

BDM 3035\_01- Big Data Capstone Project 2023F-T3\_AML 3104\_01

**Assignment - ANN & LINEAR REGRESSION**

Date of Submission: Friday, 3rd December 2023

Instructor: Prof. Ishant Gupta

Submitted by: Megha Chauhan (C0864100)

Contents

1. [Introduction 3](#_bookmark0)

[WHAT? 3](#_bookmark1)

1. [Data Preprocessing 3](#_bookmark2)
2. [Linear Regression 3](#_bookmark3)
3. [Artificial Neural Network (ANN): 3](#_bookmark4)
4. [Comparison and Analysis: 4](#_bookmark5)

[Performance Metrics 4](#_bookmark6)

[Strengths and Weaknesses 4](#_bookmark7)

[Complexity vs. Predictive Performance 4](#_bookmark8)

1. [Visualization 4](#_bookmark9)

[Linear vs. ANN Comparison 4](#_bookmark10)

1. [Conclusion 5](#_bookmark11)
2. [Challenges Encountered 5](#_bookmark12)

**1. Introduction:**

Predicting housing prices is a crucial task in real estate, guiding stakeholders in making informed decisions. This report delves into two distinct methods, Linear Regression and Artificial Neural Network (ANN), for forecasting housing prices in California using the California Housing dataset. We assess and compare these models based on key metrics, discussing their strengths, weaknesses, and the impact of complexity.

**2. Data Preprocessing:**

We loaded and explored the California Housing dataset, which includes various housing-related features. Features such as median income, housing median age, and proximity to the ocean were examined. Missing values were handled through standard imputation techniques, and the dataset was split into training and testing sets, with 80% of the data allocated for training.

**3. Linear Regression:**

We implemented a Linear Regression model using scikit-learn, training it on the dataset and making predictions on the testing set. Linear Regression, known for its simplicity, is effective in predicting numerical values. Evaluation metrics such as Mean Squared Error (MSE) and R2 Score were employed.

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

To introduce complexity and capture non-linear relationships, we employed an ANN using TensorFlow and Keras. The neural network comprised a single hidden layer with 64 neurons and a linear output layer. The model was trained on the scaled training set, and predictions were made on the scaled testing set. TensorBoard was utilized for visualizing the ANN architecture.

**5. Comparison and Analysis:**

**Performance Metrics**: Both Linear Regression and ANN achieved competitive results, with Linear Regression providing reliable predictions and ANN leveraging complexity for enhanced performance.

**Strengths and Weaknesses**: Linear Regression excels in simplicity and interpretability but may struggle with complex, non-linear patterns. In contrast, ANN captures intricate relationships but demands more data and computational resources, potentially leading to overfitting.

**Complexity vs. Predictive Performance:** The ANN, with its complexity, demonstrated improved predictive performance compared to Linear Regression. While ANNs excel at learning intricate features, the trade-off involves challenges in interpretability and increased computational requirements.

**6. Visualization:**

The scatter plot visualization comparing actual values with predictions from Linear Regression and ANN vividly illustrates their performance. It showcases how ANN captures nuanced relationships, especially in areas with non-linear patterns.

**7. Conclusion:**

In summary, both Linear Regression and ANN offer viable solutions for predicting California housing prices. Linear Regression provides simplicity and interpretability, while ANN introduces complexity to enhance predictive performance. The choice between these models depends on specific task requirements, considering factors such as interpretability, computational resources, and data nature.

**8. Challenges Encountered:**

Implementation challenges included handling missing values, scaling features for ANN, and selecting an appropriate neural network architecture. Interpreting the complex relationships in ANN highlighted the trade-off between model complexity and interpretability.