Javier Ochoa

Department of Computer Science, Western Governors University

C964: Capstone Project

Student ID: 010405717

April 10th, 2024

**Computer Science Capstone Topic Approval Form**

The purpose of this document is to help you clearly explain your capstone topic, project scope, and timeline. Identify each of these areas so that you will have a complete and realistic overview of your project. Your instructor cannot sign off on your project topic without this information*.*

*Note: You must fill out and submit this form. Space beneath each number will expand as needed.*

*Note: Any costs associated with developing the application will be the responsibility of the student.*

**INFORM INSTRUCTOR:**

Potential use of proprietary company information: (Y/**N**)

There is no potential use for proprietary company information in this project.

**ANALYSIS:**

1. Project topic and description:

The project is a machine learning based real estate price predictor. The data to train the model will be gathered from a public data website [here](https://www.kaggle.com/datasets/harlfoxem/housesalesprediction). The data set is limited to homes sold in King County, Washington, USA in 2015. The Machine learning algorithm that will be used is the linear regression function from [scikit-learn](https://scikit-learn.org/stable/). The model will train on 7 points of the data of the 21,000 recorded homes sales. Data points like price, bedrooms, bathrooms, square foot living space, square foot lot, number of floors, condition, and grade will all be used to train the machine learning algorithm. Once the model is trained users will be able to submit one or multiple fields estimate the price based on the model. The users will be able to access a locally hosted web application to enter the fields.

1. Project purpose and goals:

The project’s purpose to provide an easy to use web application that can estimate a potential price a home could be sold for in King County based on previous homes sales in 2015 given parameters such as bedroom, bathrooms, square foot, etc. The goal is to give an accurate home sale price for a home sold in King County in 2015. A multiplying factor based on the mean home increase price since 2015 (1.7 to 2.2) could be potentially added to estimate current homes prices in 2024. This web application has the potential to be supplemental tool to be used by home buyers, sellers, and appraisers.

1. Descriptive method:

Since the project utilizes historical data from 2015 regarding home sales in King County, Washington, to train a machine learning model, the descriptive method involves analyzing and understanding patterns and correlations within this data set. This analysis will guide the machine learning model to make accurate price predictions for homes based on specified parameters (bedrooms, bathrooms, square footage, square foot lot, grade, condition.).

1. Predictive or prescriptive method:

This is a predictive project. The model is predicting the price of a home based on characteristics of a home given by the user. The 7 potential fields are bedrooms, bathrooms, square foot living, square foot lot, floors, condition, and grade. The model has been trained on those 7 fields with a labeled 8th field that has the prices of the home.

**DESIGN and DEVELOPMENT:**

1. Computer science application type (select one):

* Mobile (indicate Apple or Android)
* **Web**
* Stand-alone

The application type will be a Webapp. This web interface is where users will get information on the model and be able to input 7 points of data on a single house to make a prediction on price.

1. Programming/development language(s) you will use:

The backend will by Python33 for the logic and Python3 and Sci-Kit to train the model. The front end will be connected with the web Framework Flask. The front end will be HTML with limited to no CSS.

1. Operating system(s) or platform(s) you will use:

The application will be platform for independent. The goal is to make the server run locally using the Flask built features.

1. Database Management System you will use:

There will be no database Management system besides access to .CSV files within the project files containing the data of homes sold to train the model.

1. Estimated number of hours for the following:
   * 1. Planning and design: 30 hours.
     2. Development: 70 hours.
     3. Documentation: 20 hours
     4. Total: 120 hours.
2. Projected completion date:

June 1st, 2024

**IMPLEMENTATION and EVALUATION:**

1. Describe how you will approach the execution of your project.

The approach will be starting in setting up a Flask project in IntelliJ IDEA. Creating and training a model using Ski-Kit linear regression. Then creating a front end out of html that uses Flask to connect the back end. In the end there should be two fully webpages with a properly working backend that includes two webpages with one where users can submit a form and another for the predictive result.

**This project does not involve human subjects research and is exempt from WGU IRB review!**

There are no human subjects with this project.

**STUDENT’S SIGNATURE**

Javier Ochoa

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

By signing and submitting this form, you acknowledge that any costs associated with the development and execution of the application will be your (the student's) responsibility.

**INSTRUCTOR’S SIGNATURE:**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**INSTRUCTOR APPROVAL DATE:**