Training-test split

Pre-processing of data