Hybrid AI Talent Acquisition Model: An Opinion Mining and Topic based approach

Darothi Jas
Christ University, India

Fr Dr Arun C Antony
Christ University, India

Anshul Saxena
Christ University, India
anshul.saxena@christuniversity.in

Manjari Sharma
Christ University, India

Sharad Gupta
Christ University, India

Abstract: **Artificial Intelligence models have found their usage in the human resource domain. In this paper, job reviewers' opinions on online discussion boards have been captured. The relative importance of factors has been established through an extensive literature review. First, LDA Topic modelling by adopting PCA is performed on unstructured text data has been analyzed. Second, sentiment analysis using the Li-Hu method has been employed to understand job seekers' satisfaction with job portals. The proposed model, "Hybrid AI Talent Acquisition Model," follows a novel approach to streamlining the jobseeker opinion related to online outlets.**

Keywords - NLP, Topic modelling, Sentiment Analysis, Analytics, Artificial Intelligence

I. INTRODUCTION

India is rapidly growing its business umbrella over the globe and taking over major global projects with high-tech tools and web use. There is a massive demand in the international market for Indian talent. Here internet plays an undeniable role in handshaking the prospects with the demand pool. The job portals play a significant role in acting as the perfect bridge. Still, the question arises if existing job portals like Naukri, monster, or LinkedIn contribute significantly to users' experience. User experience is measured in terms of the website content, job posts authenticity, reliability of employer, and customer service.

It has been observed that before managing a human resource pool, it is essential to select people with valuable knowledge, attitude, and relevant skills.Incorrect selection leads to mismanagement of hiring costs and consequences of a misfit at a job. Similarly, jobseeker must experience seamless transition during the entire recruitment cycle.

This research paper examines user sentiment towards India's three most popular job portals. The portals under observation here are Naukri, Monster, and LinkedIn. The study focuses on analyzing and understanding the key factors that help enhance the end-user experience while using the portal. The research is a sentiment analysis based on secondary data from multiple review websites. A total of 600 reviews have been observed to understand user reactions towards the three websites- Naukri, Monster & LinkedIn.

The critical contribution of this study is an enhanced framework titled "Hybrid AI Talent Acquisition Model (HAITAM)".This model has been proposed to track job seekers' underlying issues & pain points while using job portals on a real-time basis. The authors have used topic modelling and sentiment analysis techniques to implement the framework.Hybrid LDA, along with the Le-Hu sentiment analysis technique, was deployed as part of the Hybrid AI Talent Acquisition Model (HAITAM).

II. RELATED WORK

Existing body of work has widely applied the artificial intelligence (AI) and deep learning (DL) techniques to solve the problem in HR domain.It has benn discussed in detail in the table given in Table 1.

After reviewing the existing body of work, it has been observed that there is a scope of enriching the existing framework with advance NLP models like Topic Modelling and Sentiment Analysis to make the topic extraction and sentiment scoring techniques more effective to enhance the user's experience.

III. PHILOSOPHICAL UNDERPINNING

Job portals have changed the entire ballgame of recruitment where an effective site can help improve the brand image of a company, help in reduction of hiring cost and time, better employee candidate filtration, makes recruitment process more transparent, helps to know HR policies better, understand market better for strategic positioning, better navigation of jobs etc. Web technologies are expanding fast in today's world, and every organization relies on internet-based tools for online operations, internet sites, and promotion. Job boards are digital platforms that enable employers to advertise their openings and job seekers to develop in their careers.[1].

This study has been carried out

- To investigate the sentiment score for the existing job portals using opinion mining
- Examine the role of topic modeling in topic extraction using LDA
- To achieve the above said objectives, authors has proposed an enhanced Hybrid AI Talent Acquisition Model (HAITAM)(Fig 1) which incorporates hybrid LDA technique and Li-Hu framework to analyses the continuous data stream which is captured in the form of User Generated Content on the internet.

TABLE I.   RELATED WORK

| Author	Year	Technique/Tool	Key Findings
Ameer et al.[1]	2020	Logistic Regression and Random Forest	Used AI techniques to predict customer churn
Srinivas and Malini[2]	2020	Conceptual Study	Captured the effect of the Internet of Things (IoT), Machine learning (ML), autonomous agents and things (Robotics), Cloud computing, Cognitive computing and Artificial Intelligence (AI) in digitalizing various HR functions
Venusamy et al.[3]	2020	Artificial Intelligence (AI) grounded PC program	Utilization of chatbots in the ehancing the employee experience using gamification
Rahman, M., &Patra, A.[4]	2020	Conceptual Study	Simplicity, consumer experience, effort expectancy, subjective norm, and trustworthiness are the important determinants of behavioural Intent to use job portals in the Malaysian context, according to the research.
Saura et al.[5]	2022	LDA	Use of digital technologies and platforms based on the analysis of user-generated content (UGC) in Twitter
Wadhawan, S., & Sinha, S[6]	2019	Conceptual Study	User Satisfaction, Subjective Norms of Use, Enhanced Services, Speculative Career Opportunities, Service Quality, and Perceived Integrity are all factors that influence millennial job seekers' perceptions of job portals.
Lee & Kang[7]	2017	Dominance Analysis and LDA Topic Modeling	Application the analysis of reviews and text mining to the HR domain
Michelman[8]	2018	Big Data Analytics	Industry is using big data and analytics to gain better understanding of employee workforce
Sisodia et al[9].	2017	Random Forest, k-nearest neighbor and Naïve Bayes classifier	Identification and optimization of reasons behind employee churn
Sumathi et al.[10]	2018	Efficient decision making framework	Conversion of data from unstructured to structured struture for enhanced decision making in HR Analytics
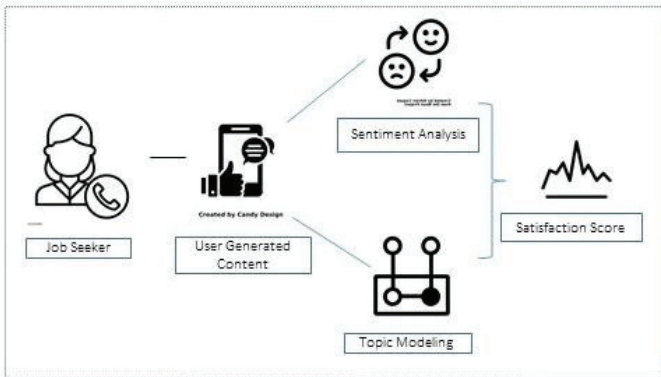
Fig. 1.   Hybrid AI Talent Acquisition Model

Sentiment Analysis –Researchers has gauged the underlying opinion from the tweets and news headlines using sentiment analysis. Initially, Sentiment score is generated based on the based on the no. of positive and negative words identified using dictionary-based approach. Sentiment score calculate based on the following equation.

Stage 1 –**Data Collection**: For this particular analysis, reviews of the were jobseekers were required so as to analyse and understand their perception towards the job portals and for the data collection web scrapping method was used to get draw reviews from various websites. Social media, review websites etc were platforms wherein we performed web scrapping method. Web scrapping is a method used to extract and process large data sets from particular URLs/websites. This technique is used to bin the sentiments into positive, negative and neutral category based on the sentiment score.

Topic Modelling: Extracting contextual meaning from the given corpus and then grouping them to make sense of the corpus analyzed is one of the main motives behind performing NLP. The topic modeling technique is used to identify and group the set of words from the given corpus. The current research Latent Dirichlet Allocation (LDA) technique has been deployed for topic generation. LDA is an unsupervised learning technique based on the generative probabilistic model where each topic probabilities provide an explicit representation of a document. The aim of LDA is to find topics a document belongs to, based on the words in it.

IV.   RESEARCH METHODOLOGY

This study is aimed towards carrying out a comparative sentimental analysis towards user feelings about the most popular job portals in Indian context. The data collected will help identify the factors that make a particular website more attractive than its competitors and what kind of sentiments do users showcase with their brand.

Data Collection: The initial step was to identify data collection sources, which were authentic and identified. Out of the most popular names, three of the most used job portals (Naukri, Monster & LinkedIn) were chosen for carrying out the comparison. Social media is one of the best instruments for gathering data about user reviews and sentiments because they spend a lot of time on it. (Bagheri& Islam, 2017). The research analysis is being carried out based on secondary data collected from from Mouth Shut, Consumer affairs, Site jabber, Indeed, Ambition box, Trust Pilot, GetApp and Quora. The sample size is 600 based on user reviews.This particular method was used to collect the data for the analysis.

Stage 2-Data Preprocessing Corpus-The data that was collected in the stage 1, would be imported to Orange application for further processing of analysis. Following are the steps initiated while analyzing the data:

Transformation: Given review data would be transformed into lowercases, and further processed by removing all the tags and URLs if found any. It also removes accents if found any.

Tokenization: Under Tokenization, we would be splitting by regular expressions and would be using or keeping only

2

words. This step splits the text document into words, sentences &bigrams. Tokenization is a key element, which breaks the unstructured data into discrete elements. Under this research analysis, the review data document is processed to remove white spaces. In there review corpus, the emoticons and other special symbols are kept as a pat of processes document.

Normalization: WordNet Lemmatization technique would be used to process the data set and extract some contextual meaning/ text out of the document. WordNet labels semantic relations among the words and thus it processes and extract meaningful contextual words out of the given text document.

Filtering: In this process we will remove the stop words in the English language. This step processes the selection of words and removes unwanted words, symbols, etc. Stop word is used to remove (and, or…in) from the text document. Inthis research analysis, it removes mathematical symbols, hashtags, special characters from the reviews data collected. Further the filtered list would be refined using TD-IDFmethod.

Keyword Extraction: Once the text pre-processing step has been completed in the stage 2 now the stage 3 begins. Using Term Frequency method, the number of appearance of a particular word token from the data set list is gauged out then its relative importance ratio when compared with other words is analyzed using TD-IDF score. This method generated score for each word token with respect to their relative importance. Word cloud displays the most frequently appearing words. Different keywords are recognized and identified related job portals which has a significant level of important when considered as the factors, which affects consumer's perception while talking about the job portals.

Stage 3-Topic Modelling with Latent Dirichlet Allocation (LDA) Process (Fig2) would be performed here for the further analysis. Now let us state the inputs and outputs: Input is the corpus, which has the text document of the review collected. Outputs would contain differentiation of the document with topic name extracted from the document and then weights are assigned to each topic as well as to words as per its importance or repetition in the text document and as per the importance ratio when compared to other topics. Topic Modelling technique is used to find sub topics in the given text corpus. Essentially pre- processed tokens will be converted into a dictionary with word index and it's counting the corpus.

Stage 4-Sentimental Analysis using Liu Hu: Sentiment Analysis predicts sentiment for each document in a corpus. Here we are choosing Liu- Hu model, which only works on English as well as Slovenian version. Sentiment Analysis forecasts the sentiment of each document in a corpus. It classifies the sentiment very well out of the processed documents. In this stage we will use Liu-Hu model which analyzes and processes our reviews into either positive or negative reviews/responds through assigning sentiment scores for each reviews processed. Liu Hu: sentiment analysis using a lexicon. The final score is the difference between the sum of positive and negative words, normalized by document length, and multiplied by 100. The final score reflects the document's percentage of sentiment difference. Liu Hu computes a single normalized sentiment score in the text (negative score for negative sentiment, positive score for positive sentiment, 0 for neutral), whereas Vader outputs scores for each category (positive, negative, neutral) and appends a total sentiment score called a compound.

V. Empirical Result and Analysis

Sentiment Analysis Sentiment analysis in a business context is a methodology for assessing whether content of a research is positive, negative, or neutral through natural language processing. Sentiment analysis has been widely used on textual data to assist organisations in analysing brand and product sentiment through customer experience and aid in better grasp at customer demands.
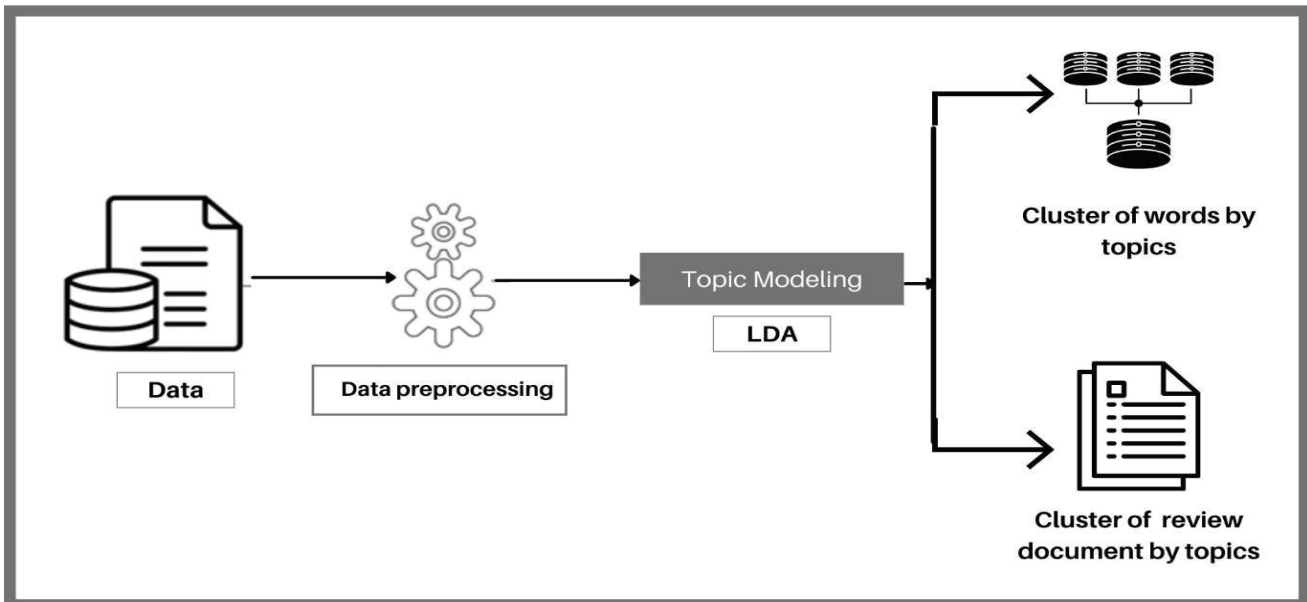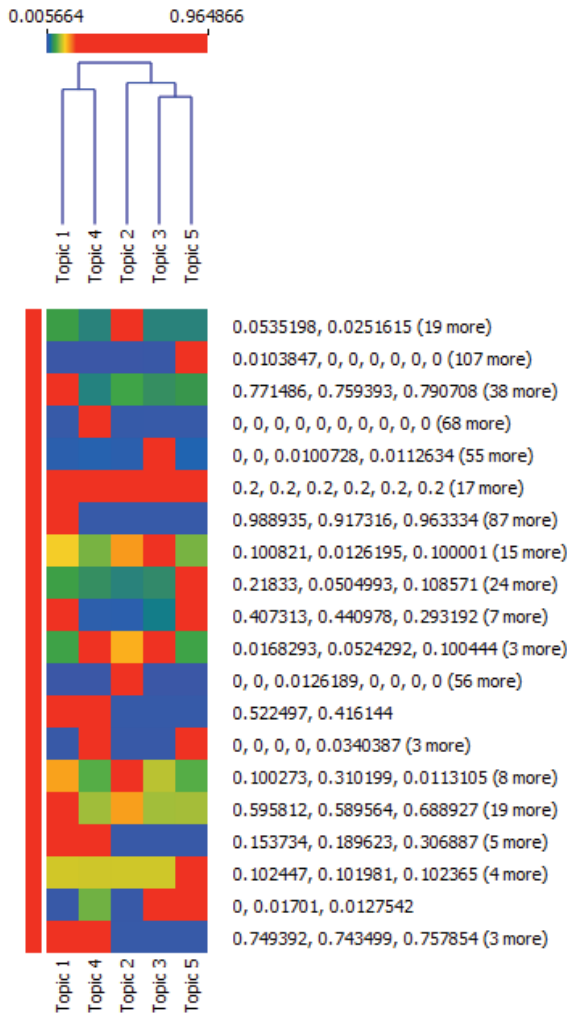

Fig. 2. Topic Modelling Process

3

Fig. 3. Violin Plot

A. Violin Charts

Violin plots are useful when observing the distribution of quantitative data. They're particularly potent (Fig 3) when analysing distributions between different factions. The density curves of each group's peaks, falls, and tails can be studied to see whereclusters are similar or distinct. Violin graphs gives range of indexes, which help us understand thepositive and negative limits

Based on the above chart, following observation table (table 2) has been prepared which show the jobseekers opinion polarity while using the job portals.

TABLE II.   SENTIMENT SCORE

Job Portal	Score Range	Positive/Negative Sentiment %
Naukri	-120 to +60	20|80
Monster	-20 to +30	47|53
LinkedIn	-30 to +70	54|46

B. Topic Modeling

Keyword Extraction: When we talk about all the three portals, the contextual image (Fig 4:Word Cloud) of how people relate to these websites are expressed through the

Fig. 4. Word Cloud

words: job, account, service, resume, professional, companies, network, site, platform, money, support, people, company services, experience, search and career. Users perceive LinkedIn as professional and useful site where their profile feature stands out. Monster has successfully built a brand image for user friendly, good service and best platform for job search.

Finally, Naukri over the years has attracted a theme of paid services, fake, fraud and lack of authenticity of jobs showcased. A topic modelling approach has been deployed to understand the underlying issues about which jobseekers are discussing about on various discussion forums online. From the keywords results and word cloud, a clear indication of user reviews has been observed. Total five (Table 3) topics based on LDA approach have been generated to highlight the discussion points. Table 3 gives an overview of all the topics listed. Themes listed are identified manually based on the factors highlighted under conceptual study listed under related work.

TABLE III.   TOPIC LIST

Topic List	Theme	Keywords
Topic 1	Account login	free,account,network,fake account
Topic 2	Job Postings	site service, customer support,fake jobs
Topic 3	Customer Support	greatful, recommend, professional
Topic 4	Social Media	supporting formation, social network, service
Topic 5	Resume	useful, information, resume service.

Based on the clustering scores depicted on heat map (Fig 5), it can be observed that topic 3&5 has more similarity in comparison to topic 1,2 &4.Using the dimension reduction technique like PCA further refinement of results can take place.

Fig. 5. Clustering Map

reveal important aspects of job seeking (topic 2) such as site service, customer support, and fake jobs. Proposed Hybrid AI Talent Acquisition Model (HAITAM) provides better insight into jobseeker's mindset using elements from sentiment analysis and topic modelling.Future work might include enhancing the framework using advance topic modelling techniques like Heretical and Hybrid-LDA and increasing the sample size from 600 to almost 1200 reviews.

VI. CONCLUSION AND FUTURE WORK

In business applications, this pilot study provides insight into which topics jobseekers deem important, as well as how they feel about it. This enables targeted product development and customer experience improvements. This pilot study contains a variety of pro, but a separate topic model into each product may reveal aspects that job seekers care about. For example, this analysis already started to Fig 4 – Heat map

REFERENCES

[1] S. P. R. a. S. M. M. Ameer, ""Human Resource Analytics using Power Bi Visualization Tool,"," in 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020.

[2] T. N. M. a. D. B. Srinivas, "Technological Transcends: Impact of Industrial 4.0 on Human Resource Functions," in Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020.

[3] N. K. R. a. M. Y. K. Venusamy, "A study of Human Resources Development through Chatbots using Artificial Intelligence," in 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020.

[4] A. P. Munjarin Rahman, "Shared Values of E- Recruitment Portal:Determinant Factors of Job- Seekers' Intention to use Job Portals," Taylor Business Review, vol. 9, no. 1, pp. 1-31, 2020.

[5] S. J.-S. D. S. P, "Exploring the challenges of remote work on Twitter users' sentiments: From digital technology development to a post-pandemic era," Journal of Business Research, vol. 142, pp. 242-254, 2022.

[6] S. Wadhawan, "Factors Influencing Young Job Seekers Perception towards Job Portals," AIMS International Journal of Management, 2018.

[7] J. a. K. J. Lee, "A Study on Job Satisfaction Factors in Retention and Turnover Groups using Dominance Analysis and LDA Topic Modeling with Employee Reviews on Glassdoor.com,"ICIS , vol. 26, 2017.

[8] P. Michelman, "7 Is Your Company Ready for HR Analytics?," in How to Go Digital: Practical Wisdom to Help Drive Your Organization's Digital Transformation," MIT Sloan Management Review, pp. 59-64, 2018.

[9] S. V. a. A. P. D. S. Sisodia, "Evaluation of machine learning models for employee churn prediction," in International Conference on Inventive Computing and Informatics (ICICI), 2017, 2017.

[10] S. Sumathi, "Data Analytics platform for intelligent agriculture," in 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018.

[11] B. P. Sudiana, "Users' Interest Assessment on Job Portal," International Journal of Web Portals, Jakarta, 2014.

[12] S. O. A. M. S. Janes O. Samwel1, "The influence of employees' recruitment on performance outcomes of family owned business in Tanzania case of Nyamagana and Ilemela Districts," African Journal of Business and Management, Mwanza, 2016.