Deliverable 3

1. Final Training Results

In the previous deliverable, I calculated the probability of each feature in training set, testing set and validation set separately, and use them to vectorize (replacing the string representation with the probability of those features) each set. This time, I calculated the probability of each feature across the whole data set(prob_x_overall) and map them to each set. However, this still gives me 100% accuracy. I also implemented the functions of calculating the recall score, precision score, which are both 1.0. Hence I can only admit that my model is somehow perfectly accurate.

2. Final demonstration proposal

- Key idea: This model is to predict a prisoner's tendency to reoffend after they are released so that the judge can decide how much sentence to give. Different factors, including ethnicity, age and type of crime they committed and so on, have within them more subtypes with different probabilities of a recidivist.

- Method: Introduction of SVM and Random Forest working mechanism based on lectures and sklearn documentations.
- Results and future research: For results there should be two parts, the first is prediction's accuracy is extremely high and reached 100%, the second is the probability distribution of different factors of recidivists.
  For future research, the dataset can be further expanded and the model can be deployed into practice as an assistive tool for legal procedures.
- References as convention.