University of Greater Manchester

NEW YORK COLLEGE
THE INTERNATIONAL COLLEGE OF GREECE

OPEN-SOURCE SOFTWARE DEVELOPMENT

"VALUCASA"

SWE6005

Module Tutor: Mr. Angelopoulos Stylianos

Student Name: Kristiana Janaqi, Spyridon Mitsis, Charalampos Trapalis

Student ID: 2228038, 2228036, 2228058

CONTENTS

INTRODUCTION

The project aims to create a website for homeowners looking to sell their properties. Sellers will be able to input details about their house, and a machine learning model will provide an estimated price based on the current market value.

VALUCASA

ValuCasa is an open-source web application that scrapes housing market data and provides sellers with a suggested market price for their property. Simply input your house details, and ValuCasa analyses similar listings to estimate a competitive price.

PROJECT MANAGEMENT

The development of this website will be divided into four key areas: **Scraping**, **Prediction**, **Backend**, and **Frontend**. Each section will follow a structured approach to ensure efficiency, accuracy, and seamless integration.

PROJECT MANAGEMENT

Team Responsibilities:

- Kristiana – Responsible for Scraping
- Spyros– Responsible for Prediction Model & Backend
- Xaris – Responsible for Frontend

PROJECT MANAGEMENT

Scraping (Kristiana)

- Collect real estate price data from websites like Spitogatos, focusing on properties in Athens.
- Use web scraping tools to automate data extraction and gather relevant property details.
- Store the data in a csv file for easy access during model training and predictions.

PROJECT MANAGEMENT

Prediction Models (Spyros)

- Data Loading & EDA
  - Load CSV data, analyse missing values & distributions.
  - Visualize correlations & feature relationships.
- Preprocessing
  - Handle missing values (KNN imputation).
  - Treat outliers (IQR).
- Modelling
  - Base Models: Random Forest, XGBoost, Gradient Boosting (hyperparameter-tuned via Optuna).
  - Ensembles: Stacking & Voting Regressors.
- Evaluation
  - Metrics: RMSE, MAE, R^2, MAPE (log + original scale).
  - Residual analysis & error distribution plots.

PROJECT MANAGEMENT

Backend (Spyros)

- Take best performing trained and pickled model and deploy it as an API.
- Framework: FastAPI, known for its speed and ease of use.
- The API will allow real-time predictions by sending requests with input data.
- Enables seamless integration with web applications and other system.

PROJECT MANAGEMENT

Fronted (Xaris)

- Build a Next.js frontend to provide a dynamic and user-friendly interface.
- Design a form for sellers to input house details, which will be sent to the API for price estimation.
- Display predicted prices along with relevant market data to help sellers make informed decisions.

PROJECT TECHNOLOGIES & TOOLS

- Scraping: Python curl_cffi, os, pydantic, json
- Prediction: Python, Scikit-learn, XGBoost, Optuna, Pandas, Matplotlib
- Backend: Python, FastAPI
- Frontend: Next.js, React, Postman
- Project Management: GitHub Projects for task tracking, GitHub for version control

BUSINESS MODEL

The business model for the platform will be based on **donations**. A **"Support Us"** button will be added to the website, allowing users to make voluntary contributions. This will help sustain the platform's ongoing development, updates, and maintenance. Donations will serve as the primary revenue stream, providing the necessary funds for improvements and long-term growth.

MIT LICENSE

- Permissive free software license
- Allows:
  - commercial use
  - modification
  - distribution, and private use
- Requires:
  - preserving copyright notice and license text

- Why MIT License?
  - Developer-friendly & business-friendly
  - Encourages open-source collaboration
  - Minimal restrictions, maximum reuse

CONTRIBUTING

- Get Started
  - Fork the repo, clone locally, and create a branch.
- Code & Commit
  - Write clear commit messages.
- Submit a Pull Request
  - Describe changes clearly in the PR template.
  - Link to related issues (e.g., Closes #123.)
- Rules
  - Be respectful! Follow the Code of Conduct
  - Keep PRs focused; avoid massive changes.

CODE OF CONDUCT

- Our Pledge
  - Maintain a harassment-free and inclusive community.
  - Treat Everyone with respect and professionalism.
- Expected Behavior
  - Show empathy and kindness.
  - Respect different opinions and experiences.
  - Provide and accept constructive feedback.
  - Prioritize community well-being.
- Unacceptable Behavior
  - Harassment, discrimination, or personal attacks.
  - Trolling, doxxing, or inappropriate language.
  - Publishing private information without consent.
- Enforcement
  - Violation result in warnings, temporary bans, or permanent bans.
  - Community leaders will review and enforce rules fairly.

SUMMARY

- Project Overview: ValuCasa is an open-source web application that estimates property prices using machine learning.
- Key features: Web scraping for housing data, ML-based price prediction, and an interactive user interface.
- Business Model: Donation-based support to sustain development.
- License: MIT License for open-source collaboration.

THANK YOU