RAJALAKSHMI ENGINEERING COLLEGE
An AUTONOMOUS Institution
Affiliated to ANNA UNIVERSITY, Chennai

25 years OF ACADEMIC EXCELLENCE

Department of Computer Science and Engineering

REAL ESTATE INVESTMENT SAFETY PREDICTION USING MACHINE LEARNING

Mrs. M. Divya M.E.
Asst.professor, dept of cse,
Rajalakshmi Engineering College

Sanjana Shree S
(2116220701245)

Problem Statement and Motivation

In the real estate sector, investors often face uncertainty when determining the safety and profitability of a property investment. Key factors like location crime rate, infrastructure quality, proximity to amenities, and property valuation can significantly impact investment decisions. However, relying solely on intuition or unstructured information can lead to poor outcomes and financial losses. This project aims to address this issue by developing a machine learning-based solution that predicts whether a real estate investment is safe or not. By leveraging data-driven insights, the model empowers investors to make informed decisions, reduce risk, and increase confidence in their investment choices.

Existing System

In the current real estate investment landscape, decision-making is largely driven by human judgment, market trends, and recommendations from agents or brokers. Investors typically rely on manual research, local knowledge, or past experiences to assess whether a property is a safe investment. This approach lacks consistency, is prone to bias, and does not utilize the full potential of available data. Additionally, there are limited intelligent tools or automated systems that can evaluate property safety based on quantifiable metrics, making the process inefficient and often unreliable.

Objectives

- To develop a machine learning model that classifies real estate investments as "Safe" or "Not Safe".
- To preprocess and analyze synthetic real estate data for model training.
- To evaluate multiple classification algorithms and select the best-performing one.
- To implement the selected model using Python libraries like scikit-learn and XGBoost.
- To build a user-friendly web interface using Streamlit for real-time predictions.
- To assist investors in making data-driven and risk-aware investment decisions.

Abstract

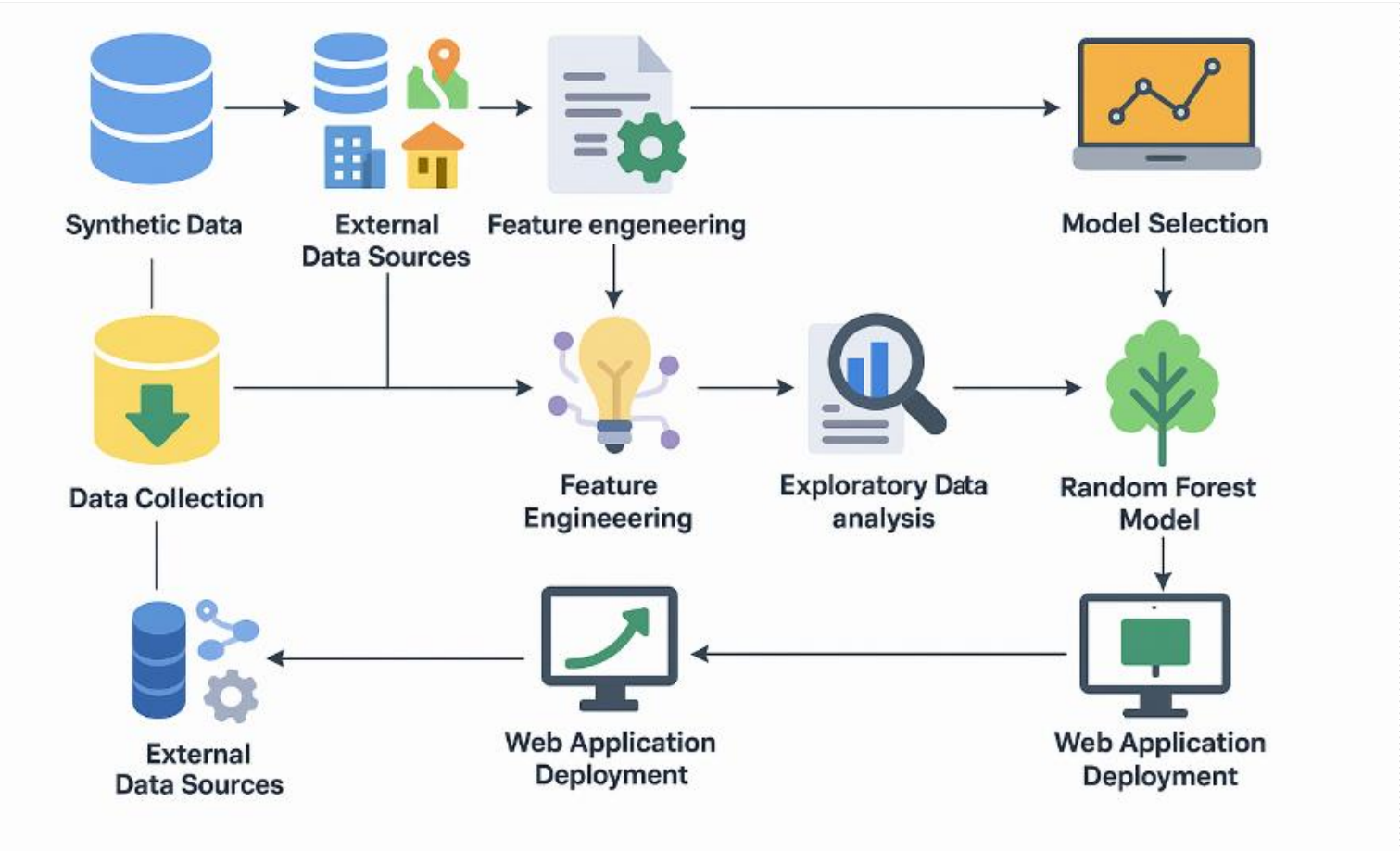
This project, **"Real Estate Investment Safety Predictor"**, uses machine learning to assess the safety of real estate investments based on property features, market trends, and location data. A synthetic dataset was analyzed using classification algorithms, with the **Random Forest model** delivering the best results. The final model was deployed via a **Streamlit web app** for real-time predictions. This tool aids investors in making informed, data-driven decisions. It demonstrates the power of AI in real estate risk analysis and offers potential for future enhancement using real-world data and economic indicators.

Proposed System

The system predicts real estate investment safety using machine learning. It involves:

- **Data Preprocessing**: Cleaning, encoding, and scaling synthetic real estate data.
- **Exploratory Data Analysis**: Understanding feature relationships and distributions.
- **Model Training**: Using classification algorithms (Logistic Regression, Decision Tree, Random Forest).
- **Model Evaluation**: Based on accuracy, precision, recall, and F1-score.
- **Best Model Deployment**: Random Forest integrated into a **Streamlit web app**.
- **User Interaction**: Real-time predictions via a simple and intuitive interface.

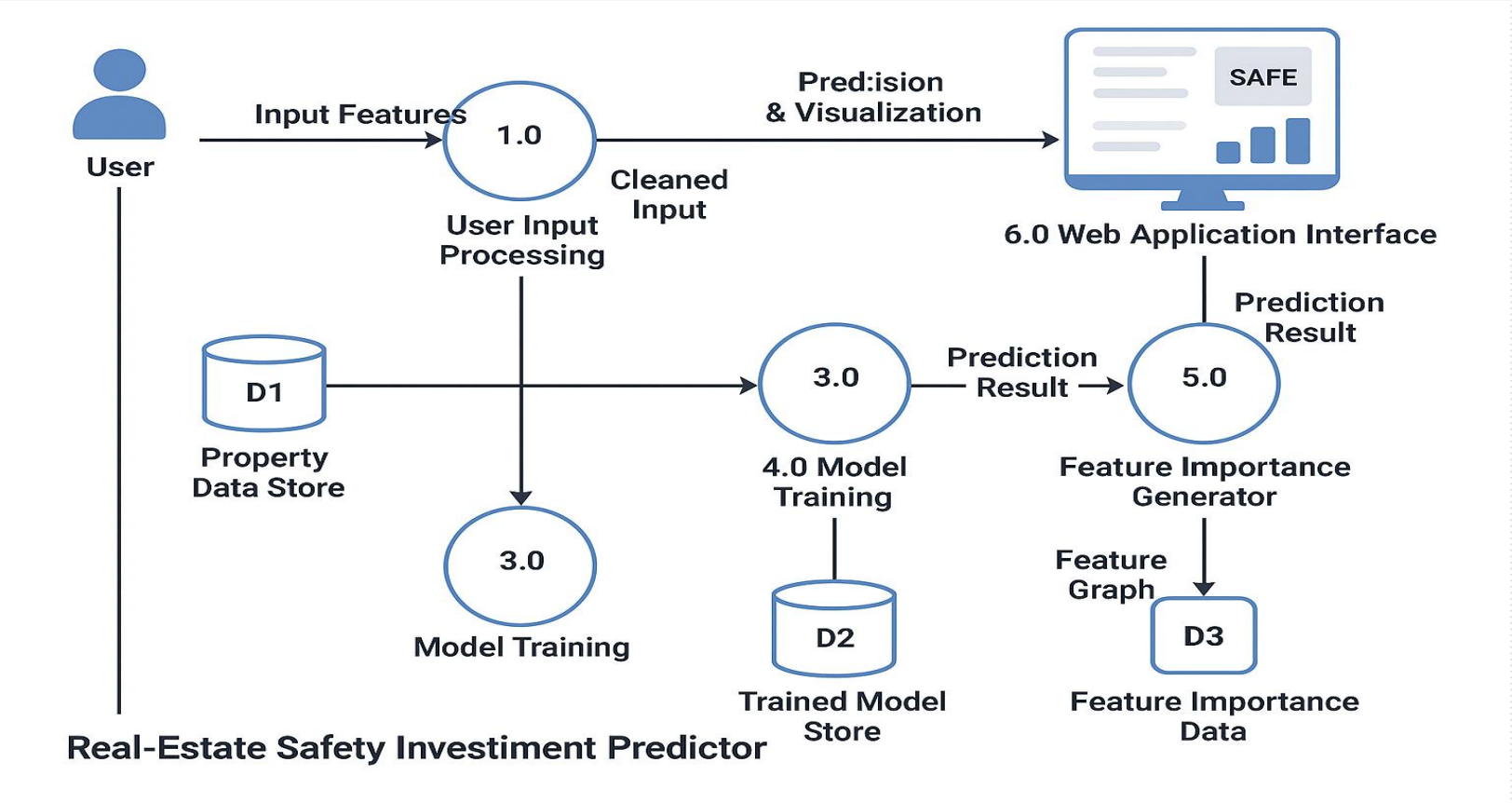
System Architecture



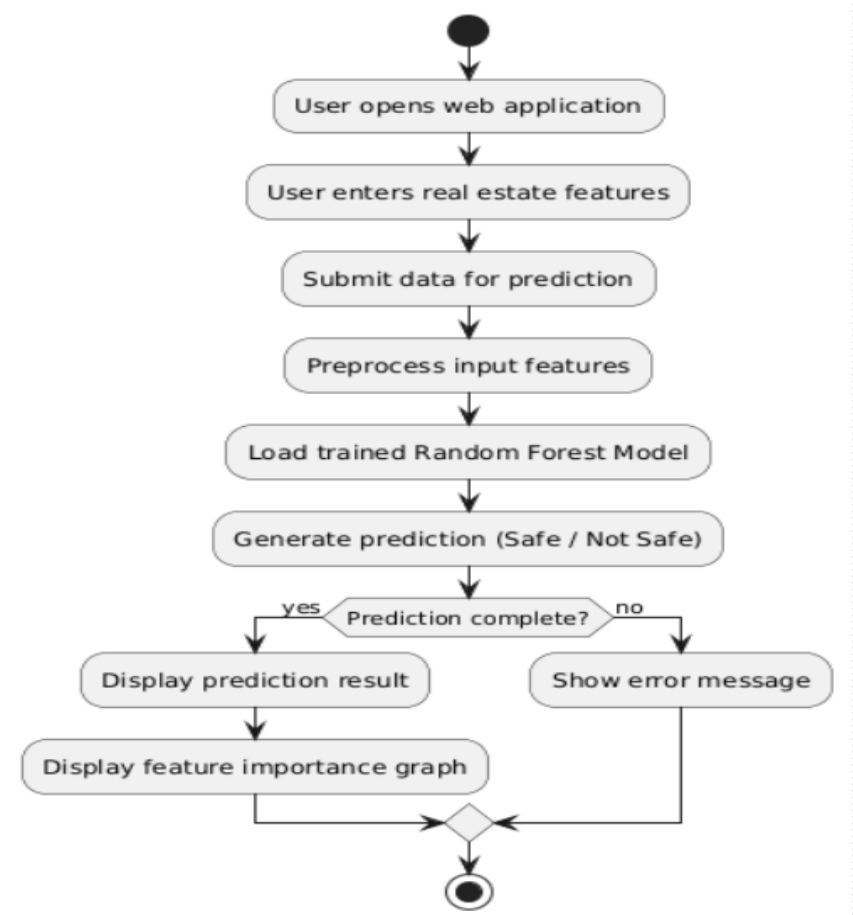
List of Modules

- **User Input Module** - Collects property-related input features from the user.
- **Prediction Engine Module** - Predicts whether the property is a safe investment using a trained model.
- **Feature Importance Visualization Module** - Displays a graph showing the most influential features in the prediction.
- **Data Collection Module** - Aggregates real estate and safety data from synthetic and external sources.
- **Model Training Module (Backend)** - Trains the Random Forest model using preprocessed and engineered features.
- **Web Application Module** - Hosts the user interface and integrates prediction and visualization features.

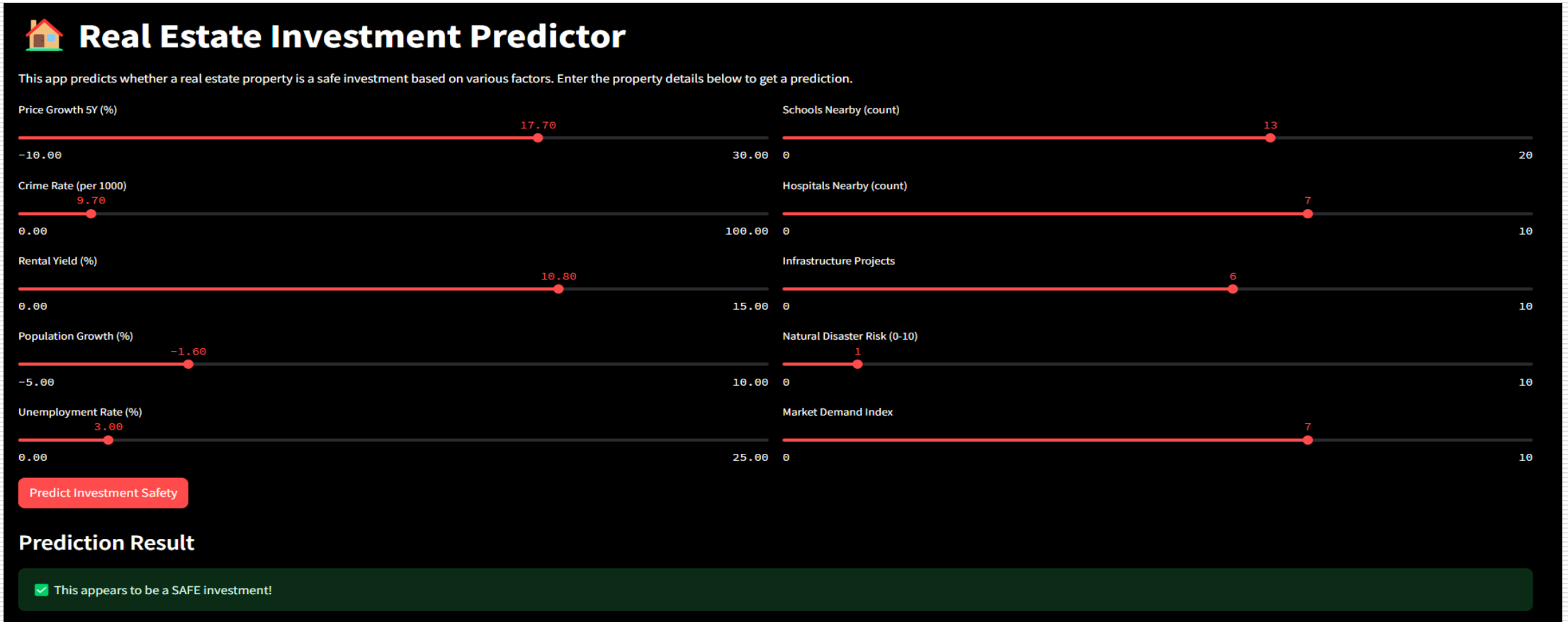
Functional Description for each modules with DFD and Activity Diagram



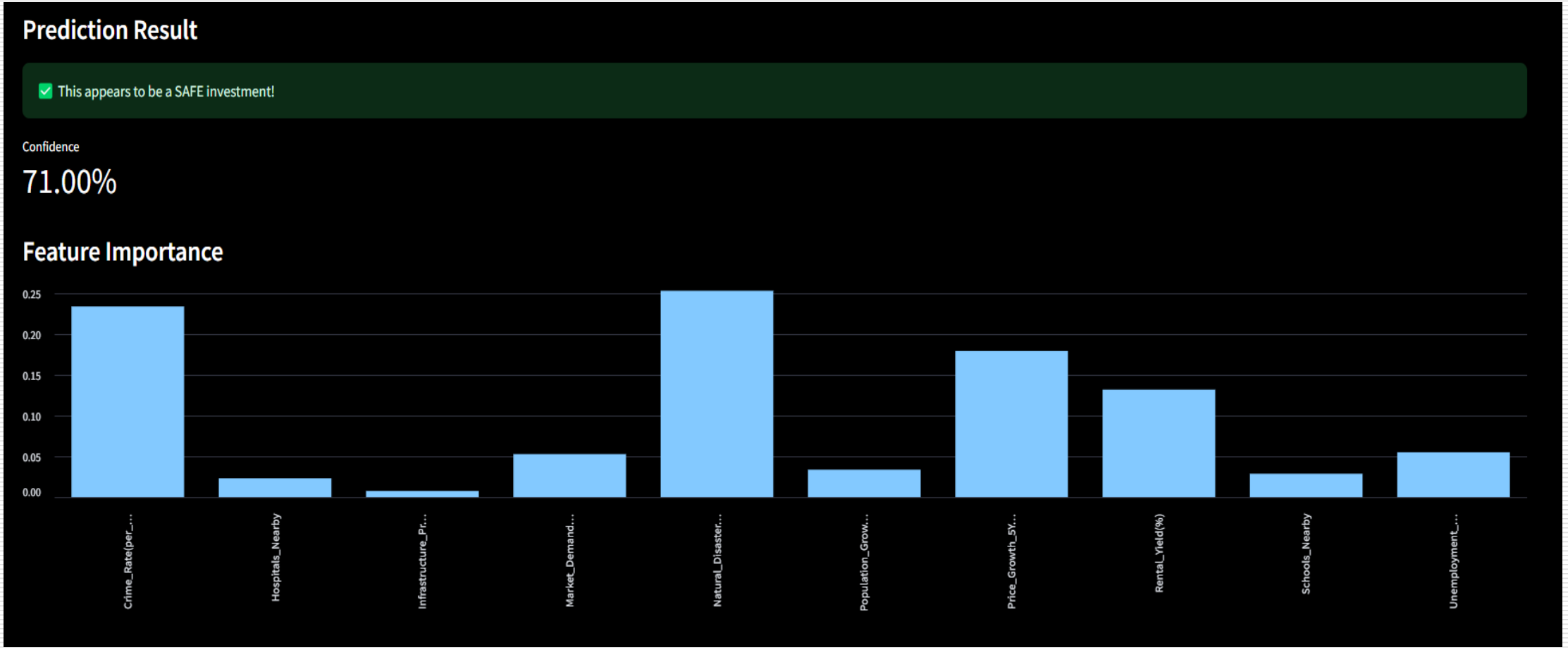
Functional Description for each modules with DFD and Activity Diagram



Implementation & Results of Module



Implementation & Results of Module



Conclusion & Future Work

The Real Estate Investment Predictor effectively analyzes multiple factors to determine if a property is a safe investment, as shown in the example result. It balances economic, demographic, and infrastructure indicators to provide reliable predictions. Moving forward, the model can be enhanced by integrating real-time data, expanding features like property-specific metrics, and improving prediction accuracy using advanced ML techniques. UI improvements, risk categorization, and explainability tools like SHAP can further increase user trust and usability. Eventually, deploying this as a full-featured web or mobile app could make it a practical tool for real estate investors.

References

- S. R. Khan and F. A. Afsar, "Real Estate Investment Decision Support System Using Machine Learning," *2021 International Conference on Artificial Intelligence (ICAI)*, Islamabad, Pakistan, 2021, pp. 1-6, doi: 10.1109/ICAI52833.2021.9445227.
- P. Sharma, A. Ghosh, and A. Tiwari, "A Data-Driven Approach to Predict Property Prices and Investment Risks in Real Estate," *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, Greater Noida, India, 2020, pp. 886-891, doi: 10.1109/GUCON48875.2020.9231253.

Thank You