**California Housing dataset**

**Objective:**

Predict housing areas using the mediam income in the California Housing dataset using machine learning models, specifically Linear Regression and Artificial Neural Network (ANN).

**1. Introduction:**

The goal is to analyze the California Housing dataset, consisting of features such as median income, house age, and geographical coordinates, to predict house prices. Two models, Linear Regression and ANN, were employed for the prediction task.

**2. Data Exploration:**

The dataset comprises 20,640 entries with eight features.

Features include MedInc, HouseAge, AveRooms, AveBedrms, Population, AveOccup, Latitude, and Longitude.

No missing values were present in the dataset.

**3. Linear Regression:**

Linear Regression was used to predict 'HouseAge' based on 'Latitude' and 'Longitude'.

RMSE (Root Mean Squared Error) was chosen as the performance metric.

**4. Artificial Neural Network (ANN):**

A simple ANN with one input layer, two hidden layers (64 and 32 units), and one output layer was implemented using TensorFlow/Keras.

The model was trained, and RMSE was used to evaluate its performance.

**5. Model Comparison:**

RMSE for Linear Regression: 12.06

RMSE for ANN: 10.87

The ANN model outperformed Linear Regression, indicating its ability to capture complex patterns in the data.

**6. Challenges Encountered:**

Data Visualization: Representing geographical data on maps within a Jupyter Notebook posed challenges. Interactive maps might require additional libraries like Plotly.

Model Interpretability: Neural networks are considered "black-box" models, making interpretation challenging compared to Linear Regression.

Hyperparameter Tuning: Fine-tuning hyperparameters for optimal model performance required time and experimentation.

**7. Recommendations:**

The ANN model demonstrated superior performance for predicting house prices based on geographical features.

Further optimization and tuning of hyperparameters could enhance the model's accuracy.

**8. Conclusion:**

The choice between Linear Regression and ANN depends on the nature of the data and the complexity of relationships.

While Linear Regression is interpretable and suitable for simpler relationships, ANN can capture intricate patterns, offering improved predictive performance in certain scenarios.

**9. Future Work:**

Explore advanced visualization techniques for geographical data.

Experiment with different neural network architectures and hyperparameters for further optimization.

**10. Acknowledgments:**

Gratitude to scikit-learn and TensorFlow communities for providing powerful tools for machine learning.

**11. References:**

California Housing Dataset

Scikit-learn Documentation

TensorFlow Documentation