A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Suzanne Cotter* |  |
| *Sbs24020* |  |
| *Machine Learning* |  |
| *CA1* |  |
| *28th April, 2024* |  |
| *28th April, 2024* |  |

**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.

Contents

[Introduction 1](#_Toc158384946)

[Chapter 1 1](#_Toc158384947)

[Chapter 1.1 1](#_Toc158384948)

[Chapter 1.1.1. 1](#_Toc158384949)

[References 2](#_Toc158384950)

# Introduction

The Ames Housing dataset is a perfect subject for this project due to the following reasons:

* It is relevant in the real world as it is about people’s homes and describes features such as size, location and amenities.
* It has a wide range of features with seventy-nine in total, both numerical and categorical which make it perfect for testing and training prediction models on.
* It is large enough and complex enough to showcase various machine learning techniques and offers challenges like missing data, categorical variables and feature engineering opportunities.
* There is a lot of information available on this dataset making it a good start for a first assignment in a Machine learning module.

Suggested possible analysis / project questions are mentioned below (this is a small, suggested, sample of

questions, other questions may be more appropriate to your project)

o What are the most important features for predicting X as a target variable?

In the case of the Ames Housing set, the target variable is obvious that it’s the house price. The most important features for predicting what the target variable will be are worked out as follows:

* Firstly clean the data. We do this by

o Which classification approach do you prefer for the prediction of X as a target variable, and why?

o How to classify the loyal and churn customers using Support Vector Machines?

o Why is dimensionality reduction important in machine learning?

The student would need to consider the following instructions (a - d) during the development of this

project.

a) Logical justification based on the reasoning for the specific choice of machine learning approaches.

b) Multiple machine learning models (at least two) using hyperparameters and a comparison between

the chosen modelling approaches.

c) Visualise your comparison of ML modelling outcomes. You may use a statistical approach to argue that

one feature is more important than other features.

d) Cross-validation methods should be used to justify the authenticity of your ML results.

You will present the findings and defend the results in the report (MS Doc) by highlighting your work. Your

report should capture the following aspects that are relevant to your project investigations.

1. 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.

(20 marks)

2. Characterization of data, pre-processing, explanation and description of techniques used for the

variation in the accuracy across three training splits (20%, 25% and 30%) using cross validation

techniques.

(30 marks)

3. What is the primary purpose of hyperparameter tuning in machine learning? Could you elaborate on

specific hyperparameter tuning techniques (e.g., GridSearchCV) applied to machine learning models

to find optimal parameters?

(25 marks)

4. Interpret and explain the results obtained, discuss overfitting / underfitting / generalisation, provide a

rationale for the chosen models and use visualisations to support your findings. Comments in Python

code, conclusions of the project should be specified at the end of the report. Harvard Style must be

used for citations and references.

# 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.]

# Conclusion

# References