

Personalized Healthcare Recommendations Report

Name: Parnika Sunda

Institution: Unified Mentor

Date: July 20, 2025

Objective

The aim of this project is to build a machine learning model that predicts whether a person is likely to donate blood again, using features such as recency, frequency, total amount donated, and time. This can support healthcare campaigns by targeting likely repeat donors and improving outreach strategies.

Dataset Overview

File: `blood.csv`

Records: 748

Columns: 5

- **Recency**: Months since last donation
- **Frequency**: Total number of donations
- **Monetary**: Total blood donated (in c.c.)
- **Time**: Time in months since the first donation
- **Class**: Target variable (1 = Likely Donor, 0 = Not Likely)

Workflow Summary

1. Data Exploration & Cleaning

- Verified dataset shape: (748, 5)
- No missing values
- Data types: all integer-based features

2. Visual Analysis


- Histograms plotted for all features to understand distributions
- Correlation heatmap showed strong relationship between Monetary, Frequency, and Time
- Created **interactive Power BI dashboard** for business-friendly presentation:
 - Class prediction pie chart
 - Feature-wise bar chart per class
 - Class filters for interactivity

3. Modeling Process

- Target: **Class** (0 = Not likely to donate, 1 = Likely donor)
- Features: Recency, Frequency, Monetary, Time

Models used:

Model	Accuracy	Notes
Logistic Regression	76%	High precision for Class 0; poor recall for Class 1
Decision Tree	63%	Struggled with class imbalance
Random Forest	75.4%	Balanced performance after oversampling with SMOTE
XGBoost	76.3%	Best performance overall with SMOTE; higher recall for Class 1

 **SMOTE** (Synthetic Minority Oversampling Technique) was applied before training Random Forest and XGBoost to fix the **class imbalance**, helping models learn from minority (Class 1) examples more effectively.

Power BI Dashboard Summary

- Dashboard titled: "**Blood Donation Likelihood Dashboard**"
- Key visuals:
 - Pie chart of predicted donor classes
 - Bar chart of feature averages per predicted class
 - Filters for class segmentation
- Layout optimized with soft backgrounds and section separators

Key Insights

- **Monetary** is the strongest predictor — likely donors tend to have donated significantly more blood
- **Time** and **Frequency** also contribute positively to Class 1 predictions
- Models without SMOTE failed to detect Class 1 properly
- After applying **SMOTE**, ensemble models like **Random Forest** and **XGBoost** achieved much better **recall** for likely donors

Future Improvements

- Add richer features like **age**, **blood type**, or **recent health history**
- Integrate the system into a **real-time donor outreach app**
- Convert Power BI visuals into a **live dashboard** with API-based updates
- Deploy a simple **Streamlit web app** to allow NGOs/hospitals to check likelihood scores