Startup and get going: by Michael Burak

Identifying factors involved in getting your startup acquired.

Overview

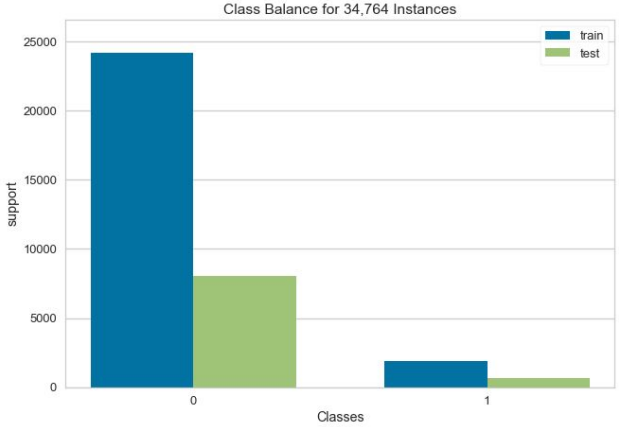
-Startups: Success, failure, and acquisition.

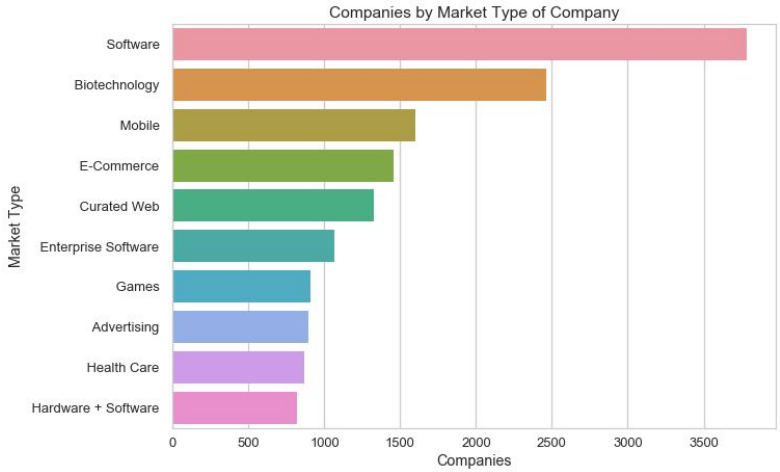
-The data: financial(ex: venture capital funding), regional(ex: what city the company is in), type of company(ex: a Software company), time based data(ex: when the company was founded), status of the company(the target of analysis, acquired/open/closed.)

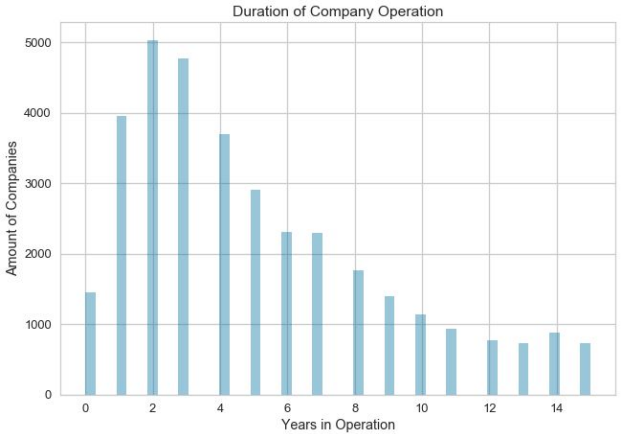
-The problem to model: building an approach to classifying what makes for an acquired startup by using tree-based modeling

The Data:

-Imbalance favoring 0(not acquired)
-Dominated by select markets
-Mostly newer startups



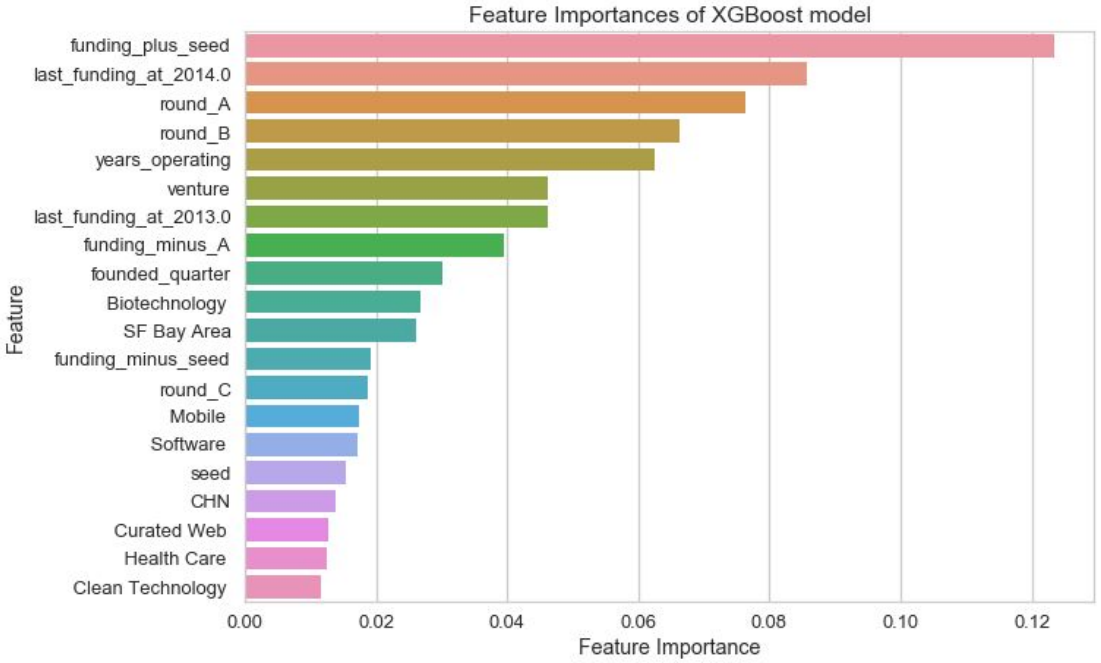




The Modeling: Looking for answers in the trees

Final Model: *F1 weighted score is 0.861*
F1 micro score is 0.828
F1 macro score is 0.628

Conclusions



Contribution[?]	Feature	Value
+0.536	funding_plus_seed	0.000
+0.344	years_operating	0.267
+0.264	round_A	0.000
+0.248	SF Bay Area	0.000
+0.230	Education	1.000
+0.158	seed	0.000
+0.117	round_B	0.000
+0.114	founded_quarter	0.000
+0.092	venture	0.000
+0.048	funding_rounds	0.000
+0.042	Curated Web	0.000
+0.041	first_funding_at_2005.0	0.000
+0.038	last_funding_at_2012.0	1.000
+0.034	round_C	0.000
+0.027	New York City	0.000
+0.022	per_round	0.000
+0.022	last_funding_at_2011.0	0.000
+0.021	debt_financing	0.000
+0.015	last_funding_at_2006.0	0.000
+0.010	angel	0.000
+0.008	first_funding_at_2007.0	0.000
+0.008	Boston	0.000
+0.004	Mobile	0.000
+0.003	round_D	0.000
+0.003	Seattle	0.000
+0.003	CAN	0.000
+0.001	Social Media	0.000
+0.000	Finance	0.000
+0.000	Games	0.000
-0.004	FRA	0.000
-0.006	GBR	0.000
-0.006	Manufacturing	0.000
-0.008	Health and Wellness	0.000
-0.009	Hardware + Software	0.000
-0.012	Health Care	0.000
-0.012	Clean Technology	0.000
-0.013	CHN	0.000
-0.029	Software	0.000
-0.040	Biotechnology	0.000
-0.098	<BIAS>	1.000
-0.138	last_funding_at_2013.0	0.000
-0.311	last_funding_at_2014.0	0.000

Future Work

-PCA for faster performance

-Genetic algorithms and Bayesian techniques for tuning models

-Clustering to create additional data columns

-More interpretability

Thank you!