Linear Regresstion hands-on implementation

Group Members:

Mr Hamza, Muzammil

AI Proposal

Section A

2023

**Project Proposal: Unveiling the Mechanics of Linear Regression: A Manual Implementation Approach**

**Project Overview:**

The objective of this project is to gain a deep understanding of linear regression by implementing the model manually without relying on existing machine learning libraries. Linear regression is a fundamental algorithm in machine learning, widely used for predicting numerical outcomes based on input features. The manual implementation will involve creating the model, training it on the "movies\_metadata" dataset, and evaluating its performance.

**Project Goals:**

1. **Manual Implementation of Linear Regression:**

- Develop a linear regression model from scratch without utilizing any machine learning libraries (e.g., scikit-learn).

- Implement the core components, including the hypothesis function, cost function, and gradient descent optimization.

2. **Dataset Selection:**

- Choose the "movies\_metadata" dataset for the project. This dataset contains various features such as budget, popularity, runtime, and revenue, making it suitable for predicting movie revenue using linear regression.

3. **Data Preprocessing:**

- Perform data preprocessing steps, such as handling missing values, scaling, and splitting the data into training and testing sets.

4. **Feature Engineering:**

- Explore and implement basic feature engineering techniques to enhance the model's predictive capabilities.

5. **Model Training:**

- Train the manually implemented linear regression model using the "movies\_metadata"dataset.

- Implement gradient descent to optimize the model parameters and minimize the cost function.

6. **Evaluation:**

- Evaluate the performance of the model on the testing dataset from "movies\_metadata" using appropriate metrics such as mean squared error (MSE) or R-squared.

7. **Comparison:**

- Compare the results of the manually implemented linear regression model with a baseline model (e.g., using a library like scikit-learn). This will help validate the correctness of the manual implementation.

8. **Documentation and Reporting:**

- Maintain clear and detailed documentation throughout the project.

- Summarize the findings and insights in a comprehensive report, including visualizations and comparisons.

**Tools and Technologies:**

- Programming Language: Python

- Libraries: NumPy for mathematical operations, Matplotlib/Seaborn for visualizations

**Expected Challenges:**

- Understanding and implementing the mathematical concepts behind linear regression.

- Ensuring the correctness and efficiency of the manual implementation.

**Conclusion**:

This project aims to deepen your understanding of linear regression by implementing the model manually without relying on existing libraries. Through this hands-on approach using the "movies\_metadata" dataset, you will gain insights into the inner workings of machine learning algorithms and strengthen your practical skills in model development and evaluation.