**Model:**

* Cleaned data and changed categorical data to numerical, didn’t solve weight problem.
* Solved unbalanced dataset with SMOTE
* Changed 6 categories of “case\_status”, out label, to binary “denied” and “certified”
* Split data into Training data and holdout testing data as the real world data; further split Training data into: training data and testing data to train model
* Wrote function to run default parameters automatically, and evaluate models with f-measure, recall and precision, and ROC and AUC graph as evaluation metrics. Used accuracy for training data and testing data to detect overfitting problem

→ tried random forest first, 90-ish accuracy, want to improve model

* Tried to use dummied method to solve weight problem in between, but 1000-ish columns, didn’t work efficiently, so discard it
* Tried GridSearchCV and randomizedSearchCV to auto tune parameters, but it torn my computer apart and it was dead for so many times, I turned back to self-tune parameter process
* Re-cleaned data and tried new dataset

→ the accuracy scores was worse than before, so go back to first version of cleaned data

→ didn’t run out SVC , tried 3 times, everytime I have to disturb kernel and restart all (find out why)

Next steps:

* Tried to use cross validation to increase accuracy
* Apply the model to the real world data
* Made slides
* Find pro and cons behind each algorithms to illustrate why I choose to try them and why some succeed and some fail
* Explain why to use roc curve and auc curve and other evaluation metrics