



change the way you see companies

Deepak (James), Justin, Saniya, and Jeff

---

---

# The Initial Problem



## What is a Good Company?

- > Initial conception was to use Glassdoor ratings data as primary source
-

---

# The *Real* Problem



- Back to the drawing board!

---

# The Revised Problem

- Will a company “succeed” and IPO or get bought out?
- Or “fail” and close?



---

# The Process

## Week 1

- Scraping data from Glassdoor
- Data exploration

## Week 2 - Modeling

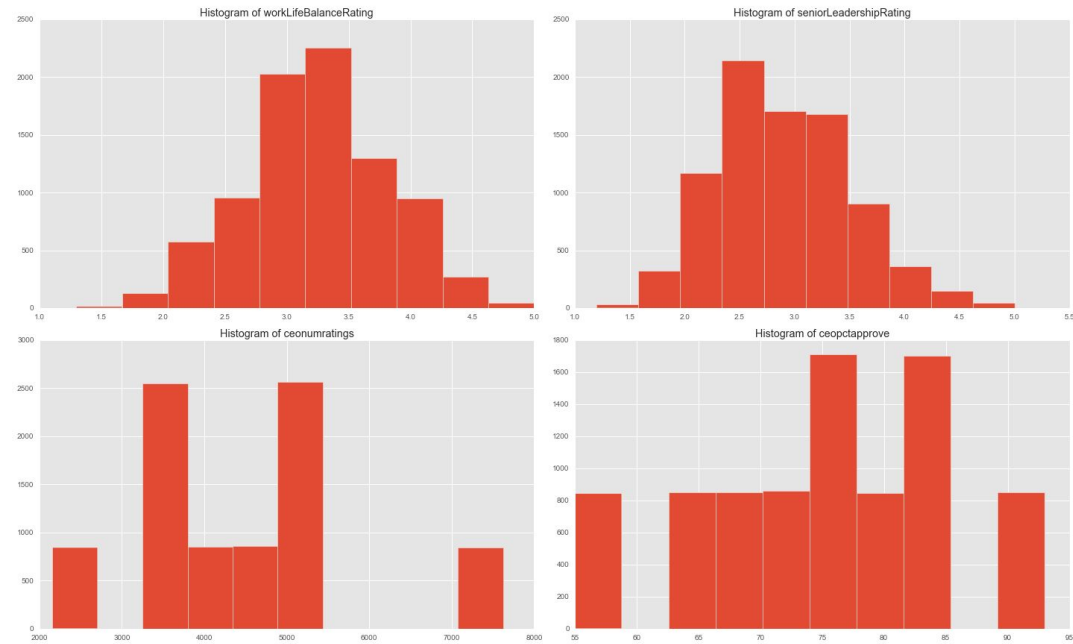
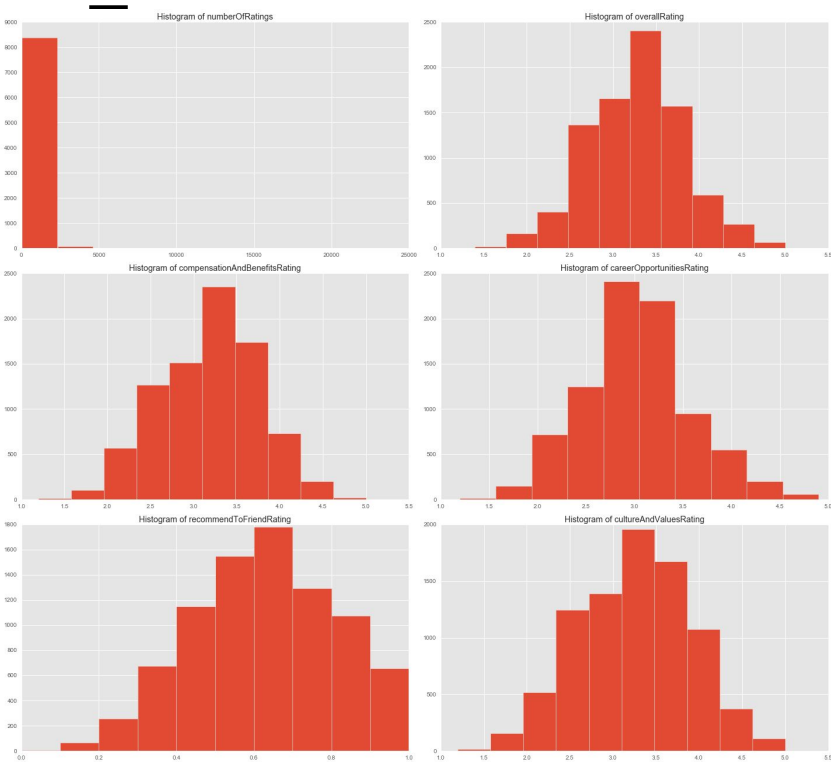
- Initial model
- “Pivot” to re-focus on data

## Week 3 - Modeling / Web dev



# The Data

- Glassdoor: top company information, reviews
  - Large number of similar ratings
  - Long tail of companies with few ratings
- AngelList: funding information
- Crunchbase
  - Funding rounds
  - Operating status
  - Everything!

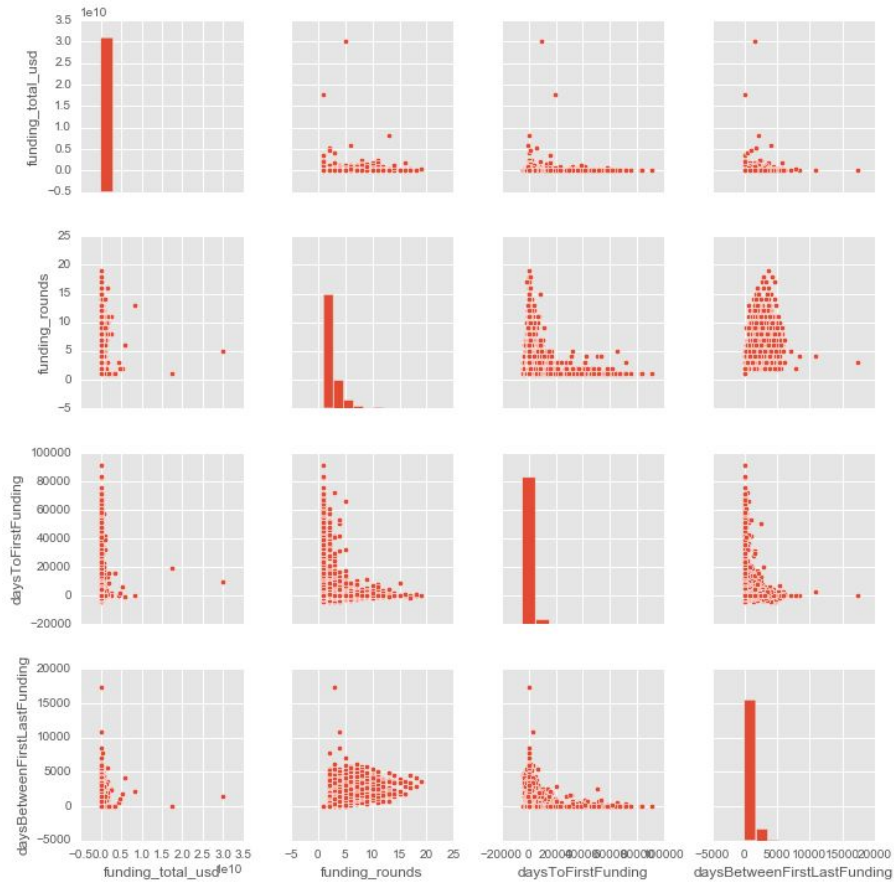


Glassdoor Histograms

# Glassdoor Pairplot







Crunchbase Pairplot

---

# Demo

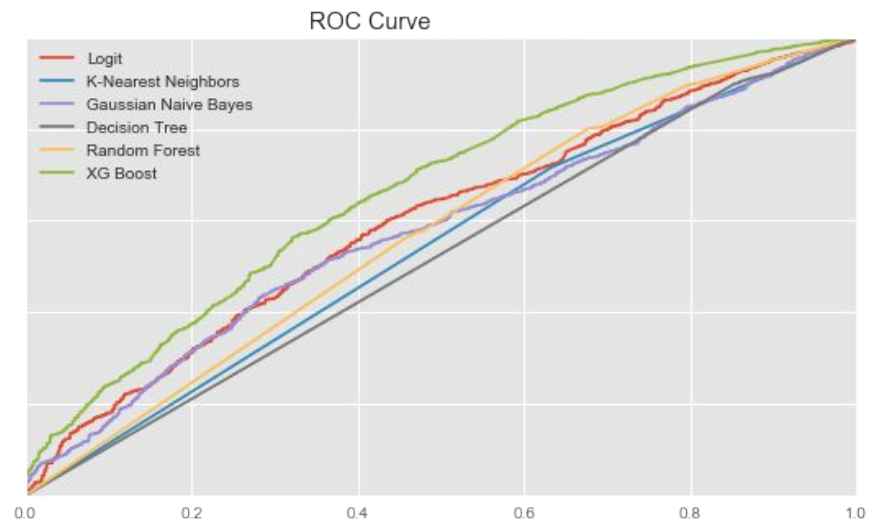
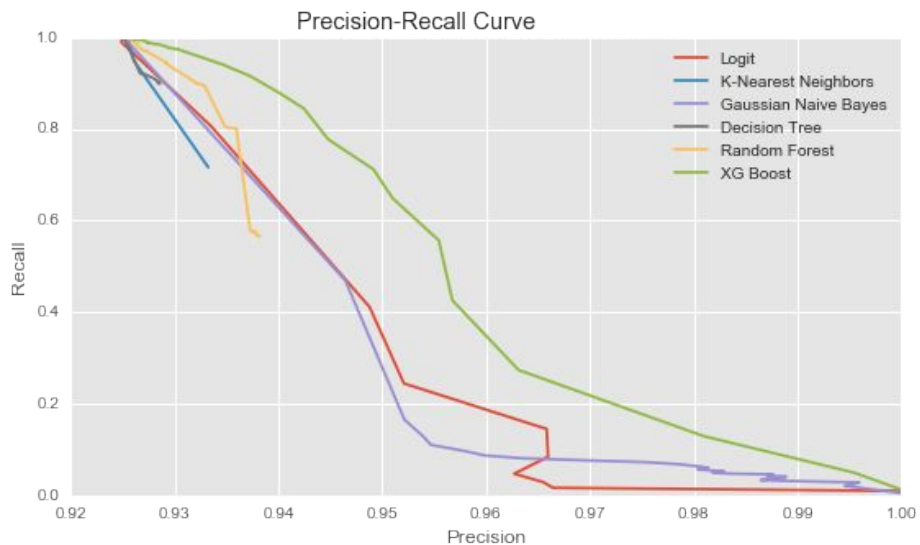
---

---

# The Models

- Logistic Regression - AUC: 0.60
  - K-Nearest Neighbors - AUC: 0.54
  - Gaussian Naive Bayes - AUC: 0.58
  - Decision Tree - AUC: 0.52
  - Random Forest - AUC: 0.55
  - **XGBoost Classifier** - AUC: 0.66
-

—



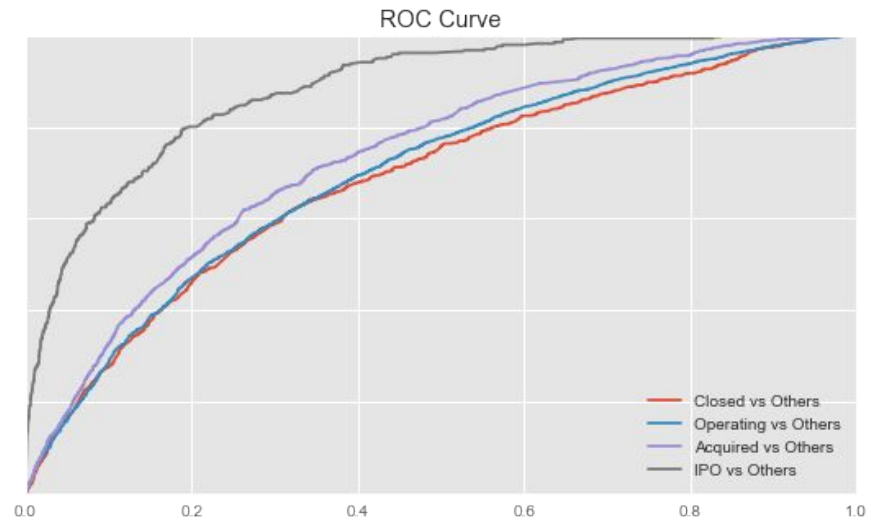
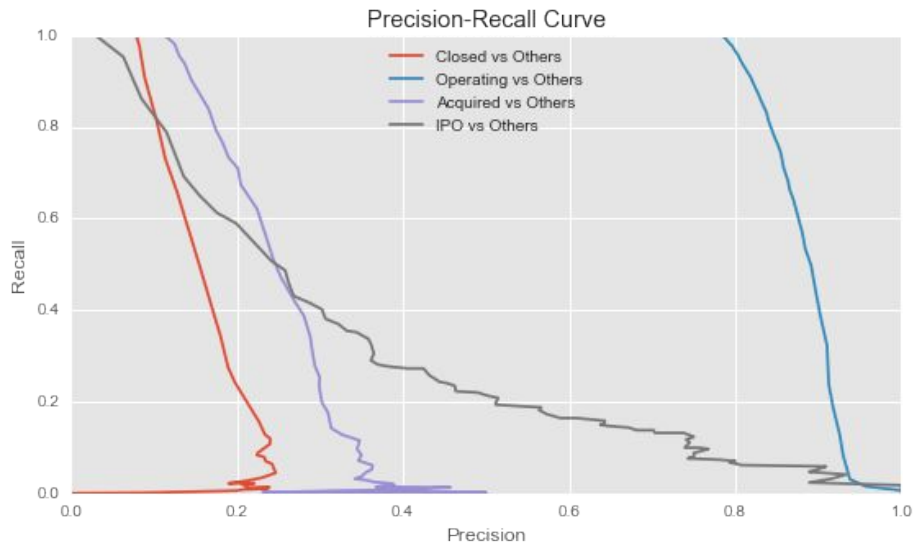
PR & ROC Curves for all models

—

---

# XGBoost AUC Scores

- Closed vs Others: 0.69
  - Operating vs Others: 0.70
  - Acquired vs Others: 0.76
  - IPO vs Others: 0.88
-



PR & ROC Curves for XGBoost

---

---

# The Findings

## Overall Importances

- Funding total
- Days to first funding
- Funding period
- Funding rounds
- VC tier

## Binary “Unimportances”

- Recommend to Friend
- CEO approval
- Senior Leadership
- Overall Rating
- Number of CEO Ratings

## Multi-classification “Unimportances”

- Big Data Analytics
  - Recruiting
  - Nonprofits
  - Fitness
  - iOS
-

---

# The Interface

- Flask
- D3
- Javascript





---

# Demo

---

---

# Next Steps

- More Features
    - Budget breakdown of the companies
    - Stock market data at company milestones
    - Marketing avenues of the company products
    - How quickly a company grew
    - More granularity in “operating”
    - Get more data on Glassdoor ratings and other features
    - Level of education of employees and founders
    - Opportunities by geography
  - Website Deployment
    - Combining classification and exploration pages
    - Checkboxes for categorical variables
    - Responsive score text color and svg elements
    - Public hosting
-

---

---



**THE END**

---