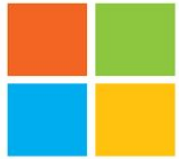
Classifying Big Tech Employee Reviews

Grace Miles ‘19, Amy Liu ‘19,
Madeleine Cheyette ‘19

Motivation & Related Work



What do Glassdoor reviews about popular big tech companies reveal?

Will reviews and ratings alone be sufficient to identify the company in question?

Scholar Research: Employee Satisfaction Correlations with Company Success & Financial Risk (Luo et al. [2016] Ji et al. [2017])

Kaggle users: numeric plots, word clouds, sentiment analysis

Data

- Current and former employee reviews on Glassdoor from Amazon, Apple, Facebook, Google, Microsoft and Netflix.
- 67,529 reviews with 17 features

	Total Reviews	Overall-Rating	Percent Current Employee
Amazon	26,430	3.58	66%
Apple	12,950	3.96	56%
Facebook	1,590	4.51	80%
Google	7,819	4.34	60%
Microsoft	17,930	3.82	63%
Netflix	810	3.41	50%

Top 10 most common company words across the text review columns by company

Amazon	Apple	Facebook	Google	Microsoft	Netflix
work	work	work	work	work	work
manag	great	peopl	great	compani	peopl
good	compani	compani	compani	good	manag
compani	peopl	great	peopl	great	compani
peopl	manag	cultur	good	peopl	great
great	benefit	manag	manag	manag	get
get	good	place	place	benefit	cultur
time	get	lot	get	lot	good
lot	retail	get	lot	get	time
hour	time	good	benefit	team	job

Methods

- Preparing the data:
  - Bag of Words
  - TFIDF
- Attempted dimension reduction
  - SelectKBest (chi2)
  - PCA
- Models
  - Baseline: MNB
  - KNN with PCA
  - Logistic Regression
  - LinearSVC

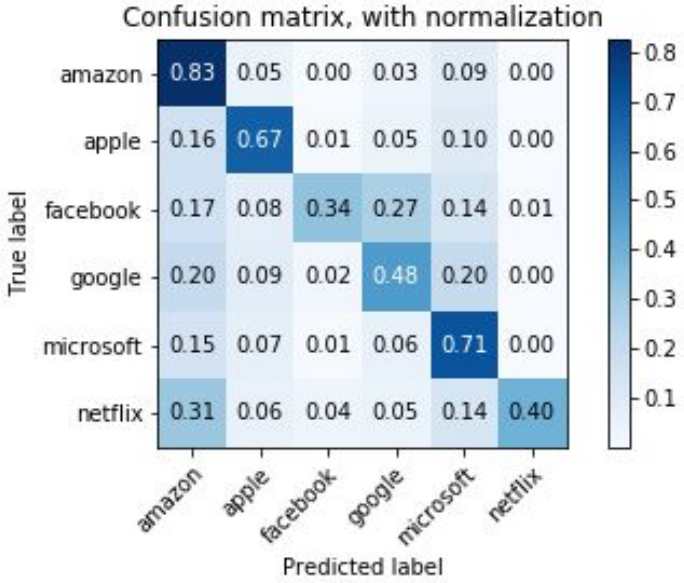
Results

- LinearSVC performed the best once the ratings data was taken into account.
- The separated vocabulary for the three text columns (summary/pros/cons) and Bag of Words representation performed the best.

	Full	**Current**	**Former**	**Vocabs**
MNB	0.68	0.68	0.68	0.70
Logistic Regression	0.71	0.72	0.71	0.72
LinearSVC	0.69	0.70	0.69	0.71

LinearSVC Deep Dive

	Precision	Recall	F1-Score	Support
Amazon	0.75	0.83	0.79	6,522
Apple	0.72	0.67	0.69	3,237
Facebook	0.49	0.34	0.40	390
Google	0.56	0.48	0.52	2,003
Microsoft	0.70	0.71	0.71	4,502
Netflix	0.73	0.40	0.52	229



Discussion

Table 4: Most Important Features by Company
(prefixes indicate words found in review subcategories: summary, pros, and cons)

Amazon	Apple	Facebook	Google	Microsoft	Netflix
cons_shred	sum_detect	sum_concentrix	cons_philanthropi	cons_redmond	cons_overdr
cons_truck	sum_mould	cons_vamp	pros_mtv	cons_nokia	cons_dvd
sum_epitom	sum_deed	cons_detect	cons_correspond	cons_intrigu	pros_reed
pros_commerc	sum_marcom	cons_idiosyncrasi	cons_mastermind	pros_strip	cons_gato
sum_wick	cons_cupertino	cons_php	sum_adword	cons_fanat	sum_nazi
pros_wiki	pros_sap	pros_fb	sum_gsx	cons_blatantli	pros_utah
cons_frugal	cons_represent	cons_symptomat	sum_quirki	cons_plethora	cons_reed
cons_pager	pros_laser	pros_coo	pros_mk	sum_bing	sum_tsr
cons_bezo	sum_nearbi	cons_meme	pros_larri	pros_inquisit	sum_vacat
pros_dfw	pros_irrat	sum_fb	cons_65k	pros_premera	cons_readili
cons_24x7x365	pros_cupertino	cons_murder	sum_pod	sum_365	pros_agent
pros_department	cons_tim	sum_obviou	sum_evalu	cons_lion	pros_deck
cons_fabric	cons_cio	pros_intersect	pros_spare	sum_suprem	sum_websit
cons_nanni	sum_patent	pros_hustl	cons_unsupport	cons_scorecard	sum_protect
pros_downturn	pros_propag	cons_mpk	cons_funni	sum_gtsc	sum_amsterdam

Discussion

- Classification exploits latent identifiers like CEO names, company buzzwords, and industry references
- High frequency of generic and dispassionate language poses challenge for multiclass classification problem
- Popular company reputations do seem to have basis in employee experience

Future Work

- More extensive deidentification of reviews
- Filtering based on position within company
- Alternative methods for combination of text and numerical ratings data

Conclusion

- Glassdoor reviews provide pertinent and novel insight into big tech company culture that can be evaluated with machine learning
- In implementing classification, we use BOW to represent data, attempt feature reduction through PCA/SelectKBest, and find that the best classification model was surprisingly Logistic Regression
- Through classification, we see the way company buzzwords and reputation influence employee experience, but broadly reviews were similar across all big tech companies
- Suggests that the corporate world may not be as divided as we would otherwise think!