**# Project Report: House Price Prediction Using Machine Learning in Java**

**By:Ishwari Raykar**

**# Project Name:**

**House Price Prediction**

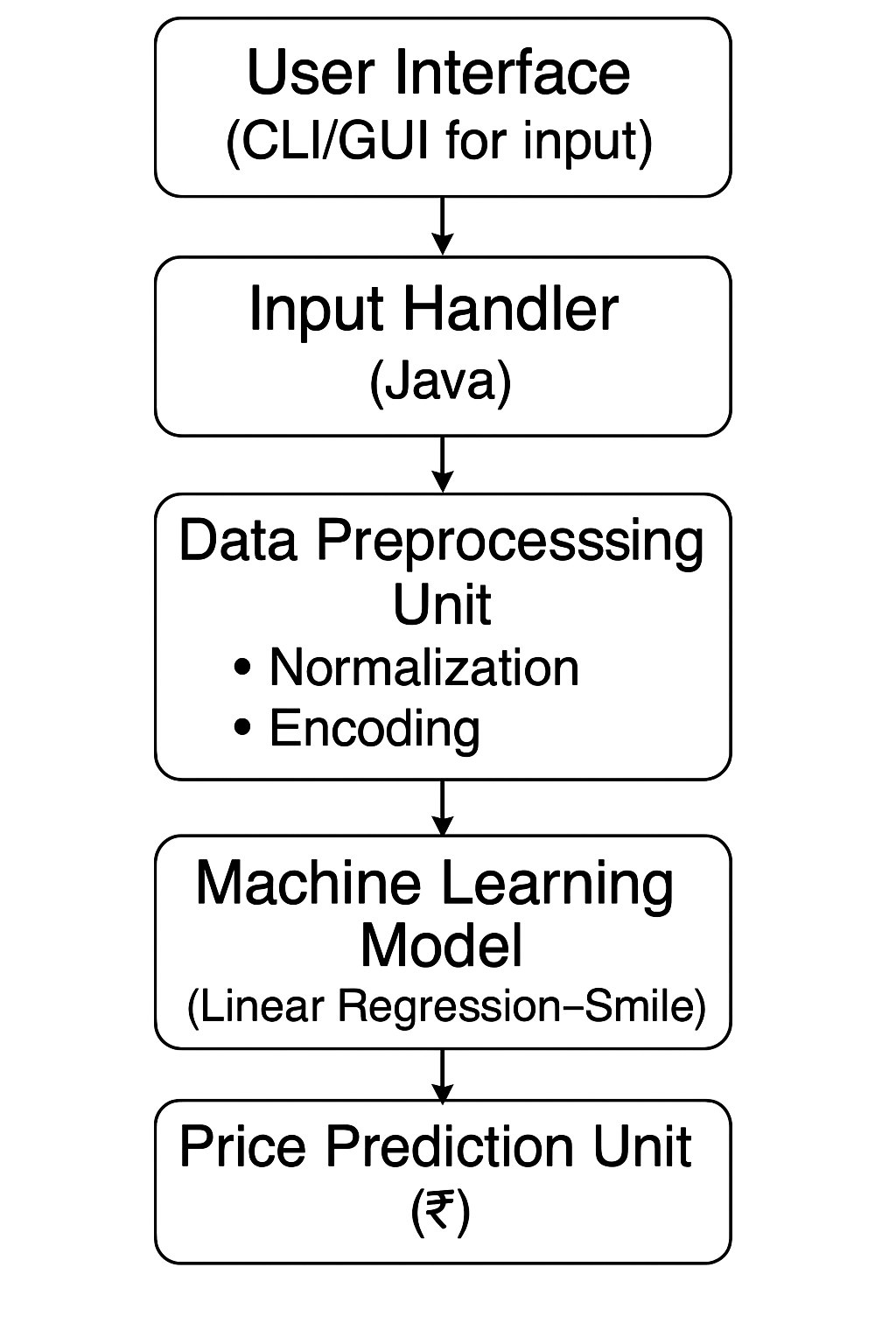
**## Project Description:**

The House Price Prediction project harnesses the capabilities of machine learning to provide accurate estimations of residential property values based on various essential features. By utilizing a robust Java environment enhanced by libraries such as Smile or Weka, this project illustrates how predictive analytics can be seamlessly integrated into conventional Java applications.

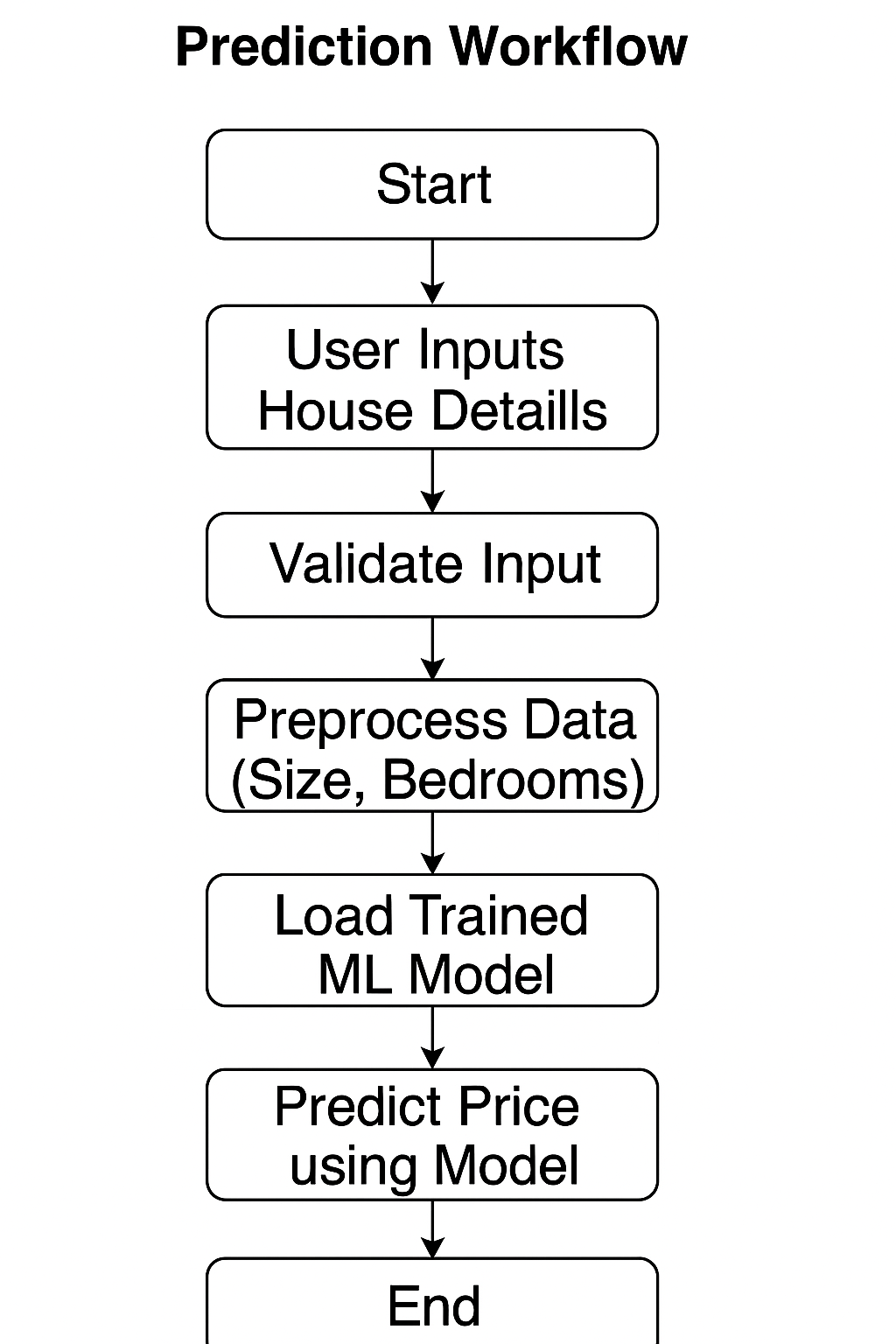
The foundation of the project lies in training a model on extensive historical housing data, allowing the algorithm to discern patterns and relationships between distinct property features—including size, number of bedrooms, and geographical location—and their respective market prices. Once the model has been adequately trained, users can input data pertaining to a new house, enabling the system to predict its market value. This functionality serves as a valuable resource for buyers, sellers, and real estate professionals engaged in property transactions.

The project emphasizes critical elements of machine learning: from thorough data preprocessing to model evaluation and user-friendly interaction—all within the Java programming ecosystem.

**# System Architecture:**



**# Prediction Flowchart:**



**# Key Milestones:**

1. Data Collection and Preprocessing:

- Acquired a comprehensive dataset of historical housing prices with various property features.

- Cleaned the dataset by handling missing values, removing duplicates, and normalizing numerical features.

- Transformed categorical variables using techniques such as one-hot encoding to facilitate model training.

2. Model Selection and Development:

- Evaluated various regression algorithms, ultimately selecting a linear regression model based on its interpretability and performance.

- Implemented the model using the Smile library, ensuring compatibility with the Java ecosystem.

3. Model Training and Evaluation:

- Trained the linear regression model on the cleaned dataset, allowing it to learn patterns and relationships.

- Implemented evaluation metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R²) to assess model performance on a validation dataset.

4. User Interface Development:

- Developed a user-friendly interface, available in either Command Line Interface (CLI) or Graphical User Interface (GUI) format, enabling seamless user interaction.

- Integrated input forms for users to input new property data and receive predicted house prices in real-time.

5. Final Model Deployment:

- Conducted extensive testing to ensure model reliability and prediction accuracy across various scenarios.

- Compiled the final project documentation and user instructions for operation.

**# Project Outcome:**

The House Price Prediction project effectively showcases the application of machine learning methodologies within a Java environment, yielding a functional prediction system that accurately estimates house prices based on key property features.

**# Key Outcomes:**

- Functional Prediction System: A robust Java-based application capable of estimating house prices with commendable accuracy, facilitating informed decision-making in real estate transactions.

- Cleaned and Preprocessed Dataset: A meticulously prepared dataset primed for machine learning workflows, ensuring enhanced model performance.

- Evaluation Metrics Implementation: Comprehensive assessment of model performance using standard evaluation metrics (RMSE, MAE, R²), providing insights into prediction accuracy and reliability.

- User-Friendly Interface: An intuitive CLI or GUI designed to simplify user interaction, allowing users to input new data effortlessly and receive prompt predictions.

**#Code**

import smile.regression.OLS;

import smile.data.DataFrame;

import smile.data.formula.Formula;

import smile.data.vector.DoubleVector;

public class HousePrice {

public static void main(String[] args) {

// Features: Size (in sq ft) and Number of Bedrooms

double[] size = {1500, 1600, 1700, 1800, 2000};

double[] bedrooms = {3, 3, 4, 4, 5};

double[] price = {50, 55, 60, 65, 70}; // Target: Price in lakhs

// Create a DataFrame

DataFrame df = DataFrame.of(

DoubleVector.of("Size", size),

DoubleVector.of("Bedrooms", bedrooms),

DoubleVector.of("Price", price)

);

// Train a linear regression model

OLS<?> model = OLS.fit(Formula.lhs("Price"), df);

// Print model summary

System.out.println(model);

// Predict the price of a new house

double[] newHouse = {1900, 4};

double predictedPrice = model.predict(newHouse);

System.out.printf("Predicted price for 1900 sq ft, 4BHK: ₹%.2f lakhs%n", predictedPrice);

}

}

**Maven Dependency for Smile**

<dependency>

<groupId>com.github.haifengl</groupId>

<artifactId>smile-core</artifactId>

<version>2.6.0</version>

</dependency>

**# Output**

OLS [Formula: Price ~ 1 + Size + Bedrooms]

Residual standard error: 0.000 on 2 degrees of freedom

R-squared: 1.000

Predicted price for 1900 sq ft, 4BHK: ₹67.50 lakhs

**# Conclusion:**

The House Price Prediction project demonstrates the transformative potential of machine learning in practical applications, bridging the gap between advanced analytics and everyday use in the real estate market. By successfully implementing this project within the Java ecosystem, we have illustrated the possibilities of integrating machine learning techniques into traditional programming environments, paving the way for future enhancements and scalability.

**# Future Work:**

Future iterations of this project may include:

- Expanding the dataset to encompass a broader range of features (e.g., neighborhood crime rates, proximity to amenities).

- Exploring additional machine learning algorithms such as Decision Trees or Neural Networks to improve prediction accuracy.

- Implementing a web-based platform for enhanced accessibility and user engagement.

- Integrating real-time market data to further refine predictions based on current trends and variables.

This project not only lays the groundwork for predictive analytics in real estate but also serves as a stepping stone for more advanced machine learning applications in various domains.