Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 4, 646-654

Artificial Intelligence Methods to understand and improve Employee Experience

[1]Ch v phani Krishna, [2]Dr.Katta Subba Rao, [3]Dr. Rajendra kumar Ganiya, [4]V Rama Krishna
[1]Professor,CSE department ,TKR Engineering College,Hyderabad, Telangana, Phanik16@gmail.com
[2]Professor, B V Raju Institute of Technology, Narsapur, Medak, Telangan, subbarao.k@bvrit.ac.in
[3]Professor, Department of Information Technology,Vignan's Institute of Information Technology, A.P, India, rajendragk@rediffmail.com
[4]Professor,Vignan's Foundation for Science, Technology and Research, Vadlamudi,vramakrishna2006@gmail.com

Abstract
The success of an organization depends on the performance of its employees. In this ultra-competitive job market, organizations are forced to quickly understand the intrinsic and extrinsic factors that concern employees. Organizations have historically depended on various surveys to understand what their workforce is thinking about different aspects of their employment. Though effective, employee surveys often follow a boilerplate template and often fail to capture the emerging socio-economic changes. Capturing employees' feedback using text allows workers to express themselves and provides excellent insight into the latent themes. This paper offers a three-step framework to understand employee experience using Indeed reviews as a proxy. We leverage Natural Language Processing techniques such as N-gram analysis, sentiment analysis, and topic modeling to bring forward latent topics of discussion.

1. Introduction:

Organizations and business leaders have always favored data-driven decision-making. The two popular dimensions of data-driven decision-making are the information at hand and the process of generating insights(Provost & Fawcett, 2013).

The process of valuing information has significantly evolved over the years. As the first commercial databases were launched in years 1968 , organizations started to think about storing data in a well-organized format(Fry & Sibley, 1976). In the early days, companies limited themselves by only storing the most business-critical data. In the next fewdecades, innovations in database storage helped organizations store much of the non-business critical information in the databases. The next significant innovation in the data storage space happened with the invention of Hadoop in early 2000's (Shvachko et al., 2010). The idea of storing the data, agnostic of the data type in its rawest format, and being able to define the metadata at a later time, changed the paradigm of data storage and retrieval. Organizations started storing non-traditional data such as reviews, video footage, job descriptions, resumes, emails, sensors, and log data.

Another dimension of data-driven decision-making is the insights generation process. Modern research overwhelmingly suggests that data-driven organizations are innovative when compared to their peers(Ghasemaghaei & Calic, 2019). There are three significant phases in generating insights – descriptive, predictive, and prescriptive analytics(Lepenioti et al., 2020). Spead sheets and charts used to be the gold standard for generating descriptive insights. They provide a straightforward way to consume and visualize information. Business intelligence and data visualization tools have provided a convenient way for decision-makers to analyze historical data. Statistics, machine learning, deep learning methods are popular methods that are helping organizations generate predictive and prescriptive insights.

But, as organizations accumulated a wealth of non-traditional data assets such as text, they started thinking about leveraging insights from this data to make critical business decisions(Markham et al., 2015; Tussyadiah & Zach, 2015). Natural language processing (NLP), a popular sub stream of Artificial Intelligence (AI), deals with exploring text data in a supervised and unsupervised setting. Topic modeling falls into the unsupervised algorithms. The goal of topic modeling is to understand the latent topics and aspects that are part of large bodies of text.

As the storage costs decreased and the innovation in the AI space increased, we seen a meteoric raise in the industry adaptation of AI and NLP. Marketing, Sales, Customer service, and Supply chain were the early adopters to this change.

The human resources (HR) function in an enterprise is critical for the performance of the whole organization. The HR function is responsible for hiring, coaching, satisfying, and promoting the right talent to fulfill the organization's short-term and long-term goals efficiently and effectively.

The role of an HR function becomes complete if it can effectively balance out the organizational needs with that of individuals. Unfortunately, for HR, there isn't a simple way to understand the comprehensive needs and sentiments of associates. At a high-level, it is critical to – "Deeply understand people and their needs", "Embrace expansive and holistic thinking", "Make the intangible tangible", "Insist on radical participation", "Iterate and experiment", "Trust and appreciate the process" (Plaskoff, 2017). However, in reality it's very difficult to capture and measure the holistic associate experience. The primary reason for this is because of the way HR gathers feedback. HR function typically relies on various survey mechanisms to collect associates' sentiments around compensation, associate experience, job satisfaction, benefits, work culture, etc. (Church & Waclawski, 2017). Though the surveys effectively measure the macro-level pulse, they fail to quantify micro-level themes that resonate with associates.

One potential solution to understand how employees are feeling, is the use of text analytics. Natural language processing (NLP), a subfield of Artificial Intelligence, aims to develop syntactic and semantic level understanding of language to develop insights (Chowdhary, 2020).

Though HR partners used descriptive analytics, they are relatively late adopters to modern technologies such as AI (Renkema, 2022). The exciting part is that the opportunity to use NLP is waiting at HR's doorstep. The traditional surveys almost always show up in a standardized format where associates are expected to choose from a predetermined list. HR rarely encourages associates to type their responses in a free-form text box. NLP techniques such as word frequency analysis, sentiment analysis, topic modeling and aspect detection can be a game changer for HRas it enables them to gather latent themes.

In this paper, we leverage publicly available employee reviews data for top technology companies to understand the topics of discussion among associates. We leverage popular NLP techniques such as n-gram analysis, sentiment analysis, and topic modeling to present latent insights. Finally, we present a framework that informs readers on when it's best to use the three NLP techniques. The primary purpose of this paper is not necessarily to discuss the corresponding NLP techniques in detail, rather we intend to present a framework to leverage NLP when understanding employee experience and motivate readers to pursue AI to solve people related problems.

2. Data:

The dataset used for this study is obtained from a Kaggle competition[1]. Company reviews for top technology companies such as Amazon, Apple, Facebook, Google, Microsoft, and Netflix were collected from popular job posting aggregator. The distribution of company reviews is summarized in Figure1.

[1]https://www.kaggle.com/datasets/fireball684/hackerearthericsson

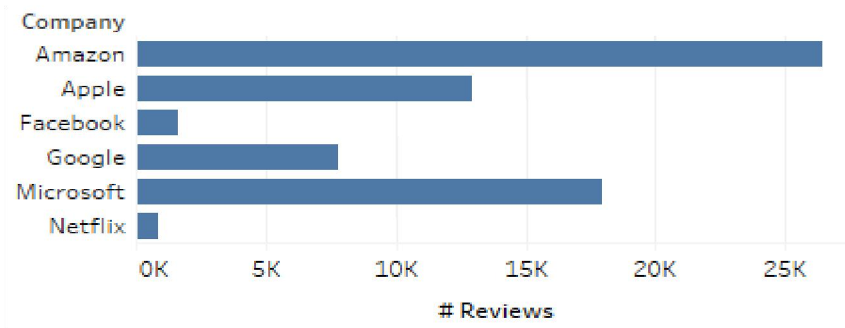

Figure 1:Number of reviews by company

For each review, Overall star rating, Pros, Cons, Review time, location, and Employee job title are collected (Figure 2).

Figure 2: Components of an Indeed review

The timeframe of the reviews ranges from 2008 to 2018. The total number of reviews by company are summarized in Figure 3. The reviews are disproportionately high for Amazon as they significantly ramped up their hiring since 2010.

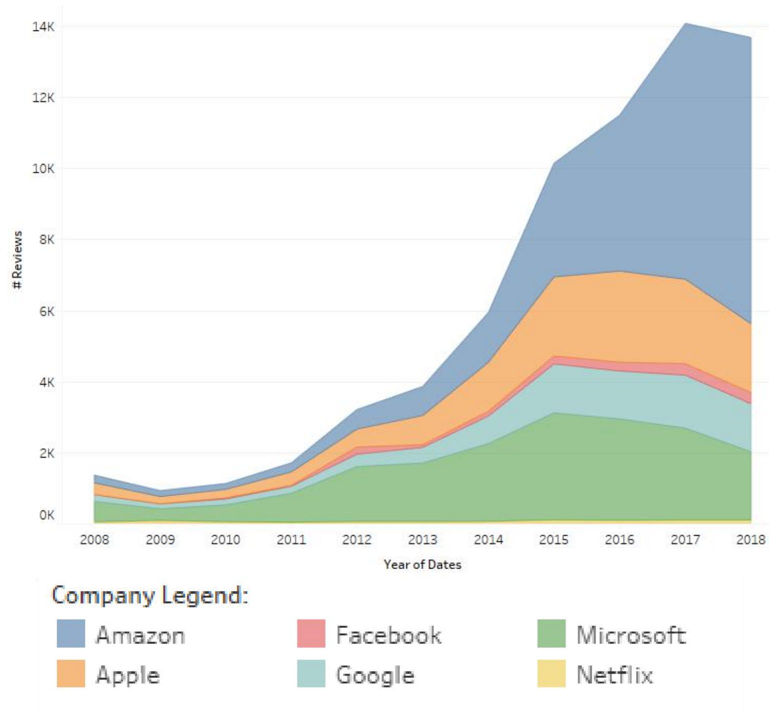

Figure 3: Distribution of reviews over the years

Along with the overall star rating for the review, reviewers are also asked to rate their company for other aspects of the job such as "Job Work/Life balance", "Compensation/ Benefits", "Job Security/ Advancement", "Management", and "Job Culture". The mean distributions of these variables for each company are summarized in Table 1.

Table 1: Mean distributions of star ratings

Company	**Job Security/ Advancement**	**Compensation/ Benefits**	**Job Culture**	**Management**	**Job Work/Life balance**	**Overall Star Ratings**
Amazon	3.60	3.69	3.53	3.17	3.01	3.59
Apple	3.42	4.00	4.10	3.45	3.35	3.96
Facebook	4.36	4.55	4.51	4.26	3.92	4.51
Google	3.96	4.36	4.35	3.83	3.98	4.34
Microsoft	3.66	3.97	3.66	3.13	3.57	3.82
Netflix	3.06	4.06	3.52	3.17	3.21	3.41

3. Introduction to Natural Language Processing:

Natural Language Processing (NLP) is a subfield of Artificial Intelligence that employs computationally sound techniques to understand, learn, and generate natural language (Hirschberg & Manning, 2015). The idea of leveraging and training machines to understand text is not new, in fact origins of NLP can be dated back to 1950s (Nadkarni et al., 2011).The reason for recognizing NLP as a sperate sub-field in AI is to put a special emphasis on the different representation and processing techniques that need to be used. Generally, there are three phases involved in an NLP implementation – text pre-processing, word-vector representation, and machine learning.

Text comes in variety of formats, and the pre-processing techniques ensure us to obtain certain levels of standardization. Stop word removal, punctuation removal, special characters removal, and stemming are the common pre-processing methods.

The process of generating word-vectors involves breaking down text to it's fundamental forms such as words/ alphabets and representing them in numeric formats called embeddings. There are several innovations in the field of word embeddings and these innovations are directly proportionate to the progress that's being made in NLP. Different methods to represent text is concisely summarized by Camacho-Collados et al.(Camacho-Collados & Pilehvar, 2018).

4. Analyzing employee reviews to generate insights:

In this section, we discuss three different techniques to understand and analyze employee reviews. We first start with word frequency and n-gram based methods, we then extract the sentiments associated with each employee review. We conclude our analysis by conducting Latent dirichlet allocation, a topic modeling technique to uncover latent topics of discussion.

4.1 N-gram and word frequency-based analysis:

An N-gram is n N-character substring of a larger sequence of characters. The characters in this case can be words, alphabets, DNA sequences, or any temporal data structure. Typically, N-gram representations slices the string into a list of overlapping sequences (Cavnar & Trenkle, 1994). An example N-gram representations of a sample review is presented in Table 2.

Table 2: N-gram representation of a sentence

N	Original sentence	N-gram representation
Unigram (1-gram)	The company provides great benefits	['The', 'company', 'provides', 'great', 'benefits']
Bi-gram (2-gram)	The company provides great benefits	['The company', 'company provides', 'provides great', 'great benefits']
Tri-gram (3-gram)	The company provides great benefits	['The company provides', 'company provides great', 'provides great benefits']

The reviewers are asked to present "Pros" and "Cons" separately in their respective text boxes. Our intention is to comprehensively understand the positives and negatives of a company, so we combined the pros and cons to form a full review. We the conducted pre-processing steps such as stop word removal, convert text to lower case, remove special characters, and lemmatize the words.

Figure 4 resents the most frequently used N-grams. Just by looking at the most frequent N-grams such as "smart people", "long hour", "great benefit", one can easily come up with general themes.

(work,)	50025	(work, life)	4918	(work, life, balance)	1850
(great,)	29369	(life, balance)	4740	(great, place, work)	936
(people,)	27412	(great, benefit)	3849	(good, work, life)	526
(company,)	26524	(smart, people)	3715	(great, company, work)	473
(good,)	26518	(worklife, balance)	3449	(great, benefit, great)	439
(benefit,)	17928	(place, work)	3080	(great, people, work)	388
(get,)	16898	(work, environment)	2760	(lot, smart, people)	382
(lot,)	16542	(good, benefit)	2305	(great, work, environment)	341
(time,)	15706	(long, hour)	2125	(good, worklife, balance)	333
(employee,)	12920	(people, work)	2101	(great, people, great)	326

a).Most frequent unigrams b) Most frequent bi-grams c) Most frequent tri-grams

Figure 3: Most frequent N-grams

We summarized the most frequent n-grams for the reviews that received an overall star rating in the bottom two box (one or two stars). The information is summarized in Figure 5. Firgure 5 presents a more clear insights into the potential problems that are expressend in the reviews as there received a very low start rating. Tri-grams such as "10 hour shift", "15 minute break", "high turnover rate" indicate that people are generally expressing their concerns around long working hours and attrition rates.

(work,)	8701	(work, life)	655	(work, life, balance)	619
(people,)	5508	(life, balance)	639	(10, hour, shift)	68
(good,)	4768	(worklife, balance)	547	(lot, smart, people)	64
(company,)	4518	(smart, people)	475	(great, place, work)	57
(get,)	4463	(good, benefit)	446	(look, good, resume)	56
(management,)	3992	(great, benefit)	372	(15, minute, break)	55
(time,)	3776	(long, hour)	353	(pay, good, benefit)	51
(manager,)	3775	(good, pay)	330	(hour, per, week)	47
(employee,)	3668	(work, environment)	319	(work, long, hour)	44
(job,)	3048	(place, work)	301	(high, turnover, rate)	43

a) Most frequent unigrams b) Most frequent bi-grams c) Most frequent tri-grams

Figure 4: Most frequent N-grams for 1- and 2-star reviews

This general framework can be followed to analyze the data at different dimensions. Reviews by -company, job title, year, country, employee status (current vs. former) are some of the ways to further slice and dice this data. The level of detail depends on the business case and organizational priorities.

4.2 Sentiment Analysis:

Sentiment analysis deals with the identification and extraction of sentiment from a given body of text using NLP techniques. Sentiment analysis can be used to understand emotions, opinions, speculations among others (Mejova, 2009). Sentiment analysis has many layers and variations. For example, there are structured, semi-structured, and unstructured sentiment analysis techniques (Medhat et al., 2014). Sentiment analysis can also be conducted at different scales. For instance, we can generate the words that imply sentiment, we can predict the polarity of the text (positive vs. negative), or we can generate multi-point scale sentiment analysis. Sentiment analysis is an active research area, and it is common to find custom sentiment analysis classifiers that are trained on specific domain or for a specific task. Since our goal is to demonstrate the applicability of sentiment analysis to measure employees' sentiment, we use a pretrained classifier called VADER. VADER uses a simple rule-based model for general sentiment analysis by constructing a gold-stand list of lexical features that indicate opinion (Hutto & Gilbert, 2014). The VADER sentiment analyzer provides the positive, neutral, negative, and compound scores for each review. A value of '-1' indicates extremely negative sentiment and a value of '+1' indicates a strong positive sentiment.

Figures 6 and 7 summarize the results from our sentiment analysis. Several inferences can be made from the results. For instance, Google has the highest mean positive scores while Amazon has the most negative scores. Similarly, from Figure 7, we can observe that mean negative and positive scores are increasing year over year. This can indicate that employees are becoming more expressive in their reviews by stating their opinions.

Company	Compound	Neg	Neu	Pos
Amazon	0.4438	0.0835	0.7046	0.2119
Apple	0.6046	0.0713	0.6834	0.2453
Facebook	0.7002	0.0525	0.7135	0.2340
Google	0.6266	0.0593	0.6779	0.2628
Microsoft	0.5883	0.0641	0.7008	0.2351
Netflix	0.5137	0.0817	0.7084	0.2099

Figure 5: Mean VADER sentiment scores by company

Like n-gram analysis, sentiment analysis has the potential to uncover insights at different dimensions.

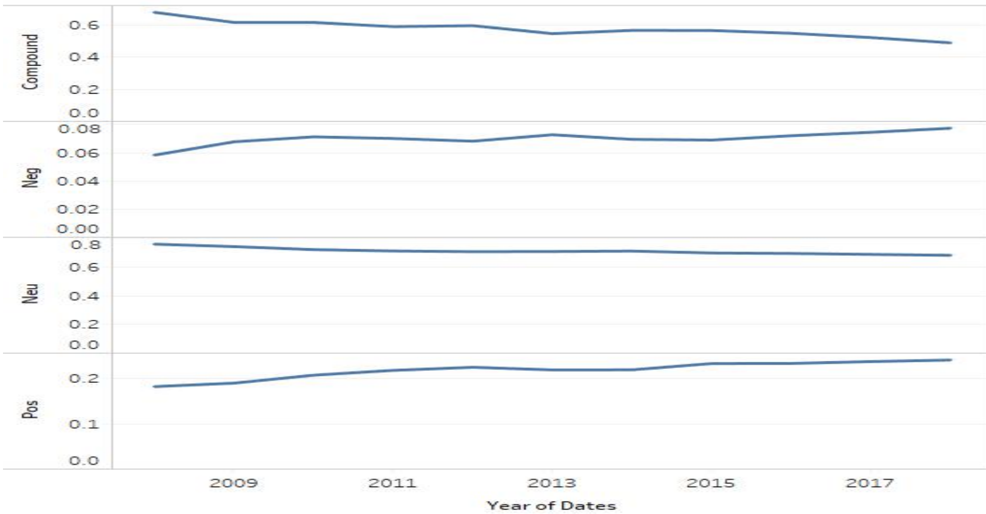

Figure 6: Mean VADER sentiment scores by year

4.3 Topic modeling:

Topic modeling technique is of the most trusted ways to explore latent topic of aspects of discussion. Topic models are especially useful when applied on a large text corpus such as reviews/ surveys as it has the power to accurately summarize information. In general, topic modeling can be thought of an unsupervised learning approach to cluster documents or reviews. Topic models help us to represent a large collection of documents in "k" topics. The topics are generally inferred by observing the key words and the corresponding weights. These topics can then be assigned to the original document.

Our research in regards to the topic modeling is largely based on the work by Madding et al. (Madding et al., 2020). Madding et al. have conducted original research on Indeed employee reviews by generating topic models using Latent Dirichlet allocation (LDA) and Attention Based Aspect Extraction. They have discovered eleven useful topics as an outcome of their analysis. Because our goal is to propel research interest in this space, we limit our experimentations to LDA. We also limit our number of chosen topics to five. The number of topics for an LDA algorithm is a hyper-parameter and can be tuned to generate course-grained or fine-grained insights. Table 3 represents the inferred topics based on the key words. These topics can then be assigned back to each individual review to conduct a thorough analysis.

Table 3: Inferred topics based on key words

Inferred Topic	**Key words**
People related	Manager, employee, people, time, management
Working hours	Hour, time, long, flexible, long_hour, free, home
Working style	Work, worklife_balance, work_environment
Tools/ Technology	Agile, teams, python, java
Benefits/ Career	Health_benefit, career_path, career_advancement, internal_politics

We also used a visualization technique called LDAvis (Sievert & Shirley, 2014). This interactive technique helps us to select a topic and visualize the top-30 salient terms for each topic. The size of the topic bubble represents the prevalence of that topic among the reviews.

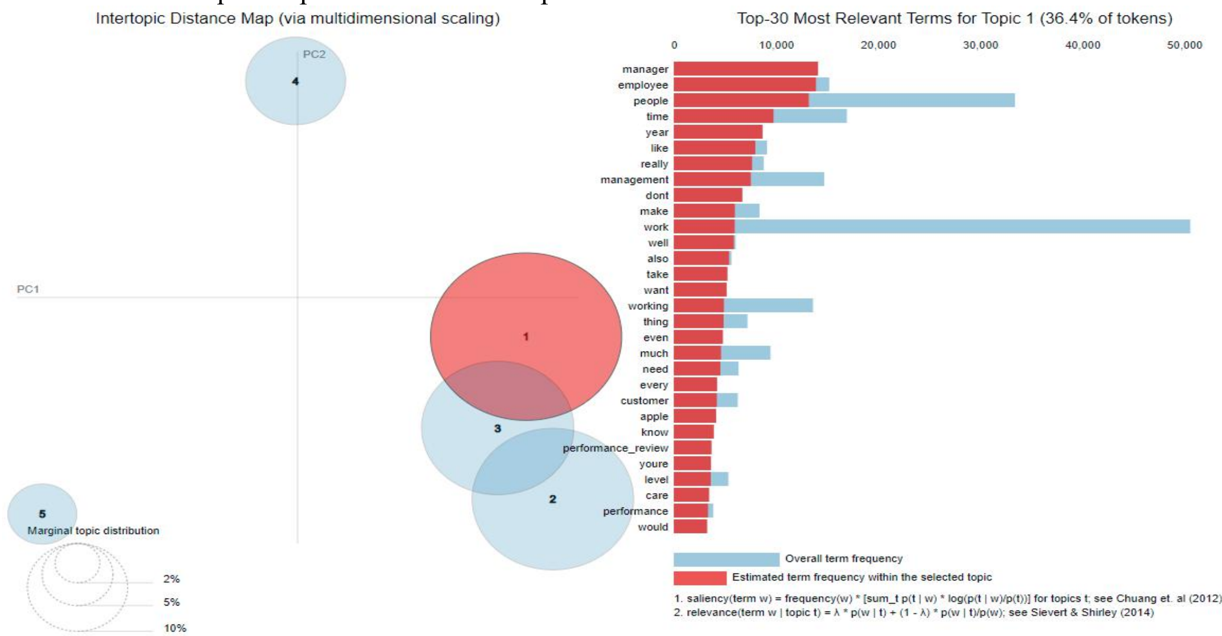

Figure 7: LDAvis, a visualization technique to interact with topic models

5. Conclusion:

We conducted a thorough analysis of the Indeed reviews from top technology companies with an intention to present text-based reviews as an excellent alternative to canned employee surveys. The three techniques employed in this study have different use cases, and they also offer varying degrees of implementational complexities. We found that N-gram-based word frequency analysis is best suited when the HR partners have limited time/ resources and want to get their employees' pulse quickly. Bi-grams and tri-grams were most helpful when we analyzed the data at different dimensions.

On the other hand, Sentiment analysis provides a different flavor to **measuring** your employee experience. This is simply not possible to do with traditional employee surveys. Sentiment analysis techniques have the power to significantly transform your employee listening strategy as several established sentiment classifiers can be readily applied to the reviews.

Finally, the LDA-based topic models are most valuable when the HR leaders want to understand emerging topics. Topic models provide a competitive advantage to HR teams as they can move from reactive to proactive approaches because of the hidden insights generated by topic models. However, it takes skill and experience to finesse topic models.

References:

1. Camacho-Collados, J., & Pilehvar, M. T. (2018). From word to sense embeddings: A survey on vector representations of meaning. *Journal of Artificial Intelligence Research*, *63*, 743–788.
2. Cavnar, W. B., & Trenkle, J. M. (1994). N-gram-based text categorization. *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, *161175*.
3. Chowdhary, K. (2020). Natural language processing. *Fundamentals of Artificial Intelligence*, 603–649.
4. Church, A. H., & Waclawski, J. (2017). *Designing and using organizational surveys*. Routledge.
5. Fry, J. P., & Sibley, E. H. (1976). Evolution of data-base management systems. *ACM Computing Surveys (CSUR)*, *8*(1), 7–42.
6. Ghasemaghaei, M., & Calic, G. (2019).

Does big data enhance firm innovation competency? The mediating role of data-driven insights. *Journal of Business Research*, *104*, 69–84.
7. Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261–266.
8. Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225.
9. Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, *50*, 57–70.
10. Madding, C., Ansari, A., Ballenger, C., & Thota, A. (2020). Topic Modeling to Understand Technology Talent. *SMU Data Science Review*, *3*(2), 16.
11. Markham, S. K., Kowolenko, M., & Michaelis, T. L. (2015). Unstructured text analytics to support new product development decisions. *Research-Technology Management*, *58*(2), 30–39.
12. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113.
13. Mejova, Y. (2009). Sentiment analysis: An overview. *University of Iowa, Computer Science Department*.
14. Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, *18*(5), 544–551.
15. Plaskoff, J. (2017). Employee experience: the new human resource management approach. *Strategic HR Review*.
16. Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, *1*(1). https://doi.org/10.1089/big.2013.1508
17. Renkema, M. (2022). Consequences of artificial intelligence in human resource management. In *Handbook of Research on Artificial Intelligence in Human Resource Management*. Edward Elgar Publishing.
18. Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010). The hadoop distributed file system. *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, 1–10.
19. Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70.
20. Tussyadiah, I., & Zach, F. (2015). Hotels vs. peer-to-peer accommodation rentals: Text analytics of consumer reviews in Portland, Oregon. *Peer-to-Peer Accommodation Rentals: Text Analytics of Consumer Reviews in Portland, Oregon (April 10, 2015)*.