After data pre-processing the data frame consists of 36275 rows and 27 columns. This dataset is split two times. Initially, data is separated into 85% train data with (30833, 27) and 15% into validation data with (5442, 27). Secondly, validation is split into two parts i.e., validation (80% of (5442, 27)) and test data (20% of (5442, 27)). Overall, below is the structure of the data before applying sampling techniques.

***Train data:*** (30833, 26)

***Train target:*** (30833,)

***Validation data:*** (4353, 26)

***Validation target:*** (4353,)

***Test data:*** (1089, 26)

***Test target:*** (1089,)