ONSTRUCTING INTELLIGENCE

Procore Technologies – Machine Learning
Capstone 2016-2017
Vivek Patel – Team Lead
Nicole Moghaddas – Team Scribe
Lauren Dumapias, Deanna Pham, Charlie Getzen

PROBLEM/MOTIVATION

Make analytic predictions for problems faced in the construction industry

Focused on predicting latency of schedule on a construction project

Determined for future projects whether they will be on time, late or very late

Over the remaining lifespan of the project, we will make predictions about project budget/safety

TECHNOLOGIES

Utilized the machine learning libraries available through the Python community

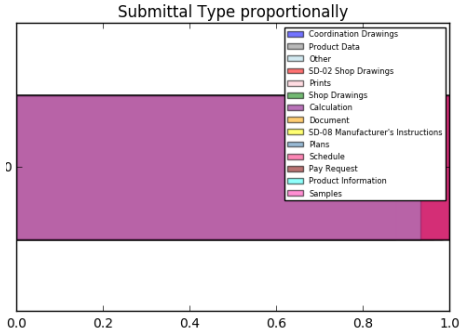
Used iPython notebooks to organize and visualize our data

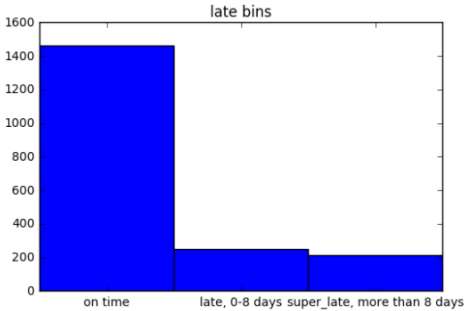
Pros

- Fluid workflow
- Quick and information-dense results

Cons

- Difficult to get all dependencies installed
- Difficult to collaborate on a single notebook in real-time





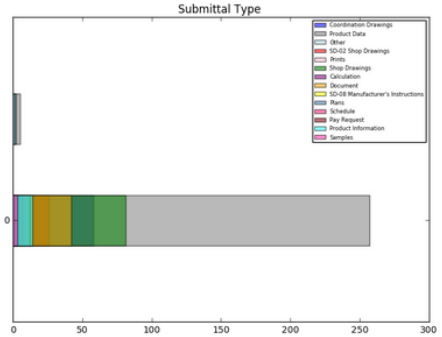
SOLUTION DESIGN (Part I)

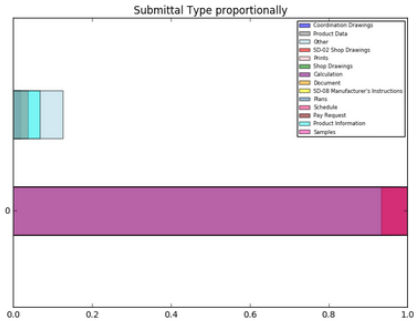
Considered multiple classification schemas (gaussian naive bayes model, nearest centroid classifier, k-nearest neighbors, logistical regression, random forest)

- Random forest is main focus as of now

Ensure that data is useful in making predictions by arranging it into buckets

- Quantized continuous data into discrete data to be able to analyze it





SOLUTION DESIGN (Part II)

Used k-fold cross-validation to determine which classification method is best

- Get an accurate sense of how well our model did
- Run training data on different chunks to get accuracy of model, to avoid overfitting and generalize
- Run training on same data frame for multiple classifiers
- Run training data on a classifier, using different parameters
- Classifier that returns a result that's most realistic/most accurate is the one we choose to use

Future

Move from basic classification into regression

Add budget prediction / budget score

Add safety prediction / safety score

PROTOTYPE DEMO

- Demonstrate iPython notebook capabilities
  - Pull in data frame
  - Instantiate variables/classes
  - Plot/visualize data
- Demonstrate training/testing of large data set
  - Predict whether company’s future will be on time, late or very late