**Individual reflection report – Paul O’Donovan (22208104)**

# **1. Personal contribution (500 words)**

I collaborated on this project with my teammate throughout the entire process. We worked together on all facets of the project, from model training and evaluation through data investigation and cleaning.

In terms of designing a solution, my teammate and I collaborated to comprehend the issue of target feature prediction utilizing the supplied dataset. We spoke about the variables and the data and developed a strategy on how to handle the problem.

During the feature selection phase, we worked together to determine which variables could affect the desired feature and which variables could be disregarded.

We both contributed to the implementation of the Python and scikit-learn library code used to load, purge, and alter the data during the software development process. Additionally, we collaborated to build the scripts that generated the visualizations and evaluation metrics as well as to train and assess the various models.

We both studied carefully about the COVID-19 epidemic and the information gathered by the CDC throughout our investigation into the issue and the subject area. Additionally, we collaborated to find potential bias or confounding causes in the dataset and develop solutions for them.

We both worked on cleaning and converting the dataset throughout the data preparation stage to make sure it was appropriate for use in our models. We worked together to put these tactics into practice after discussing various methods for dealing with missing data and irregularities in the categorical features.

We both contributed to creating, analysing, and debating the categorization evaluation metrics for each model during the evaluation process. Additionally, we collaborated to develop a clear and thorough report summarizing our findings and recommendations.

We worked closely together on the project by exchanging screens and debating issues. To overcome obstacles and accomplish our goals, we were able to draw on one other's talents and knowledge.

Implementing the code for feature transformation and data cleaning, creating some of the visualizations, and participating in the discussion and analysis of the outcomes were all personal efforts I made to the project.

I want to highlight the important role that my teammate played in the success of our project. Their extensive knowledge of statistics and exceptional skills in model interpretation greatly contributed to the creation and assessment of our predictive models. It was our combined efforts that led to the successful outcome of the project.

I believe that my personal contributions helped to ensure the overall success of the project by contributing to the development of accurate and effective predictive models, and by helping to communicate our results and insights in a clear and concise manner. Additionally, we were able to ensure that our work was well-coordinated and that we were able to take advantage of each other's strengths and skills.

Overall, I believe that our collaborative efforts and the unique strengths and skills brought by each member of the team were key to the success of the project.

# **2. What did you learn from the project? (500 words)**

Using a public dataset and the data science methodology, I developed and assessed prediction models for a real-world issue as part of this project. I obtained expertise in particular areas such as feature selection, model training, model evaluation, and data cleaning and exploration.

I gained knowledge of how to deal with missing values, categorical features, and the conversion of continuous features to numeric representations during the data cleaning step. Additionally, I discovered how to divide the dataset into training and test sets using the train\_test\_split function in scikit-learn.

I gained knowledge on how to use correlation matrices, boxplots, and stacked bar charts to represent relationships between characteristics throughout the data exploration phase. Based on these visuals, I also learnt how to recognize promising traits for prediction.

I gained information about how to select a subset of promising features for prediction during the feature selection process and how to support my decisions with visuals and domain expertise.

I had the opportunity to acquire valuable knowledge and skills in training and interpreting linear regression, logistic regression, and random forest models using scikit-learn. I spent time researching and studying the key concepts and techniques necessary to effectively develop these models.

In addition to gaining experience with the model training process, I also developed my ability to assess the effectiveness of our models. This involved learning how to utilize classification assessment metrics, such as confusion matrices, accuracy, precision, recall, and F1 scores. I made sure to thoroughly evaluate the performance of each model we developed, considering these metrics, and adjusting our approach as necessary.

Overall, the project was successful in meeting its goals and yielding reliable and significant findings. There were, however, certain areas where it could have been done better.

The datasets utilized in this study weren't perfect and had several restrictions that made working with them difficult. For instance, some of the characteristics had missing values, necessitating imputation or deletion of partial entries. Additionally, several category features were represented inconsistently, necessitating standardization.

Despite these obstacles, we were able to advance in our investigation of feature correlations and the selection of potential features for prediction.

The method for choosing features needed improvement as well. Based on visuals, we were able to pinpoint several interesting features for prediction.

In conclusion, this project gave participants invaluable practical experience using data science methods to solve actual problems. I gained knowledge in using scikit-learn to manage data cleaning, data exploration, feature selection, model training, and model evaluation. I also gained knowledge about how to evaluate model performance using classification assessment metrics and model coefficient interpretation. In the future, I would like to sharpen my data preparation and feature selection abilities and investigate more sophisticated modelling approaches to enhance model performance.

**3. Anything else related to this project (500 words)**

I considered this assignment to be a worthwhile educational opportunity that gave me the chance to use the theories and methods I had learned in class to solve a genuine problem. I was able to better comprehend the practical components of data science, such as data cleaning, feature selection, model training, and evaluation, because the project's hands-on nature helped to cement these ideas in my mind.

Additionally, the assignment allowed me the chance to collaborate with a team member, which was a worthwhile experience in and of itself. Working together allowed us to better utilize our own skills and strengths, resulting in higher-quality work than if we had worked separately.

The project's biggest problem was handling the data cleaning procedure. Some missing data and erroneous representations that needed to be handled required our attention. In order to determine the most effective method for cleaning and processing the data, we had to experiment with various strategies.

Choosing the best technique for each case was one of the main hurdles in the data cleaning process. We had to experiment to determine the optimal method because there were occasions when one method performed better than another. This necessitated a great deal of trial and error and rigorous evaluation of the trade-offs between various strategies.

Our evaluation of the prediction models we created presented another difficulty. While it was very simple to train the models, it was more challenging to analyze the findings and choose the best model. We had to make decisions on which model was most suited for the task at hand after carefully weighing the trade-offs between several evaluation indicators.

For instance, we evaluated our models’ using measures like accuracy, precision, recall, and F1 score. Nevertheless, there were instances when these measurements clashed, and we had to decide which indicator was most crucial for our specific issue. This needed us to have a thorough understanding of the problem we were attempting to answer as well as the domain.

Despite these challenges, I was able to gain a better understanding of the data cleaning process and model evaluation. This will be invaluable as I continue to work on real-world data science problems.

Overall, I considered this project to be a worthwhile learning opportunity that assisted me in honing crucial data science skills like problem-solving, cooperation, and communication. I am more assured in my ability to use the theories and methods I have studied in class to solve problems in the real world, and I am eager to learn more about this subject.