Jefferson Veliz

Professor Avinash Jairam

Data mining for business analytics

August 5th, 2023

Lab #4

In a world increasingly shaped by data, this lab merges from the compelling idea that we can harness data's potential to predict health statuses whether someone is obese, overweight, normal weight, or underweight. By leveraging the power of basic attributes like age, gender, height, weight, and BMI, we delve into a realm where data-driven insights directly impact individual well-being. By using decision tree and random forest algorithms, I aimed to gain hands-on experience in machine learning.

The Decision Trees algorithm was employed to perform multi-class classification, aiming to predict individuals' health statuses based on attributes such as age, gender, height, weight, and BMI. This report presents a comprehensive evaluation of the model's performance, with a specific focus on the weighted average metrics and overall accuracy. The weighted average precision for the Decision Trees model is 0.98, indicating a high level of accuracy in the model's predictions across all health status categories, including "Normal Weight," "Obese," "Overweight," and "Underweight." This precision score underscores the model's ability to make accurate predictions and minimize false positives. The weighted average recall is 0.97, suggesting that the Decision Trees model effectively captures instances from all health status categories. This indicates the model's proficiency in identifying true positives and minimizing false negatives.The weighted average F1-score is 0.97, showcasing a balanced blend of precision and recall for all health status categories. This demonstrates the model's ability to achieve a harmonious trade-off between accurate positive predictions and comprehensive category coverage.The overall accuracy of the Decision Trees model is 0.97. This signifies that the model correctly classified 97% of all instances in the dataset, reflecting its general capacity to predict individuals' health statuses accurately.

The Random Forest algorithm was employed for multi-class classification, aiming to predict individuals' health statuses based on attributes such as age, gender, height, weight, and BMI. This report presents a comprehensive analysis of the model's performance, with a primary focus on the weighted average metrics and overall accuracy.The weighted average precision, recall, and F1-score metrics for the Random Forest model are all perfect, with a value of 1.00. This indicates exceptional accuracy across all health status categories, including "Normal Weight," "Obese," "Overweight," and "Underweight." The model's accuracy reflects its precise predictions across the dataset.The overall accuracy of the Random Forest model is also perfect, reaching 1.00. This implies that the model successfully classified every instance in the dataset with precision, showcasing its ability to accurately categorize individuals' health statuses.

Feature importance analysis extends its impact beyond the realm of machine learning. It has the potential to deliver tangible benefits to healthcare professionals and doctors in several ways: Weight and BMI are key indicators of an individual's overall health. Their high feature importance scores emphasize their central role in predicting health statuses accurately. Doctors can benefit from this by paying particular attention to weight and BMI when assessing patients. These attributes can serve as early warning signs of potential health issues, allowing doctors to recommend suitable interventions or lifestyle changes.