E-Portfolio link: <https://simbi007.github.io/e-Portfolio/>

Machine Learning E-Portfolio link: <https://simbi007.github.io/e-Portfolio/Machine%20Learning%20Page.html>

**Reflective Piece: Machine Learning Module Resubmission**

**Introduction**

The Machine Learning module provided me with an in-depth exploration of the theories, tools, and practices which are essential to machine learning. It covered aspects such as model selection, implementing models using Python, and evaluating their performance. Additionally, we engaged in collaborative discussions that focused on the ethical, social, and legal challenges faced by machine learning professionals today. During the module, I participated in a team project whereby we analysed an Airbnb dataset and I completed an individual project as well. As Dutta (2023) highlights, self-reflection is crucial for students as it allows them to critically evaluate their learning experiences. In this reflective piece, I will discuss the module's content, my personal experiences, the skills I have developed, and how they will support my future academic and professional growth.

**Module Details**

The module was divided into multiple units, each addressing an aspect of machine learning. In the first unit, we discussed the impact of the Fourth Industrial Revolution (Industry 4.0) on various sectors. I also learned that over the past 20 years, the emergence of Big Data has significantly influenced machine learning adoption across fields like astronomy, economics, and biology (Qiu et al., 2016).

The second unit introduced Exploratory Data Analysis (EDA), a critical step in any data science project. EDA enables data scientists to identify anomalies and select appropriate features before developing predictive models. Following this, units three and four focused on correlation and regression techniques using Python, and how they can be used to identify features which strongly influence the target variable. We applied Pearson correlation, linear regression, multiple linear regression, and polynomial regression models to real-world datasets, observing how small changes in data could impact relationships between variables.

Units five and six introduced cluster analysis, an unsupervised machine learning technique. Clustering algorithms group observations based on similarities, revealing hidden patterns that are crucial for data exploration (IBM, 2025). As an e-portfolio task, I calculated the Jaccard coefficients for three individuals based on their pathological test results, this coefficient measures the similarity or diversity between samples sets (Lee, 2025).

In unit six, my team and I completed a collaborative project on analysing an Airbnb dataset to answer a business analytics question. Units seven through ten focused on Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs). I learned about how ANNs are inspired by the human brain, how backpropagaion helps adjust weights during training, and how CNNs are particularly effective for image classification tasks. We also discussed the ethical and legal implications of AI applications, such as AI-generated creative writing.

Finally, in units eleven and twelve, I prepared an individual presentation on an image classification task using CNN and the MNIST dataset, and I read about the potential challenges and opportunities in machine learning.

**My Experience**

Overall, I found the module to be instrumental in my development as a data professional. It provided a foundation in concepts such as ethics, EDA, and the structure and training of neural networks. One of the highlights for me was learning how to select the right machine learning model based on the problem at hand, and how to evaluate model performance effectively.

The team project was a sort of centrepiece of the module experience. Working with my team members to analyse the Airbnb dataset strengthened my collaborative and communication skills. Although there were challenges — for example, some members were unable to attend meetings regularly — we overcame these obstacles by maintaining clear meeting protocols and dividing tasks fairly.

On a personal note, the biggest challenge I faced was relocating to another country during the module. This caused delays in my coursework and eventually led to me having to resubmit of both my team project and individual presentation. Although it was quite stressful, this experience has taught me time management, and adaptability, which are vital skills not just in life but in a professional environment.

**Skills Gained and Future Applications**

Throughout the module, I developed a wide range of skills. I gained experience in Python programming, EDA, regression techniques, clustering, and the implementation of neural networks. Additionally, I now feel confident using tools like Jupyter Notebook for hands-on data analysis and model building. In addition, the module helped me improve my critical thinking and problem-solving abilities.

Working in a virtual team setting enhanced my collaboration, communication, and project management abilities. Understanding how to organize and allocate tasks, and document meetings, will be useful in future professional roles, particularly in remote work environments.

**Conclusion**

The Machine Learning module has been an integral part of my academic journey. It provided me with a basis in machine learning concepts, practical experience with Python programming, and an understanding of the ethical, social, and legal implications of AI technologies. Through collaborative projects, individual assignments, and personal challenges, I developed both technical and soft skills.

I also think that this reflective exercise aligned with Dutta’s (2023) views on the importance of self-reflection, has allowed me to appreciate how much I have grown over the course of the module. I am confident that the knowledge and skills that I gained will not only help me in my academic progress but will also prepare me professionally in the rapidly evolving fields of data science and machine learning.

**References**

Dutta, S, et.al, (2023). Reflection and peer assessment to promote self-directed learning in higher education, *Journal of Educational Research and Reviews.* 11*(3)*, DOI: [10.33495/jerr\_v11i3.23.111](http://dx.doi.org/10.33495/jerr_v11i3.23.111)

IBM. (2025). What is clustering? Available at: <https://www.ibm.com/think/topics/clustering> [Accessed: 25 April 2025]

Lee, S. (2025).Understandingthe Jaccard Index for Enhanced Data Analysis Available at: <https://www.numberanalytics.com/blog/understanding-jaccard-index-data-analysis#:~:text=Overview%20of%20the%20Jaccard%20Index,union%20of%20the%20sample%20sets>. [Accessed: 25 April 2025]

Qiu, J, et.al, (2016). A survey of machine learning for big data processing , *EURASIP Journal on Advances in Signal Processing.* 67, DOI: <https://doi.org/10.1186/s13634-016-0355-x>