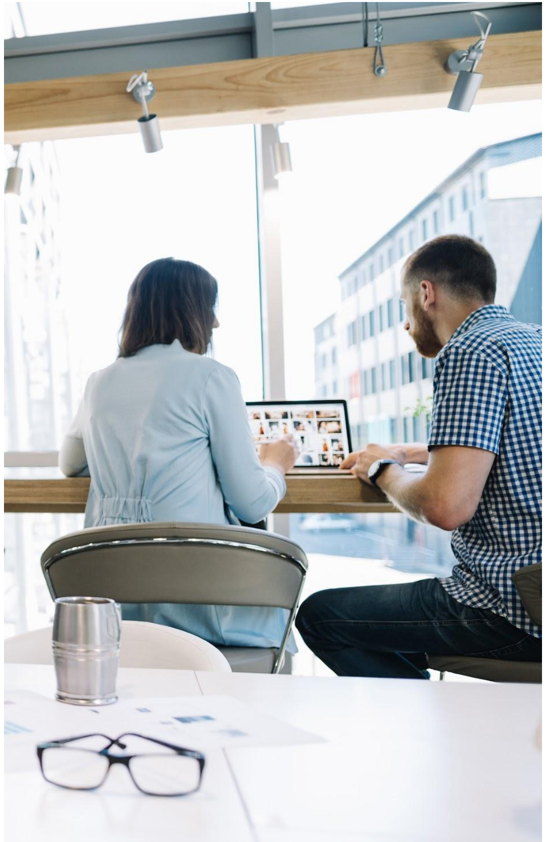


Employee Review Monitoring

- Marvin Fong

Introduction

Commentary

Commentary: This Great Resignation Wave is painful and frustrating for employers

The string of resignations are straining employers who have had to keep firms afloat, NUS Business School's Wu Pei Chuan says.

https://www.channelnewsasia.com/commentary/great-resignation-wave-quit-find-job-employer-boss-pay-mental-health-2386761

HOW WE WORK

The Great Resignation: How employers drove workers to quit

https://www.bbc.com/worklife/article/20210629-the-great-resignation-how-employers-drove-workers-to-quit

Quitting is just half the story: the truth behind the 'Great Resignation'

https://www.theguardian.com/business/2022/jan/04/great-resignation-quitting-us-unemployment-economy

Problem Statement

Being a staff in our firm's data analytics and insights team, we were notified by our HR department that our employee retention rate has reached an all-time low this year and that the HR department do not have enough manpower to read through quarterly employee feedbacks made by all 10,000 employees in the firm.

Our team has been tasked with automating the process to identify positive and negative reviews made by staffs. Giving HR a brief overview on the ratio of positive to negative reviews and within these reviews, the firm is able to know what positive aspects had the firm done well and what pain points which are affecting retention rate. In doing so, department managers and HR could quickly gain actionable insights.

Objective

Through this project, we hoped to save man-hours of HR staffs or managers that had to read through tens of thousands of lengthy reviews and manually decipher what actions needs to be taken as soon as possible.

Losing talents within the organization is very costly

Contents

01
Data Acquisition & EDA

Data Acquisition

- Kaggle dataset: Employee review data from Glassdo

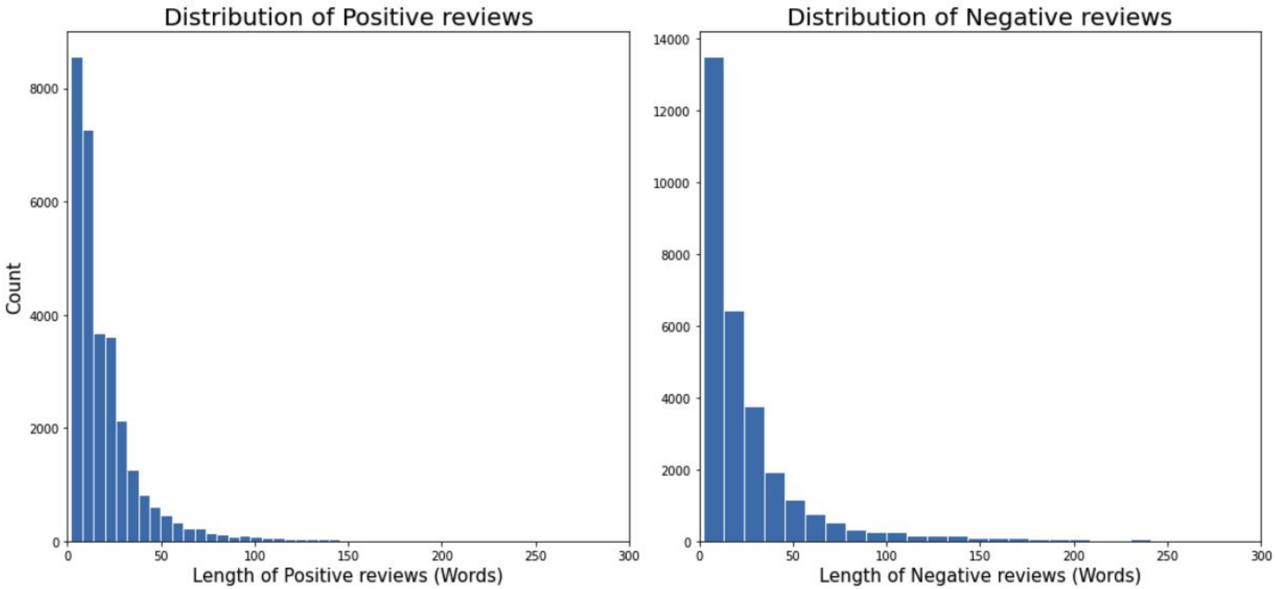
- 30,300 observations & 17 features

- Notable columns: ‘Place’, ‘summary’, ‘positives’,
  ‘negatives’,
  ‘advice to management’,
  ‘overall score’

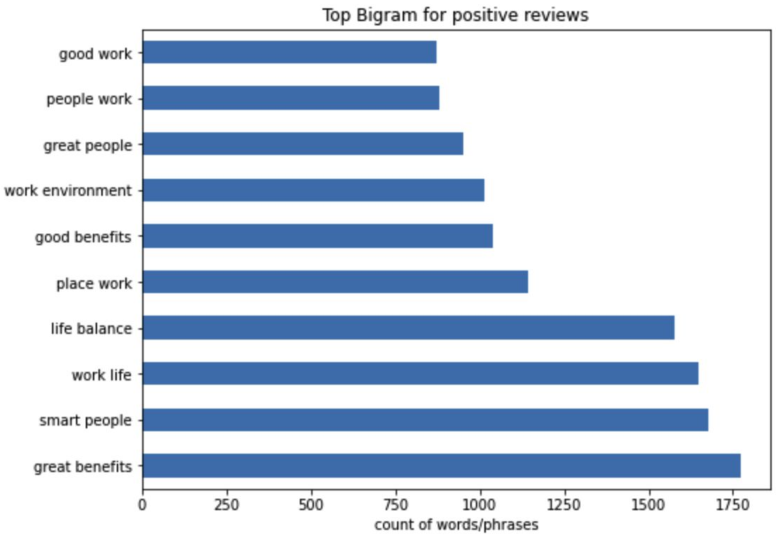
glassdoor

Place	location	date	status	job_title	summary	positives	negatives	advice_to_mgmt	score_1	score_2	score_3	score_4	score_5	score_6	overall
rtup_1	NaN	Dec 11, 2018	Current Employee	Anonymous Employee	Best Company to work for	People are smart and friendly	Bureaucracy is slowing things down	NaN	4.0	5.0	5.0	4.0	5.0	0	5.0

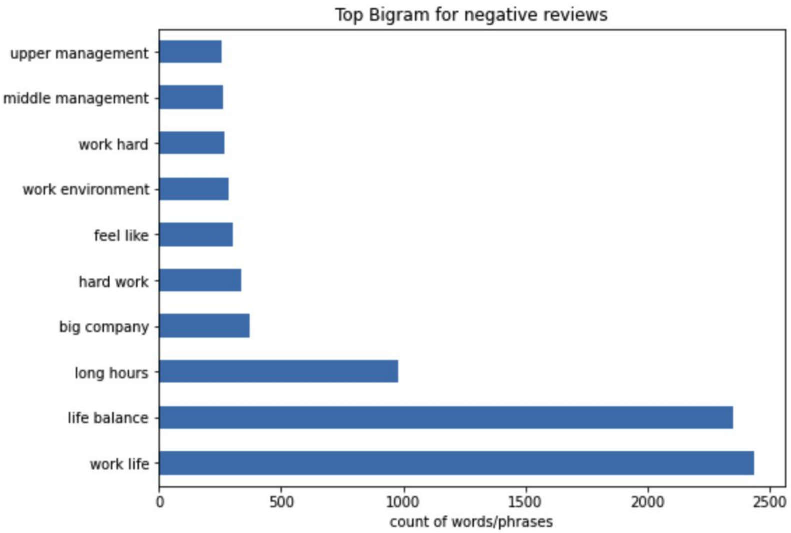
EDA


Distribution of Positive reviews
Count
8000
6000
4000
2000
0
0
50
100
150
200
250
300
Length of Positive reviews (Words)
Distribution of Negative reviews
14000
12000
10000
8000
6000
4000
2000
0
0
50
100
150
200
250
300
Length of Negative reviews (Words)

EDA

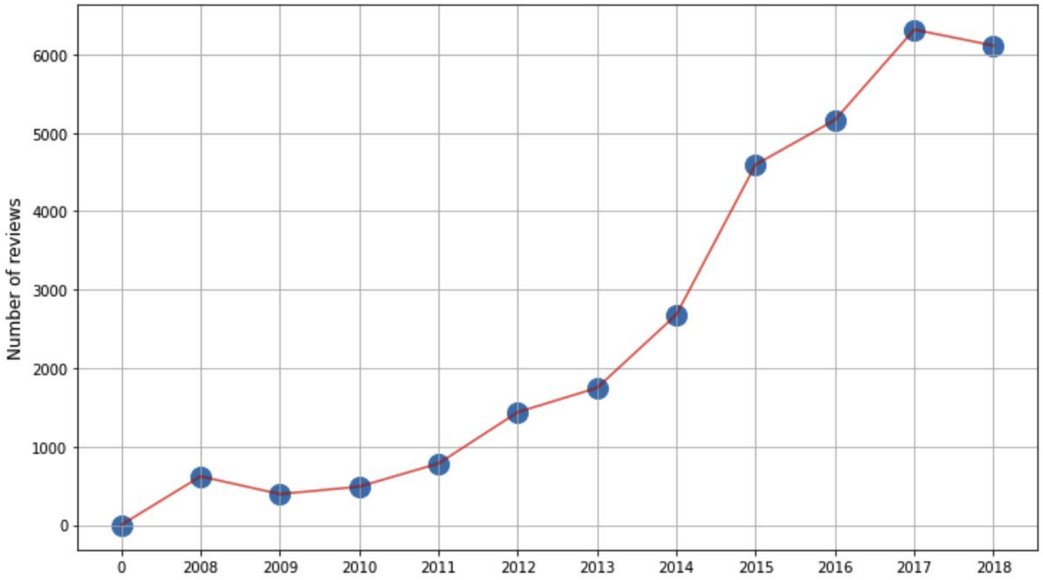


Top Bigram for positive reviews
good work
people work
great people
work environment
good benefits
place work
life balance
work life
smart people
great benefits
0
250
500
750
1000
1250
1500
1750
count of words/phrases



Top Bigram for negative reviews
upper management
middle management
work hard
work environment
feel like
hard work
big company
long hours
life balance
work life
0
500
1000
1500
2000
2500
count of words/phrases

EDA

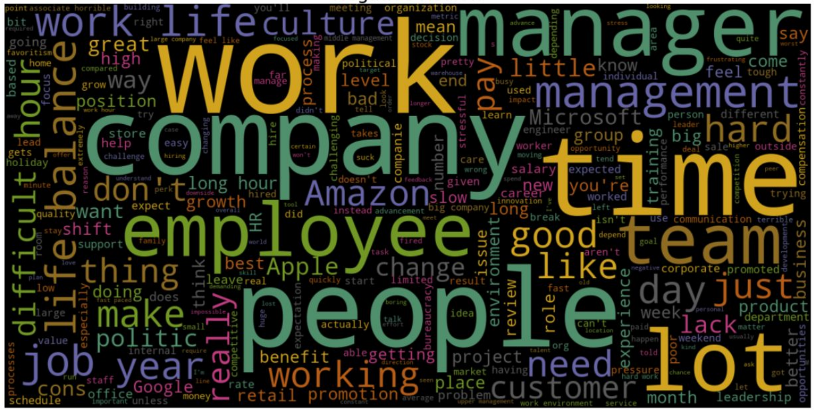


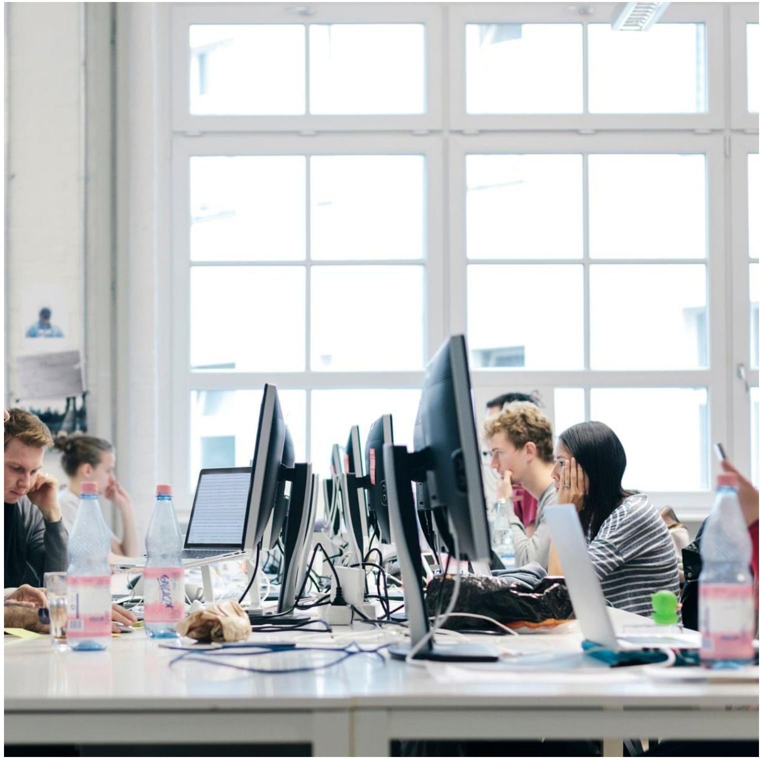
Number of reviews
6000
5000
4000
3000
2000
1000
0
0
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018

EDA

Positive Review

Negative Review





02 Preprocessing & Sentiment Analysis

Preprocessing

- Columns related to reviews:
  ‘ summary’,
  ‘positives’,
  ‘negatives’
  ‘advice_to_mgmt’

- Dropped all Glassdoor review scores as we will not be relying on them for sentiment analysis

- Combined the 4 columns related to reviews

- Fill NaN values with blanks

summary	positives	negatives	advice_to_mgmt
Best Company to work for	People are smart and friendly	Bureaucracy is slowing things down	NaN
Moving at the speed of light, burn out is inev...	1) Food, food, food. 15+ cafes on main campus ...	1) Work/life balance. What balance? All those ...	1) Don't dismiss emotional intelligence and ad...

Preprocessing

- Removed all non-english reviews using spaCy Language Detector.

- This is done because sentiment analysis technique might give inaccurate results for non-english reviews. Furthermore, we are only interested in classifying english reviews.

Preprocessing

Example of reviews classified correctly as non-english:

3461	Apple Goede verdiende baan. Leuke mensen. Goed nagedacht over hoe mensen goed kunnen werken alleen om met een groep. Moeilijk een baan te krijgen. Je moet al een goeie opleiding of carriere hebben .	nl
3462	Très bonne entreprise Bonne ambiance bonne collaboration J'ai beaucoup apprécié travailler dans cette entreprise nb d'heure de travial C'était dur de gérer vie privée et vie professionnelle.	fr
3463	ok geht so aber es kann besser sein zu wenig Essen Gehalt nich hoch genug	de

Example of reviews classified incorrectly:

6108	SDE pretty great company, come and join us no no no no no no	pt
6138	great jobs excellent environment awesome people great environment none none none none none great place	fr

Sentiment Analysis

TextBlob :

- Sentiment function of TextBlob returns two properties - ‘Polarity’ and ‘Subjectivity’

- Polarity is a float which lies in the range of [-1, 1] where 1 means positive statement and -1 means negative statement.

- Subjectivity is also a float with range [0, 1] and sentences refer to personal opinion, emotion / judgment.

Sentiment Analysis

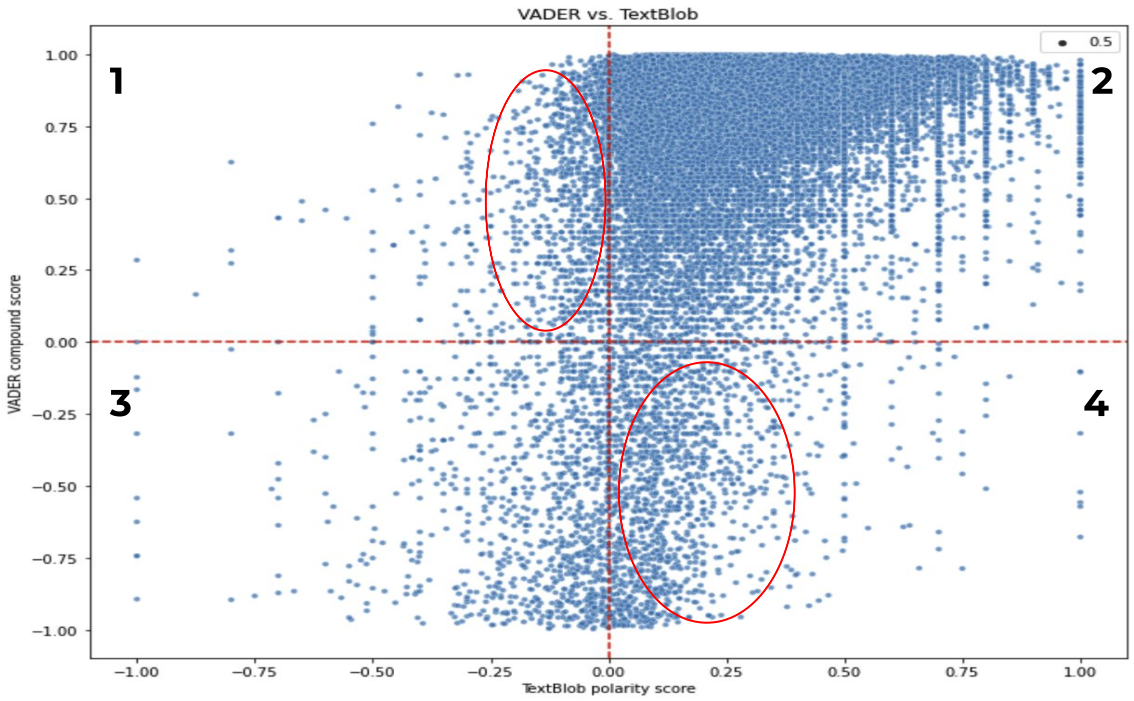
VADER Lexicon

- Study shows that VADER performs as good as individual performs as good as individual human raters at matching ground truth.

- The reason behind this is that VADER is sensitive to both Polarity and Intensity (how positive or negative is the sentiment) of emotions.

- The **compound score** is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). This is the most useful metric if you want a single unidimensional measure of sentiment for a given sentence.

TextBlob & VADER Lexicon results comparison



03 Modeling & Model Evaluation

Machine Learning Models

Steps :

- Preprocessing
    - Changing ‘Neutral’ class to ‘Negative’
    - Text normalization via Lemmatization
    - Adding frequently appearing words in N-grams and WordCloud into existing Sklearn stopwords package
    - Train-validation-split
    - SMOTE using imblearn pipeline (1 : 90%, 0 : 10%)

Machine Learning Models

Steps :

- Modeling :
  - Multinomial Naive Bayes, Logistic Regression, SVM
  - With CountVectorizer and TfidfVectorizer

- Evaluation metrics :
  - ***Specificity** (accurately predict negative reviews to give insights into organizational problems)
  - Precision (correctly classifying positive reviews out of all actual positive reviews)
  - Accuracy

*optimized

Deep Learning Models

Steps :

- Preprocessing
    - Changing ‘Neutral’ class to ‘Negative’
    - Text cleaning (remove punctuations & convert to lowercase)
    - Upsample minority class (1 : 90%, 0 : 10%)
    - Train-validation-split

Deep Learning Models

Steps :

- Modeling :
    - RNN model using Keras Tensorflow
    - Added embedding layer to create vector representation
    - Added SpatialDropout1D to spread out weights (i.e. regularization)
    - Added LSTM, 1x hidden layer and output layer

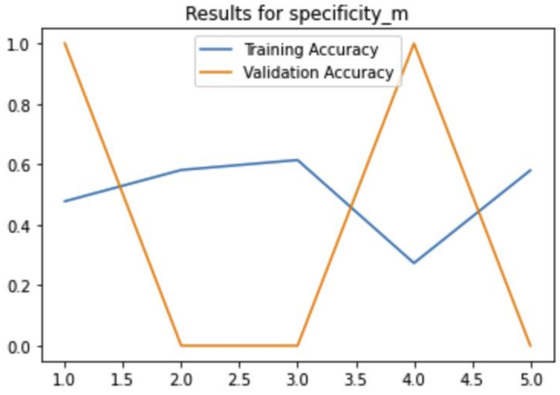
- Evaluation metrics :
    - ***Specificity** (accurately predict negative reviews to give insights into organizational problems)
    - Precision (correctly classifying positive reviews out of all actual positive reviews)
    - Accuracy

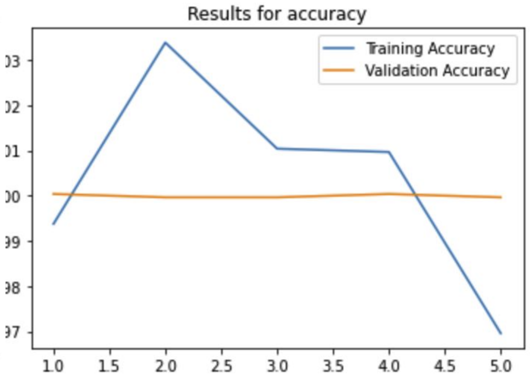
*optimized

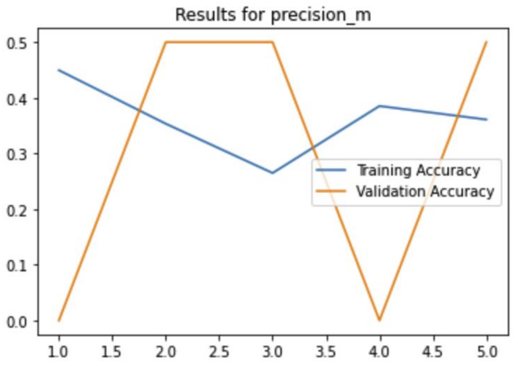
ML Model Evaluation

	Best CV (Specificity)	Precision	Accuracy	Cohen Kappa
SVC - TVEC	1.0	1.0	0.098	0.001
LR - TVEC	**0.754**	**0.974**	**0.909**	**0.560**
MNB - TVEC	0.608	0.957	0.845	0.349
SVC - CVEC	0.587	0.957	0.804	0.285
LR - CVEC	0.573	0.958	0.908	0.499
MNB - CVEC	0.502	0.946	0.871	0.349

DL Model Evaluation







04 Topic Modeling (LDA)

Latent Dirichlet Allocation

LDA

- Topic modeling is a method for unsupervised classification of documents, which finds some natural groups of items (topics) even when we’re not sure what we’re looking for

- Here, we want to specifically look at all negative reviews to turn them into quick and accessible insights for management / HR and find out what aspects should the organization be focusing on improving

- Best parameters: n_components = 5
  learning_decay = 0.7
  batch_size = 64

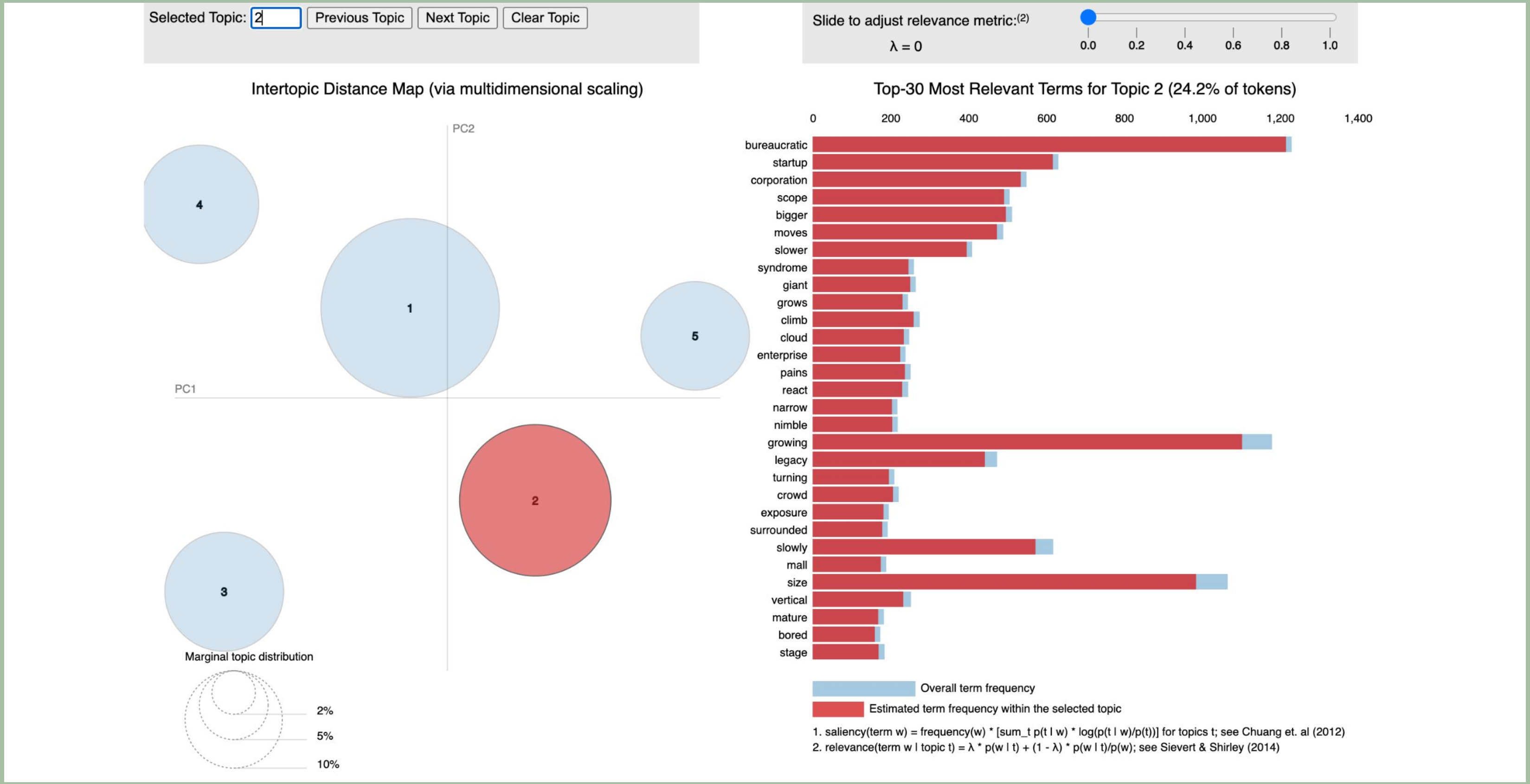
Latent Dirichlet Allocation

Package used to visualize LDA results:

- pyLDAVis is a great tool to interpret individual topics and the relationships between the topics.

- On the left-hand side of the visualization, each topic is represented by a bubble. The larger the bubble, the more prevalent is that topic. The distance between two bubbles represents the topic similarity.

- The right-hand side shows the top-30 most relevant terms for the topic you select on the left. The blue bar represents the overall term frequency, and the red bar indicates the estimated term frequency within the selected topic. So, if you see a bar with both red and blue, it means the term also appears at other topics.


Selected Topic: 2
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 0
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
4
1
5
2
3
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 2 (24.2% of tokens)
0
200
400
600
800
1,000
1,200
1,400
bureaucratic
startup
corporation
scope
bigger
moves
slower
syndrome
giant
grows
climb
cloud
enterprise
pains
react
narrow
nimble
growing
legacy
turning
crowd
exposure
surrounded
slowly
mall
size
vertical
mature
bored
stage
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

05 Future Improvements

Recommendations for further improvement

Areas for improvement:

- Try other estimators like KNN, XGBoost and RandomForest as well as other text normalization techniques such as Snowball or Lancaster Stemming.
- Further fine-tuning of RNN-LSTM model to improve training and validation scores.
- Try other deep learning models such as HuggingFace-BERT model using PyTorch.
- Using other pre-trained models from HuggingFace to predict sentiments of non-english reviews such as German and French.
- Deployment of models into an app/interface that could upload quarterly reviews in csv format and gives an output to visualize the results for stakeholders.

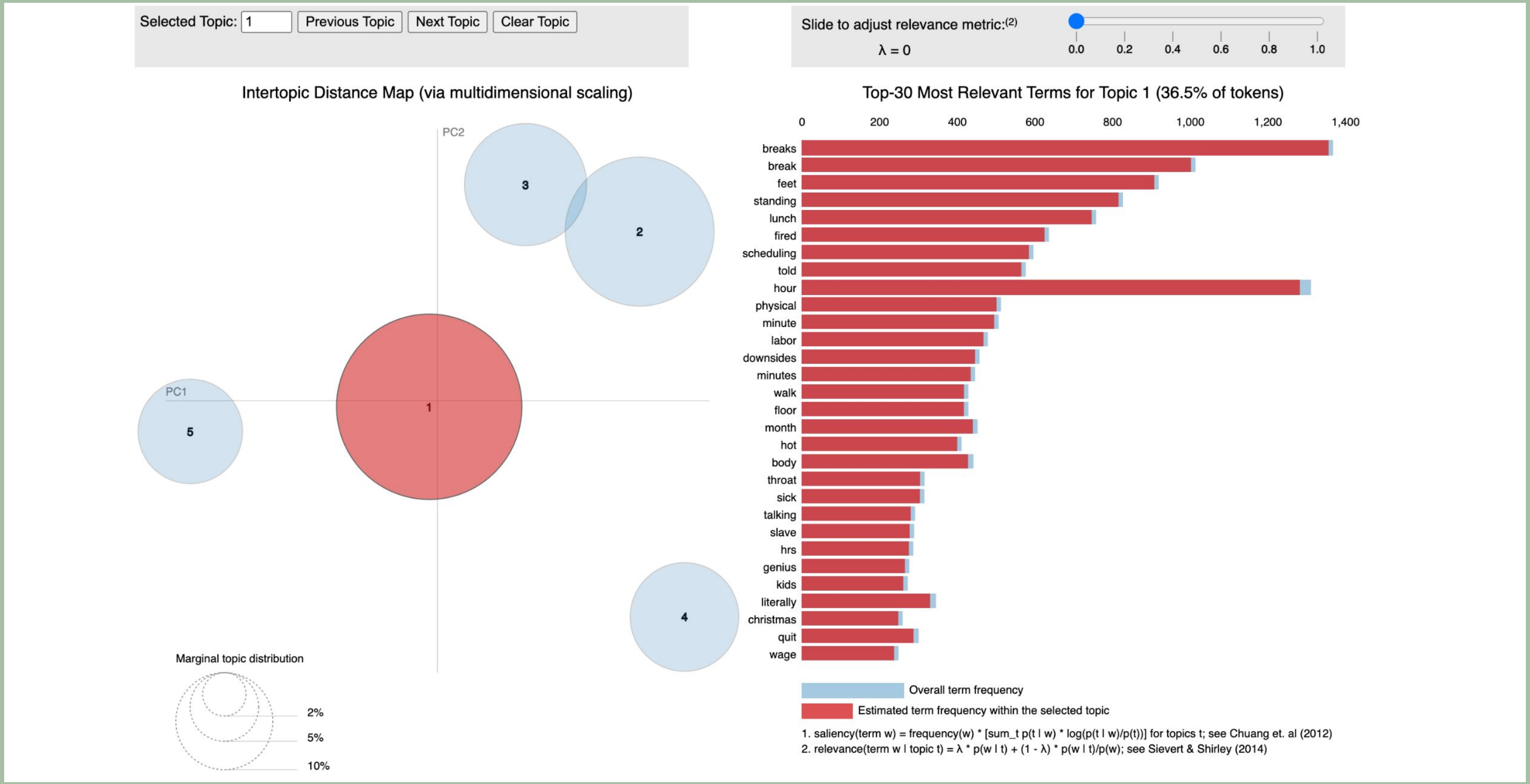
06
Future Deployment / Test Set

Predictions of negative reviews with test set

Big Tech Never have to skip a paycheck. Google ha an over abundance of technician and under-abundance of thinkers.	0
Engineer salary wa fine but in line with other company work wa boring. I disliked it.	0
pocket of incompetence Overall, Google pay well and shower people with perks. There are pocket of excellence. The company is extremely profitable. There are deep, deep pocket of dysfunction and incompetence at Google. Do your due diligence on the team and manager you're considering joining. There are bozo at every company, and Google is no exception to this rule. Learn to strategy.	0

Predictions of positive reviews with test set

great place to grow! Great health benefits. Many internal job opportunities.Very smart people. You work on the world's biggest problems. Red Tape. Tough to negotiate internal politics. Big company so hard to reach the top.	1
An ocean of opportunity diverse set of people and problem they solve. a never-give-up attitude. sometimes feel like a behemoth trying to get nimble. the inter team rivalry tends to get unhealthy align workgroups and incentive better to create cohesion rather than friction internally	1
Tech Gaint Equip it employee wid huge salary :D High qualification required to join it Good Going....try searching new and bigger talent apart 4rm jst considering academic	1



Selected Topic: 1
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 0
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 1 (36.5% of tokens)
0
200
400
600
800
1,000
1,200
1,400
breaks
break
feet
standing
lunch
fired
scheduling
told
hour
physical
minute
labor
downsides
minutes
walk
floor
month
hot
body
throat
sick
talking
slave
hrs
genius
kids
literally
christmas
quit
wage
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Thanks!

Do you have any questions?