Point system:

Calculating a cyclist's performance involves assessing several key metrics, providing a comprehensive view of their fitness and capabilities. Among the essential parameters we evaluate are speed, heart rate, distance covered, calories burned, cadence, and power output. These values represent an average for a single user, collected and stored within a Bigquery Database.

To distill these metrics into a standardized measure, we've devised a point system, with a maximum score of 120, allocating 20 points to each category. It's important to note that this scoring system isn't universally applicable; it's tailored to each individual's unique attributes and goals.

The process of converting these metrics into points involves taking the daily average values and scaling them to fit the 20-point range. This approach ensures that each aspect of a cyclist's performance contributes equally to their overall score. By summing up these points, we obtain a holistic view of their daily performance.

What makes this system particularly valuable is its potential to forecast future performance based on historical data. Utilizing predictive models, we can anticipate a cyclist's future metrics, factoring in variables such as workout duration for a specific day. This predictive capability enables us to offer valuable insights and recommendations to enhance their training regimen and achieve their fitness objectives effectively.

Analysis on the Results:

Utilizing our meticulously crafted point system, we conducted a comprehensive analysis employing various machine learning models, including Linear Regression, Support Vector Machine (SVM), K-Nearest Neighbors, Random Forest, and the formidable Neural Network. Our objective? To predict a crucial value for future dates and years with precision and insight.

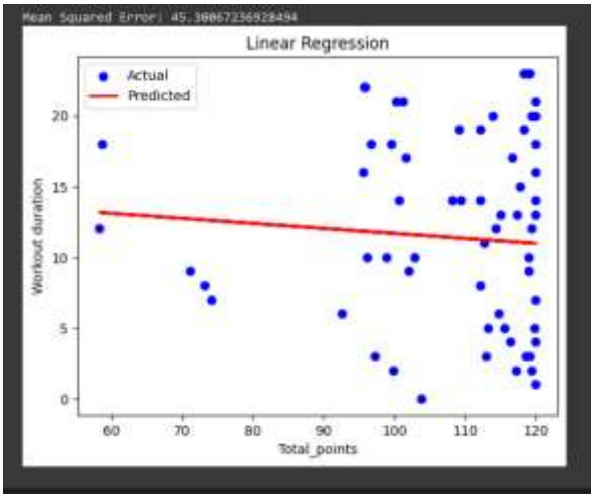
1) **Linear regression:**

   We've adopted this particular model for two compelling reasons. Firstly, we're keen to observe how our data responds to both fundamental models and more advanced ones. This approach allows us to gain insights into the nuances and complexities of our dataset.

   Secondly, the nature of our data necessitates the use of either a regression or classification model. These models are aptly suited to handle the type of information we're working with, providing us with valuable predictions and categorizations.

   Our initial foray into employing a Linear Regression model has yielded promising results. The Mean Squared Error (MSE) value of 45.30 is noteworthy as it falls below the threshold of 50. In statistical terms, this signifies a decent level of accuracy in our predictions. Additionally, when examining the accompanying visual representation, we note that a select number of data points closely align with the linear regression line. This alignment is indicative of the model's quality, as it's generally considered effective when data points cluster closely around the regression line.

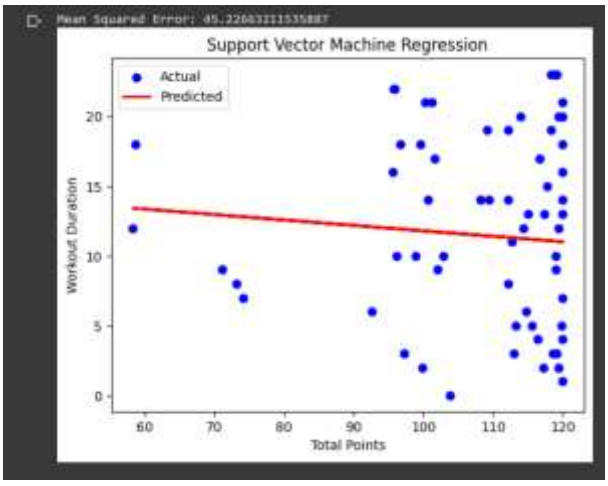
In essence, our choice of model demonstrates a pragmatic and data-driven approach, allowing us to better understand our data's behavior and extract valuable insights to inform future decisions and optimizations.



2) Support vector Machine (SVM):

Support Vector Machines (SVMs) emerge as a powerful tool when tackling classification tasks that involve non-linear data structures. Their effectiveness shines through in applications such as cyclist behavior prediction or route recommendation, where the inherent complexity of the data demands a versatile and accurate approach.

Our exploration into SVMs has yielded compelling results. Notably, the Mean Squared Error (MSE) value of 45.27 is quite commendable, mirroring the quality of predictions achieved by the Linear Regression model. In this context, an MSE of this magnitude is indicative of solid model performance. It's worth highlighting that, akin to Linear Regression, the graphical representation of the SVM results reveals a similar pattern, with data points closely mirroring the predicted line. A stronger alignment of data points with the predicted line would have further reinforced the model's fitting capability.

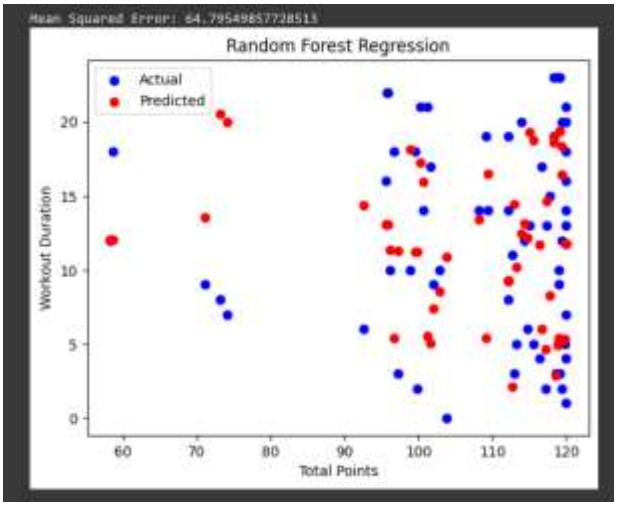


3) Random-Forest Regression:

Decision trees and random forests offer valuable capabilities in classification tasks, making them particularly apt for predicting a cyclist's decision to stop at a specific point or continue along their route. Additionally, they excel at feature importance analysis, providing insights into the most influential factors driving these decisions.

Our exploration into random forests, however, has yielded results distinct from the previous models. The Mean Squared Error (MSE) for the Random Forest model stands at 64.80, marking it as the highest among the selected models. This higher MSE suggests that the model's predictions deviate further from the actual data points compared to the other models.

The visual representation of the data reinforces this observation, as a significant proportion of the predicted data points appear noticeably distant from the actual data points. This divergence implies that the chosen Random Forest model might not be the most suitable for our current dataset, as it struggles to capture the underlying patterns and relationships.



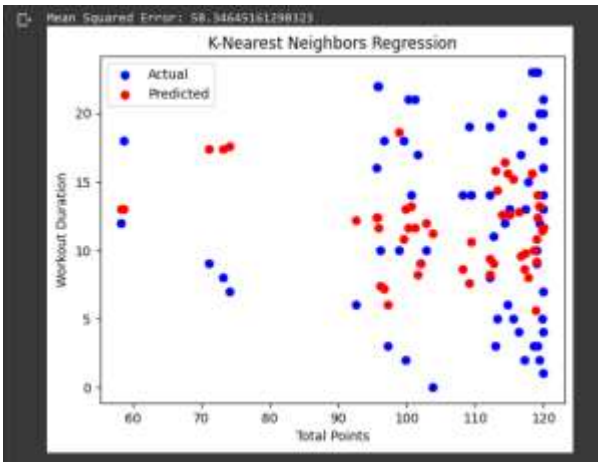
4) K-Nearest Neighbour (KNN):

K-Nearest Neighbors (KNN) can be valuable for cyclist data analysis by aiding in cyclist safety predictions, route recommendations, behavior analysis, and bike theft prevention. By considering historical data and relevant factors, KNN can help identify accident-prone areas, suggest safe routes, classify cyclist behavior, and predict bike theft risks, making it a versatile tool for enhancing cyclist safety, convenience, and overall cycling experience.

Upon evaluating the results, we find that the Mean Squared Error (MSE) stands at 58.37, ranking it as the second highest among the models we've implemented. This MSE value signifies that the model's predictive accuracy is somewhat lower compared to other models we've explored.

A closer examination of the graphical representation reinforces this observation. It becomes evident that a significant portion of the predicted data points is notably distant from the

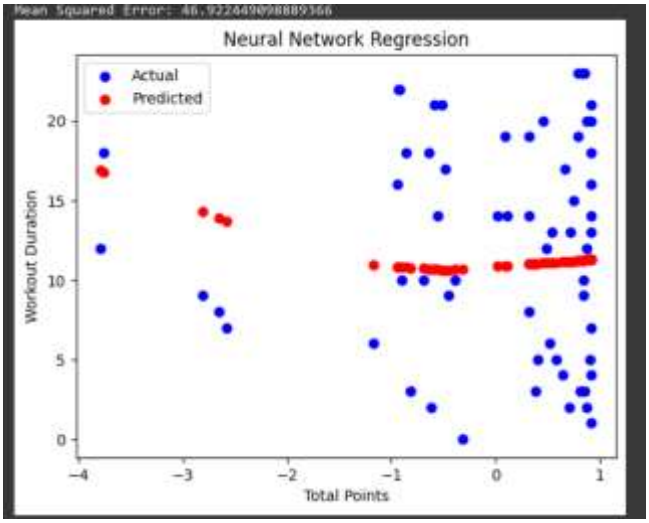
actual data points. Interestingly, the predicted values themselves tend to cluster closely together. This Contrast of data points implies that the chosen model may not be the best fit for our dataset in its current state.



5) Neural Network:

Neural networks are a powerful tool for point prediction tasks using cyclist data, allowing us to forecast continuous variables like cyclist speed, travel time, or distance traveled. The process begins with data collection, encompassing factors such as weather conditions, road types, and cyclist attributes, followed by rigorous preprocessing, including data cleaning and feature scaling. After splitting the dataset into training and validation sets, relevant features are selected and engineered to capture meaningful patterns. The neural network architecture, which can be feedforward, convolutional, or recurrent depending on the nature of the data, is chosen with consideration to layer structure, units, and activation functions. Training the neural network on the training dataset allows it to learn the underlying relationships in the cyclist data, with adjustments made during validation to optimize performance. For our dataset we have Feedforward method.

In this training process, a neural network model was trained for 100 epochs on a dataset, and the loss (error) was tracked at each epoch. The loss started at a high value of 172.19 and gradually decreased over time, converging to a final value of 48.62. The loss value represents how well the model's predictions match the actual target values, with lower values indicating better performance. In this case, it suggests that the model's predictions improved significantly during training. After training, the model's performance was evaluated using Mean Squared Error (MSE), which measures the average squared difference between predicted and actual values. The MSE was calculated to be 46.92, indicating the overall accuracy of the model's predictions.


Mean Squared Error: 46.92264909888893
Neural Network Regression
Actual
Predicted
Workout Duration
Total Points
20
15
10
5
0
−4
−3
−2
−1
0
1