COMP303.2-ADVANCE PYTHON PROGRAMMING COURSE PROJECT

Submitted By:

- Kaan Akkök
- Melih Gülbay

Submitted To:

- Dr. Ali Cihan Keleş

Report Title:

- A Machine Learning-Based Istanbul House Price Prediction System with Data Visualization and User Interface

Project Title:

- Istanbul House Price Prediction and Visualization System

Group Members:

- Kaan Akkök
- Melih Gülbay

ABSTRACT

This project implements a house price prediction system for Istanbul, utilizing multiple machine learning models and an interactive interface. It uses five regression algorithms (Linear Regression, Random Forest, Support Vector Regression, Gradient Boosting, and XGBoost) to predict prices based on location, size, and room configuration. Advanced visualization features such as geographic heat maps and statistical distributions are included. The system processes data across Istanbul's 39 districts, providing predictions and market analysis. Performance evaluation shows Random Forest with the highest accuracy ($R^2 > 0.85$), while interactive visualizations help explore trends and prices effectively.

INTRODUCTION

The Istanbul real estate market's complexity and size present challenges for accurate price prediction and analysis. This project tackles these challenges by combining machine learning-based price prediction with interactive visualization tools. The system aids real estate professionals and buyers in understanding market trends and making informed decisions.

Primary Objectives:

- Implement multiple machine learning models for price prediction
- Create an intuitive interface for data input and visualization

- Develop interactive geographic visualizations of price distributions
- Provide detailed statistical analysis of market trends
- Enable comparative analysis of different districts and property types

DESIGN

The system architecture follows a modular design pattern with four main components:

1. **Machine Learning Module (models.py)**:
    - Implements five different regression models
    - Handles data preprocessing and feature engineering
    - Provides model evaluation and cross-validation
    - Manages model persistence and loading
2. **Visualization Module (gui_visualization.py)**:
    - Creates interactive plots and charts
    - Implements geographic visualization using Folium
    - Provides various statistical visualizations
    - Handles dynamic updates and user interactions
3. **Prediction Interface (gui_prediction.py)**:
    - Manages user input collection
    - Handles real-time price predictions
    - Provides confidence intervals and comparative statistics
    - Displays relevant market insights
4. **Metrics Display (gui_metrics.py)**:
    - Shows model performance comparisons
    - Displays evaluation metrics
    - Provides interactive performance visualizations

IMPLEMENTATION

The implementation utilizes Python with several key libraries:

1. **Machine Learning Implementation**:
    - scikit-learn for traditional ML models
    - XGBoost for gradient boosting

- Pandas for data manipulation
- NumPy for numerical computations

2. **User Interface Development**:
   - Tkinter for the main GUI framework
   - Matplotlib for statistical visualizations
   - Folium for interactive maps
   - Seaborn for advanced statistical plots
3. **Data Processing Pipeline**:
   - Feature engineering for categorical variables
   - Data standardization for applicable models
   - Cross-validation implementation
   - Performance metric calculation
4. **Visualization Features**:
   - Interactive district-wise price maps
   - Statistical distribution plots
   - Comparative analysis tools
   - Performance metric visualizations

CONCLUSION

The implemented system successfully combines machine learning capabilities with interactive visualization tools to provide a comprehensive solution for house price prediction and market analysis in Istanbul. Key achievements include:

1. **Model Performance**:
   - Successful implementation of five different regression models
   - Achieved high prediction accuracy with Random Forest model
   - Effective handling of geographic and categorical features
2. **Visualization Capabilities**:
   - Interactive geographic visualization of price distributions
   - Comprehensive statistical analysis tools
   - User-friendly interface for data exploration
3. **System Usability**:
   - Intuitive user interface for price prediction

- Detailed market insights and analysis
- Interactive exploration of real estate data