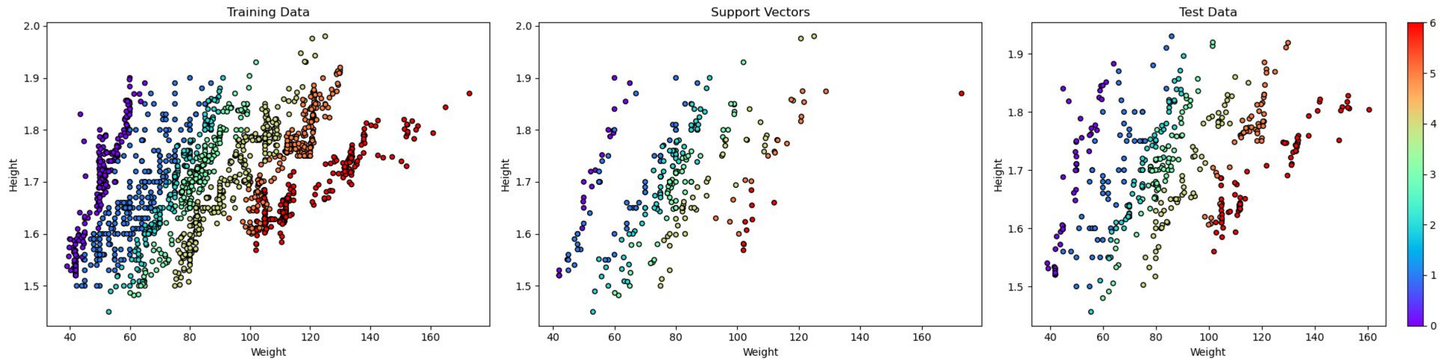
Comparative Analysis of Obesity Prediction Models: Linear Regression, SVM, and MLP

Sie Hendrata Dharmawan, Brendan Liu, Joseph Martinez

DARTMOUTH ENGINEERING

Introduction

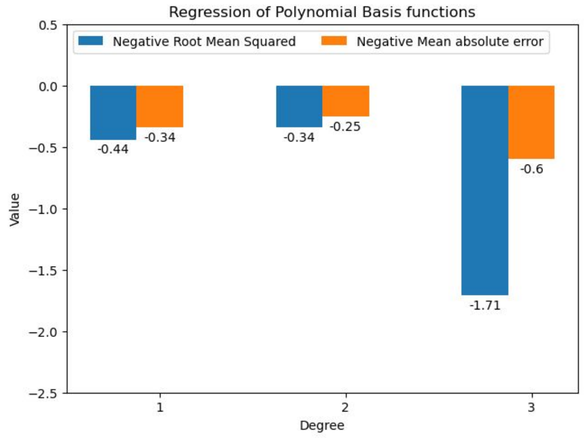
This project explores the application of machine learning techniques to predict obesity levels based on eating habits and physical condition data from individuals in Mexico, Peru, and Colombia. Using a dataset of 2,111 instances with 16 features, we implement and compare three machine learning approaches: Linear Regression, Support Vector Machines (SVM), and Multi-Layer Perceptron (MLP) neural networks. The dataset classifies individuals into seven categories ranging from "Insufficient Weight" to "Obesity Type III." Our research aims to determine which predictive model performs most effectively in estimating obesity levels based on lifestyle factors, potentially offering insights for targeted intervention strategies in public health.

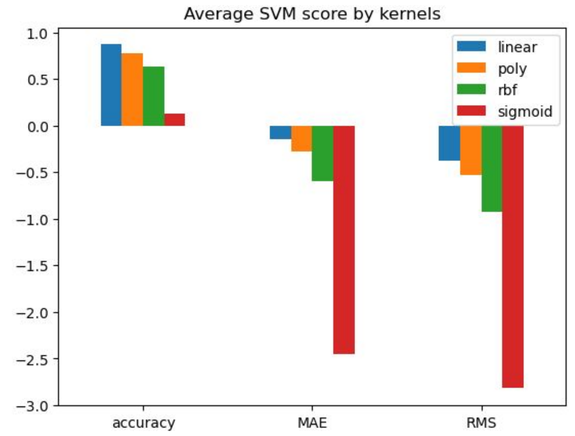


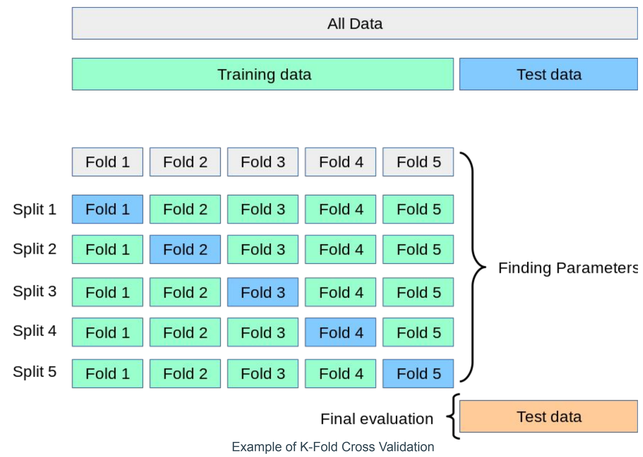
Methodology

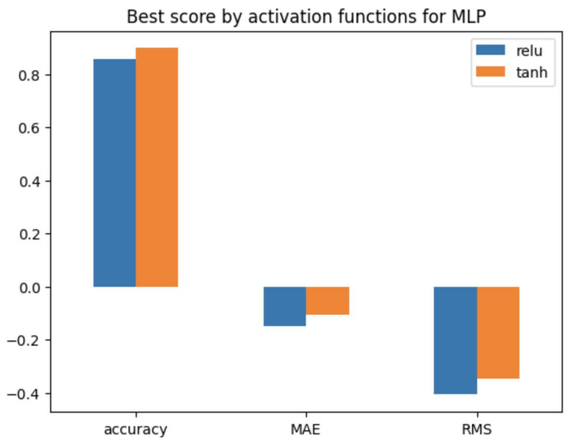
All models were evaluated using 10-fold stratified cross-validation (StratifiedKFold). We assessed model performance using three distinct metrics:

1. Accuracy : Proportion of correctly classified instances
2. Negative Root Mean Squared Error (RMSE) : Measuring the square root of the average squared differences between predicted and actual values
3. Negative Mean Absolute Error (MAE) : Measuring the average absolute differences between predicted and actual values









Conclusion

The SVM model, configured with a linear kernel, regularization parameter C=100.0, and gamma set to 'scale', achieved an accuracy of 96.03%, outperforming both the Linear Regression baseline and the best MLP model (90.99% accuracy).

According to the confusion matrix to the right, all of the predictions are within 1 label away from the target label