Analyzing Glassdoor Reviews

Unveiling Job Satisfaction Factors Through DEI, Work-Life Balance, and more using ML

Dina Trac, Antoine Lavoie, Sanjana Kotha, Angela Ma, Eric Lu

TABLE OF CONTENTS

01

Research Question

Which company features are most important to employees?

Or more specifically, what are some of the biggest factors for employees that make or break their work experience?

Environmental Factors

Location, Company Outlook

Company Values

DEI, Benefits & Opportunities

Management Values

Training, Salary

Work Culture

Work life balance, Culture Values

Who are the Stakeholders?

Prospective employees

Current employees interested in learning how their company values differ in satisfaction factors

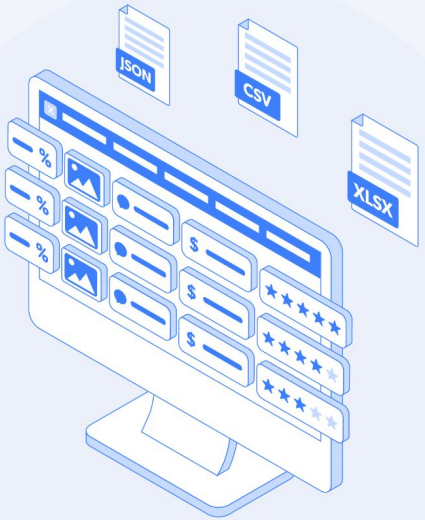
Companies interested in reinforcing or changing their workplace culture

How does this benefit them?

Prospective employees will make better informed decisions about potential job opportunities

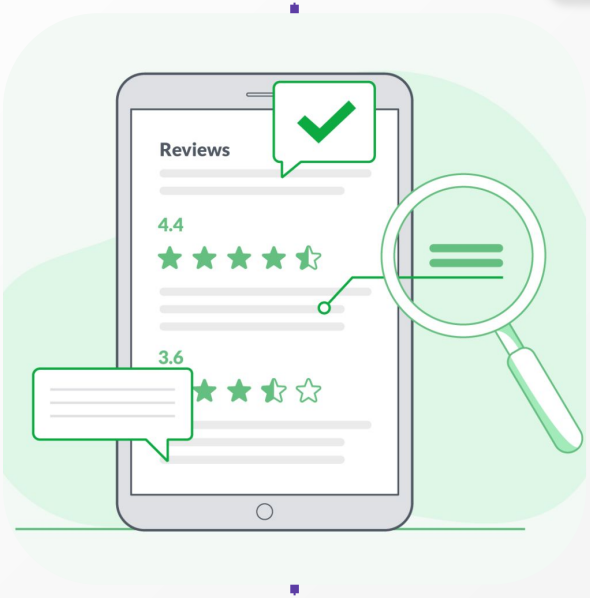
Current employees will use these findings to assess how their personal job satisfaction aligns with the overall company sentiment

Companies with positive reviews can reinforce the factors that contribute to high job satisfaction to improve overall workplace culture



02

Datasets



Glassdoor Job Reviews Dataset

Overview:

This dataset contains job descriptions and features rankings for various industries.

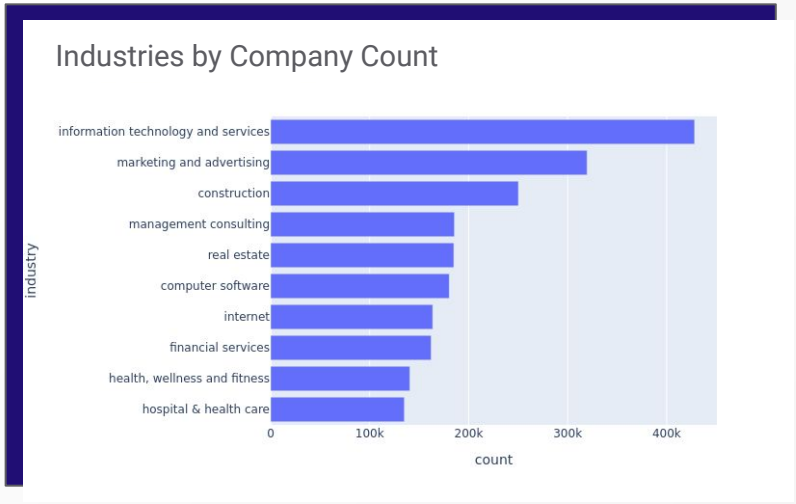
For each company in the dataset, various criteria are included such as work-life balance, income, company culture, etc.

Data Preprocessing:

We removed rows with null values in the Diversity & Inclusion column, dropped the location column, and mapped the dataset to the “7+ Million Company Dataset”.

Source: Kaggle

7+ Million Company Dataset



Dataset Overview:

This dataset contains notable information from more than 7 million companies, including industry, number of employees, and headquarters.

Dataset Use:

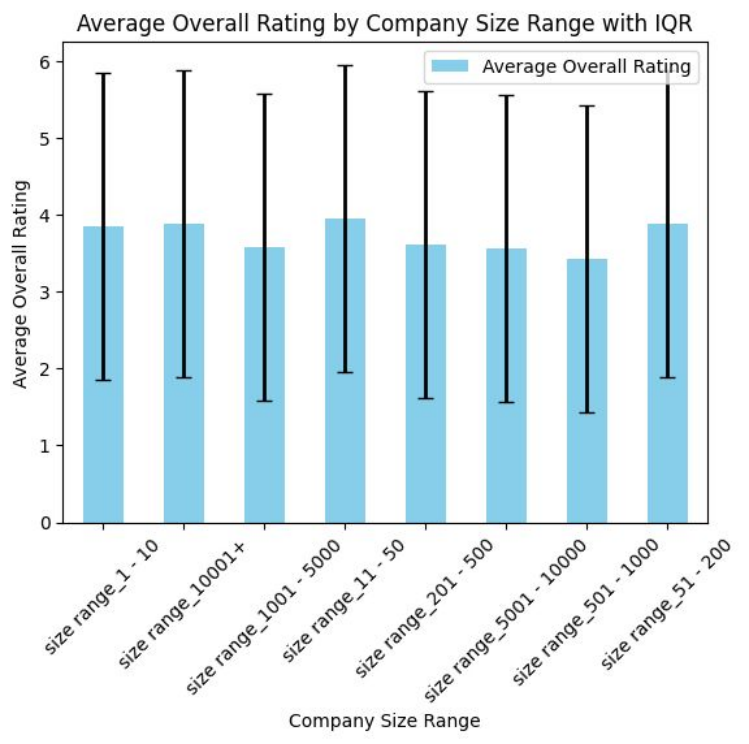
We joined this dataset with the Glassdoor job reviews to get more contextual information about companies reviewed, hoping to gain more insights on satisfaction by company-related features.

Feature Engineering:

One-hot encoded: industry, size range, Imputed null values

Source:

People Data Labs 2019 Global Company Dataset on Kaggle



Size Range vs. Rating

Due to the vast range of size in the dataset, we decided to **bin the company size** together to **reduce overfitting**, which resulted in a **30% increase in accuracy** for logistic regression, from 30% to **60.59%**.

Larger companies with size ranges **(1001-5000, 5001-10000, 10000+)** tend to have lower average overall ratings, while smaller companies with size ranges **(11-50, 51-100, 201-500)** tend to have higher average overall ratings.

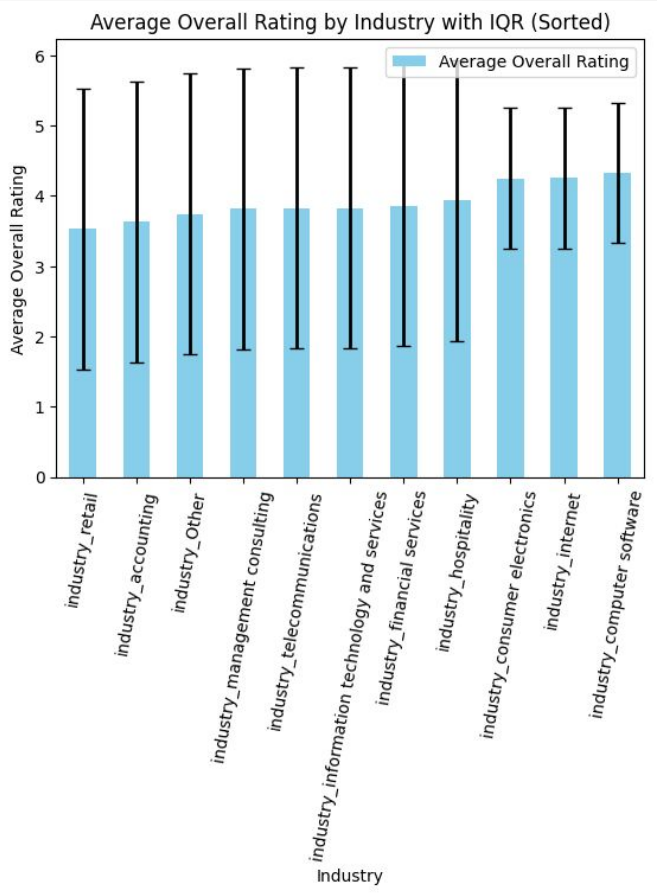
This implication could be that prospective employees working for startups, in comparison to big tech companies, have a **better general outlook** of their company on Glassdoor.

Industry vs. Rating

Tech Industry Companies, such as Consumer Electronics, Internet, and Computer Software, tend to have a **smaller IQR**, but the **highest average overall rating**.

Service Industry Companies, such as IT, Consulting, and Telecommunications, had a much **larger IQR**, but the **median overall rating**.

Other Industries such as **Retail and Accounting** had the **lowest average overall rating**.



VADER Dataset

Dataset Overview:

This dataset maps words, phrases, and emoticons to polarity scores validated by humans, and meant to decompose texts for sentiment analysis.

Dataset Use:

We used this dataset on the employee reviews (headline, pros, and cons), to gain insight into the polarity and sentiment of each review. We used VADER because it is typically used for social media sentiments, which are also written by humans.

Feature Engineering:

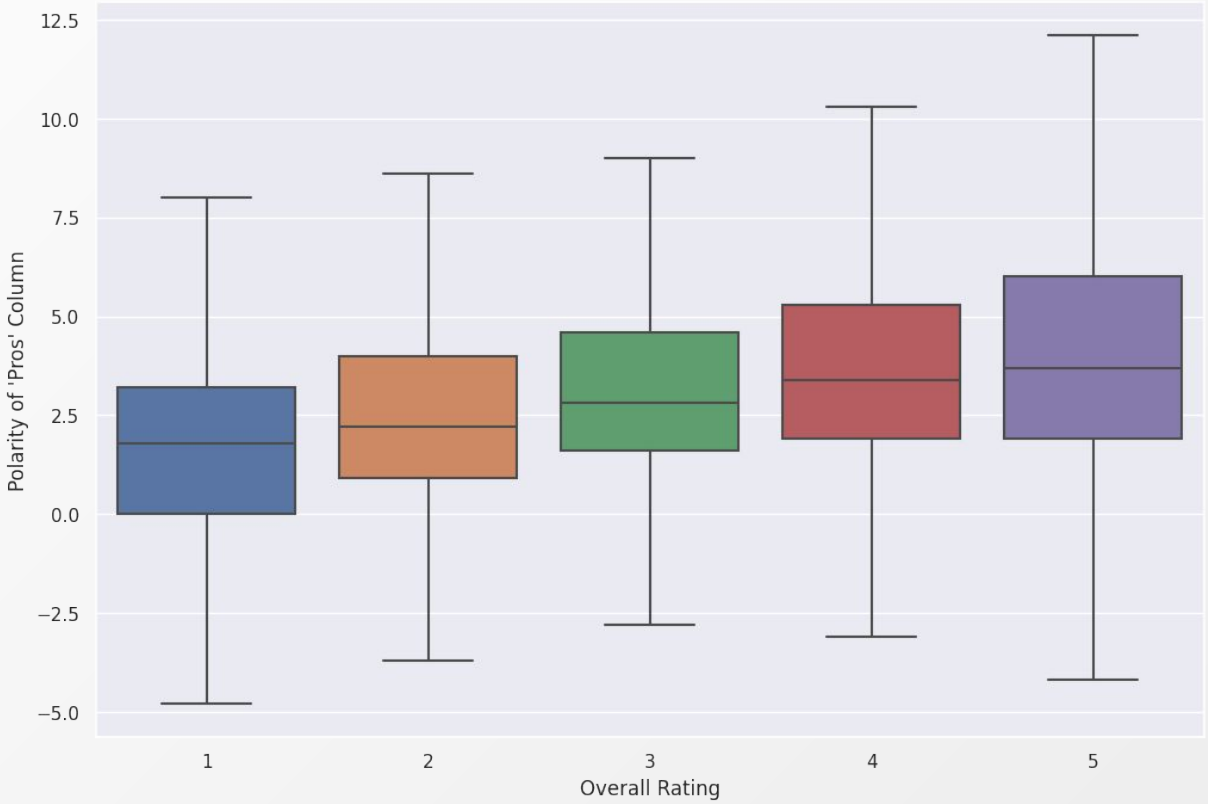
Regex string manipulation

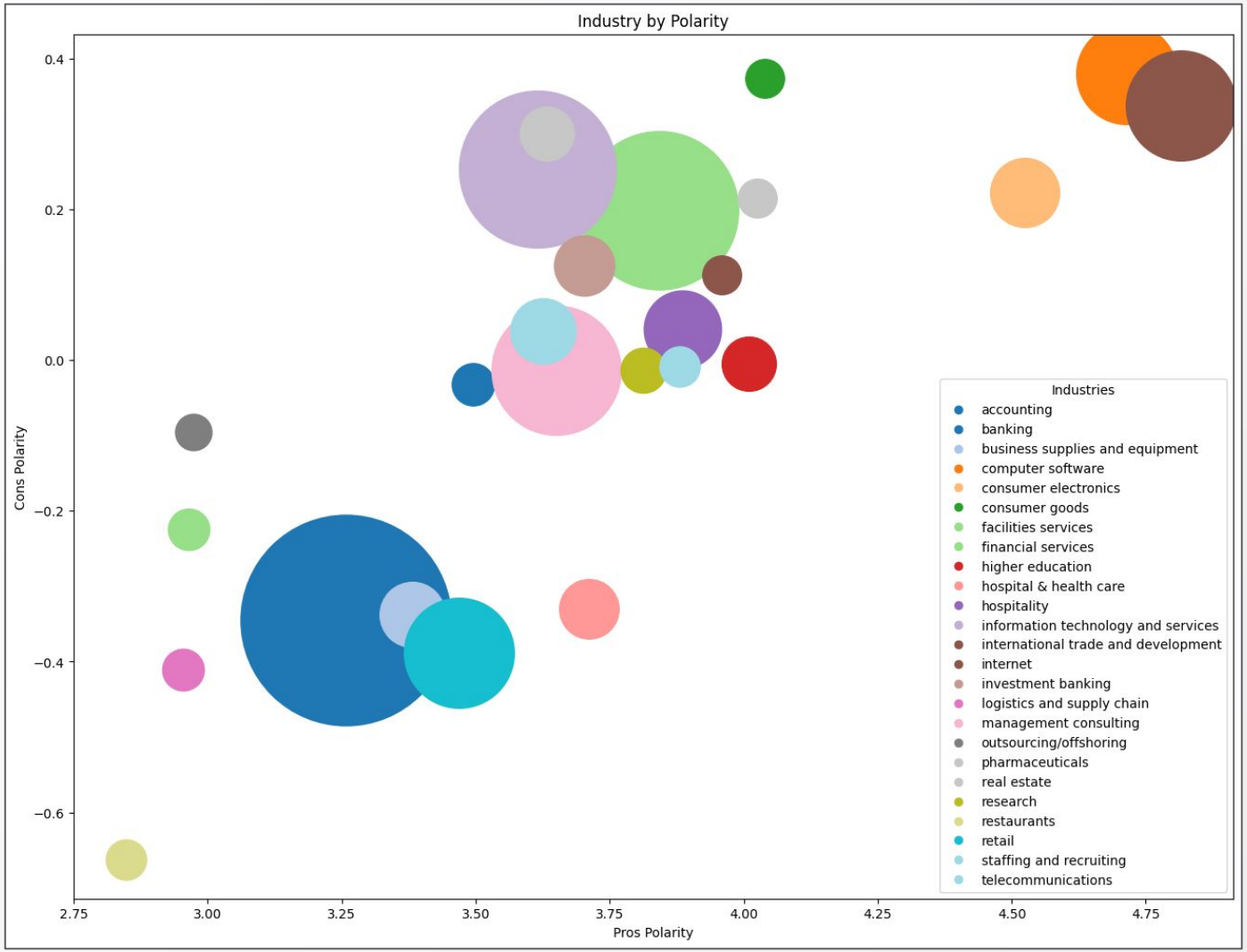
Source:

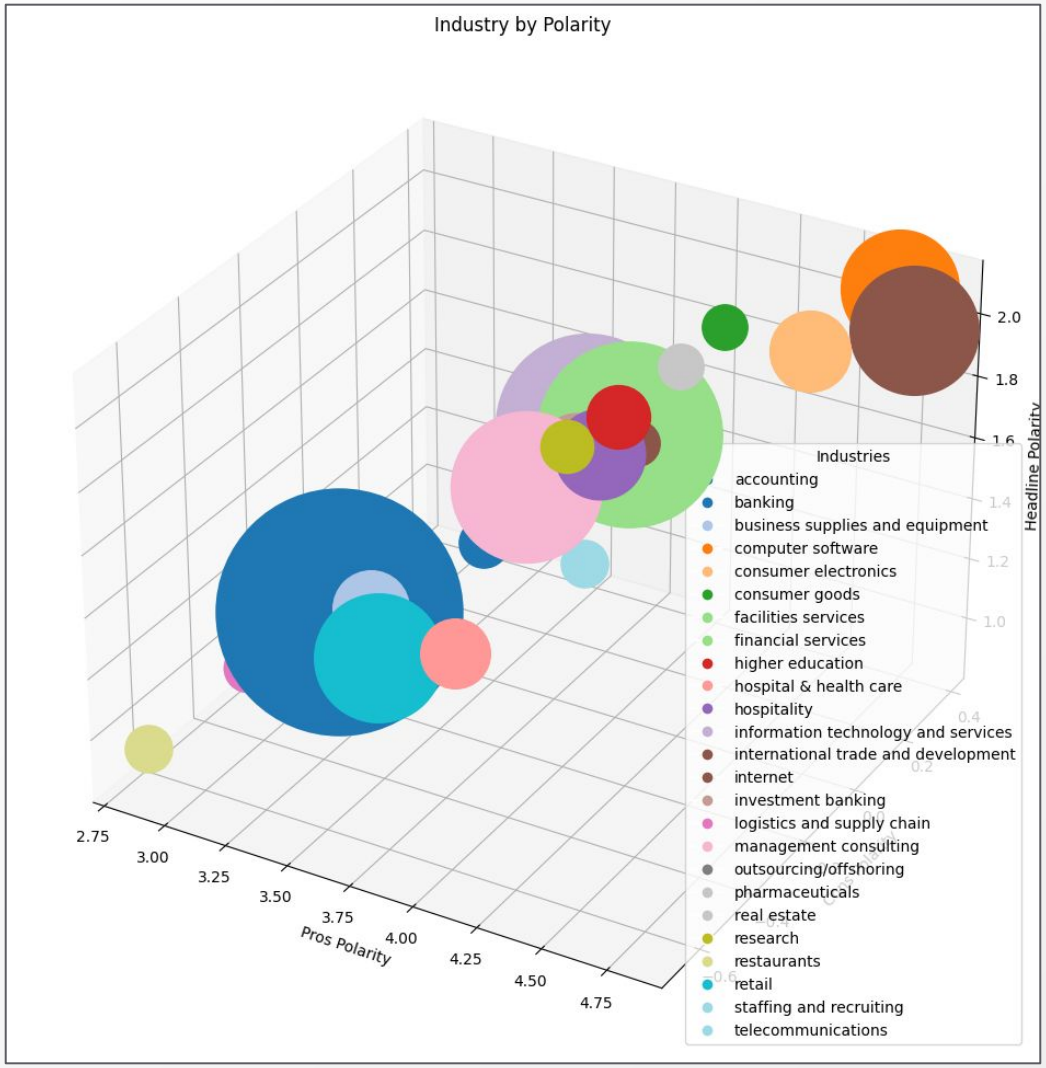
vaderSentiment on Github

VADER Polarity Score of “Pros” Column vs. Overall Rating

This shows the trend of overall rating increasing with polarity of the “pros” column.




Industry by Polarity
Cons Polarity
Pros Polarity
0.4
0.2
0.0
−0.2
−0.4
−0.6
2.75
3.00
3.25
3.50
3.75
4.00
4.25
4.50
4.75
Industries
accounting
banking
business supplies and equipment
computer software
consumer electronics
consumer goods
facilities services
financial services
higher education
hospital & health care
hospitality
information technology and services
international trade and development
internet
investment banking
logistics and supply chain
management consulting
outsourcing/offshoring
pharmaceuticals
real estate
research
restaurants
retail
staffing and recruiting
telecommunications


Industry by Polarity
Pros Polarity
2.75
3.00
3.25
3.50
3.75
4.00
4.25
4.50
4.75
Headline Polarity
2.0
1.8
1.6
1.4
1.2
1.0
0.4
0.2
Industries
accounting
banking
business supplies and equipment
computer software
consumer electronics
consumer goods
facilities services
financial services
higher education
hospital & health care
hospitality
information technology and services
international trade and development
internet
investment banking
logistics and supply chain
management consulting
outsourcing/offshoring
pharmaceuticals
real estate
research
restaurants
retail
staffing and recruiting
telecommunications

03

Feature Engineering Methods

Keyword Analysis Methods

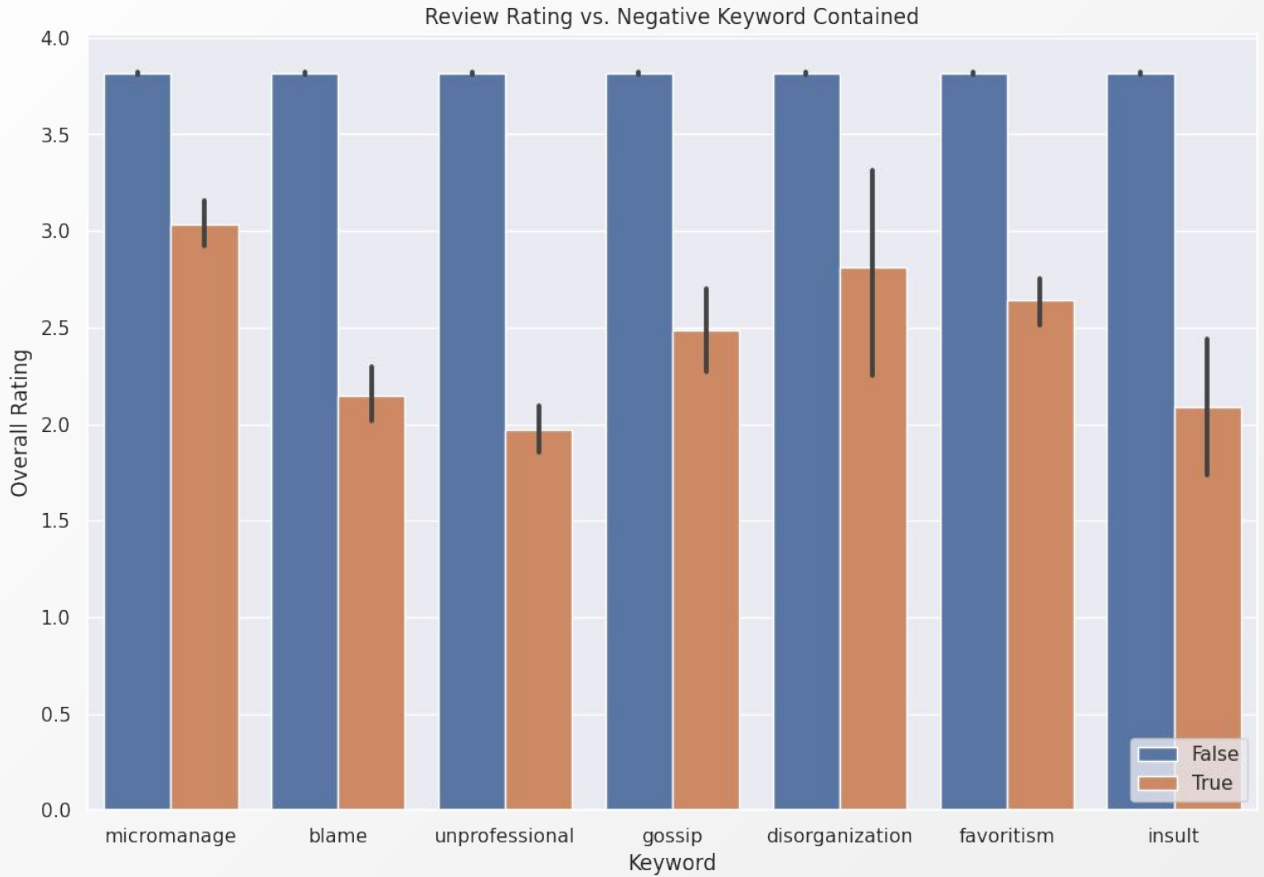
When directly looking at the textual reviews of the dataset, we noticed that the ‘headline’, ‘pros’, and ‘cons’ columns contained text we could use to extract keywords. We chose keywords based on our intuition and online research.

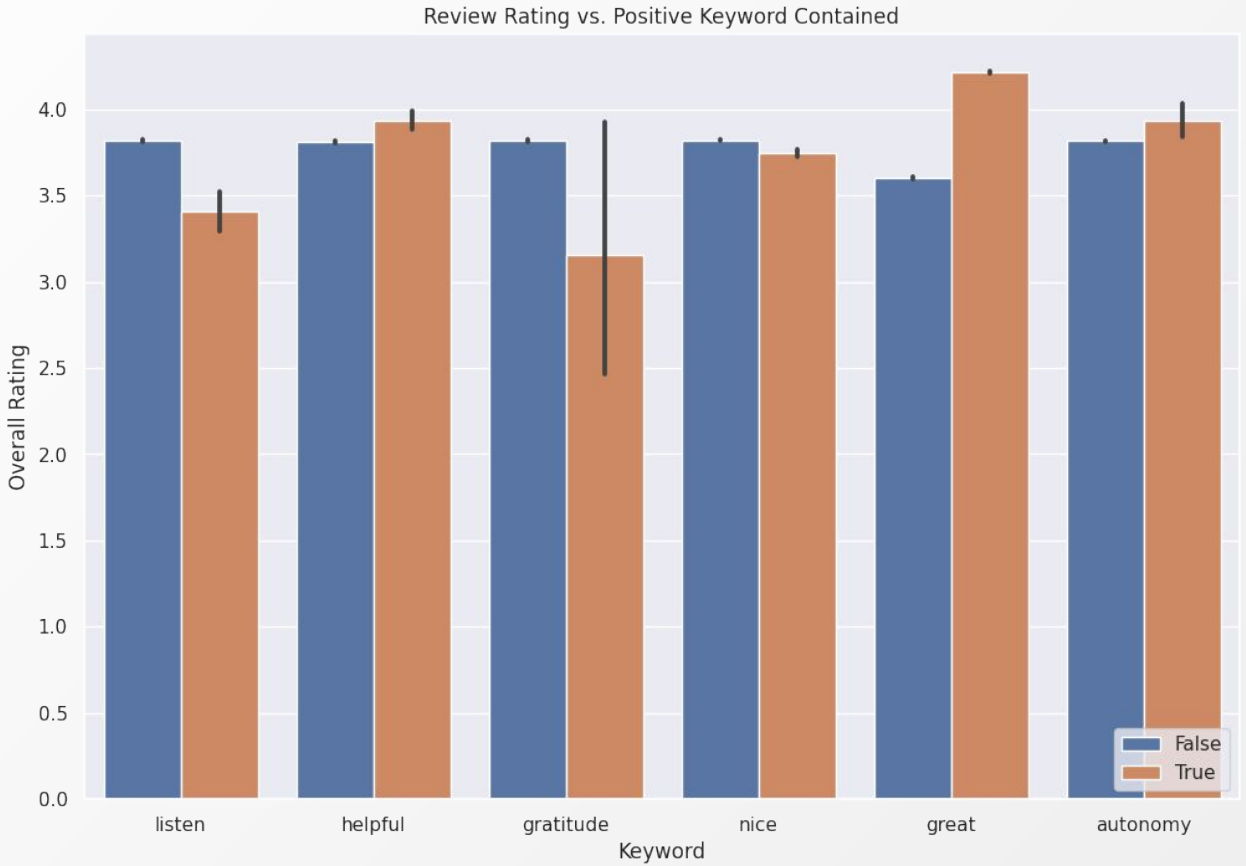
Some of the keywords we explored in negative reviews were:

- “Micromanage”
- “Unprofessional”
- “Disorganization”
- “Gossip”

Some of the keywords we explored in positive reviews were:

- “Helpful”
- “Gratitude”
- “Listen”
- “Great”


Review Rating vs. Negative Keyword Contained
Overall Rating
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
micromanage
blame
unprofessional
gossip
disorganization
favoritism
insult
Keyword
False
True


Review Rating vs. Positive Keyword Contained
Overall Rating
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
listen
helpful
gratitude
nice
great
autonomy
Keyword
False
True

Keyword Analysis Takeaways

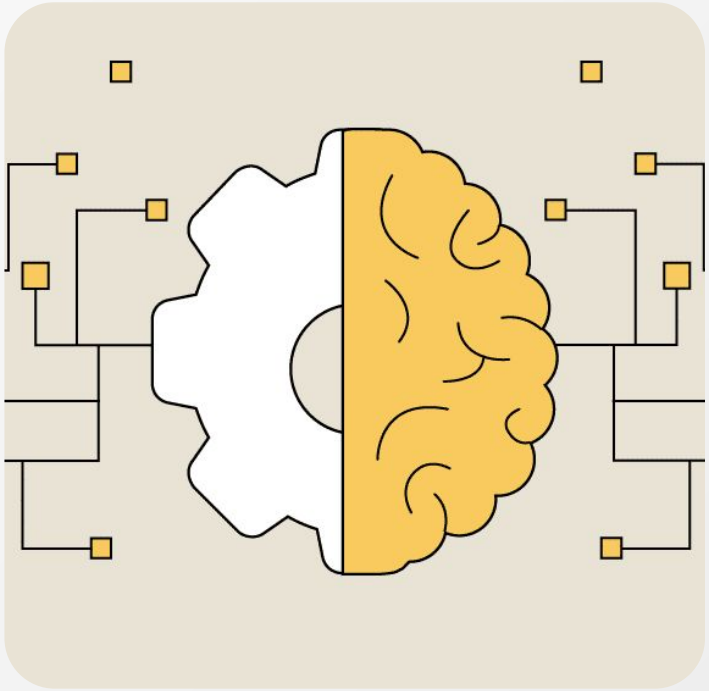
The negative keywords tested seem to be potentially informative features to use when training the model

- “Micromanage”
- “Unprofessional”
- “Blame”
- “Gossip”
- “Disorganization”
- “Favoritism”
- “Insult”

Key Takeaway: The negative keywords used when training the model could potentially be characteristics of workplace environments that lead to negative impressions of a job. From what we’ve observed, negative job reviews typically surround the notion of compensation, while positive job reviews typically surround non-quantifiable features, such as Work-Life Balance.

04

ML Methods



General Strategy

Model Testing

- Splitting dataset: training, validation, and test sets. (Hidden Test Set)
- Identifying key features of interest
- Intuitive evaluation of performance
- Initial small scale tests on various models
- Compute accuracy on validation set

Model Optimization

- Analyze features from a perspective of overfitting (i.e. One-Hot Encode/LabelEncoder)
- Validation set evaluation, trial and error
- Grid Search CV for parameter optimization
- Even more cross-validation
- Focused on accuracy as KPI for models

NLP (BERT)

- Apply BERT-based NLP model for further polarity and text-based analysis
- NLP Model implemented in Transformer library predicting a score (1-5) based on a text review
- 2 million downloads in last 30 days

Name of Model	Parameters	Validation Accuracy
Logistic Regression	`solver = 'liblinear'`	60.83

K-Nearest Neighbor (KNN)	`metric = 'manhattan'`
	`n_neighbors = 15`
	`weights = 'distance'`
		66.14

Decision Tree	`criterion ='entropy'`
	`max_depth = 10`
	`min_samples_leaf = 1`
	`min_samples_split = 10`
		65.31

Random Forest	`max_depth = None`
	`max_features = 'auto'`
	`min_samples_leaf = 1`
	`min_samples_split = 5`
	`n_estimators = 200`
	`random_state = 42`
		68.39

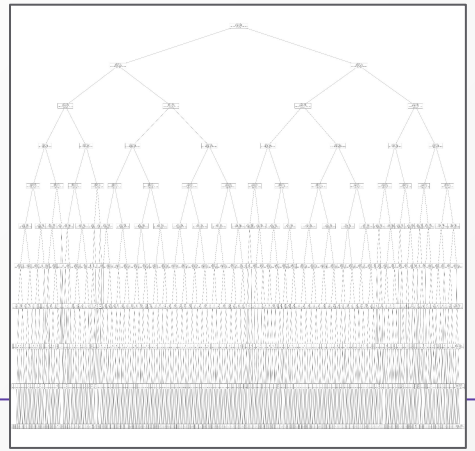
Neural Network	`hidden_layer_sizes = (50,)`
	`activation= 'relu'`
	`solver = 'adam'`
		66.48

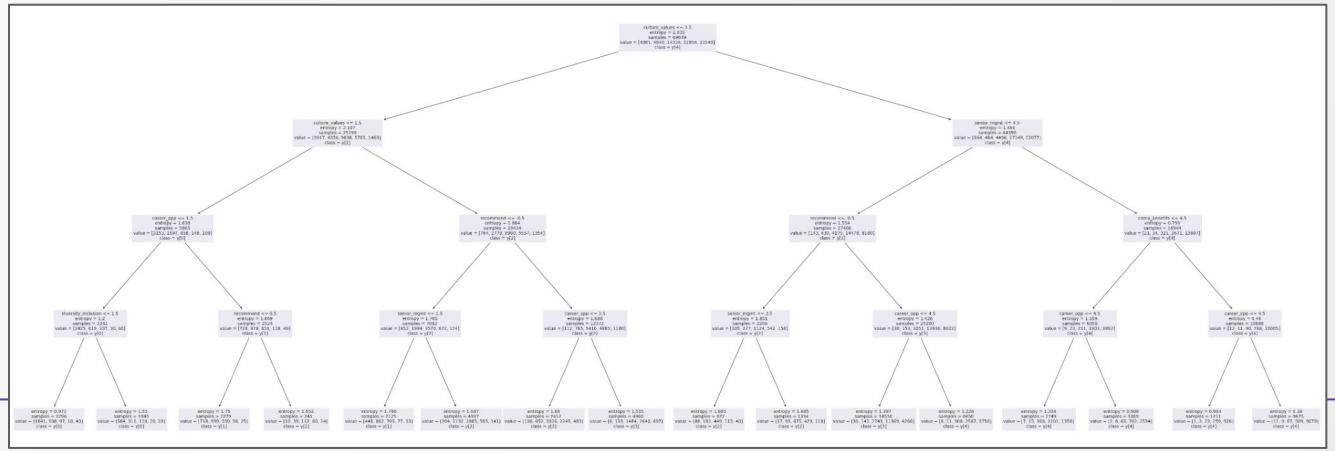
Decision Tree

Interpretability: Decision Tree favored for its interpretability, crucial for understanding model decisions.

Decision Tree Strengths: Simplicity, ease of visualization, and understanding.

Decision Tree Weaknesses: Generalizability (small data changes result in large inaccuracies), more complex models lose interpretability





Feature Importances:

	Feature	Importance
1	culture_values	0.095896
5	senior_mgmt	0.083422
9	pros_polarity	0.082211
3	career_opp	0.079898
10	cons_polarity	0.076312
4	comp_benefits	0.068482
11	headline_polarity	0.068343
0	work_life_balance	0.056116
12	current employee estimate	0.052274
2	diversity_inclusion	0.051432
6	recommend	0.048958
8	outlook	0.033157
7	ceo_approv	0.030847
13	years of experience_1	0.013523
18	years of experience_null	0.012984
15	years of experience_3	0.010475
46	country_united kingdom	0.009615
26	industry_Other	0.009506
47	country_united states	0.009424
27	industry_accounting	0.008971
30	industry_financial services	0.008059
16	years of experience_5	0.007568
32	industry_information technology and services	0.006544
37	country_Other	0.006522
49	size range_10001+	0.006092
53	size range_5001 - 10000	0.005443
34	industry_management consulting	0.005426
17	years of experience_8	0.004631
35	industry_retail	0.004601
14	years of experience_10	0.004232
33	industry_internet	0.004222
28	industry_computer software	0.003951
50	size range_1001 - 5000	0.003684
31	industry_hospitality	0.003165
36	industry_telecommunications	0.002645
41	country_germany	0.002357
29	industry_consumer electronics	0.002304
43	country_jordan	0.002072
52	size range_201 - 500	0.001599
54	size range_501 - 1000	0.001480
55	size range_51 - 200	0.001425
45	country_switzerland	0.001418
40	country_france	0.001363
48	size range_1 - 10	0.001067
44	country_netherlands	0.000936
19	micromanage	0.000844
51	size range_11 - 50	0.000811
38	country_belgium	0.000683
24	favoritism	0.000668
42	country_ireland	0.000611
21	unprofessional	0.000482
39	country_denmark	0.000481
20	blame	0.000370
22	gossip	0.000270
25	insult	0.000067
23	disorganization	0.000063

Random Forest

Feature Importance: Random Forest provided valuable insights on feature importance in predicting ratings.

Random Forest Strengths: Better handling of large feature sets and robustness to overfitting.

Neural Network Limitations: Despite high complexity, didn't yield significantly better results; issues with interpretability and overfitting. Random Forest is a happy medium between NN and Decision Tree.

Language	Accuracy (exact)	Accuracy (off-by-1)
English	67%	95%
Dutch	57%	93%
German	61%	94%
French	59%	94%
Italian	59%	95%
Spanish	58%	95%

firm	overall_rating mean	size	nlp rating mean	size
Salesforce	4.333333	48	4.208333	48
American-Express	4.346154	26	4.115385	26
Google	4.383333	60	4.083333	60
Marriott-International	4.340426	47	4.063830	47
CBRE	3.758621	29	4.034483	29
Microsoft	4.069767	86	3.988372	86
Apple	4.121622	74	3.972973	74
SAP	4.395349	43	3.860465	43
IBM	4.013393	224	3.799107	224
Deloitte	3.745174	259	3.675676	259
EY	3.662577	163	3.668712	163
Morgan-Stanley	3.629630	27	3.666667	27
PwC	3.647799	159	3.654088	159
Oracle	3.666667	126	3.634921	126
KPMG	3.575540	139	3.633094	139
J-P-Morgan	3.862385	109	3.623853	109
Pizza-Hut	3.457143	35	3.600000	35
Barclays	3.700000	40	3.600000	40
NHS	4.060606	33	3.575758	33
Tesco	3.553846	65	3.553846	65
Citi	3.800000	70	3.514286	70
Vodafone	3.750000	56	3.500000	56
HSBC-Holdings	3.788462	52	3.461538	52
BT	3.407407	27	3.407407	27
McDonald-s	3.352459	244	3.401639	244
Goldman-Sachs	3.743590	39	3.384615	39
GlaxoSmithKline	3.392857	28	3.214286	28
BNY-Mellon	3.187500	32	3.093750	32

NLP (BERT)

Review Scoring: NLP rating on the entire review (headline, pros, and cons) gives further insight into how employees rate their company based on the sentiment of the review itself.

NLP Strengths: Offers another feature beyond overall rating. Perhaps some employees rate their companies highly, but didn't have strong sentiment in the actual review.

NLP Limitations: Pre-trained NN model (uninterpretable), average exact accuracy (67%), expensive computation, marginally different context (product reviews vs. job reviews)

05

Results and Conclusions

Feature Importance Trends

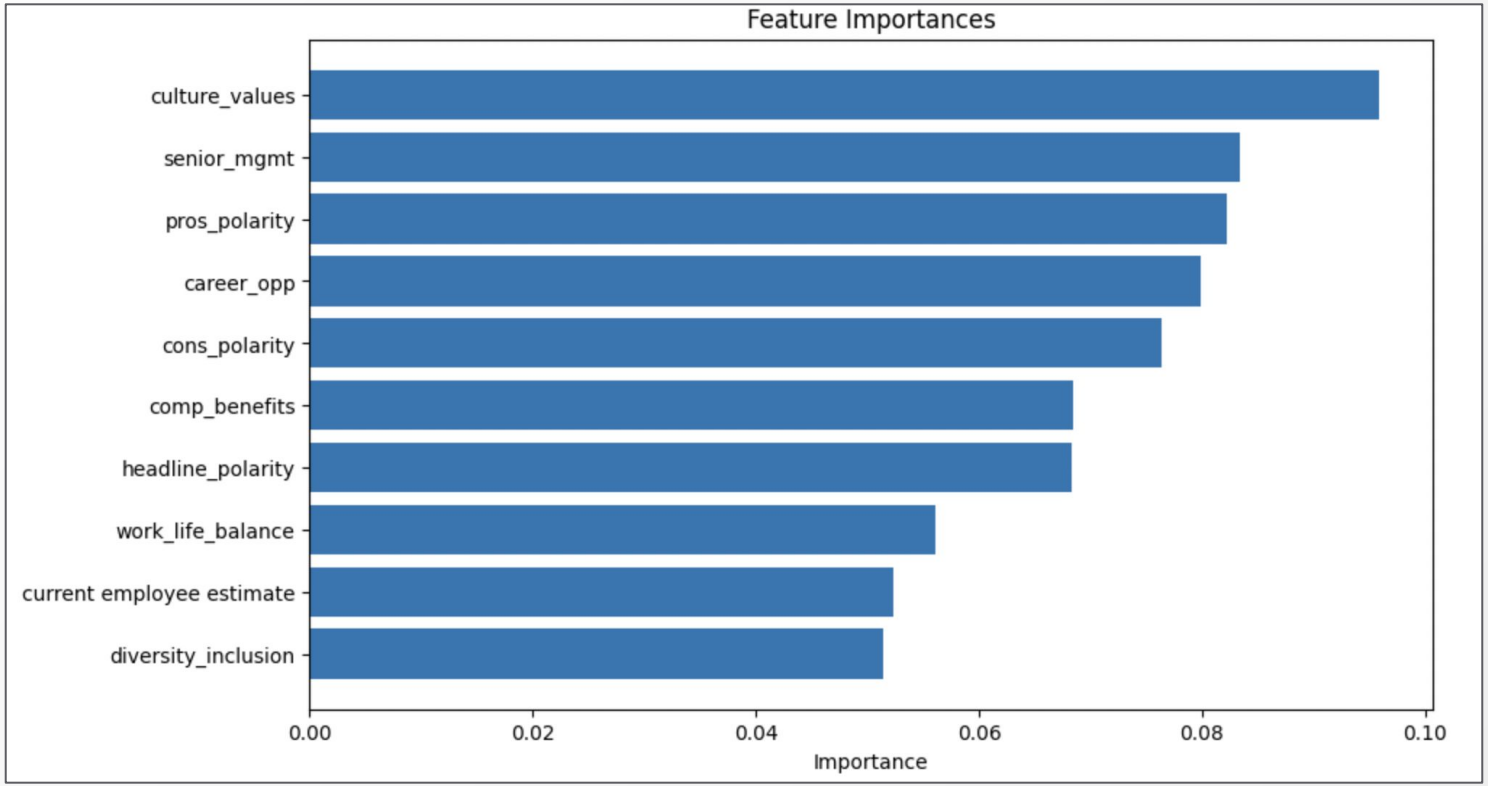
By computing the **Feature Importance** using a **Random Forest Classifier**, we observed that the features with higher importance were the original features from the Glassdoor Dataset, which were Culture Values, Senior Management, Career Opportunities, Company Benefits, and Work Life Balance.

Some of the features with higher importance that were not included in the original dataset were the **“Pros”, “Cons”, and “Headline”** Polarity Scores, which derived from the VADER Dataset.

Some of the countries with higher importance included the **United Kingdom** and the **United States**.

The words correlated to **negative** reviews from the “Cons” column had the **lowest importance**, which clashed with our initial hypothesis.

Feature Importance Trends (Top 10)



Feature Importance Analysis

The **“Pros”** and **“Cons” Polarity Scores** had higher importance than the **“Headline” Polarity Score**, indicating **emotional biases** over the textual analysis, profoundly shown in the keywords used.

“Ceo Approval” is one of the lowest feature importance, suggesting that the general outlook of the CEO does not reflect as much as we thought on the overall rating of the company.

“Size Range” trends show that small sized companies (1-10, 11-50, 51-200) have the highest average overall rating with (11-50) range having the highest median, 75%, and max. Mid-sized companies and large companies tended to have lower overall ratings with 10001+ bin being an outlier with high overall rating.

The **UK and US** having higher importance than other countries suggest that companies residing in these areas tend to have better **Work Life Balance** and overall **Company Culture**, confirming our initial intuition.

Many of the words that we thought correlated with the “Cons” column ended up having little to no effect on the overall rating, having the **lowest importance** out of all other features.

Implications of Project and Results

The findings of this project can be used by companies to inform **strategic-decision making processes**. For instance, based on our feature importance analysis, focusing on addressing specific issues raised in the "Cons" section may have a more significant impact on overall employee satisfaction and Glassdoor ratings than emphasizing CEO approval.

Furthermore, tailoring **HR and management strategies** based on company size could lead to more effective **employee engagement** and satisfaction.

Because we found that companies in the UK and the US have better **work life balance** and company culture, companies outside of these two countries could use this information to compare themselves to global standards, identifying areas for improvement in work-life balance and company culture to enhance overall ratings.

The unexpected results in **word correlation** emphasizes the need for continuous refinement of textual analysis models.

Other Techniques Explored

TSNE

- We tried using t-SNE for analyzing Glassdoor reviews to categorize words and **see what industries have those words associated with them most,** primarily through clustering.
- Technique: Process the reviews to extract relevant words or phrases. Then apply t-SNE to these high-dimensional text vectors to reduce the dimensions. This places similar words (in terms of context or usage) closer together in the reduced space.
- Ultimately were **unsuccessful due to limited computational power.**

Neural Networks

- This model initially produced extremely poor accuracy on the validation and testing set. This was likely due to **poor parameter selection, with our validation accuracy going from 24% to 66% after optimizing parameters**. Still this model suffered from the problem of overfitting, as the model had excessive sensitivity to the training data, **prohibiting the model to generalize new, unseen data.**

Job Titles

- This column/feature was **cursed by variation**. We originally thought that management and seniority would impact reviews, however, it was very difficult to group the vast **amount of unique values or interpret their structure within a company**.

Imputing Null Values

- We tried imputing null values in the Diversity & Inclusion column to salvage missing data, but this approach led to a significant reduction in accuracy due to average bias. That said by dropping null values the dataset size drastically decrease from ~800,000 to ~100,000 rows.

Is this project ready to be used?

While the findings we obtained offer valuable insights, more refinements are still necessary before it can be declared completely ready for real-world application

- One way of enhancing this project's applicability is to try integrating real-time data. The sentiments of an employee can change within short periods of time and company dynamics can also evolve, so a model that adapts and integrates real-time data would be more robust
- We could also take job role and more industry-specific variables into consideration as additional features to refine the model's accuracy and relevance to specific organizational contexts.
- Evaluate potential biases in the dataset and model to ensure it is fair and ethical.
- Utilize word embeddings (ex. Word2Vec) to capture semantic relationships between words. This can help identify subtle nuances in language that traditional sentiment analysis may overlook.
- We could also analyze how sentiments change over time, and whether or not there are specific periods of time with significant shifts in sentiments (seasons of the year, COVID, holidays).
- Cluster reviews based on similarities in feedback. This can help identify groups of employees with similar experiences, allowing for more targeted interventions to address specific concerns.

THANKS!
Any Questions?

Resources

https://www.kaggle.com/datasets/peopledatalabssf/free-7-million-company-dataset/data

https://www.kaggle.com/datasets/davidgauthier/glassdoor-job-reviews/data

https://github.com/cjhutto/vaderSentiment

https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment?text=I+love+you

https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270

https://ds100.org/

Concepts and Ideas are pulled largely from the course material of Data 144: Instructor Zachary Pardos