A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Marty Wickham |
| *Student Number* | sbs24069 |
| *Module Title* | Machine Learning |
| *Assessment Title* | CA1 |
| *Assessment Due Date* | 28/04/2024 |
| *Date of Submission* |  |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Abstract

[NOTE: This section is designated for the abstract. Abstracts are not assigned page numbers and should precede the table of contents. If an abstract is unnecessary for your work, please delete this page.]

Attention: All notes must be removed from the document before submission!!

Contents

[Introduction 1](#_Toc158384946)

[Chapter 1 1](#_Toc158384947)

[Chapter 1.1 1](#_Toc158384948)

[Chapter 1.1.1. 1](#_Toc158384949)

[References 2](#_Toc158384950)

[NOTE: The table of contents above has been included for your convenience. To refresh the table, simply click on it, then select 'Update Table' using the mouse. You can choose to update either the page numbers exclusively or the entire table as needed.]

# Introduction

# Chapter 1

## Chapter 1.1

### Chapter 1.1.1.

A screenshot of a computer

Description automatically generated

[NOTE: For the table of contents to function properly, you must use the correct headings for all your chapters and subchapters.

**Heading 1:** This is the main heading and should be employed for the primary title or chapter. For example: CHAPTER 1.

**Heading 2:** Use Heading 2 as a subheading. For instance: Chapter 1.1.

**Heading 3:** Heading 3 provides a more detailed breakdown, such as Chapter 1.1.1.

By adhering to this hierarchical structure, you ensure an organized and effective document outline, enhancing readability and navigation. However, you are not forced to use all 3 headings, usually heading 1 and 2 are sufficient.

The remainder of your text should be written using a normal font.]

# Machine Learning CA1

## Introduction

A clear introduction, motivation, a description of the problem domain, and an explanation of how the project's goals are justified using Prediction / Classification algorithms.

Offer insights into housing prices, do they provide value for money?

The aim of this project is to address the issue of housing affordability. What can people afford to buy? What are the most important features that are required to predict housing prices? A guide on what to buy. Help people to negotiate sales price. Provide prospective sellers with valuations for their homes.

Creating a model to provide an approximation of housing prices based on buyer requirements and desired property features for prospective homeowners. The aim of this model will be to provide with the aim of informing their decision-making process as to

Using 2 machine learning models

Regression prediction

Data analysis and preprocessing

## Introduction:

For this project, I have chosen to develop and deploy two machine learning models that will hopefully be capable of predicting accurate property prices, based on universal features that inherent in all properties. Ireland is currently in the middle of a serious housing crisis with house prices becoming increasingly more expensive, and “tens of thousands of young people locked out of the housing market” (cite), with many young people feeling like the dream of owning a home has never been so far out of reach. The overarching goal of this project will be to develop a model that will be capable of providing prospective homeowners with an increased measure of control in their housing purchases, assisting in identifying properties that satisfy their requirements, while also meeting their budgetary restraints.

A use case for these models could be to help buyers identify housing which is overpriced and the most financially economic options available to them. The model could help prospective homeowners to identify regions where they are most likely to avoid

An understanding of the valuations in the housing retail market, considering the unpredictable trends driven external factors such as scarcity and inflation, could provide valuable guidance for buyers and policy makers. First-time buyers in particular, could benefit from a tool which could give them reasonably accurate estimates based on their individual needs and preferences, within a specified area, empowering negotiation exchange, saving/mortgage strategies and narrowing of the decision-making process based on means and or requirements. This tool could further guide those selling or renovating properties to choose their asking prices and renovation investments based on trends elucidated from a ML market valuation analysis. In today's dynamic real estate landscape, accurate valuation of residential properties is pivotal for various stakeholders, including homebuyers, sellers, investors, and policymakers.

Predictive modelling techniques implementing Prediction and Classification algorithms represent have emerged as indispensable tools for navigating this complex domain, especially for the average individual outside of the housing industry . This project seeks to demonstrate the utility of these algorithms to ameliorate these challenges and provide actionable insights in the domain of property valuation and investment decision-making.

Motivation:

Understanding the intricate interplay between property features and market dynamics is essential for making informed real estate decisions. Whether it's purchasing a home, optimizing investment portfolios, or crafting urban development strategies, the ability to predict house prices with precision can unlock a wealth of opportunities and mitigate risks. By leveraging Prediction and Classification algorithms, we aim to delve into this domain, unravelling hidden patterns, and empowering stakeholders with actionable intelligence.

Problem Domain:

The real estate market represents a multifaceted ecosystem influenced by a myriad of factors, ranging from property characteristics and location to economic trends and regulatory policies. Navigating this landscape requires a nuanced understanding of the dynamics at play and the ability to discern meaningful insights from vast amounts of data. Our project focuses on leveraging advanced predictive modelling techniques to analyse and interpret these complexities, ultimately facilitating more informed decision-making processes.

Justification of Goals:

The goals of this project are justified by the inherent value proposition offered by Prediction and Classification algorithms in addressing real-world challenges within the property valuation domain. By harnessing these algorithms, we aim to provide stakeholders with robust tools for predicting house prices, identifying investment opportunities, and optimizing portfolio management strategies. Through rigorous analysis and validation, we seek to demonstrate the efficacy and relevance of our approach in enhancing decision support capabilities across diverse applications within the real estate industry.

## Data Selection

To develop models capable of predicting property prices, I decided to use the “House Sales in King County, USA” dataset. This dataset contains house sale prices for King County, Seattle and includes homes sold between May 2014 and May 2015. My reasons for choosing this data are as follows:

* It is a dynamic dataset which provides a wide variety of property features to analyse.
* It is a very complete dataset that provides feature value for all observations.
* It is an exceptionally large dataset providing ample data on which to train the machine learning models.
* The dataset is well established as a reliable source evaluating for simple regression models ()

## Data Preprocessing

The dataset was imported as a Pandas data-frame object and analysed utilising the data-frame classes built in functionality. Firstly, I used the .head(20) method to give me a brief overview of what kind of data each feature contained, then I utilised the .describe() method to provide me with a numerical summary of all values contained in each feature. The .info() method allowed me to identify any categorical features and if they were any instances of NULL values. The results of this exploratory analysis revealed that the dataset is very clean containing no NULL values and only 1 categorical feature, date. Once I had identified that there were no NULL values to be cleaned or categorical features to be converted, I dropped the unnecessary features ‘id’ and ‘date’ from the data frame. After examining the data further, I noticed that there seemed to be a high number of 0 values in the column ‘yr\_renovated’. After performing a value count on this feature, I discovered that there were 20699 values of 0, so I decided to remove this column from the dataset as well. At this stage basic pre-processing of the data is complete and I could now move onto feature selection.

## Test-Train Split

## Hyperparameter Tuning

The purpose of hyperparameter tuning in machine learning is to identify the hyperparameters of a model which yield the highest accuracy.

# References