**Reflective Report on Portfolio 4**

## 1. Introduction

In this portfolio, I aimed to analyze housing prices using a dataset containing various features such as area, bedrooms, bathrooms, and more. The primary objective was to identify significant factors that impact housing prices and to build predictive models to accurately estimate them. Throughout the process, I explored different machine learning models, including Linear Regression, Polynomial Regression, KNN Regressor, and Artificial Neural Networks (ANN). This report reflects on my experience and the key learnings from working on this portfolio.

## 2. Progress and Learning Journey

From the start of this unit, my skills and understanding of data analysis and machine learning have significantly improved. Initially, I was only familiar with basic Python programming and simple data manipulation techniques using Pandas. However, through this portfolio and other exercises, I learned to build and evaluate complex models, implement clustering techniques like K-Means, and utilize Neural Networks with TensorFlow/Keras for predictive modeling.

The use of Jupyter Notebooks in the development of this project was instrumental in organizing my code, visualizations, and documentation in a coherent manner. I also became familiar with using Google Colab for cloud-based execution, which was helpful in accessing the dataset stored on my Google Drive.

In the future, I am interested in using these skills to build more complex applications, such as recommendation systems or real-time data analytics platforms, and potentially exploring cloud deployment of these models for web-based applications.

## 3. Discussion Points on Portfolio 4

### Discussion Point 1: Why I Chose the Dataset

I selected the housing dataset because it is a well-known example in the domain of regression analysis, making it ideal for experimenting with different models and algorithms. The dataset contained a mixture of categorical and numerical data, providing a good opportunity to practice data preprocessing techniques such as handling missing values, one-hot encoding for categorical features, and scaling for certain models.

### Discussion Point 2: Why I Chose the Machine Learning Models

I chose a variety of machine learning models, including Linear Regression, Polynomial Regression, KNN Regressor, and an ANN for this portfolio:

Linear Regression: I started with Linear Regression as a baseline model due to its simplicity and interpretability. It allowed me to quickly evaluate the relationships between features and the target variable (house prices).
Polynomial Regression: I extended this to Polynomial Regression to capture non-linear relationships, improving the model's performance on more complex patterns in the data.
KNN Regressor: I tested KNN because of its simplicity and effectiveness in certain datasets where local patterns may be significant. I performed parameter tuning (testing different values for 'k') to find the optimal setting for the model.
Artificial Neural Networks (ANN): Finally, I used an ANN to explore how deep learning models perform in regression tasks. The flexibility and adaptability of ANNs allowed me to achieve better performance in terms of reducing mean squared error (MSE).
These models were suitable for my analysis because they allowed me to compare different approaches (linear, non-linear, local, and deep learning) and determine which performed best under various conditions.

## 4. Insights and Conclusion

From my analysis, I found that area, number of bedrooms, and number of bathrooms are significant predictors of house prices, which aligns with intuitive expectations. The Linear Regression model performed reasonably well but was outperformed by Polynomial Regression and the ANN model, which captured more complex relationships and interactions among features.

The clustering analysis using K-Means also provided valuable insights into the segmentation of houses based on their features. I identified three clusters representing different types of houses (e.g., small, medium, and large), which further validated the variations in pricing patterns.

Overall, the results were consistent with my expectations, and the machine learning models effectively highlighted key factors influencing housing prices.

## 5. Additional Point: Challenges and Improvements

One challenge I faced was finding the optimal number of clusters in K-Means. I used the Elbow Method to identify the optimal value for 'k,' but determining the exact point where the WCSS curve flattened required careful observation. Another challenge was in tuning hyperparameters for the ANN, as different combinations of layers, neurons, and activation functions resulted in varying performances.

To improve my approach, I would like to explore Grid Search for hyperparameter tuning in future projects and experiment with other clustering algorithms like DBSCAN to better handle potential noise in the data.