A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name: Cristhian Elson Pereira Macedo* |  |
| *Student Number: 2024104* |  |
| *Module Title: Machine Learning (10 ETCS)* |  |
| *Assessment Title: CA1 Project* |  |
| *Assessment Due Date: 21st April 2024 23:59* |  |
| *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.



**House Predict Prices**

**Cristhian Elson Pereira Macedo**

**2024104**

Higher Diploma in Science in Data Analytics for Business

Machine Learning (MLDA)

Lecturer: Dr. Muhammad Iqbal

CCT College Dublin

Dublin, Ireland

2024



**SUMMARY**

[**1.** **Introduction** 4](#_Toc163335473)

[**References** 5](#_Toc163335474)

# **Introduction**

Text.

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.
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.
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?
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.

# **References**

Central Statistics Office (2024). Data Residential Property Price Index December 2023 - Central Statistics Office. [online] www.cso.ie. Available at: https://www.cso.ie/en/releasesandpublications/ep/p-rppi/residentialpropertypriceindexdecember2023/data/ [Accessed 16 Apr. 2024].

Cereniyim (2019). Distribution plot with mean and median. [online] GitHub Gist. Available at: https://gist.github.com/cereniyim/df1c8646411cc34eb9720a79400a200f [Accessed 14 Apr. 2024].

Data to Fish (2024). 5 ways to apply an IF condition in Pandas DataFrame - Data to Fish. [online] datatofish.com. Available at: https://datatofish.com/if-condition-in-pandas-dataframe/.

Müller, A.C. and Guido, S. (2017). Introduction to machine learning with Python : a guide for data scientists. Beijing: O’reilly.

Sanchhaya Education Private Limited, G. (2021). Select rows that contain specific text using Pandas. [online] GeeksforGeeks. Available at: https://www.geeksforgeeks.org/select-rows-that-contain-specific-text-using-pandas/ [Accessed 19 Apr. 2024].

The Matplotlib Development Team (2024). matplotlib.pyplot.axvline — Matplotlib 3.8.4 documentation. [online] matplotlib.org. Available at: https://matplotlib.org/stable/api/\_as\_gen/matplotlib.pyplot.axvline.html#matplotlib.pyplot.axvline [Accessed 14 Apr. 2024].

The Pandas Development Team (2024a). pandas.DataFrame.isna — pandas 1.3.2 documentation. [online] pandas.pydata.org. Available at: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.isna.html.

The Pandas Development Team (2024b). pandas.DataFrame.value\_counts — pandas 2.1.0 documentation. [online] pandas.pydata.org. Available at: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.value\_counts.html#pandas.DataFrame.value\_counts.

The Pandas Development Team (2024c). pandas.isnull — pandas 2.2.2 documentation. [online] pandas.pydata.org. Available at: https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.isnull.html.

The Pandas Development Team (2024d). pandas.Series.str.split — pandas 1.3.0 documentation. [online] pandas.pydata.org. Available at: https://pandas.pydata.org/docs/reference/api/pandas.Series.str.split.html.