Reflection:

To accomplish this in this problem, our goal was to estimate the weight of an individual based on two input variables; height and gender. Here's a detailed reflection on each part of the solution and possible improvements:

Data and Feature Selection:

The dataset provided only two primary predictors: Height and Gender. Though these are logically related factors for weight prediction and may serve reasonably well for weight prediction, there are other issues such as age, exercise level, and body composition, for example, that are not sufficiently captured if we rely only on these factors.

This research transformed the Gender aspect into a nominal feature with 0 meaning male and 1 denoting female. Although this encoding is simple, gender influence on weight may differ greatly and this may not requite the underlying structural disparity between males and females.

Model Choice:

Linear Regression was used as the next model as this is one of the simplest models which directly provide relation with the parameters used. Training the model was done with Height and Gender: and the output for weight was obtained with MSE of 101.24. Although this is a quite logical first action, the linear regression supposes a strong and direct connection between the objectives and the features, which does not have to be true for such a type of data.

The results, with regard to MSE, imply that while the model does find some relationship between height, gender and weight, the results could still be more accurate.

Model Performance:

Hence the MSE of 101.24 indicates that the model under/over estimated the weight by about 10 when evaluated against the square root of MSE. As you can see, this is quite good but it could be made significantly smaller still by adding more features typical to this type of document.

Hence the rather high MSE shows that height and gender alone cannot be used as proper predictors. Weight also depends on other factors such as; lifestyle, exercise regime, diet, and genetic predisposition none of which is captured in this dataset.

Improvements:

Feature Engineering: Further expansion of the features like age, exercise, diet even BMI would probably enrich the model. Some of these factors have close associations with weight and might assist with the variation that is not respected presently.

Non-Linear Models: Because of this it could mean that the relationship between height, gender and weight is non-linear, using techniques such as Polynomial Regression, Random Forest or even Neural Networks may show improved results.

Data Scaling: We could also try whether a pre-processing of height, using for instance, the scaling or normalization would yield better result if we test with other algorithms such as K-Nearest Neighbors (KNN) or Support Vector Machines (SVM).