🔍 Stroke Prediction System using Logistic Regression

Final Project Documentation

📘 Acknowledgement

We take immense pleasure in presenting our project, "Stroke Prediction System using Logistic Regression", and would like to express our sincere gratitude to everyone who contributed to its completion.

We are grateful to our institution for providing us with the necessary infrastructure, labs, and learning environment. Our heartfelt thanks go to our project guide for their consistent support, expert advice, and patience throughout the development of this system.

We would also like to thank our friends and families for their continuous encouragement and motivation during this journey.

📑 Table of Contents

1. Project Overview

1.1 Introduction

Stroke is a serious medical condition that can lead to long-term disability or death. It occurs when the blood supply to part of the brain is cut off. Early prediction and prevention are critical in reducing stroke-related mortality.

This system leverages **Logistic Regression** and a web interface built with **Flask** to allow users to input personal and medical information and instantly receive a risk prediction. If a stroke risk is detected, a warning is issued along with helpful resources.

1.2 Scope and Objective

- **Scope**: Predict the likelihood of a stroke based on user input and display educational content if risk is high.
- **Objective**:
    - Perform real-time risk prediction using a trained model.
    - Educate users with information like FAST symptoms.
    - Provide immediate alerts (video, modal popup) for high-risk cases.

1.3 Modules and Description

👤 User Module

- **Input Form**: Enter data such as age, gender, health conditions, etc.
- **Validation**: Client-side validations to ensure input integrity.
- **Prediction**: Display Likely / Not Likely result with visual aids.

🔍 **Model Input Features:**

1. Gender
2. Age
3. Hypertension
4. Heart Disease
5. Ever Married
6. Work Type
7. Residence Type
8. Average Glucose Level
9. BMI
10. Smoking Status

1.4 Existing vs Proposed System

Existing System	Proposed System
Manual risk analysis	ML-based automated prediction
No real-time warning	Modal + Video alert for high-risk
Lack of user guidance	FAST image + stroke education built-in
Desktop only	Mobile responsive web app

---

2. Project Lifecycle

2.1 Model Used: Waterfall Model

- Requirements Analysis

- Design
- Implementation
- Testing
- Maintenance

---

3. Project Design

3.1 Architecture

- Frontend: HTML, CSS, Bootstrap 5
- Backend: Python (Flask)
- ML Model: Logistic Regression + SMOTE
- Data Flow: User Input → Preprocessor → Model → Result View

---

4. Implementation

4.1 Technologies Used

- Python 3.9
- Flask Framework
- Scikit-learn, Pandas, Numpy
- HTML5, CSS3, Bootstrap
- Joblib (for model export)
- SMOTE (to handle class imbalance)

4.2 Frontend Highlights

- Client-side validation for age, glucose, BMI
- Responsive design with Bootstrap
- Video background for "Likely"
- FAST image + modal education section

4.3 Preprocessing

- Categorical columns encoded using OneHotEncoder & OrdinalEncoder
- Numerical columns passed directly
- Final input: 18 feature columns
- Preprocessing object saved with model

4.4 Model Building

- Multiple models tested: Logistic Regression, Random Forest, XGBoost, LightGBM
- Logistic Regression with SMOTE gave highest **recall**

Final model stored as dictionary:

python
CopyEdit

```
{
  'model': lr_with_smote,
  'preprocessor': column_transformer,
  'encoded_cols': [...],
  'numeric_cols': [...]
}
```

-

4.5 Flask Integration

- Routes:
  - `/` → Input form
  - `POST /` → Accept form, transform, predict
- Output:
  - Likely / Not Likely with probability
  - Custom visuals for each case

---

5. Testing

5.1 Levels of Testing

- Unit Testing: Each route, function
- Integration Testing: Model + Web together
- System Testing: Frontend to backend flow
- User Acceptance Testing: Usability tested with users

5.2 Validation Criteria

- Age must be >10 and <110
- All fields are mandatory
- Inputs are range-validated and sanitized

5.3 Test Case Example

Test	Input	Expected Output
Age Validation	Age = 5	Error: Age must be >10

Stroke Case	Smoker + Heart Disease	Prediction: Likely + Warning Modal

6. Advantages & Limitations

✅ Advantages

- Accurate ML-based stroke risk prediction
- Real-time alerts and symptom guidance
- Easy, mobile-friendly web access
- High recall from SMOTE balancing

⚠️ Limitations

- Prediction based on limited features
- Not a replacement for clinical diagnosis

7. Conclusion

This project successfully integrates machine learning with a user-friendly web interface to predict stroke risk. By combining clinical features with logistic regression and strong UI/UX, it serves as a powerful tool for early intervention and public awareness.