Abstract

After recent technological advancements in astronomical surveys, modern astrophysics is concerned with the study and characterization of distant objects such as galaxies, stars and quasars. Obtaining the optical spectrum and consequently deriving the redshift could instantly classify these astronomical sources but as long as spectroscopic observations are not available for many galaxies and the process of measuring the shift can be time-consuming and infeasible for large samples, a machine learning approach could be applied to determine the redshifts of galaxies from their photometric colors. In the current manuscript, by using the flux magnitudes from the Sloan Digital Sky Survey (SDSS) catalog, a database of color indices acting as an approximation for the spectrum is created. These color indices (u-g, g-i, r-i and i-z) are considered as our input features and a subset of sources containing spectroscopic redshifts were chosen as the training dataset. As the final step, a decision tree algorithm was designed to obtain a rather accurate estimation of the redshifts and then its evaluation procedures were investigated. Limitations of astronomical surveys which often lead to imaging a large number of faint galaxies, necessitated the requirement of sophisticated ML algorithms which can simplify the process of using the data to inform our view on understanding the universe.