A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name:* | Kevin Scully |
| *Student Number* | 24275 |
| *Module Title* | Machine Learning |
| *Assessment Title* | CA1 |
| *Assessment Due Date* | 24th November 2024 |
| *Date of Submission* | 18th November 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 to Problem Description, Motivation, and Objectives 3](#_Toc182663127)

[Characterization and Cleaning of Dataset, Training and Testing of ML Models 3](#_Toc182663128)

[Purpose of Hyperparameter Tuning and Application of Hyperparameter Tuning Technique 3](#_Toc182663129)

[Interpretation of Results, Code Description, Comments, Conclusions, Citations, and References 4](#_Toc182663130)

[References 5](#_Toc182663131)

# Introduction to Problem Description, Motivation, and Objectives

A current major issue for the Irish government is the shortfall in supply of social housing to meet the ever-increasing demand. The lack of social housing has a direct impact on homelessness figures, increasing the number of families living on the street. (Clarke, et al., 2024) Therefore, social housing is a key pillar of current economic policy.

This report will aim to help the government departments make decisions on the construction of social housing by using machine learning models to analyse the following objectives:

1. To identify which feature is the most correlated with the number of social housing units constructed. (Correlation Matrix)
2. To predict the number of social housing units constructed based on the feature identified in objective 1. (Linear Regression / Random Forest)

The dataset used for this analysis was the ‘Social Housing Construction Status Report Q2 2022’. This provided data on the construction of social homes within Q2 2022. The information within the data includes the funding programme, the local authority, the scheme/project name, the number of units constructed, the approved housing body, and whether the construction was on-site or completed. (Department of Housing, Local Government and Heritage, 2022)

# Characterization and Cleaning of Dataset, Training and Testing of ML Models (250)

*“Characterization of data, pre-processing, explanation and description of techniques used for the variation in the accuracy across three training splits (10%, 15% and 25%) using cross validation techniques.*

*An excellent characterization and cleaning of the dataset that summarizes all details from source to fields. An excellent accuracy obtained based on the training and testing of ML models using three logical splits. Cross-validation is used to test the generalizability of the model and it should justify the results in an excellent way.”*

# Purpose of Hyperparameter Tuning and Application of Hyperparameter Tuning Technique (250)

*“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?*

*Describes the purpose of hyperparameter tuning with clarity and describes hyperparameter tuning techniques with clarity, though may have some minor gaps or less detailed explanation.”*

# Interpretation of Results, Code Description, Comments, Conclusions, Citations, and References (250)

*“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 assignment should be specified at the end of the report. Harvard Style must be used for citations and references.*

*An excellent interpretation and explanation of the results, code description, comments, conclusions, citations, and references based on problem specification and objectives. The results clearly state that the models are neither overfitted nor underfitted. An excellent defence is provided.”*

# References

**There are no sources in the current document.**