Personalized Healthcare Recommendations

QUESTIONS & ANSWERS

1. What is the primary objective of this project?

Answer:
To build a machine learning model that predicts whether a person is likely to donate blood again based on behavioral variables such as Recency, Frequency, Monetary contribution, and Time.

---

2. What dataset is used in this project?

Answer:
The *blood.csv* dataset, also known as the Blood Transfusion Service Center Dataset, containing 748 rows and 5 columns.

---

3. What do the features Recency, Frequency, Monetary, and Time represent?

Answer:

- **Recency:** Months since the last donation
- **Frequency:** Total number of previous donations
- **Monetary:** Total volume of blood donated
- **Time:** Months since the first donation

---

4. What is the target variable in this dataset?

Answer:
Class — a binary label where

- **1 = likely to donate again**,

- **0 = unlikely to donate**.

---

5. Are there any missing values in the dataset?

Answer:
No, the dataset is completely clean and contains zero missing values.

---

6. Why is Recency an important feature?

Answer:
Recency reflects donor engagement. People who donated recently are more likely to donate again.

---

7. Why was the Random Forest model chosen?

Answer:
It performs well on tabular data, handles nonlinear patterns, prevents overfitting, and provides high accuracy.

---

8. How was the dataset split for training and testing?

Answer:
An **80:20 stratified split** was used to maintain class proportions.

---

9. Why was StandardScaler used?

Answer:
To normalize the numerical features and improve model performance by putting all features on the same scale.

---

10. What evaluation metrics did you use?

Answer:
Accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC score.

---

11. What was the model's task classification type?

Answer:
A **binary classification** task predicting donor vs. non-donor.

---

12. How does Frequency affect donor likelihood?

Answer:
Higher frequency indicates consistent past donations, making future donations more likely.

---

13. What role does Monetary play in prediction?

Answer:
It reflects total donated blood volume; higher values usually correlate with frequent donors.

---

14. Why is Time a relevant feature?

Answer:
It shows how long a donor has been affiliated with the donation program, helping differentiate new vs. long-term donors.

---

15. What is the significance of feature importance in the model?

Answer:
It indicates which features influence predictions the most and helps validate model transparency.

---

16. How does the personalized recommendation system work?

Answer:
It converts model predictions into actions:

- Prediction = 1 → Encourage donor
- Prediction = 0 → Send awareness and motivation messages

---

17. What real-world problem does this project solve?

Answer:
It helps blood banks predict future donor participation, improving supply chain management.

---

18. How was deployment implemented?

Answer:
Using a REST API built in **Flask** and a user-friendly web interface using **Streamlit**.

---

19. Can this model be improved further?

Answer:
Yes, by adding demographic/medical features, tuning hyperparameters, or trying advanced algorithms like XGBoost.

---

20. What ethical considerations apply to this project?

Answer:
Ensuring data privacy, preventing bias, obtaining donor consent, and maintaining transparency in predictions.