**Project Title: H1N1 & Seasonal Flu Vaccine Prediction**

**Design Document**

**C-DAC, Bangalore**

**Version 1.0**

**09-02-2025**

**Prepared By:**

Abhishek Jaydeo Bhure

Kaname Hariom Venkatrao

Shubhi Paliwal

Shiva Gupta

Kokate Neha Vitthal

## **Index**

1. Introduction
2. System Overview
3. Architecture Design
4. Data Design
5. Machine Learning Models
6. Risks & Constraints
7. Future Enhancements
8. Conclusion

### **Introdution**

**1.1 Project Overview**

This project aims to predict how likely individuals are to receive their H1N1 and seasonal flu vaccines using machine learning models. Various algorithms such as Support Vector Machine (SVM), Logistic Regression, Random Forest, Multi-Layer Perceptron (MLP), and Linear models have been employed to improve prediction accuracy.

**1.2 Purpose**

The purpose of this document is to provide a detailed design overview of the H1N1 & Seasonal Flu Vaccine Prediction system, including its architecture, data processing pipeline, model selection, and implementation details.

**1.3 Scope**

This project focuses on analyzing individual characteristics and behaviors to predict vaccine uptake. The models are trained on features such as demographics, health status, and behavioral factors.

**1.4 Intended Audience**

This document is intended for data scientists, machine learning engineers, software developers, and project stakeholders involved in vaccine prediction research and development.

**1.5 References**

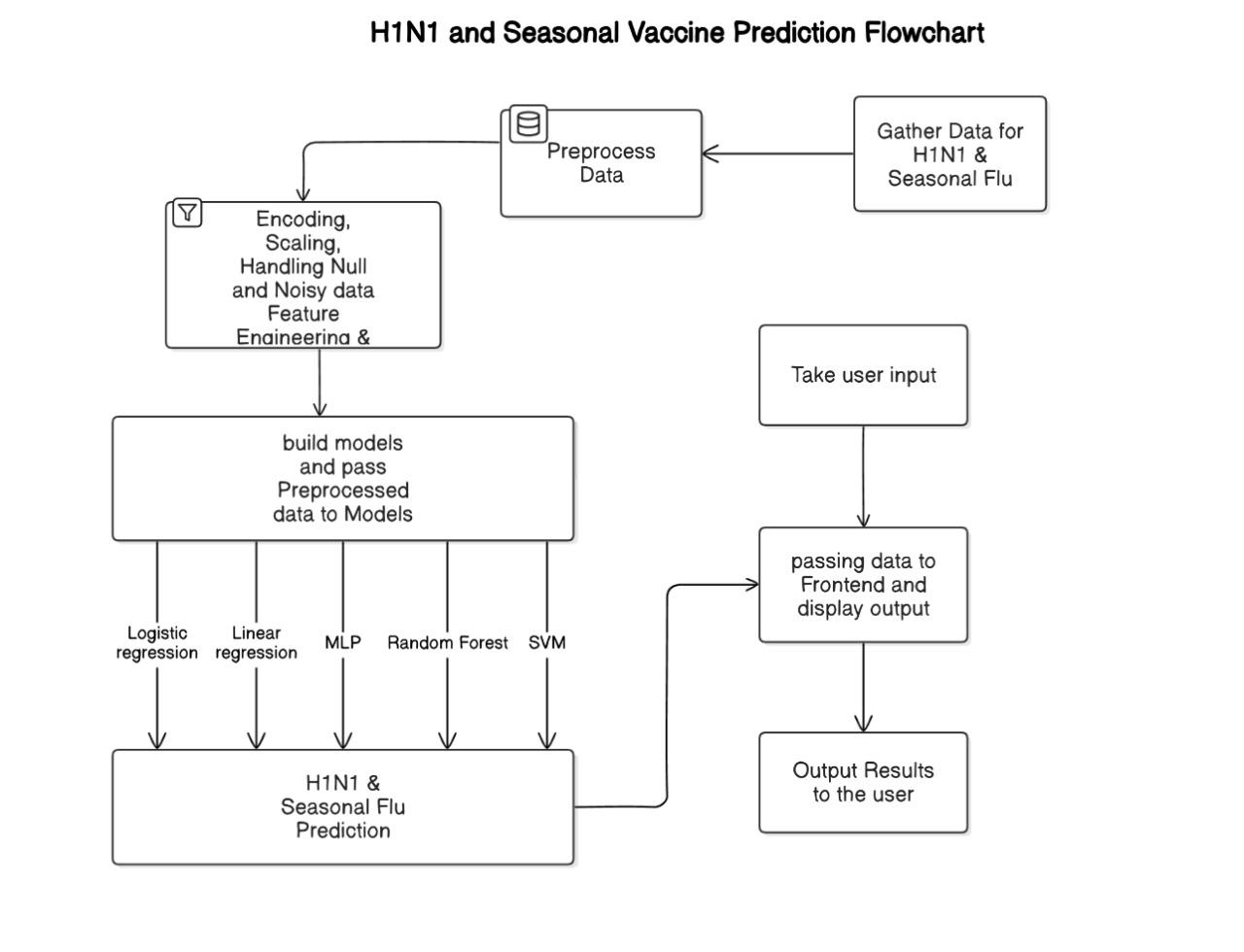
* Functional Requirement Specifications (FRS)
* Machine Learning Documentation
* Research Papers on Vaccine Prediction

**2. System Overview**

The system processes demographic and behavioral data to predict vaccine uptake probabilities. It uses multiple machine learning models to generate predictions and selects the best-performing one based on evaluation metrics.

1. **Architecture Design**

#### 3.1 Architecture Diagram



3.2 Components

* **Data Preprocessing**: Handles data cleaning, encoding categorical features, and normalizing numerical features.
* **Feature Engineering**: Identifies the most relevant features for improving model performance.
* **Model Training & Selection**: Trains various models and selects the best one based on accuracy, precision, recall, and F1-score.
* **Prediction & Evaluation**: Uses the trained model to predict vaccine uptake and evaluates performance.

### **4.Data Design**

### **4.1 Data Sources**

* Publicly available vaccine uptake datasets
* Health survey data containing demographic and behavioral information

4.2 Data Preprocessing

* Handling missing values
* Encoding categorical variables
* Feature scaling for numerical attributes

#### 4.3 Database Schema

| | Field Name | Data Type | Description | | --- | --- | --- | | respondent\_id | Integer | Unique Identifier | | h1n1\_concern | Float | Concern level for H1N1 | | doctor\_recc\_h1n1 | Integer | Doctor recommendation for H1N1 | | income\_poverty | Categorical | Income category | | age\_group | Categorical | Age category | | h1n1\_vaccine | Integer | Target variable (1: Vaccinated, 0: Not Vaccinated) | | seasonal\_vaccine | Integer | Target variable (1: Vaccinated, 0: Not Vaccinated) | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

### **5. Machine Learning Model**

#### 5.1 Model Selection

We implemented and evaluated the following models:

1. **Support Vector Machine (SVM)**
2. **Logistic Regression**
3. **Random Forest**
4. **Multi-Layer Perceptron (MLP)**
5. **Linear Regression-based Model**

#### 5.2 Model Evaluation

Each model was evaluated based on:

* **Accuracy**
* **Precision & Recall**
* **F1-score**
* **ROC-AUC Score**

#### 5.3 Best Model Selection

* The best model was chosen based on its performance across multiple metrics. Hyperparameter tuning was performed to optimize accuracy.

### **6. Risks & Constraints**

* Data imbalance affecting prediction performance
* Privacy concerns with sensitive health data
* Computational cost of training deep learning models

### **7. Future Enhancements**

* Implementing deep learning models for better accuracy
* Deploying the model as a web-based application for real-time predictions
* Expanding dataset to include more demographic and behavioral factors

### **8. Conclusion**

This document provides a structured approach to designing the H1N1 & Seasonal Flu Vaccine Prediction system. By leveraging multiple machine learning techniques, the system aims to provide reliable vaccine uptake predictions and support public health decision-making.

### **Appendix**

#### A. Acronyms & Definitions

* **FRS** - Functional Requirements Specification
* **SVM** - Support Vector Machine
* **MLP** - Multi-Layer Perceptron
* **ROC-AUC** - Receiver Operating Characteristic - Area Under Curve

#### B. References & External Links

* Research Paper on Vaccine Prediction
* Machine Learning Documentation

#### C. Design Considerations

* Scalability & Performance Testing
* API Integration for Deployment