Yash Parikh

Abstract

This research investigates three machine learning models and a comparison was conducted using different performance measures to determine which algorithm would effectively predict protein classifications based on residue count, structure of molecular weight, and protein sequences. Decision Trees, Random Forests and Extra Trees models are applied on a structural protein sequences dataset. This dataset also contains other features for extraction methods and details on protein structures. Based on the experiments that were conducted on these models, it was demonstrated that Extra Trees model had comparable results but marginally better than the Decision Trees and Random Forests models.