

LEVERAGING MACHINE LEARNING ALGORITHMS TO GAIN INSIGHTS INTO THE MINDSETS OF IT PROFESSIONALS IN MUMBAI

Dr. Kavitha Venkatachari

Dean - School of Ai and Future technologies, Analytics, Universal Ai University, Karjat City,
Kushiwali, PO Gaurkamath, Vadap, Raigad, Mumbai, Maharashtra

Objective of the study: The objective of this study is to employ HR analytics, utilizing machine learning algorithms and descriptive statistical methods on unstructured datasets, to gain insights into the perceptions of employees regarding the organization. The focus is on understanding employee satisfaction levels and discerning factors that influence their mindset, ultimately aiding HR departments and decision-makers in optimizing Return on Investment (ROI) from human capital.

Methodology: In this study, the author utilizes R programming software to comprehensively comprehend and forecast the determinants of job satisfaction. Employing advanced machine learning techniques, specifically text analytics and sentiment analysis, the research aims to predict patterns and discern sentiments and opinions concerning the organization.

Research design and approach: The author distributed open-ended questionnaires to mid-level employees within the IT sector, collecting and analysing a total of 250 responses for this study. The data exclusively pertains to Mumbai, with study limitations acknowledged regarding sample size and employee category. The author sourced data from both multinational IT corporations and national-level companies, recognizing that the study's outcomes may exhibit variations based on diverse sectors and employee hierarchies.

Findings and recommendations: The study reveals that IT employees experience notable work pressure, with challenges related to their ideas being acknowledged by superiors. However, the younger generation expresses contentment with the provided facilities and the overall work environment. Subsequently, the collected data will undergo training and testing processes. Through this, a model will be developed, allowing for the evaluation of methods to predict the probability of an employee either staying or leaving the company.

Keywords: *HR analytics, Employee job satisfaction, text analytics, R programming, Sentiment analysis*

Introduction: Human Resource Analytics, a transformative frontier in contemporary business, has gained substantial momentum in India, reshaping the conventional contours of human resource management. As the intricate tapestry of India's business landscape undergoes rapid metamorphosis, propelled by globalization, technological innovation, and a competitive talent market, the strategic adoption of HR analytics emerges as a pivotal catalyst for organizational success. In this dynamic milieu, where the workforce is diverse and the demand for skilled

professionals is escalating, HR analytics provides a sophisticated, data-driven approach to navigating the complexities of talent management. The adoption of HR analytics signifies a shift from traditional, intuition-based decision-making to a future-oriented, evidence-based paradigm. It heralds a new era where HR professionals leverage advanced data analytics, statistical models, and cutting-edge technology to glean valuable insights into workforce dynamics, employee engagement, talent acquisition, and overall organizational efficacy. Strategic workforce planning stands out as one of the primary advantages conferred by HR analytics in the Indian context. By analysing historical data and predicting future trends, organizations can strategically align workforce capabilities with overarching business goals, ensuring a timely and effective deployment of talent. Talent acquisition, another critical facet, is revolutionized through HR analytics, enabling organizations to identify optimal recruitment channels, analyze candidate profiles, and streamline the hiring process for efficiency and diversity. Moreover, the heightened competition for skilled professionals necessitates a nuanced focus on employee engagement and retention. HR analytics empowers organizations to assess employee satisfaction, pinpoint key drivers of engagement, and predict factors influencing attrition, allowing for the implementation of targeted strategies to enhance the overall employee experience and curtail turnover.

Performance management, learning and development, diversity and inclusion, and workforce health and well-being are additional domains where HR analytics manifests its transformative impact. The objectivity brought to performance management allows organizations to identify high-performing employees, recognize areas for improvement, and tailor development programs to enhance overall productivity. Continuous learning, essential in a rapidly evolving business landscape, benefits from HR analytics through the assessment of training program effectiveness, identification of skill gaps, and personalization of learning experiences. In the realm of diversity and inclusion, HR analytics facilitates the measurement of diversity metrics, progress tracking, and the identification of areas for improvement, fostering an inclusive workplace culture that values diversity across all dimensions. Furthermore, the emphasis on employee well-being, a cornerstone of organizational health, is enhanced through HR analytics, allowing organizations to assess well-being, identify stressors, and implement initiatives promoting a healthy work-life balance. However, the transformative potential of HR analytics in India is not without its challenges. The journey towards data-driven HR practices necessitates a cultural shift within organizations, the development of robust technological infrastructure, and the cultivation of a skilled workforce capable of interpreting and leveraging data insights effectively. Concerns related to data privacy and ethical considerations underscore the imperative of responsible data use in the pursuit of optimal human resource management. As India positions itself as a global economic powerhouse, the adoption of HR analytics emerges as a strategic imperative, reshaping the contours of talent management and organizational success. This paradigm shift towards data-driven HR practices not only propels India into a future where human capital is harnessed with precision but also serves as a testament to the transformative power of analytics in the intricate domain of human resource management. It is in this convergence of technology, strategy, and

human capital that the future of work in India is being meticulously crafted, paving the way for a new era of organizational excellence and talent optimization.

Purpose of data mining techniques: In the realm of structured and unstructured data, analysts predict and segregate file patterns, employing meticulous methods for insightful analysis. Extracting relevant patterns and behaviors from unstructured datasets is a complex process, necessitating the application of machine learning algorithms. Data mining proves pivotal, especially in predicting scenarios like bankruptcy and identifying potential loan defaulters. The application of association mining unveils buyers' purchase patterns in retail sales data, aiding decision-makers in identifying correlations among seemingly unrelated products commonly purchased together. Market basket analysis sheds light on retail dynamics, exposing the frequency of sales for specific items and revealing customer-preferred product pairings. In the financial sector, particularly banking, data mining techniques are instrumental in identifying fraud transactions through debit cards by pinpointing abnormal data indicative of potential data entry errors. The multidimensional nature of data, foundational in data warehousing and mining systems, adds depth and sophistication to analytical pursuits, playing a pivotal role in discerning intricate data patterns. The integration of machine learning algorithms, data mining, and multidimensional data analytics collectively elevates the analytical landscape, offering valuable insights into patterns, trends, and behaviours across diverse datasets.

Structured and unstructured data: In an organized file structure, fields are stored with fixed lengths, exemplified by student enrolment numbers, phone numbers, Aadhaar card numbers, and PIN codes. Student names, divided into first, last, and middle names, are stored as text or string formats for accessibility, storage, analysis, and interpretation. Creating tables in Ms Access or Structured Query Language involves defining fields in a design view. For instance, a database administrator creating a train ticket reservation system designs tables for trains, passenger details, and payment information. Each table requires a primary key, which serves as a secondary key for connecting with other tables. In the passenger details table, unique keys for passengers, including first name, last name, middle name, age, gender, and travel details, are stored.

Background of the study: Job satisfaction, regarded as an individual's emotional contentment with their work, serves as a crucial motivational factor in professional life. This concept delineates the relationship between individuals, ranging from low-level workers to high-level blue-collar professionals, and their employers. This research centres on employing statistical techniques to explore job satisfaction, with a specific focus on leveraging data mining methods to understand the satisfaction levels of IT sector employees. Many IT professionals' express dissatisfaction, citing monotonous tasks leading to boredom and stress. Additionally, the ever-evolving nature of the IT field, introducing new concepts and themes, becomes a source of frustration for some employees, impacting their overall happiness. Job satisfaction, a prominent subject in organizational behavioural research, reflects an employee's emotional attitude toward their work, influenced by factors such as the work environment and organizational characteristics.

Establishing a high-performance and well-compensated work system is identified as a key driver for achieving both organizational productivity and heightened job satisfaction among employees.

Review of literature:

In Lahoti and Ramteke's (2014) research, various mining techniques and their applications in functional areas are explored. Cluster analysis unveils similar groups and outliers, while data mining tools predict valuable patterns from digital records in fields like finance, marketing, and business. Business intelligence tools and artificial intelligence models aid in understanding human behaviour and changing attitudes. Data mining tasks encompass anomaly detection, association rule formation, clustering, classification, and regression. Techniques such as neural networks, genetic algorithms, and decision trees are employed for analysis. The use of public and private clouds, particularly in education through Educational Data Mining, facilitates efficient data storage and complex query resolution.

In Kumari and Pandey's (2011) research, the focus lies on measuring job satisfaction's impact on employee productivity and organizational growth. Satisfied employees contribute to increased productivity, fostering organizational competitiveness. The study delves into diverse sectors, exploring the relationship between scalability, various factors, and job satisfaction. Utilizing the C4.5 algorithm, the research involves preprocessing datasets for analysis. C4.5, a decision tree algorithm, generates rules by removing redundancy and transforming data into a normalized form. The study employs supervised learning techniques, commonly applied in neurosurgery analysis and data churn in the telecommunications sector, to decipher the intricacies of job satisfaction across industries.

Jantan, Yusoff, and Noh (2012) emphasize the crucial role of human capital in maintaining organizational competitiveness. Their research employs the Support Vector Machine (SVM) algorithm, specifically the Sequential Minimal Optimization (SMO) technique, to match individuals with suitable roles, addressing optimization challenges. The study utilizes academic performance data from Malaysian secondary school teachers, employing a four-level classification with 14 attributes. While acknowledging the need for enhanced accuracy, the authors suggest the incorporation of other SVM algorithms like Grid Search and Gabriel graph. The research concludes that the novel SVM classification in Human Resource Management (HRM) holds promise for advancing talent management activities.

Gap analysis: The literature review reveals diverse applications of data mining techniques, emphasizing sentiment analysis through social media platforms like Twitter to gauge customer satisfaction and emotions. Notably, studies employ algorithms such as naïve Bayes, C4.5, and J4 for analysis, with tools like Weka and SPSS commonly utilized. In this research, open-source R is employed to assess IT employees' job satisfaction and sentiments, uncovering novel patterns in their satisfaction levels.

Research design and Research approach/methodology applied: This research employs association rule mining, clustering, sentiment analysis, and text mining to identify patterns in unordered data, offering valuable insights into employee mindset. Through techniques like sentiment analysis and text mining, the study analyses the sparsity, word cloud, and sentiments (positive, negative, neutral) expressed toward the organization. The unstructured questionnaire, distributed through Google Forms, garnered 275 responses from 300, shedding light on employees' sentiments and opinions. The study aims to explore the analysis of unstructured or semi-structured data using data mining and differentiate between sentiment analysis and text analytics.

Purpose of the study:

- Explore the application of data mining techniques in the analysis of unstructured or semi-structured data, such as text files.
- Investigate the distinctions between sentiment analysis and text analytics.

Hypothesis:

- Utilizing text analytics, we aim to predict and identify patterns.
- Does a correlation exist between job satisfaction and work timings, as perceived by employees?
- Can we establish a connection between satisfaction and the company/department through word cloud analysis?
- Is there an association between satisfaction, experience, and future outlook from an employee's standpoint?

Limitations of the study:

- The experimental study's limited population restricts the generalizability of results to other states or companies.
- A broader participant pool representing diverse employee stages would enhance the study's comprehensiveness.
- Non-disclosure of personal information by some respondents due to busy schedules and job security concerns.
- The study's scope is confined to IT companies specifically in Mumbai.
- The questionnaire is exclusively distributed among middle-level managers.

Text mining, also known as text analytics or text data mining, involves extracting valuable information and patterns from unstructured textual data. Various algorithms and techniques are employed in text mining to analyse and process large volumes of text. Naïve Bayes algorithm: Utilized for classification tasks, it calculates the probability of a text belonging to a particular category based on word frequencies. Support Vector machines (SVM): An algorithm for both classification and regression tasks. SVM can be applied to categorize texts into predefined classes' means clustering: A clustering algorithm that groups similar texts together based on their content. It helps discover patterns and themes in the data. Latent semantic analysis: LSA identifies the

relationships between words and topics within a large corpus. It is effective for extracting the underlying structure in textual data. Decision trees can be used for text classification. They create a tree-like model based on features extracted from the text. Tokenization, stemming, and lemmatization are common NLP techniques used in text mining to preprocess and normalize text data. LDA is a popular algorithm for topic modelling, helping identify the main themes or topics present in a collection of documents. Word embeddings represent words in a continuous vector space, capturing semantic relationships. These embeddings are useful for tasks like sentiment analysis and document similarity. Beyond Naive Bayes and SVM, other classification algorithms like Random Forest, Logistic Regression, and Neural Networks can be applied to categorize text data. Specifically designed for determining sentiment in text, VADER and similar algorithms assign sentiment scores to words and phrases.

Procedure to extract data from the text:

1. Load the data sets
2. Save the text in a text file format.txt

Install all packages

- TM
- Corpus
- Gg plot
- Clor brewer

Load the R package for text mining and then load text file into R package

- Data is converted into a corpus file

Pre processing

- Special characters are removed from the text
- Stop words are removed from the text file.
- Stop words means it won't make any sense to the output.
- Ending words like 'ing', 'es', 's' should be removed.
- White spaces must be removed from the text file.

Stage 1

- The text file is converted into a matrix format
- Then transpose matrix
- Measure the sparsity of the matrix
- Least and the most occurring words are found.
- The word having highest frequency (many times) are measured.
- Plot the frequency

Stage 2

- Superior subordinate
- Progression
- Opportunity
- Involvement

Associations reveal that work satisfaction is closely linked to interactions with supervisors and higher-ups. Employee contentment often stems from supportive bosses who actively engage in decision-making processes. The data provides valuable insights for HR to gauge industry perceptions, organizational involvement, and sentiment polarity among employees.

Find_Assocs_1(dtm, c("satisfaction" , "company"), corlimit= 1.0)

- | | |
|--------------------------|------------------|
| • Motivation | respect |
| • Workplace | remuneration |
| • Training | relaxation |
| • Timing | recognizes |
| • Supportive | promotional |
| • Supervision | participative |
| • Superior subordinates' | friendship |
| • Salary | responsibilities |

During the association analysis, the machine correlates the variables, revealing insights from IT sector employees. The findings indicate that the workforce perceives the company as supportive, providing motivation, promotions, and training. This perspective is particularly prominent among the younger, tech-savvy generation adaptable to the latest technologies.

Conclusion: In conclusion, the study sheds light on the varying mindsets among employees, emphasizing the positive sentiment predominantly observed in the younger generation. However, it highlights concerning issues within the middle-level management, particularly those with over a decade of experience, struggling with work pressure and a lack of involvement in decision-making processes. The findings underscore the importance of addressing the concerns of this managerial segment to enhance retention and overall job satisfaction.

The study suggests that proactive measures, informed by sentiment analysis, could enable HR and management to discern employee opinions regarding superiors, organizational aspects, and productivity. By delving into the sentiments expressed, organizations can pre-emptively identify potential challenges, fostering a more supportive and engaging work environment. The application of statistical tools provided valuable insights into the significance of relationships between specific variables. Nevertheless, the research advocates for the integration of advanced data mining methods capable of accommodating multiple variables in diverse data structures to comprehensively gauge employee satisfaction levels.

Ultimately, the paper serves as a crucial resource for management, offering a genuine depiction of employee sentiments and concerns. By leveraging sentiment analysis and data mining techniques, organizations can proactively address issues, improve workplace dynamics, and create strategies that resonate with the diverse needs of their workforce. This approach facilitates a more informed and strategic decision-making process, enhancing overall employee satisfaction and contributing to the long-term success and sustainability of the organization.

Implications: The implications of the study's output are substantial for organizational management and human resources. Firstly, the positive sentiment among the younger workforce signals a positive workplace culture. However, the challenges faced by the experienced middle-level managers underscore the need for targeted interventions to address work pressure and enhance their involvement in decision-making. The findings highlight the importance of tailoring strategies to different employee segments, recognizing the unique needs of each group.

HR departments can leverage sentiment analysis to gain deeper insights into employee opinions, enabling the identification of potential areas for improvement. Proactive measures, such as mentorship programs and stress management initiatives, could be implemented to support middle-level managers. Additionally, fostering a collaborative decision-making environment can enhance overall job satisfaction. Overall, the study's implications advocate for a nuanced and personalized approach to employee engagement and satisfaction, ultimately contributing to a more harmonious and productive workplace.

References:

- Anup Arvind Lahoti, Prof. P. L. Ramteke (2014),” Data Mining technique its needs and using its applications”, International journal of computer science and mobile computing, ISSN 2320-088X, Vol. 3, Issue. 4, April 2014, pg. no. 572-579.
- Geeta Kumari and K. M. Pandey (2011),” Job Satisfaction in Public Sector and Private Sector: A Comparison”, International Journal of Innovation, Management and Technology, Vol. 2, No. 3, Pg. No. 222-228, June 2011.
- Hamidah Jantan, Norazmah Mat Yusoff and Mohamad Rozuan Noh (2012),” Intelligent DSS for Talent Management: A Proposed Architecture using Knowledge Discovery Approach”, 6th International Conference on Ubiquitous Information Management and Communication, DOI: 10.1145/218475.2184857, February 2012.
- Maurizio Pugno and Sara Depedri (2009), “Job Performance and Job Satisfaction: An Integrated Survey”, Discussion Paper No. 9, Universita Delgi Studi Di Trento, Department Of Economics, 2009.

Web references:

- <http://searchbusinessanalytics.techtargert.com/definition/t-ext-mining>.

- <http://guides.library.duke.edu/content.php?pid=383688&sid=3143978>.
- http://en.wikipedia.org/wiki/Stop_words.
- <http://en.wikipedia.org/wiki/Stemming>.
- <http://consultingblogs.emc.com/manjunathasubbarya/archive/2011/03/04/stemming-and-lemmatization.aspx>.
- <http://www.techopedia.com/definition/13698/tokenization>.
- <http://en.wikipedia.org/wiki/Tokenization>
- The Text Mining Handbook—Advanced Approaches in Analyzing Unstructured Data – Ronen Feldman, James Sanger.
- Tapping into the Power of Text Mining – Weiguo Fan, Linda Wallace, Stephanie Rich, and Zhongju Zhang.
- A Survey of Text Mining Techniques and Applications -- Vishal Gupta, Gurpreet S. Lehal.
- Unsupervised cleansing of noisy text – Danish Contractor, Tanveer A. Faruque, Venkata Subramaniam.