Reflection

The Introduction to Artificial Intelligence module helped me build both the theoretical and practical knowledge on AI and how to create AI models from a given dataset. While the lectures dived more into the theoretical knowledge, the coursework helped me to implement those into practical. We were a team of 2 where our idea was to build classifier models that can predict shopper's intention on e-commerce website.

Coding:

Once the dataset was taken from the Kaggle website, I have performed different data analysing technique to learn about the dataset. I have checked whether the dataset contained null values and what should be the best case to do, whether I should remove the rows containing null values, fill them or remove them. I have created several graphs such as histogram, count plot and pie charts to visualize and analyse the dataset and understand the target column ("Revenue"). As stated in the report our dataset was imbalance, so first I created a model (Random Forest Classifier) testing how it performs on imbalanced dataset and evaluated its results. Even though the model accuracy was quite high, but it lacked in precision, recall (mostly) and f1 score which indicated how poor the model was in unbalanced dataset. So, I have researched different data balancing methods such as under-sampling and oversampling to balance the dataset. I have tried under-sampling technique and saw a significant change in the result (increased precision, recall and F1 scores) of the model, and further discussing with my teammate we have decided to work with under-sampled dataset.

Report:

I have started the report describing the dataset and the problem domain, I have given a brief description of each column of the dataset. Then I wrote about how the dataset was analysed, what graphs were used, what was done with missing value and why. I also explained about how the imbalanced dataset was dealt and why we had to balance. On the result section I have talked about the differences between the confusion matrix of the unbalanced and balanced models. On the evaluation section I have explained how we lost some accuracy scores on the balanced dataset but gained good percentage of precision, recall and f1 score. And why the models of balanced dataset are more accurate despite scoring lower accuracy score than the unbalanced dataset. And finally, the conclusion part was about the difficulties faced.

Conclusion:

Overall, I have learnt all the key aspects of how to create AI models step by step. I have learnt how data in analysed, how to prepare the data, how to split the data for training and testing, I have learnt about different types of models that can be created and their pros and cons. Also, how to evaluate the models created and further tune them to get the best outcomes.