ANALYSIS & RESULTS:

During my data preparation, I identified and removed columns with high missing values from the taxi dataset, such as 'airport\_fee'. Further, I merged this cleaned dataset with weather data, aligning it by 'pickup\_date' and removing irrelevant weather columns. To tackle the issue of outliers, which could skew analysis, I employed the interquartile range (IQR) method, defining outliers as points beyond 1.5 times the IQR from the first and third quartiles. This helped maintain the dataset's integrity by focusing on typical trip data, crucial for accurate model predictions.

After deducting outliers, a model is designed to predict ‘trip\_timing’ using ‘trip\_distance’ in a single-layer neural network with an SGD optimizer. This initial attempt resulted in a high Mean Squared Error (MSE) of 1,798,551,936.00 and a Mean Absolute Error (MAE) of 4,262.57. To enhance performance, I refined the dataset further, limiting trip distances to between 1.5 and 4 miles. This adjustment significantly improved the model's performance, achieving an MSE of 5.28 and an MAE of 1.93 across all optimizers including SGD, Adam, and RMSprop.

I enhanced my neural network model by incorporating an extended feature set that included 'trip\_distance', 'PULocationID', 'DOLocationID', 'time\_taken', 'tavg', and 'pres'. I also focused on removing outliers specifically in 'time\_taken' and set a learning rate of 0.001 for all optimizers. This adjustment improved the model's accuracy considerably, resulting in a consistent performance across all utilized optimizers. For the SGD optimizer, the Mean Squared Error (MSE) on the test data was 1.1962 and the Mean Absolute Error (MAE) was 0.9408. Similarly, using the Adam optimizer, I achieved an MSE of 1.1964 and an MAE of 0.9403. The RMSprop optimizer also mirrored these results with an MSE of 1.1965 and an MAE of 0.9404, demonstrating robust and uniform improvement in model predictions across different optimization techniques.

In the latest model iteration, I expanded the neural network by incorporating layers with 512, 256, and 128 neurons, finishing with a single output neuron for predictions. The training and testing datasets were normalized using StandardScaler, which helped standardize input feature scales, enhancing model performance. I utilized three optimizers like SGD, Adam, and RMS, all set at a learning rate of 0.001, and implemented an early stopping mechanism with a patience of 5 to prevent overfitting. The training stopped early across all optimizers due to non-improvement in validation loss. The SGD optimizer concluded training at the 9th epoch with a Mean Squared Error (MSE) of 1.2809 and a Mean Absolute Error (MAE) of 0.9582. Adam and RMSprop optimizers showed similar performance, stopping at the 13th epoch, with Adam achieving an MSE of 1.1963 and MAE of 0.9376, and RMSprop slightly higher with an MSE of 1.2098 and MAE of 0.9425. These results highlight the effective learning capability and robustness of the network across different optimization techniques.

In the advanced phase of model development, I employed a deep neural network architecture with layers consisting of 1024, 512, 256, 128, and 64 neurons, each trained with SGD, Adam, and RMSprop optimizers at a learning rate of 0.001. StandardScaler was used to normalize the data for consistent feature scaling. Early stopping was set with a patience of 5 to prevent overfitting. The SGD optimizer concluded at the 7th epoch with a Mean Squared Error (MSE) of 1.2497 and a Mean Absolute Error (MAE) of 0.9555. Adam, achieving the best performance, ran all 25 epochs and resulted in an MSE of 1.1886 and an MAE of 0.9363. RMSprop stopped at the 10th epoch with an MSE of 1.2708 and an MAE of 0.9559. These results highlight the critical balance between model complexity, optimizer choice, and training strategy to optimize performance and prevent overtraining.

|  |  |  |  |
| --- | --- | --- | --- |
| **Neural Networks** | **Optimizer** | **Validation MSE** | **Validation MAE** |
| Single Layer (1 input value) | SGD | 5.2848 | 1.9325 |
| Single Layer (1 input value) | Adam | 5.2826 | 1.9309 |
| Single Layer (1 input value) | RMSprop | 5.2831 | 1.9317 |
| Single Layer (7 input value) | SGD | 1.1962 | 0.9408 |
| Single Layer (7 input value) | Adam | 1.1964 | 0.9403 |
| Single Layer (7 input value) | RMSprop | 1.1965 | 0.9404 |
| Multi layered (512-256-128 Neurons) | SGD | 1.2809 | 0.9582 |
| Multi layered (512-256-128 Neurons) | Adam | 1.1963 | 0.9376 |
| Multi layered (512-256-128 Neurons) | RMSprop | 1.2098 | 0.9425 |
| Multi layered (1024-512-256-124-61 Neurons) | SGD | 1.2497 | 0.9555 |
| Multi layered (1024-512-256-124-61 Neurons) | Adam | 1.1886 | 0.9363 |
| Multi layered (1024-512-256-124-61 Neurons) | RMSprop | 1.2708 | 0.9559 |

The advanced neural network configurations, particularly the model with layers configured as 1024-512-256-128-64 neurons and optimized with Adam, performed best, achieving an MSE of 1.1886 and an MAE of 0.9363. This highlights the effectiveness of the Adam optimizer in handling complex models and datasets, significantly enhancing prediction accuracy and demonstrating the model's robustness in accurately forecasting taxi trip durations.