Natural Language Processing Models of INDEED Employer reviews

RECOMMENDATIONS FOR EMPLOYERS

FLATIRON SCHOOL

Phase 4 Final Project

Authors: Elena Kazakova

Cohort: DS02222021

Instructor: James Irving

Outline of the presentation

- Summary
- Business Problem
- Data
- Final Model
- Results
- Conclusions

Summary

Analysis of Employer Rated Reviews

- **INDEED.com job search site employer reviews**
- **Boulder most popular employers**

The Resulting Models Provide Insight into

- What **factors** affect employees' **satisfaction with a company**
- **How to spot a negative review** and react quickly

Business Problem

- Analysis of answers to open-ended questions:
  - **Most informative**
  - **Most difficult to analyze**
- The NLP (Natural Language Processing) approach is a logical answer to this problem
- The goals of this project:
  - **Develop predictive classification model of employee reviews**
  - **Gain insight into factors affecting satisfaction/dissatisfaction of employees**

Data

INDEED job search site:

121620 records

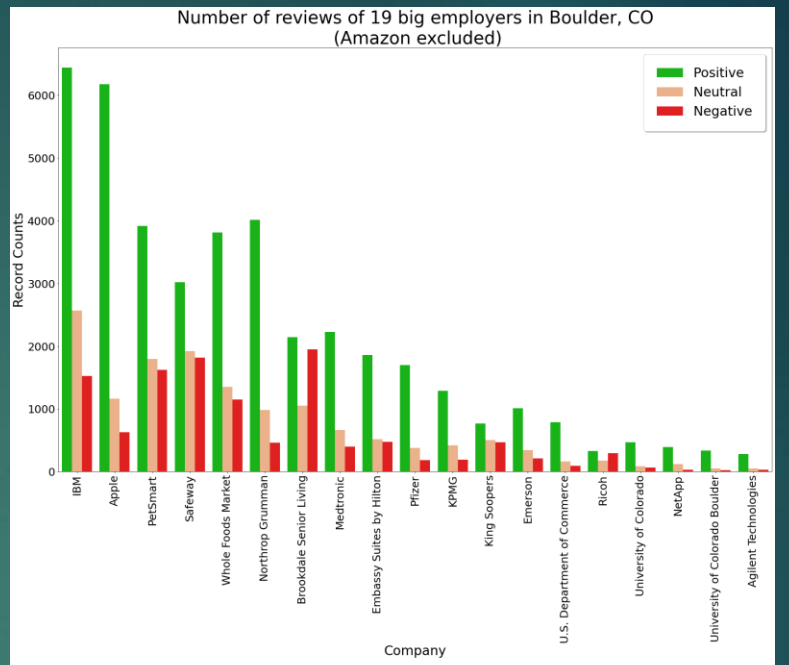
46 Boulder area employers

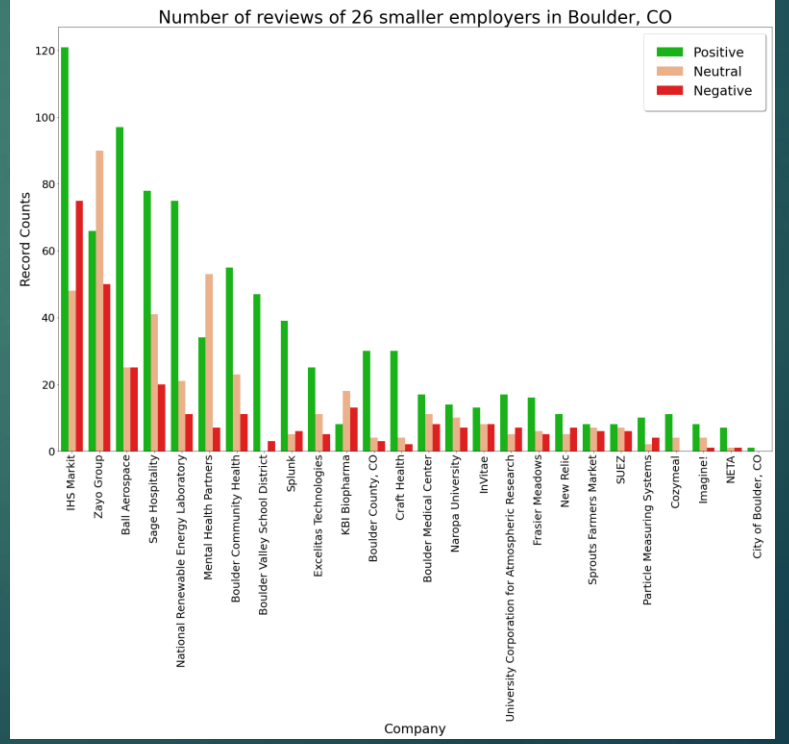
Positive, Negative and Neutral reviews

Employers list includes but is not limited to:

- IBM
- Apple
- PetSmart
- Safeway
- Whole Foods Market
- King Sooper
- Brookdale Senior Living
- Medtronic
- Pfizer
- University of Colorado Boulder

(Full list of employers is available in Appendix)

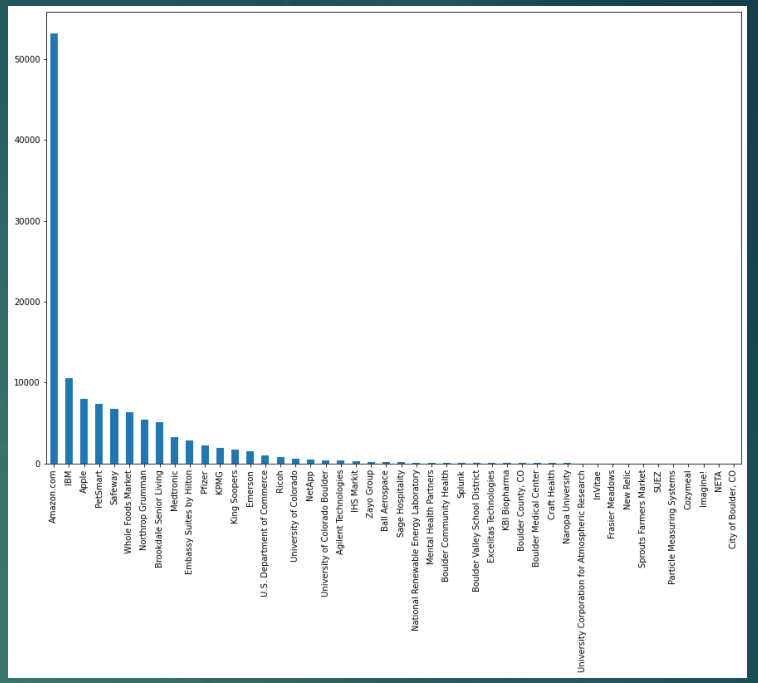


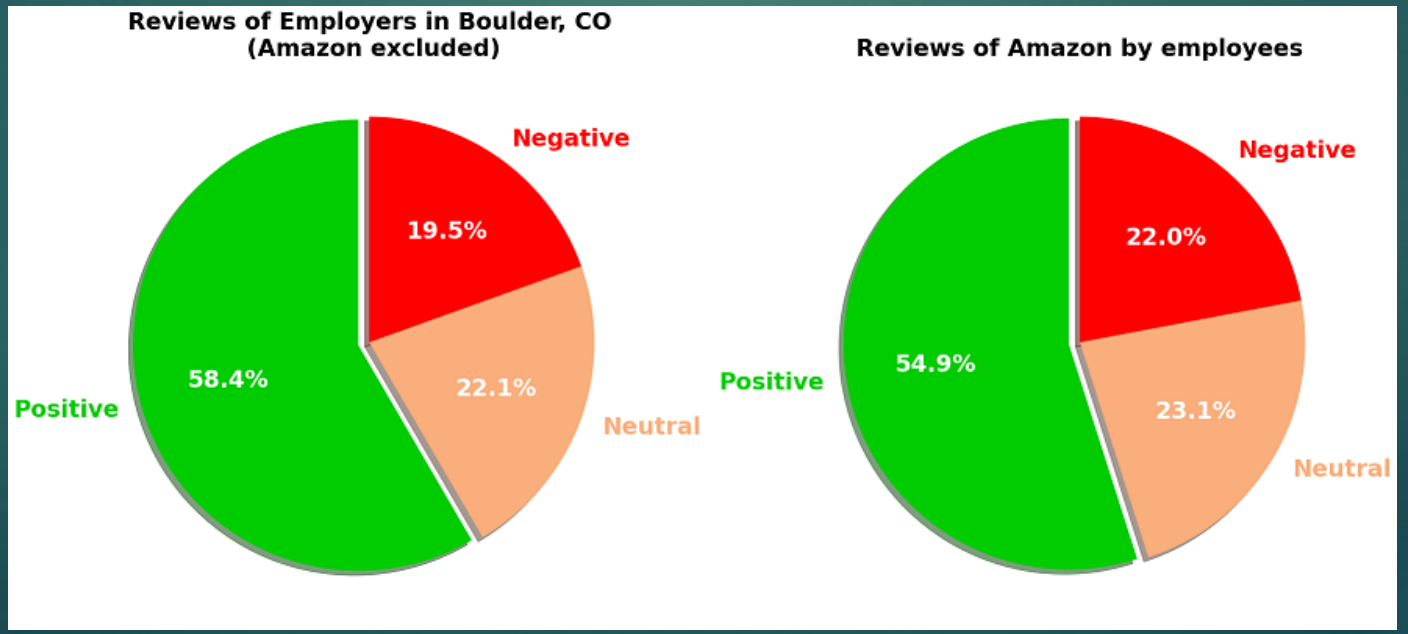


Data (continued)

Amazon reviews: 44% of all reviews used

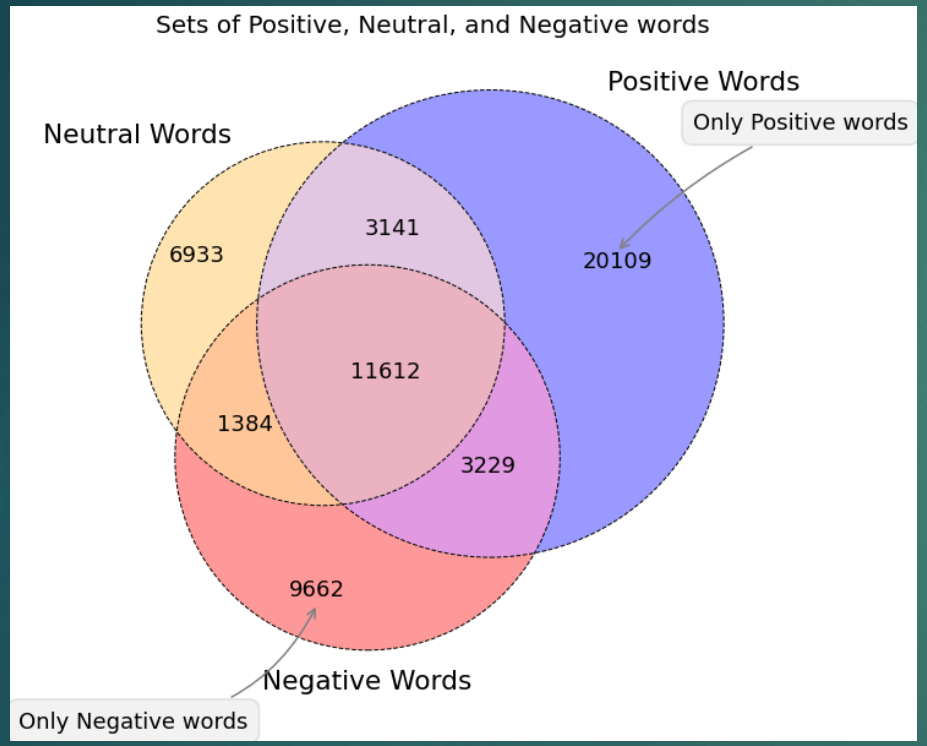
The split between Amazon's classes slightly different



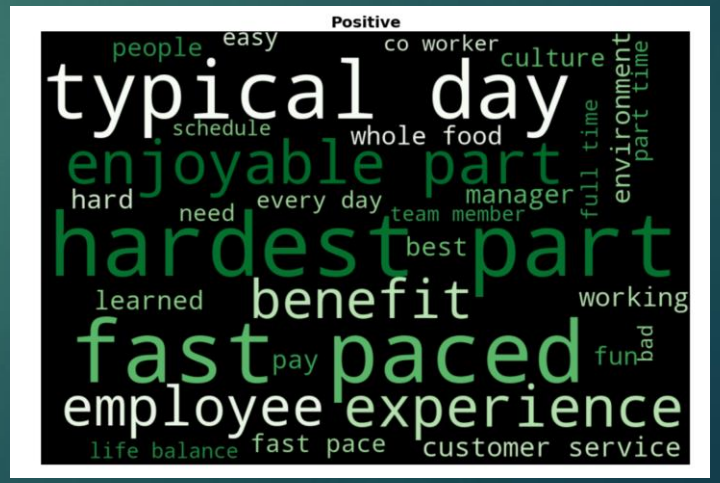


Data (continued)

Words usage in Negative, Positive and Neutral reviews



Frequency of words usage

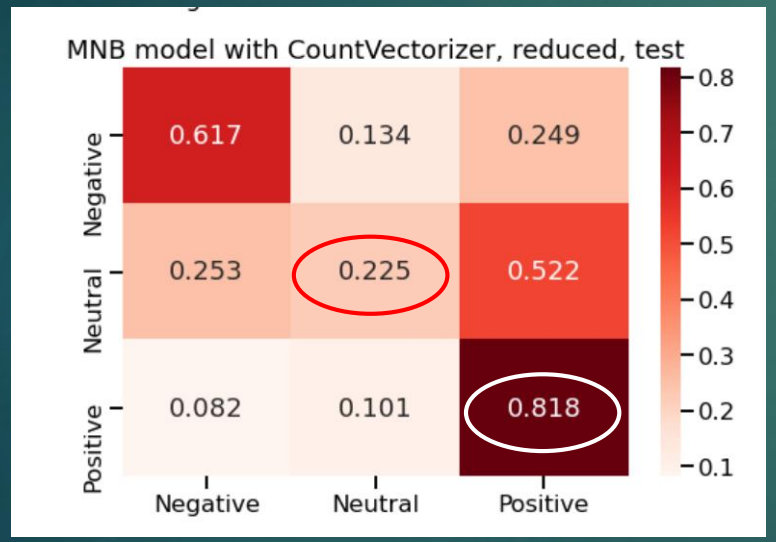


- Significant overlap between words used in classes
- Positive words > Negative words > Neutral words

Final Models

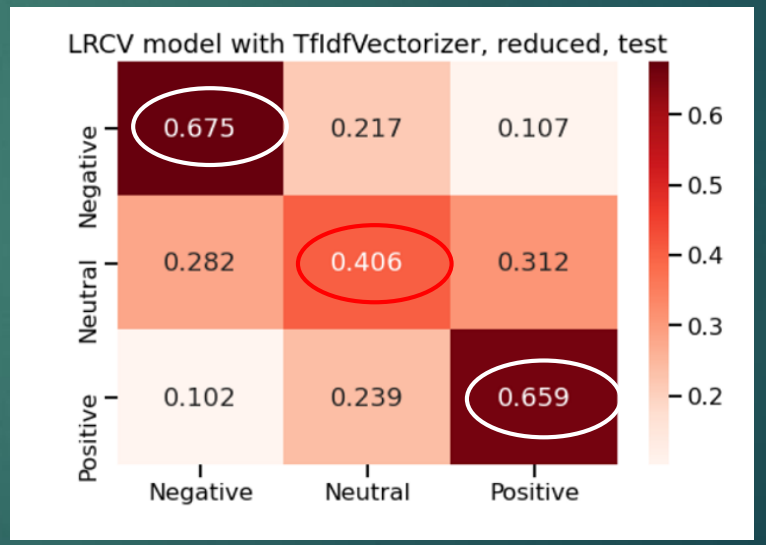
- Two models with reasonably good predictive ability
- Original set - **46962** unique words, final set- **406** unique words
- Better at predicting positive and negative reviews than neutral ones

Model #1



- Better at identifying positive reviews
- Worse at identifying neutral reviews

Model #2

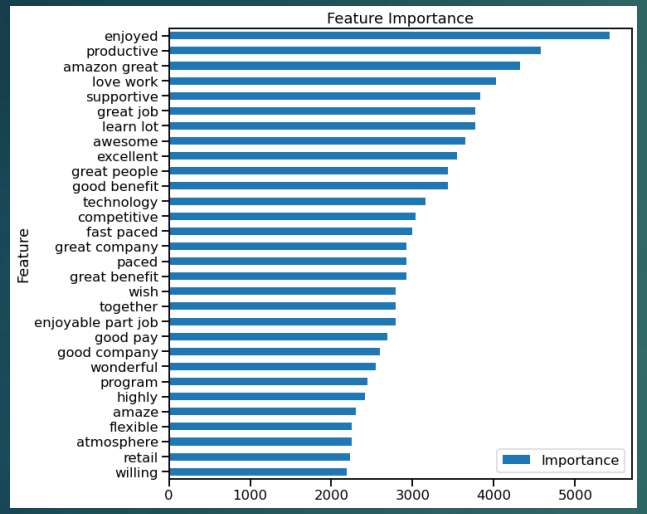


- Better balanced
- Worse at identifying positive reviews

Final Models (continued)

Importance of words for correct predictions: EMOTIONALLY CHARGED words

Model #1

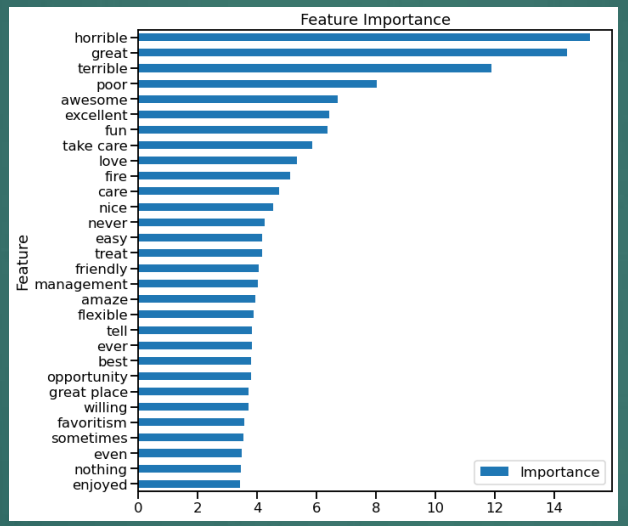


Skewed toward positivity

Enjoyed
Productive
Supportive
Awesome

Great
Wonderful
Excellent
Wonderful

Model #2



Balanced

Horrible
Terrible
Poor

Great
Awesome
Excellent

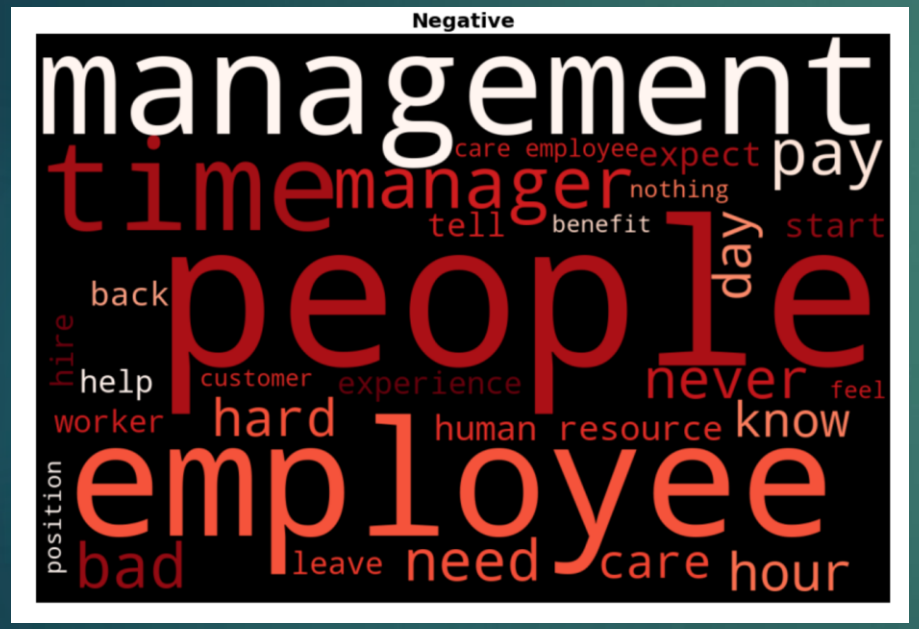
Examples of Neutral reviews

Had a **horrible** experience working there. I got sick while working there, took a medical leave to take care of myself after working all through peak season, then got terminated because they are horrible with communication and resolving issues. Would never recommend anyone to work there.

IBM has been a great place for me to learn and grow. The job variety was awesome.
This allowed me to make **excellent** contributions to every team to which I belonged.
This opportunity still exists, but is usually in the low cost countries.

Final Models (continued)

The most influential factors affecting employees' opinion of a company:



Words **manager/management/manage**

```
Enter a word to check: manager
Ratio of negative reviews with the word "manager":0.152
Ration of positive reviews with the word "manager":0.079
Ration of neutral reviews with the word "manager":0.11

Enter a word to check: management
Ratio of negative reviews with the word "management":0.237
Ration of positive reviews with the word "management":0.125
Ration of neutral reviews with the word "management":0.172

Enter a word to check: manage
Ratio of negative reviews with the word "manage":0.364
Ration of positive reviews with the word "manage":0.205
Ration of neutral reviews with the word "manage":0.27
```

Words **pay and benefits**

```
Enter a word to check: pay
Ratio of negative reviews with the word "pay":0.159
Ration of positive reviews with the word "pay":0.097
Ration of neutral reviews with the word "pay":0.131

Enter a word to check: benefit
Ratio of negative reviews with the word "benefit":0.064
Ration of positive reviews with the word "benefit":0.1
Ration of neutral reviews with the word "benefit":0.09
```

Conclusions:

Limitations of the model:

- **Relatively simple and might somewhat lack predictive power**
- **Computational resources are a limiting factor**
- **NLP multiclass classification models are less precise than binary classification models**

Additional analysis suggested:

- **Convert the problem to a binary classification problem by**
  1. removing neutral reviews
  2. Original dataset reduction to reviews rated 1 and 5 only
- **Use different techniques to assign classes labels (Clustering)**
- **Explore Theme Extraction and Context Analysis models**

Recommendations

1. Test these models built in this project internally
2. Use only anonymous and privacy-proof surveys
3. Auto-alert system to identify negative reviews
4. Positive reviews - source of information about company's strengths

Thank you!

Email: e.v.kazakova@gmail.com

GitHub: @sealaurel

LinkedIn: https://www.linkedin.com/in/elena-v-kazakova/

Appendix: Employers in the dataset

1. Ball Aerospace
2. University of Colorado Boulder
3. Boulder Valley School District
4. Boulder Community Health
5. Boulder County CO
6. Splunk
7. Naropa University
8. Amazon
9. Brookdale Senior Living
10. IBM
11. NETA
12. Apple
13. City of Boulder CO
14. Medtronic
15. Emerson
16. NetApp
17. Zayo Group
18. Whole Foods Market
19. Particle Measuring Systems
20. Boulder Medical Center
21. Sage Hospitality
22. University of Colorado
23. Mental Health Partners
24. Agilent Technologies
25. Frasier Meadows
26. Cozymeal
27. IHS Markit
28. National Renewable Energy Laboratory
29. University Corporation for Atmospheric Research
30. Sprouts Farmers Market
31. KPMG
32. New Relic
33. Excelitas Technologies
34. Northrop Grumman
35. KBI Biopharma
36. Ricoh
37. Embassy Suites by Hilton
38. U.S. Department of Commerce
39. InVitae
40. Pfizer
41. SUEZ
42. PetSmart
43. Craft Health
44. Imagine!
45. King Soopers
46. Safeway