Individual Reflection

The group projects of the Machine Learning module consisted of two assignments:

- *Assignment 1:* An analytical data report for Airbnb based on 2019 Aribnb New York dataset
- *Assignment 2:* building a neural network for object detection using the CIFAR-10-Object Recognition image dataset.

During the module, I gained comprehensive understanding of different machine learning algorithms and the underlying statistical concepts and data analysis. My previous knowledge of statistics, data analysis and data visualization made it easier for me to follow especially the first chapters of the module and the group assignment 1.

The additional reading and lecture casts were an excellent way to exploring machine learning concepts like correlation and regression, clustering and neural networks.

While I felt comfortable with exploratory data analysis during Assignment 1, focusing on an analytical data report for Airbnb, I encountered a learning curve during Assignment 2, which involved building a neural network for object detection using the CIFAR-10 dataset. I felt that being part of a group and undertaking a comprehensive assignment, in our case building and training a neural network from scratch was the best possible way of learning a new complex topic, as it allowed me to practically apply the theoretical knowledge from all units directly.

Our team was working from different geographic areas and time zones (and fully remote) which of course posed some challenges in terms of coordination and timely completion of the assignments, but was managed well overall by all group members.

For assignment 1, we decided to work asynchronously on the initial data exploration. After our first attempt we discovered that the variables we were exploring (price and availability) showed almost no correlation, so we shifted our approach. We then looked more closely at the correlation of price and availability.

My contributions of clustering analysis in Jupyter notebook of neighborhood and price was then the foundation for the further analytical report.

In short, my most valuable learning points from group assignment 1 were the following:

- Data preprocessing: removing null values and irrelevant columns streamlined the dataset, enhancing the efficiency and accuracy of subsequent analyses. This experience emphasized the foundational role of data preprocessing in ensuring the reliability of analytical insights and model predictions.
- Iterative model improvement: continuous improvement through iterating on existing solutions and approaches. Despite achieving respectable R2 scores, there remained opportunities for enhancement through feature selection and incorporation of additional variables
- Feature Importance: Analyzing SHAP summary charts revealed that variables such as room type and location exerted significant influence on listing prices. This understanding highlighted the intricate interplay between various features and their collective impact on predictive outcomes.

In contrast, assignment 2 presented some challenges for me, as neural networks were a new concept for me. Especially since some team members have had extensive prior professional experience in the subject due to their work. I therefore felt like I had to individually catch up by extensive self-study of neural networks. I enrolled in an online bootcamp on AI with a focus on CNN and computer vision on Knowledgehub to add to the knowledge and resources provided by my studies. I also found that some explainer videos on Youtube helped me to grasp concepts better by adding visual concepts.

Despite initial difficulties, I felt like I was at all times able to contribute to the problem-solving sessions with team members. What helped me most besides my dedicated self-study on the subject was reading, discussing and engaging with the contributions from my more experienced team members.

My main learnings from assignment 2 were: The iterative nature of model development. Despite starting with a basic Artificial Neural Network (ANN) architecture, the project progressed through various iterations, each aimed at improving model performance and addressing specific challenges such as overfitting. Additionally, the project highlighted the importance of leveraging established architectures like VGG and adopting best practices in deep learning, such as utilizing ReLU activation functions and the potential for significant performance gains in image classification by refinements.

Reflecting on the group projects in hindsight, I recognize that my learning was not limited to technical knowledge but also brought increased self-awareness about

working in a team with mixed experience. The single most important take-away for me is the certainty that by actively engaging with team members and further learning in my free time, I was able to not only contribute in a meaningful and fruitful way to both team assignments, but also managed to actively shape the outcome and direction with providing valuable points and perspectives.

After the end of the module, I felt confident in explaining and using the machine learning concepts introduced in the module. Applying them hands-on in a group project was key for me in gaining this holistic understanding. By continued reading of studies, academic papers and articles I continue to build upon this knowledge and try to gain more in-depth insights into machine learning applications in different areas.

In my current job, I recently took on the opportunity to work on a team project for fine-tuning a large language model to perform question-answer and text summaries in Vietnamese for the vocational training sector. The experiences gained during the group work of this module really helped in my understanding of project work-flows, breaking down the project into sub-tasks and of course the technical machine learning knowledge.