CCT College Dublin

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | Machine Learning |
| **Assessment Title:** | CA1 Project - Individual Report |
| **Lecturer Name:** | Dr. Muhammad Iqbal |
| **Student Full Name:** | Izaias de Oliveira Gomes Junior |
| **Student Number:** | 2023232 |
|  |  |
|  |  |
|  |  |
|  |  |
| **Assessment Due Date:** | 26th Nov 2023 |
| **Date of Submission:** | 26th Nov 2023 |

**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.

# Individual Report

This was a challenging project and I wish I had more time to study and improve the results. Salvador and I have been working together at the same hotel for the last one year and a half and we are good team players. We both decided that we did not want to work with a Covid-19 dataset but with crime or transport. Salvador has already worked with logistics and I am interested in climate change and solutions to mitigate it. That is why this dataset was perfect for us, uniting transport and climate change.

I tried to start working on this project in the first week from its release and I encountered my first challenge: Github. When I was looking for a dataset I found some interesting crime datasets. As I was finding them I was saving them in my Github repository. I am not sure what I did wrong but my repository was not working after I put too many files in it and deleted them at the same time without committing. I spent some days trying to fix and working on my project until I gave up and created a new one. I sent an email to Dr. Muhammad Iqbal explaining the situation.

As we learnt in the class about the Management Methodologies, the most common one and most used is the CRISP-DM. It is easier to understand because most of it we already do unconsciously. As written above, we eliminated one of the three subjects that we could choose. I was hoping to find a transport dataset because it is more interesting for me and gladly I found this one on the UCI ML Repository. I did some research about it to understand the business problem and why it would be important to predict the number of rented bikes. All companies aim to optimise their process, maximising profit and minimising costs. In our particular case it would be interesting to build a model to predict the number of bikes rented so the right amount is available at peak times and necessary maintenance can be done in quiet times.

During the Exploratory Data Analysis I concluded that the most difficult part would be the models, once the dataset had no missing values, the outliers were important for the analysis (peak hour) and all the variables would be important to explain the model. I tried to test as many models as I could and that would be better if I had more time to spend on each of them trying to improve their performance. The lack of time was not due to laziness but due to the busy daily life that we students have. Even though I could not perform everything that I wanted I am glad with the results because I learned so much with all this practice. For the next CA I need to make an effort and do more but mostly focus on what is important to not waste time trying to do everything.

I really want to improve my code technique to work faster and automatise some processes such as the tuning of parameters for the test and train split and models. The best model of our project was ANN and I could not make the code to find the best parameters work, so I did it manually which is so time consuming and prone to error. Salvador did most of the visualisation part and it was really good, that is another thing that I need to improve.