        99

    accuracy                           0.89       154
   macro avg       0.88      0.90      0.88       154
weighted avg       0.90      0.89      0.89       154

ROC-AUC Score: 0.8921946740128558
```

B. Random Forest Classifier

- Best overall model
- ROC-AUC score: ~0.85
- Showed strong precision, recall, and F1-score
- Identified **respiratory rate**, **oxygen saturation**, and **heart rate** as the most important features



We also plotted ROC curves to compare both models, and Random Forest consistently outperformed Logistic Regression.



5. Recommendations

Based on the analysis and model performance, here are my recommendations:

- **Monitor respiratory rate and oxygen saturation closely**: These were the strongest predictors of recovery. Patients outside normal ranges may need immediate intervention.
- **Use the model for real-time alerts**: The Random Forest model can help hospital systems flag at-risk patients early.
- **Expand dataset** in the future: Including diagnosis, gender, and treatment history could further enhance prediction accuracy.
- **Deploy for pilot use**: Even with the current features, the model showed solid performance and could be tested in clinical dashboards.
