

Predictive Talent Sourcing via Deep Linguistic Processing

Regina Rejith ¹, Dr. A Sharmila ²

¹Research Scholar at JAIN (Deemed-to-be University)

Email ID : reginarejith2@gmail.com

²Associate Professor at JAIN (Deemed-to-be University)

Email ID : a.sharmila@jainuniversity.ac.in

ABSTRACT

Predictive Talent Sourcing Using Deep Linguistic Processing is an advanced approach that helps in finding, attracting, and choosing the best candidates by utilizing artificial intelligence, natural language processing, and machine learning techniques. Traditional recruitment methods often rely on human review and keyword searches, but these approaches can overlook qualified candidates because of differences in language, resume formats, and terminology. Deep linguistic analysis enables more accurate and effective talent identification by evaluating the meaning of words, their relevance to the situation, how people communicate, and the professional abilities suggested in the text of resumes, job postings, and professional profiles on social media platforms. An advanced approach enables organizations to assess how well a candidate is likely to perform, align with the company culture, and succeed in their role by considering their experience, abilities, and behavioral traits, rather than just focusing on listed job requirements and credentials. Predictive models help in finding passive candidates and uncovering new talent pools by monitoring individuals' online presence and professional activities across various platforms. In addition, using deep linguistic processing enables the customization of candidate engagement initiatives and employer branding by developing specifically designed communication strategies. In today's digital age, businesses that use predictive talent sourcing strategies gain a competitive advantage in attracting the best candidates

Keywords *Predictive Talent Sourcing, Deep Linguistic Processing, Artificial Intelligence, Natural Language Processing, Recruitment Analytics, Talent Acquisition, Machine Learning, HR Technology*

INTRODUCTION:

Predictive analytics involves using past data and advanced tools such as statistical methods and machine learning techniques to forecast what might happen in the future. When we hire, we may use data to gain a better understanding of past hiring experiences and identify the qualities that contributed to the success of certain candidates. By doing this, we can apply our analysis to select individuals who are likely to succeed in a role and exclude those who do not have the required skills. It enables the collection of various types of data, such as resumes, interview outcomes, workplace behavior, and patterns that explain why individuals choose to either remain or depart from their positions. By identifying these patterns, it becomes simpler to assess an employee's potential for success, recognize their strengths and areas for improvement, and determine how well they might integrate into an existing team. In summary, the hiring process becomes more informed, leading to the creation of teams that achieve a greater level of performance. Predictive analytics is transforming the way companies approach hiring and is reshaping the future of recruitment by introducing new methods for selecting employees. It enables a clearer understanding and more effective decisions that are based on strong evidence instead of assumptions. The success of future recruitment strategies depends on organizations that effectively blend the use of

data with a people-focused approach. With the use of predictive analytics, businesses can enhance their recruitment processes and foster a genuinely inclusive and diverse workplace that allows individuals to succeed. The use of artificial neural networks has transformed natural language processing because of deep learning techniques that mimic how the human brain works, such as the backpropagation method. Deep learning has produced remarkable outcomes across many fields by creating highly abstract representations through the layered processing of input data. Natural language processing is a field within deep learning that combines the ability to understand human language with the self-learning capabilities of neural networks. By employing advanced models like Transformers and LSTMs, it became feasible to automate the detection of complex patterns in texts – a task that was previously unsolvable using traditional rule-based methods. This integration was essential for continued progress in machine translation, sentiment analysis, text generation, and led to the creation of the large language models used today. In many ways, this marked the start of a new era where language understanding became more intelligent and closely resembled human comprehension.

3. Research Questions

3.i) RQ1: How can advanced linguistic analysis enhance the precision of identifying promising individuals?

3.ii) RQ2: To what extent does sourcing with a focus on data loss prevention compare to traditional resume screening methods?

4. Research Objective

4.i.)The research objective is to examine language patterns within job market candidate data in order to identify indicators of potential high performance.

4.ii) To assess how well language-based prediction methods work compared to conventional approaches used for sourcing.

4.iii) To determine which linguistic features are strong indicators of performance and how well they fit.

5. Motivations

I must confess that this subject interests me greatly and motivates me to continue with further academic exploration, as it plays a crucial role in my Ph.D. research. The study of this topic enhances my understanding of fundamental concepts and rekindles my interest in acquiring new knowledge. As a talent expert, I cannot ignore the most recent developments in technology and how these innovations are being effectively used in real-world recruitment scenarios. This greatly assists me in staying aligned with the changing employment market and making well-informed choices about talent management.

6. Review of Literature

6. i) Mankolli, Emiliano & Bushati, Senada (2023) conducted a study in “Candidate Engagement Success Prediction Using Machine Learning and Natural Language Processing Techniques” using the methodology “New Method” on selecting candidates on multiple parameters, efficiency, and accuracy. This study showed that applying the natural language processing method to data for candidate’s profiles increases overall accuracy. The XGBoost method implementation shows the highest accuracy, enhancing the candidate's experience in predicting the candidates suitable for the job.

6. ii) Mankolli, Emiliano. (2022) made a study on “Reducing the complexity of candidate selection using Natural Language Processing” focused to language processing, which havenderstand the concepts 3 major challenging terms Artificial intelligence, machine learning and natural language processing, which have impacted the processing of large data for candidates without any limitation. The approach was a hybrid process that had 2 combinations of toolcandidate, XGBoost and BERTopic. Time and memory usage, along with similarity of context, provide higher accuracy and selection of candiates that fit the open vacancy.

6. iii) Raja Kishore Babu, C. H., et.al., (2025) studied on the Personalized Job Search with AI: A Recommendation System Integrating Real Time Data and Skill Based Matching. The authors focused on understanding the optimizing career matching and making the job search easier using the Natural Language Processing and Large Language Models. RapidAPI job search to filter location based results, optimise search efficiency, job to candidates, and real-time recommendations has been highlighted in this research paper. Finding and

suggestions by the authors were on the skill gaps, job ranking and professional profile integration, which are some of the futuristic AI-driven developments that can be seen in the job market for improvements and revolutionary capability.

6. iv) Maciej Wach and Iwona Chomiak-Orsa (2021) conducted a study of a mining comany using the decision making process using the predccictiove analytics of a investmenet protfolio. Among the multiple approaches used to predict the budget deviation, the best applied approach was discussed in their research paper, “The application of predictive analysis in decision-making processes on the example of a mining company’s investment projects.” and the study said the structured project data on understanding the key attributes.

6. v) Kumar, Vaibhav & L., M. (2018). Predictive Analytics: A Review of Trends and Techniques. Focused on studying the predictive analytics in statistical and analytical techniques. Predictive analytics is a term mainly used in statistical and analytics techniques. This term is rooted to classical statistics to predicts the future by analyzing current and historical data. The future behavior of variables can be predicted using the models of predictive analytics. A score is assigned by mostly predictive analytics models and a higher score indicates the higher likelihood of occurrence of an event and a lower score indicates the lower likelihood of occurrence of the event. Historical and transactional data patterns find out the solution for many business and science problems by identifying the risk and opportunities for every individual customer, employee or manager of an organization. this paper, presented a review of process, techniques and applications of predictive analytics.

6. vi.) Jamarani, A., Haddadi, S., Sarvizadeh, R. et al. Made a study on Big data and predictive analytics: A systematic review of applications. The authors researched on its potential harnessed for predictive analytics, a sophisticated branch that anticipates unknown future events by discerning patterns observed in historical data. Various techniques from modeling, data mining, statistics, artificial intelligence, and machine learning were employed to analyze available data to extract discriminative patterns for predictors.

This study focused to analyze the research approaches on Big Data Predictive Analytics (BDPA) on published articles since 2014 to 2023. In this article, authors fully concentrated on predictive analytics using big data mining techniques, reviewed 109 articles on Systematic Literature Review (SLR). Based on the application and content of published articles they introduced taxonomy including seven major categories of industrial, e-commerce, smart healthcare, smart agriculture, smart city, Information and Communications Technologies (ICT), and weather. The authors also touched upon benefits and weaknesses of each approach, potentially important changes, and open issues, in addition to future paths. These compiled SLR not only extended on BDPA’s strengths, open issues, and future works but also detects the need for optimizing the insufficient metrics in big data applications, such as timeliness, accuracy, and scalability, that enable organizations to apply big data to shift from

retrospective analytics to prospective predictive if fulfilled.

6. vii.) Zhang Z. (2020). Predictive analytics in the era of big data: opportunities and challenges. *Annals of translational medicines* was a study that used exceptional medicine in large volume to generate from electronic healthcare records, wearable devices and insurance companies. After performing various clinical studies on few scenarios with large volume of data using big data analysis, the risk factors, effectiveness, future analytical accuracy was streamlined. These case control study designs identified risk factors, effectiveness and statistical modeling under different designs. .

6. viii.) Bokonda, Loola & Khadija, Ouazzani Touhami & Souissi, Nissrine. (2020) came together to research on Predictive analysis using machine learning: Review of trends and methods. The study collected data from 3 scientific database, 30 papers were reviewed.

The authors came with a purpose to help anyone who wanted to use the predictive analysis methods in choosing the best Machine learning methods basis their field of application. After reviewing the existing data on the Machine learning methods available, the authors suggested the best Machine learning methods for Education was DT and ANN, building field suitable was DT, LR and RF. Botnay with DT, social science on ANN and RF, and finally Medicine with RF.

7. Description of the methodology

The information used is gathered from publicly accessible data about the candidates, including details about their employment history found on online job boards, professional networking platforms, resume databases, and recruitment websites. It also includes additional data sets like the Kaggle resume dataset, GitHub profile dataset for technical roles, and NLP benchmark corpora related to job opportunities. Additional information has been obtained from existing literature and benchmark performance metrics that are used to identify individuals with high potential.

8. Expected Outcomes

8.i.) Improved ability to predict which candidates are likely to excel in their roles.

8.ii.)Faster recruitment process with reduced workload for recruiters.

8.iii.)Identification of key language patterns associated with top performers.

8.iv.)Development of a scalable, automated system for sourcing talent.

9. Finding and Conclusion

Based on the use of secondary data, the results show that top IT talent exhibits distinct language patterns in their resumes, job profile pages, and social media interactions. Their profiles make use of more action-oriented verbs, include specific measurable achievements, use precise terminology related to their fields of expertise, and highlight effective methods for addressing challenges. Linguistic analysis based on technical portfolios, such as GitHub README files and repositories, provides more

reliable and detailed information about technical skills, ownership, and collaboration compared to analyzing text from resumes.

In addition, recruiting based on linguistic characteristics is more reliable and includes a wider variety of skills compared to traditional methods that rely on keywords. Benchmarking natural language processing datasets and performance metrics from literature enable the adjustment of prediction models without requiring direct access to an organization's performance data. However, variations in the completeness of profiles and the accuracy of self-reported information may reduce the precision of linguistic models when applied in public settings, indicating that these models are better suited for use as screening and prioritization tools rather than for making final hiring decisions.

Deep linguistic processing offers a fresh method for predicting talent by analyzing language patterns that are linked to performance, which traditional hiring methods might overlook. This study shows that models based on DLP produce better rankings of candidates and promote a more active approach to hiring

REFERENCES

- [1] SAS Institute. Predictive Analytics. SAS. Available at: SAS Predictive Analytics
- [2] Talent Management Institute. Predictive Analytics in Recruitment: A Data-Driven Approach to Hiring and Retention. Available at: Predictive Analytics in Recruitment
- [3] Natural Language Processing. Available at: Natural Language Processing Source
- [4] Deep Linguistic Processing. Available at: Deep Linguistic Processing Source
- [5] Mankolli, E., & Bushati, S. (2023). Candidate Engagement Success Prediction Using Machine Learning and Natural Language Processing Techniques. *Proceedings of the Conference on Computer Science and Communication Systems*, 431–435. <https://doi.org/10.1109/CSCS59211.2023.00074>
- [6] Mankolli, E. (2022). Reducing the Complexity of Candidate Selection Using Natural Language Processing. 1–4. <https://doi.org/10.1109/IWSSIP55020.2022.9854488>
- [7] Raja Kishore Babu, C. H., Sreerala, T., Kumar, R., & Sumith, S. (2025). Personalized Job Search with AI: A Recommendation System Integrating Real Time Data and Skill Based Matching. *International Journal of Engineering Research & Technology (IJERT)*, 14(4). ISSN: 2278-0181. Available at: IJERT Article
- [8] Wach, M., & Chomiak-Orsa, I. (2021). The Application of Predictive Analysis in Decision-Making Processes on the Example of Mining Company's Investment Projects. *Procedia Computer Science*, 192, 5058–5066. <https://doi.org/10.1016/j.procs.2021.09.284>
- [9] Kumar, V., & L., M. (2018). Predictive Analytics: A Review of Trends and Techniques. *International Journal*

of Computer Applications, 182, 31–37.
<https://doi.org/10.5120/ijca2018917434>

[10] Jamarani, A., Haddadi, S., Sarvizadeh, R., et al. (2024). Big Data and Predictive Analytics: A Systematic Review of Applications. *Artificial Intelligence Review*, 57, 176. <https://doi.org/10.1007/s10462-024-10811-5>

[11] Zhang, Z. (2020). Predictive Analytics in the Era of Big Data: Opportunities and Challenges. *Annals of*

Translational Medicine, 8(4), 68.
<https://doi.org/10.21037/atm.2019.10.97>

[12] Bokonda, L., Ouazzani Touhami, K., & Souissi, N. (2020). Predictive Analysis Using Machine Learning: Review of Trends and Methods. <https://doi.org/10.1109/ISAECT50560.2020.9523703>.