**Blood Donation Prediction Report**

**1. Introduction**

Blood donation shortages pose a critical issue for healthcare systems. This project aims to predict whether a donor will donate blood again based on historical donation data. By leveraging **machine learning**, we provide an automated prediction system to help blood banks and hospitals manage donations more efficiently.

**2. Dataset Overview**

The dataset originates from the **Blood Transfusion Service Center in Hsin-Chu City, Taiwan** and consists of **748 donor records** with the following attributes:

* **Recency (months)** – Number of months since the last donation
* **Frequency (times)** – Total number of past donations
* **Monetary (c.c. blood)** – Total blood volume donated
* **Time (months)** – Number of months since the first donation
* **Target (0 or 1)** – Whether the donor gave blood in March 2007 (1 = Yes, 0 = No)

**3. Data Preprocessing**

1. **Renamed columns** for clarity.
2. **Checked for missing values** (none found).
3. **Standardized numerical features** using StandardScaler.
4. **Handled class imbalance** using **SMOTE** (Synthetic Minority Over-sampling Technique).
5. **Split data** into **80% training and 20% testing sets**.

**4. Model Training and Selection**

**4.1 Models Evaluated**

* **Logistic Regression** (Baseline model)
* **Random Forest Classifier**
* **TPOT AutoML** (Automated model selection)

**4.2 Best Model Selection**

After running multiple models, **TPOT AutoML** identified the best-performing model:

MLPClassifier(RobustScaler(input\_matrix), alpha=0.0001, learning\_rate\_init=0.01)

* **Cross-validation score:** 80.6%
* **Final test accuracy:** 74.67%

**5. API Deployment**

To integrate the model into a real-world application, we deployed it using **Flask**. The API:

* Accepts donor data via a **POST request**.
* Loads the trained model to **predict donation probability**.
* Returns predictions as JSON responses.

**5.1 Running the API**

python app.py

The API is accessible at **http://127.0.0.1:5000/predict**.

**6. Recommendations for Improvement**

1. **Increase dataset size** to improve model generalization.
2. **Fine-tune hyperparameters** of the MLPClassifier.
3. **Deploy the API using Docker** for cloud deployment.
4. **Integrate a frontend UI** for user-friendly interaction.

**7. Conclusion**

This project successfully demonstrates the use of **machine learning** for predicting blood donations. The deployment of an **API** allows healthcare organizations to make real-time predictions, improving blood supply management. Future improvements include expanding the dataset, refining the model, and deploying on cloud platforms.