Reflections on My Machine Learning Journey

This reflective piece explores my experiences and learning outcomes during a machine learning module. It will include my thoughts on the technical challenges I faced, the collaborative efforts within my team, and how these experiences contributed to my personal and professional growth.

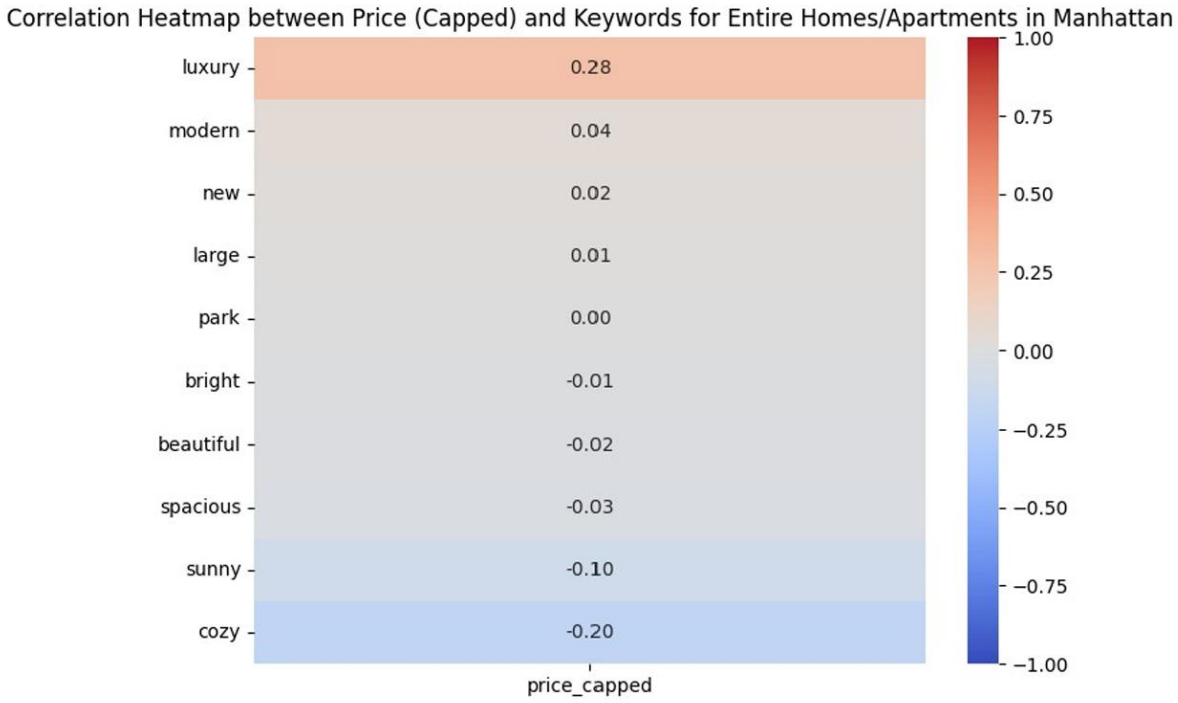
In the team project, our group of five was assigned the task of conducting a business analysis of Airbnb using a data science approach. The primary objective was to formulate and answer an insightful analytical question using the provided dataset (Dgomonov, 2019). After individually analysing the data and proposing potential questions, we collaboratively refined our ideas and agreed on the final question: “What are the primary factors that affect the listing price of properties on Airbnb, and how can these insights help hosts achieve the most value from their listings?”

My primary responsibilities included cleaning the dataset and analysing how listing names—such as word count and specific terms—impact pricing. Initially, I encountered challenges due to inadequate data preparation, which hindered my analysis. This experience reinforced the importance of thorough data cleaning as a foundation for effective analysis (Ridzuan and Zainon, 2019). Through this task, I gained a clearer understanding of how to identify and interpret correlations within data and their significance in uncovering meaningful patterns. This process also enhanced my ability to critically evaluate relationships and validate them through exploration.

At the start, I felt overwhelmed as this was my first group project at university, and I was unfamiliar with my team members. I worried about unequal participation and the

potential for missed deadlines. However, as the project progressed, I observed my teammates' reliability and commitment, which eased my concerns. Early on, I felt hesitant to share ideas but gradually gained confidence as I contributed suggestions and offered feedback.

One significant challenge I faced was submitting one of my tasks 12 hours late. While this delay did not impact the project overall, it made me feel uncomfortable, as I feared letting the team down. This reinforced the importance of managing my workload effectively and maintaining open communication with teammates (Gordon, 2023). Feedback from my colleagues was invaluable, particularly their suggestions to make my visualisations more accessible to those unfamiliar with the dataset. This helped me understand the importance of clarity in presenting findings.



The project also introduced me to Google Collaboratory, a tool essential for sharing work with the team (Google, n.d). This enhanced my technical skills and prepared me

for collaborative environments in future professional settings. Overall, this experience deepened my understanding of teamwork, improved my communication and time management skills, and enriched my academic and personal development.

Another challenge I faced during this module was building my e-portfolio. Initially, I struggled to find a suitable template that was simple, clear, and easy to navigate. After identifying a template I liked, I attempted to replicate it by inspecting the HTML code of the original site and creating an index file in my GitHub. However, the page I built looked nothing like the desired template, leaving me frustrated. After extensive research, I discovered that HTML alone was insufficient; I also needed the corresponding CSS file to control the visual presentation. Fortunately, I found the required CSS on GitHub and forked it to recreate the template successfully (GitHub, n.d).

This experience reminded me of a course I had previously completed on SheCodes (n.d), where I learned HTML and CSS. Although I had not used these skills since completing the course, this project allowed me to refresh my knowledge. I even experimented by creating a "back" button for my e-portfolio, which I thoroughly enjoyed. It demonstrated that no knowledge is ever truly lost and highlighted how non-machine-learning-related skills can enhance the presentation of my work.

```
<style>
  .back-button {
    display: inline-block;
    background-color: white;
    color: #006699;
    text-decoration: none;
    padding: 5px 10px; /* Reduced padding for a smaller button */
    font-size: 12px; /* Smaller font size */
    border: 1px solid #006699; /* Thinner border */
    border-radius: 5px;
    cursor: pointer;
    transition: background-color 0.3s, color 0.3s;
    margin: 15px 0; /* Adds space above and below the button */
  }
  .back-button:hover {
    background-color: #006699;
    color: white;
 }
</style>

<div class="button-container">
  <a href="https://dzervenes.github.io/" class="back-button">Back</a>
</div>
```

To build my e-portfolio, I had to complete several tasks, including correlation and regression analysis, k-means clustering, perceptrons, gradient cost functions, and convolutional neural networks (CNNs). Of these, CNNs were my favourite and the area where I learned the most. Initially, I found CNNs daunting, but I now feel confident building them and enjoy the process. Overfitting posed a significant challenge, as regularisation techniques I implemented, like dropout layers and data augmentation, often fell short (Hernández-García and König, 2018). Hyperparameter tuning was also demanding, as each change required careful consideration due to the long training times (Li et al., 2018). These challenges taught me patience and the importance of strategic experimentation.

```python
# First convolutional layer
model.add(Conv2D(32, (3, 3), activation='relu', padding='same', input_shape=(32, 32, 3)))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.2))

# Second convolutional layer
model.add(Conv2D(64, (3, 3), activation='relu', padding='same', kernel_constraint=MaxNorm(3)))
model.add(MaxPooling2D(pool_size=(2, 2)))
```

While CNNs became a strength of mine during this module, other areas like perceptrons remain opportunities for further learning. I understand their function but have not yet built anything practical using them, leaving me uncertain about applying them to real-world problems. I aim to deepen my understanding of perceptrons by applying them in future projects, focusing on practical use cases like classification tasks (Brownlee, 2019).

Collaborative discussions with my peers were highly enriching. Each member brought unique perspectives shaped by their interests and backgrounds, making the conversations engaging and insightful. I particularly valued their viewpoints on real-world applications, challenges, and ethical considerations, which often highlighted aspects I hadn't considered. Sharing my ideas and receiving feedback was equally rewarding, giving me a more rounded understanding of the topics and improving my ability to communicate complex concepts effectively. These exchanges broadened my knowledge and fostered meaningful intellectual dialogue.

This module was intense but transformative, greatly enhancing my confidence and professional readiness. It has already opened exciting opportunities, as I havee been invited to collaborate on a new project at work. The project involves developing a chatbot for train conductors to assist with real-time problem-solving. While I will not be

a part of the technical development team, I will work closely with them to ensure the chatbot aligns with user needs and provide input on data structuring and usability testing.

This new role is an exciting opportunity to observe the lifecycle of a machine learning application from a practical perspective. It will allow me to bridge the gap between technical teams and end-users, applying some of the collaborative skills and insights I gained during this module. Furthermore, this project represents a significant step forward in my career progression, as it positions me to take on more interdisciplinary roles in the future.

References:

Brownlee, J., (2019). *How to Implement the Perceptron Algorithm From Scratch in Python*. Available at: https://machinelearningmastery.com/implement-perceptron-algorithm-scratch-python/ (Accessed 27 January 2025).

Dgomonov (2019) *Airbnb listings and metrics in NYC, NY, USA*. Available at https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data (Accessed 18 Nov. 2024).

GitHub, (n.d.). *Quick Portfolio by evanca*. Available at: https://github.com/evanca/quick-portfolio (Accessed 27 January 2025).

Google, (n.d.). *Google Collaboratory*. Available at: https://colab.research.google.com/ (Accessed 27 January 2025).

Gordon, R., (2023). *Why It Is Necessary to Improve Team Communication*. American Public University, 12 July. Available at: https://www.apu.apus.edu/area-of-study/business-and-management/resources/why-it-is-necessary-to-improve-team-communication/ (Accessed 27 January 2025)

Hernández-García, A. and König, P., (2018). *Data augmentation instead of explicit regularization*. *arXiv preprint arXiv:1806.03852*. Available at: https://arxiv.org/abs/1806.03852 (Accessed 27 January 2025)

Li, L., Jamieson, K., Rostamizadeh, A., Gonina, E., Hardt, M., Recht, B., and Talwalkar, A., (2018). *A System for Massively Parallel Hyperparameter Tuning*. *arXiv preprint arXiv:1810.05934*. Available at: https://arxiv.org/abs/1810.05934 (Accessed 27 January 2025)

Ridzuan, F. and Zainon, W.M.N., (2019). A Review on Data Cleansing Methods for Big Data. *Procedia Computer Science*, 161, pp.731-738. Available at: https://www.researchgate.net/publication/338348131_A_Review_on_Data_Cleansing_Methods_for_Big_Data (Accessed 27 January 2025).

SheCodes, (n.d). *SheCodes: Learn Coding Skills Online*. Available at: https://www.shecodes.io/ (Accessed 27 January 2025)