

Who is going to be the next Airbnb

...

Using Crunchbase Data to Predict Startup Outcome

Classes



Raised Money
\$18.2M



Acquired\$
90M



IPO
\$33B



Dead
\$118.5M

Data

Person Data

Company Data

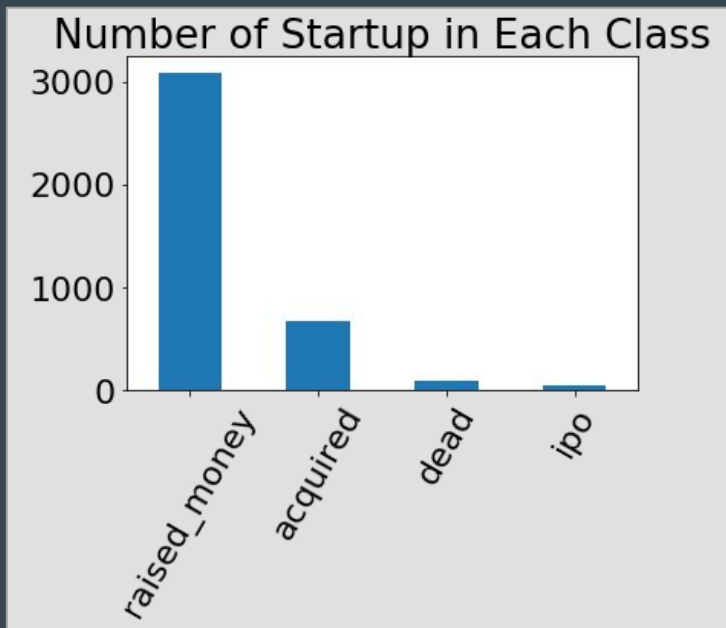


Company Data



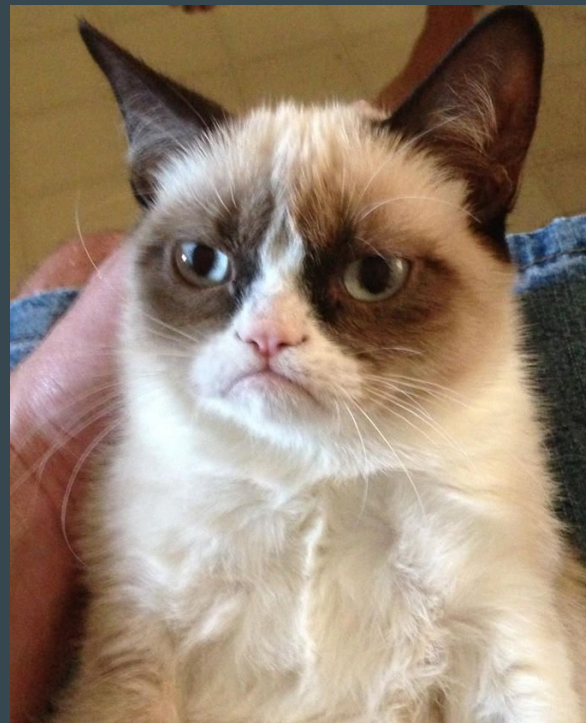
Person Data

Imbalanced Data



106,802 companies -> 10232 companies (clean data)

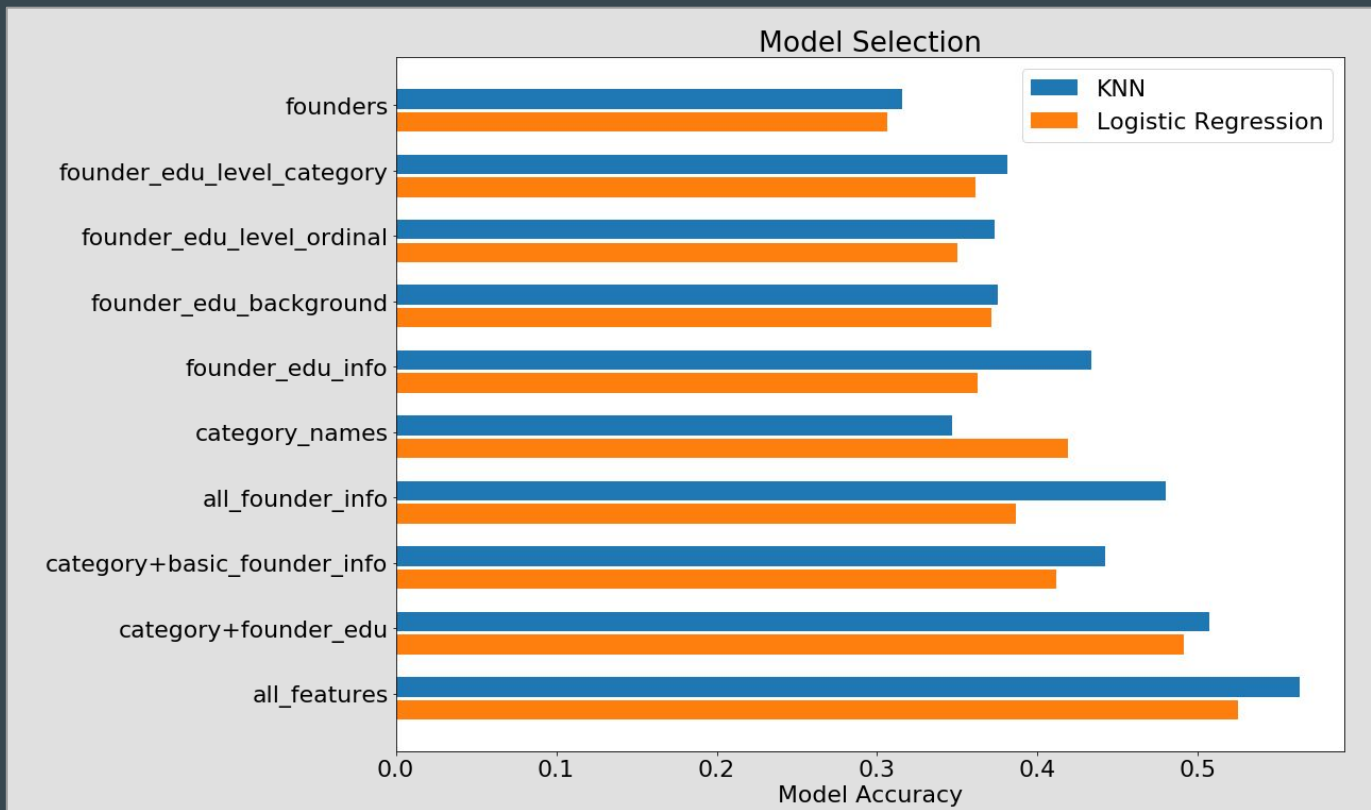
-> 2112 training ; 315 testing (balanced data)



Features

- Company Info
 - Categories (category)
 - Number of founders (cardinal)
 - Number of female founders (cardinal)
 - Number of male founder (cardinal)
- Founder Info
 - Nth startup (cardinal)
 - Founder Education Degrees (category and ordinal)
 - Founder Education Background (category)

Model Selection



Test Accuracy: 51.29%

	precision	recall	f1-score	support
acquired	0.62	0.25	0.35	137
dead	0.07	0.67	0.12	9
ipo	0.15	0.60	0.24	10
raised_money	0.51	0.41	0.45	128
avg / total	0.54	0.35	0.39	284

- Acquired:
 - + Public Relations, Consulting, or Advertising (~1)
 - - Legal and Education(~ -2.5)
- IPO:
 - + Network Hosting, Security, Biotech (~2)
 - - Public Relations (~ -10)
- Deadpooled:
 - - Security, Public Relations (~ -4)
 - + Legal, Female Founders (~ 0.5)

Tree?

Feature Importances:

founder Education Degrees

#founders_with_liberal_arts_background

category_web

category_network_hosting

That's it

Two more presentation then we're done!!

Appendix

