

Wine Quality Data Analysis Using Different Methods

Abstract

This project focuses on improving the predictive accuracy of wine quality classification by applying a blend of statistical and machine learning techniques on the Wine quality dataset. The analysis begins with individual applications of statistical quality control (sqc) to identify and manage variations in the data, and principal component analysis (pca) to reduce dimensionality and remove redundancy. Subsequently, a combined approach (sqc + pca) is proposed to enhance the data quality and feature representation. Various machine learning algorithms, including Random Forest, Support Vector Machine, and Logistic Regression, are trained on the original and transformed datasets. The performance of each model is evaluated using confusion matrices and classification metrics. Results indicate that the combined sqc + pca approach yields significant improvements in model accuracy, with the Random Forest algorithm delivering the best performance. This study demonstrates the effectiveness of integrating statistical preprocessing with dimensionality reduction to boost classification outcomes in wine quality prediction tasks.