High Level Design Document

Introduction

This High Level Design (HLD) document outlines the architecture and core components for **StackVote - Ensemble Learning Playground**. StackVote is a Streamlit-based web application enabling users to experiment with stacking, blending, and voting ensemble classifiers. The app allows dataset uploads, model selection, parameter tuning, and visualization of ensemble versus individual model performance.

---

1. System Architecture Overview

Architecture Description:
StackVote is a modular, client-server web application built with Streamlit. The frontend provides an interactive UI for user input and visualization, while the backend handles data processing, model training, and evaluation using scikit-learn and related libraries.

Main System Components:

Component	Role
User Interface	Streamlit-based UI for data upload, model selection, and visualization
Data Manager	Handles dataset upload, validation, preprocessing (pandas, numpy)
Model Manager	Manages base/ensemble model selection, parameter tuning
Training Engine	Trains individual and ensemble models (scikit-learn)
Evaluation Engine	Computes metrics, generates performance reports
Visualization Module	Renders plots and comparison charts in the UI

---

2. Component Interactions

Sequence Step	Interaction Description
1	User uploads dataset and selects models via UI
2	Data Manager validates and preprocesses data
3	Model Manager configures base and ensemble models based on user input
4	Training Engine fits models on processed data
5	Evaluation Engine computes metrics and passes results to Visualization Module
6	Visualization Module displays results and comparisons in the UI

---

3. Data Flow Overview

Data Flow Step	Source	Destination	Description
Dataset Upload	User Interface	Data Manager	User uploads CSV; data is validated
Preprocessed Data	Data Manager	Model Manager	Cleaned data passed for model configuration
Model Configurations	Model Manager	Training Engine	Model/parameter selections sent for training
Trained Models/Results	Training Engine	Evaluation Engine	Models evaluated on test/validation data
Metrics/Visuals	Evaluation Engine	Visualization Module	Results visualized in UI

---

4. Technology Stack

Layer/Function	Technology/Framework
Web UI	Streamlit
Data Processing	pandas, numpy
Machine Learning	scikit-learn
Visualization	matplotlib, seaborn, Streamlit built-ins
Language/Runtime	Python 3.x

---

5. Scalability & Reliability

- **Stateless Design:** Each user session is independent; no persistent backend state.
- **Resource Management:** Model training is limited to in-memory datasets; suitable for small/medium datasets.
- **Security:** File uploads are validated; no code execution from user input.
- **Extensibility:** Modular design allows easy addition of new models or visualization types.
- **Deployment:** Can be hosted on Streamlit Cloud or similar platforms for multi-user access.

End of Document