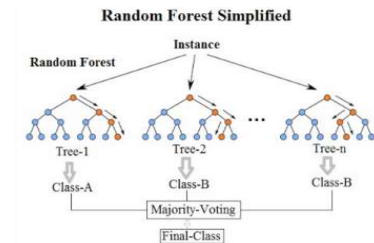
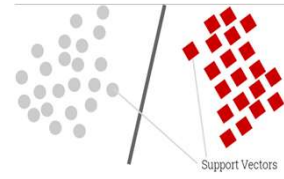
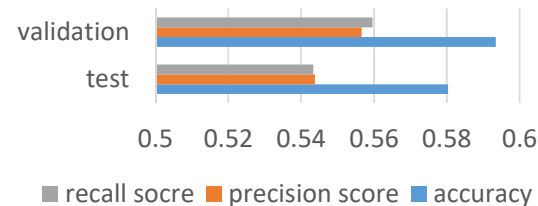


## Zilong Wang

- **Linear Support Vector Machine (LinearSVM)** tries to find a line that “best” separates two sets of data into two classes. In my case, it tries to separate all the records into recidivists and non-recidivists.
- As there is no hyper-parameter to tune in LinearSVM, I deployed default LinearSVM model by sklearn.
- **Random Forest** is good for multiclass classification as it uses predictions from multiple sub-trees and uses the majority vote of the class as the final class.
- For this model, I deployed RandomForestClassifier Model by sklearn, and since we can tune hyper-parameters, I ran RandomizedSearchCV and GridSearchCV to find the best ones.

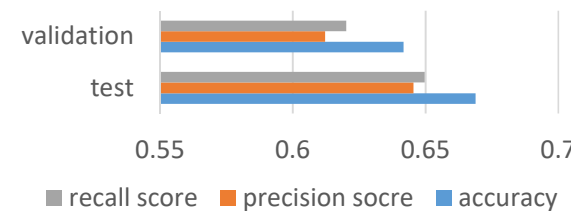


## LinearSVM

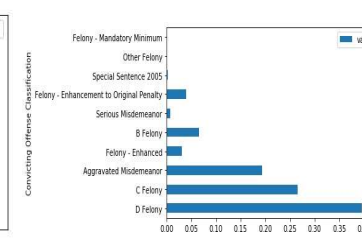
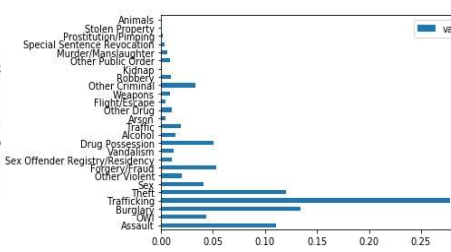
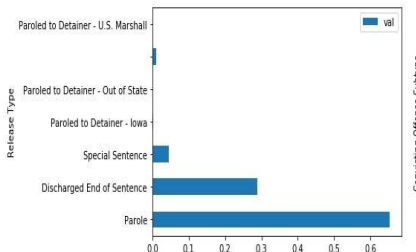
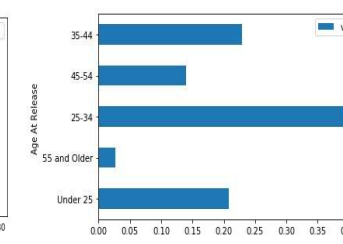
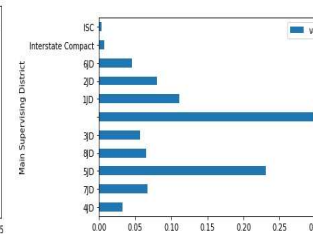
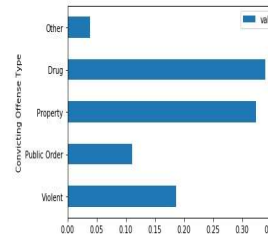
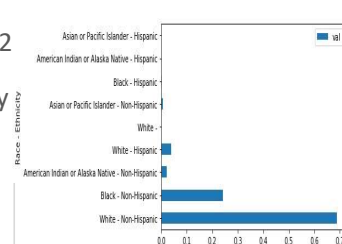


The result showed an interesting pattern that in LinearSVM the performance is better on validation set but in Random Forest the performance is reversed. However, **generally Random Forest outperforms LinearSVM.**

## RFC with optimal hyperparameters



- Below are the probability distribution of some important labels of recidivists.



Based on the results, I found that it is relatively difficult to accurately predict whether a prisoner will reoffend after being released using Machine Learning models, hence the judge should only regard the prediction as a reference, but not rely on it. However, as far as I know, these kinds of models are prevailing among legal systems, so this needs judges' attention and they should avoid blindly trusting the models.

Julia, A. Luis, C. et Thomas, T.  
Prediction algorithm for crime  
recidivism, 2015,  
retrieved from  
[http://cs229.stanford.edu/proj2015/250\\_report.pdf](http://cs229.stanford.edu/proj2015/250_report.pdf).

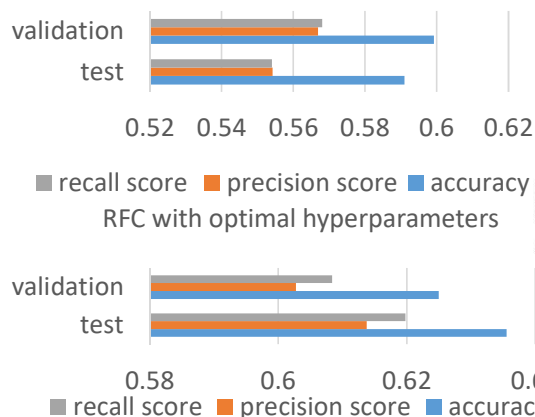
McGill Artificial Intelligence Society,  
Lecture 4&5, 2019, retrieved from  
[https://docs.google.com/presentation/  
d/1B786WM6XWdts-  
QL5B01zAngLNiKl1jwDkirv7dYMCM/e  
dit#slide=id.g62fee3643f\\_0\\_62](https://docs.google.com/presentation/d/1B786WM6XWdts-QL5B01zAngLNiKl1jwDkirv7dYMCM/eidit#slide=id.g62fee3643f_0_62)

Many prisoners reoffend after they are released, which poses a security hazard and often requires huge amount of time, money and labors to make up for the damage caused by recidivism.

This project aims to give an accurate prediction of a criminal's tendency to reoffend which judges can take into account when deciding one's sentence and release.

With supervised learning trained on historical data from prison, algorithms can learn to distinguish recidivists from non-recidivists based on different labels and predict whether a person tends to reoffend.

After learning about the racial bias existed in sentence giving, I decided to try whether removing the label 'Race' will affect the result.



Surprisingly, performance in LinearSVM improved but dropped in Random Forest. Nonetheless, the improve in LinearSVM showed that without the racial bias, the model can make better prediction. Hence, racial bias does affect models.