Data preparation:

1. Specialization is one-hot encoded.
2. Degree is one-hot encoded.
3. Age is calculated by subtracting DOB from the current date.
4. Standard Scaler is used to normalize the data.
5. Correlation is checked and highly correlated features (> 0.9) are removed(none found in the data set).

Data analysis:

1. Mean, standard deviation, min, max and quartiles is calculated for each feature.
2. Correlation is checked for each variable.
3. PCA is used to visualize the data along two principal components.

Experiments:

1) Test size:
    a) 90-10: 71%
    b) 80-20: 71%
    c) 70:30: 72%

    This shows that accuracy is almost constant with the size of the training set, so probably for the data set to get better accuracy a change in model will fare better.
2) PCA is used to reduce features to 15 accuracy is still around 71%.
3) Specialization was removed and tested still accuracy around 71%.