

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

2

- Business Problem
- Data
- Final Model
- Conclusions
- Recommendations



Business Problem

3

- Making your employees happy is a smart business move
- 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

4

INDEED job search site:

121620 records

46 Boulder area employers

Positive, Negative and Neutral reviews

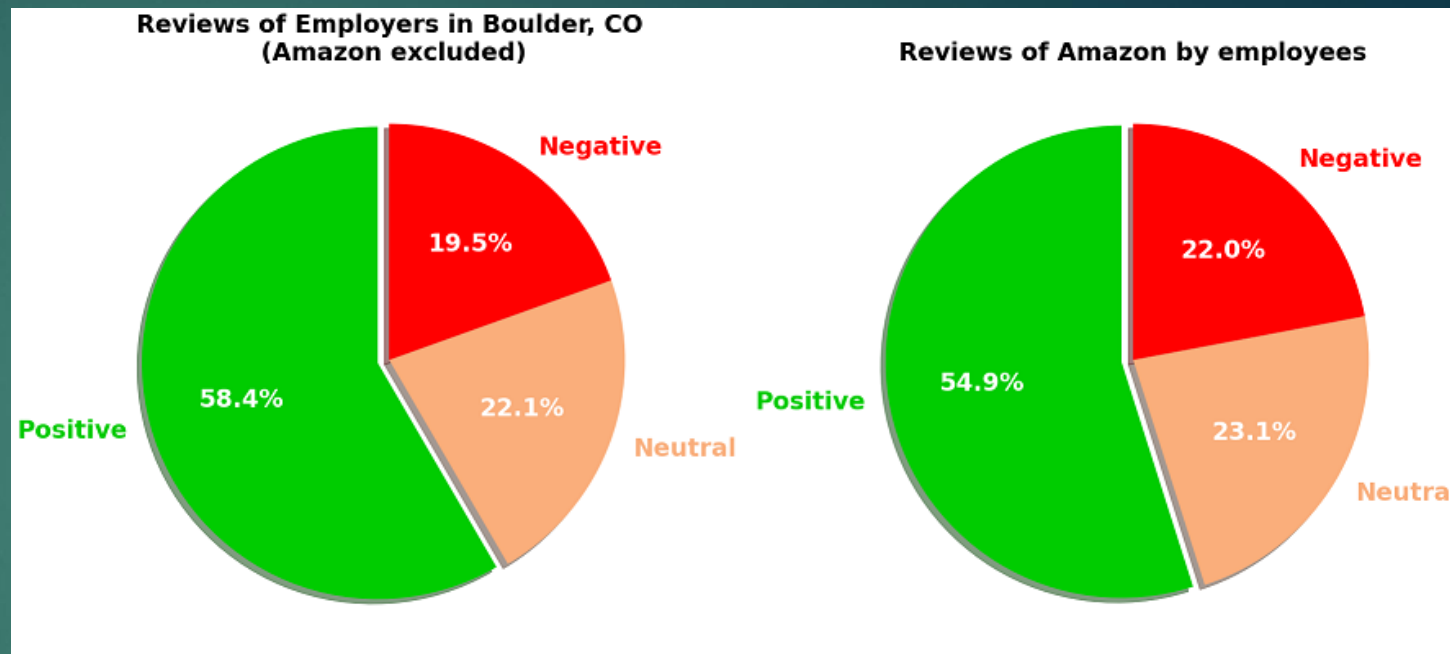
More Positive reviews

44% from Amazon

Employers list includes but is not limited to:

- IBM
- Apple
- PetSmart
- Safeway
- Whole Foods Market
- Pfizer
- University of Colorado Boulder
- ...

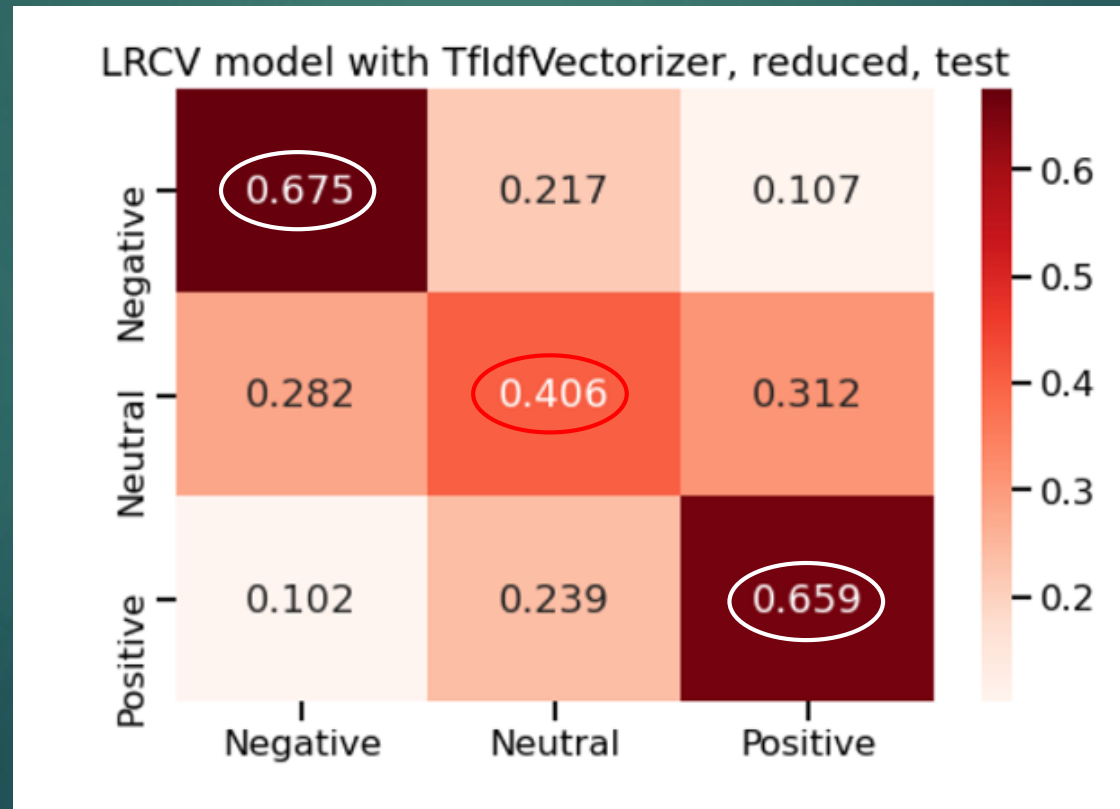
(Full list of employers is available in Appendix)



Final Model

5

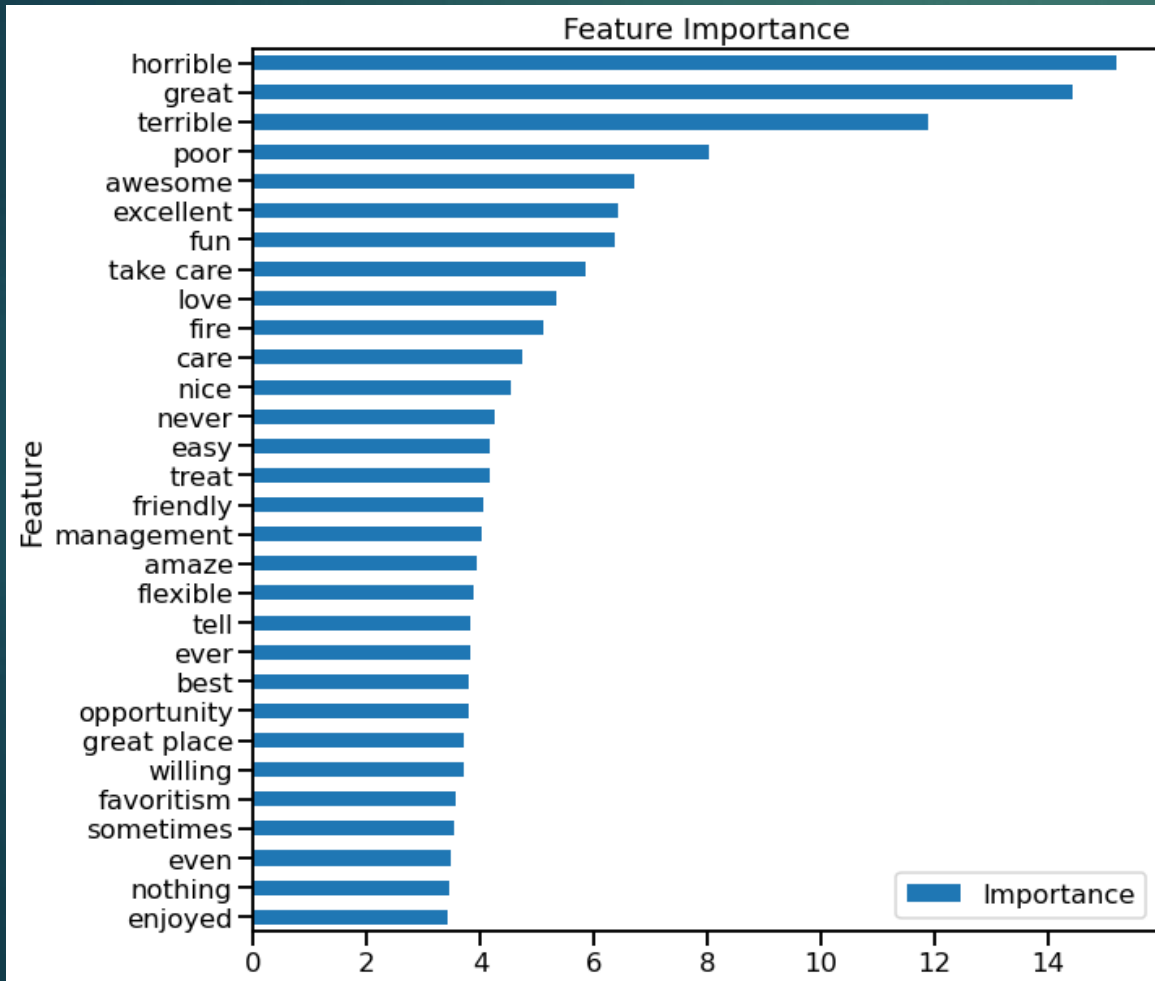
- Best model is the most balanced and the least over-fitted
- Original set - **46962** unique words, final set- **406** unique words
- Better at predicting positive and negative reviews than neutral ones



Model Insights

6

Importance of words for correct predictions: **EMOTIONALLY CHARGED** words



Positive and negative words important to the model

Horrible
Terrible
Poor

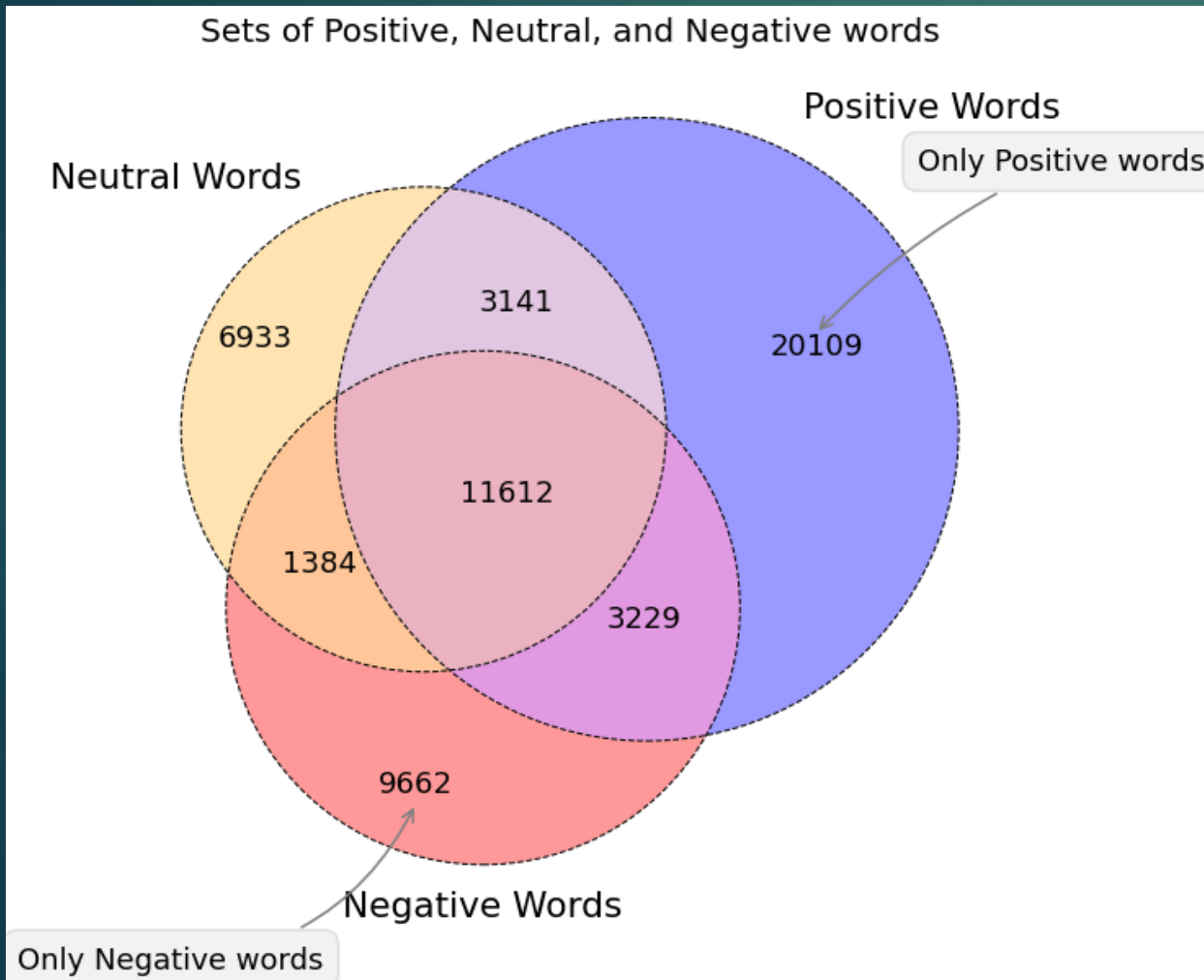
Great
Awesome
Excellent



Word Overlap Across Sentiments

7

Words usage in Negative, Positive and Neutral reviews



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



Reasons for Negative Reviews

8

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

Frequency of words usage in Negative reviews



Words **manager/management/manage**:

- Used ~twice more frequently in negative reviews
- Indicative of importance for dissatisfaction

Word **pay**:

- Also used ~twice more frequently in negative reviews

Poor management and low pay are

- equally important
- lead to:
 - Dissatisfaction
 - Disengagement
 - Lower productivity
 - Higher absenteeism



Reasons for Positive Reviews

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

Frequency of words usage in Positive reviews



Words **benefit:**

- Used ~twice more frequently in positive reviews
- Indicative of importance for satisfaction

Phrases **fast-paced/typical day/employee experience:**

- **Used more frequently in positive reviews**

Variety of tasks, fast-pace and good benefits

- Important to being satisfied with a job
- lead to:
 - *Satisfaction*
 - *Engagement*
 - *Higher productivity*
 - *Less absenteeism*
 - ***Benefits your bottom line***



Conclusions:

10

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

11

1. Test these models built in this project internally
2. Auto-alert system to identify negative reviews
3. Negative reviews – takeaways
 - Improve management style
 - Consider bringing up salaries up to industry standards
4. Positive reviews – takeaways
 - Variety of job tasks, Fast-paced environment and Good Benefits lead to
 - *Satisfaction*
 - *Engagement*
 - *Higher productivity*
 - *Less absenteeism*
 - ***Benefits your bottom line***



Thank you!

12

Email: e.v.kazakova@gmail.com

GitHub: @sealaurel

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



Appendix: Employers in the dataset

13

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

