Software Requirements Specification (SRS)

# Naive Bayes Classifier

## 1. Introduction

The Naive Bayes Classifier project aims to develop a system that predicts the presence or absence of diabetes in patients based on health-related attributes. This document outlines the requirements for the system.

## 2. System Overview

The system consists of the following components:

Data loading and preprocessing module: Loads and preprocesses the dataset.

Training and testing module: Divides the dataset into training and testing subsets, trains the Naive Bayes classifier, and generates predictions.

Evaluation module: Measures the performance of the classifier using accuracy score.

Visualization module: Provides visualizations for classification results.

## 3. Functional Requirements

### 3.1 Data Loading and Preprocessing Module

Load the dataset from a CSV file.

Handle missing values appropriately.

Perform any required data preprocessing.

### 3.2 Training and Testing Module

Split the dataset into training and testing subsets.

Implement the Gaussian Naive Bayes algorithm.

Handle both categorical and continuous features.

### 3.3 Evaluation Module

Compare predicted labels with actual labels using accuracy score.

### 3.4 Visualization Module

Generate visualizations to aid in understanding classification results.

## 4. Non-functional Requirements

Performance: The system should process the dataset efficiently.

Usability: The system should be user-friendly or provide clear instructions.

Reliability: The system should handle exceptions gracefully.

Compatibility: The system should be implemented using Python with required libraries.

## 5. Constraints

The dataset must be in CSV format.

The dataset should contain relevant attributes for predicting diabetes.