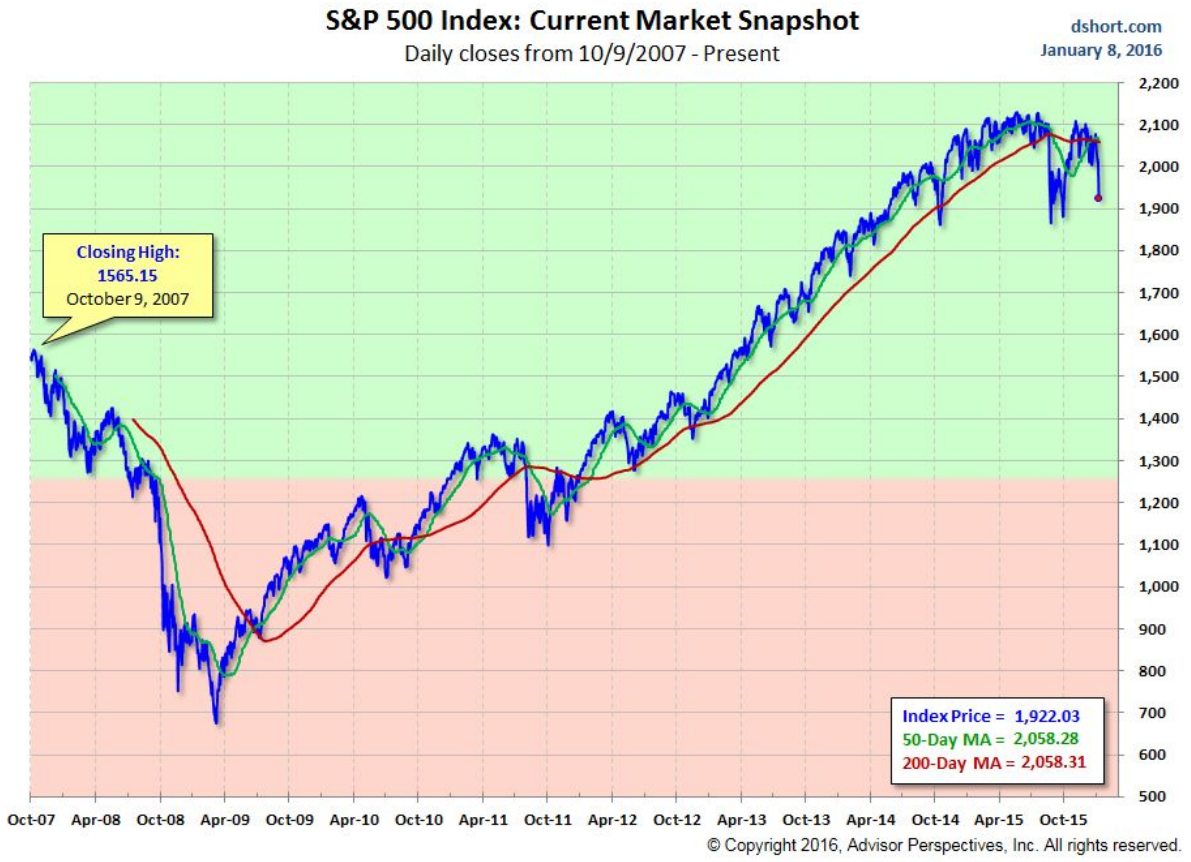
What job posting tells about a company?

Team #8

Background

Began in 2007 with a crisis in the subprime mortgage market in the USA, and developed into international banking crisis with collapse of Lehman Brothers.

After that, the market bounded back in a long period. However, different company apparently performs differently.



Which companies outperforming peers?

As we mentioned, after the historical financial crisis, some companies grow faster and some companies grow slower. This, we suspect, could be reflected by its need in job market. For example, if a company hires more, than it probably having more order to make and cash flow in its pocket.

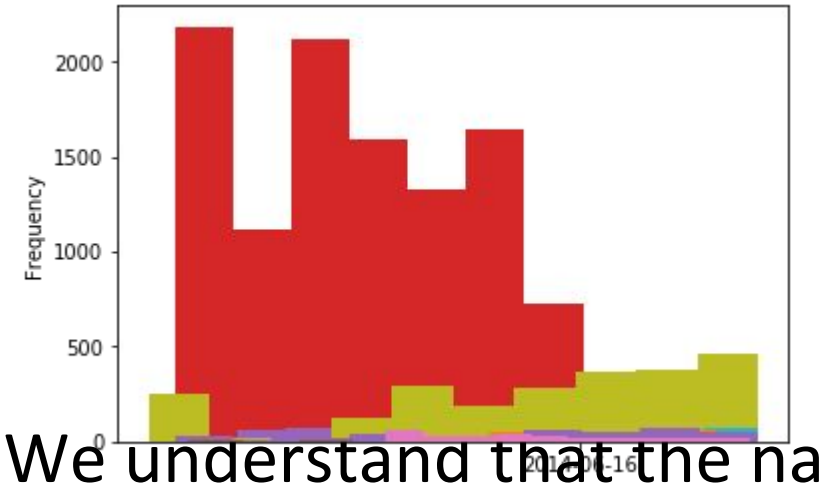
It is an extremely interesting question to ask that how does it all relating to the performence in stock market.

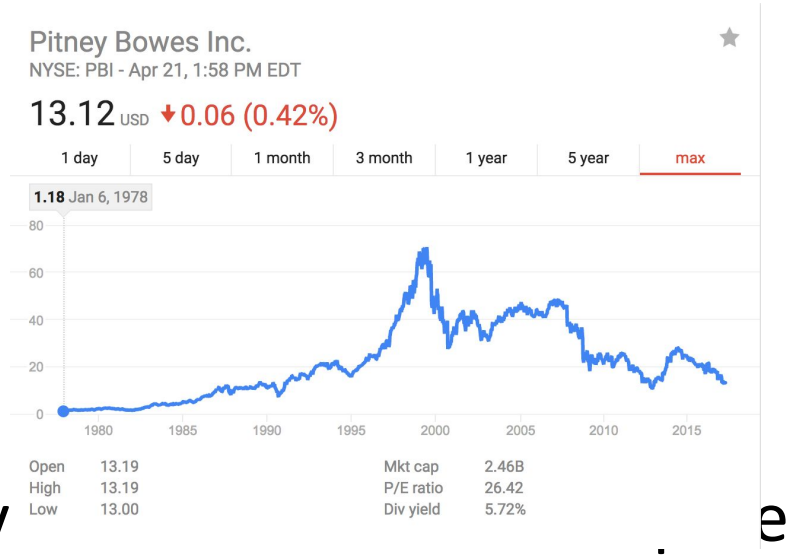
A bit of insight

An example that jobs correlates to a company's performance in stock market: Customer Service.

All technology company hires customer service people. However, it is not very profitable and is an indication of poor product quality. For example, you probably never call Google's customer service simply because you don't need.

Pitney Bowes Inc., for instance, has big customer service hiring across years. Showing in the red. However, its stock price tanks since the last financial crisis.





We understand that the nature of Pitney [redacted] es more customer services. But that does undermines the usefulness of this feature.

What we want to predict

Our goal is not to predict the stock price. On the contrary, we are predicting how a company is performing against other company.

Therefore, our predictive outcome is going to be whether or not a company is within the top 25% among other company in terms of the return over time period we are interested in.

The return is defined by the rate stock price in year t increase from the stock price in year t-1. We do realize the stock market is volatile. So we are interested in relatively long term return in terms of year.

We first select company that has continues job posting history more than 5 years so that we have enough data to use for each of them. This narrows the number of company to 70.

Furthermore, we realize that company in different field tend to have totally different employment structure. Thus, we select the industry of technology within 70 firms. This bring us to the final 17 companies we want to study.

Diebold Corp	Sanmina SCI	WebMD	Blackbaud
Mantech International Corporation	Pitney Bowes	EFI	OSI Systems
NextGen Healthcare	Acxiom Corporation	CommVault	RealPage
ACI Worldwide	ShoreTel Inc.	Sykes Enterprises Incorporated	Nortek
Engility Corporation			

Feature Engineering

One key to solve our challenge is to produce powerful feature.

We believe that job posting record tells a lot about how a company is performing. Therefore, all features we are generating is directly or indirectly for the job record. (Within in mind that, all companies are already in one industry)

Approach

1. We group job titles into 10 major categories:

   Service, Medical, Traffic, Business, Finance, Assistance, Engineering

   Labor, Law, Other
2. On top of that, we computed the ratio of each major job category within each year by each company as feature. We also use the total number of jobs posted within each time period by each company.

   The intuition is we are not only interested in how many job a company posting, but also what kind of job it is posting. By factoring in this way, we can separate the impact of these two factors.

Approach

3. As we talked in previous slide, we are interested in how a company performing against its peers. Therefore, we average across the whole industry and input that as additional feature into our model.

Intuitively, this feature tells the model what's other companies' status so that it will learn to compare against each other.

Model

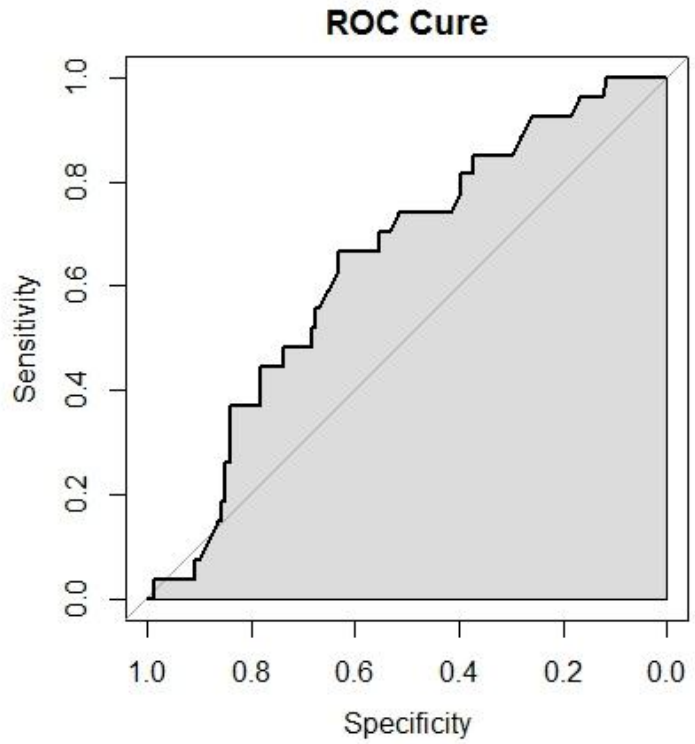
We build a simple random forest model to predict a whether company is a rockstar in financial market or not across different year.

This simple model works surprisingly well. We will report our finding in the following slides.

Results

We run different machine learning algorithms over our dataset including: logistic regression, SVM, and random forest. The results of ROC (AUC score) are shown below:

Model	AUC score
Logistic	0.56
SVM	0.59
Random Forest	0.66



Note that we are selecting company 1 of 4. This result is surprisingly good.

Conclusion

From our experiment, we show that it’s able to generate predictive features from the job record. This is especially noteworthy considering we intentionally avoid using any provided economical and geological information. This new feature could be useful for stock selection in general.