1)(a). Training data set provided is of good quality. It is nicely presented by using panda dataframes.

(b). There is no missing data. All the column in the data contains numerical values so no hot encoding required for the data.

c). No scaling required, no outliers , well presented.

2). Data is well presented using pandas dataframe. No as such data preprocessing required because as mentioned in above answer all the numerical data of all columns is within the same range so no scaling and normalization is required. Since the data is all numeric so no one hot encoding required.

3). The Supervised learning model used for classification is K-Nearest Neighbors. The main reason for using KNN is the size of the test and train data which is pretty small. Specially KNN take a lot of time for predicting for test data, but because of its smaller size it yielded good result in significant less time. Simple to implement,flexible to feature/distant choices.It can handle well multi-class cases. http://www2.cs.man.ac.uk/~raym8/comp37212/main/node264.html.

Also the data is pretty nicely scaled and normalized which makes it well suited for KNN.