Volume 1 Issue 2

SASHAKT: A Job Portal for Women using Text Extraction and Text Summarization

Jaspreet Kaur, Pragati Verma, and Sanyuktaa Bajoria*

Department of Computer Science and Engineering, JSS Academy of Technical Education, Noida, Uttar Pradesh, India 201301

Abstract

SASHAKT is a job portal designed specifically for women, utilizing text extraction and summarization techniques to provide a user-friendly and efficient job search experience. The portal extracts relevant information from job postings and summarizes it, allowing women to quickly identify job opportunities that align with their skills and qualifications. Additionally, the portal utilizes text classification algorithms to identify and filter out job postings that may be discriminatory or biased toward women. This study presents the development and implementation of SASHAKT, including a detailed description of the text extraction and summarization techniques used and the text classification algorithms implemented to detect discriminatory language. The study also presents the results of user testing and evaluations of SASHAKT, highlighting its effectiveness in improving the job search experience for women. The results of this study demonstrate that SASHAKT can help increase women's representation in the workforce by providing them with a more efficient way to find job opportunities that align with their skills and qualifications. Furthermore, the study also highlights the potential for similar text-based approaches to be applied to other areas of job search and career development for underrepresented groups such as people with disabilities and minority groups. Overall, the study concludes that SASHAKT is an innovative solution that addresses the need for a more inclusive job search experience for women by utilizing natural language processing techniques.

Keywords: Text Summarization; Text Extraction; Text Localization; Text Detection; Natural Language Processing

1 Introduction

The job search process can be challenging and time-consuming for many individuals, particularly women, who may face discrimination and bias in the workforce. Women are underrepresented in many industries and often face barriers to career advancement. According to a study by McKinsey, women are less likely to be promoted to management positions and are underrepresented in industries such as technology, finance, and manufacturing [1–4]. In the last twenty-five years, India has experienced significant socioeconomic transformation. India has experienced rapid economic growth, structural economic shifts accompanied by high urbanization rates, increased educational attainment, and declining fertility rates. However, there has been a gradual and persistent decline in women's economic activity [5]. Female economic empowerment and workforce participation are particularly significant issues in India, and the country ranks 139th out of 144 in terms of gender equality in economic participation and opportunity. Gender inclusion in the hiring process is critical because low female labor force participation impedes economic growth. India's gross domestic product (GDP) could increase by 27% if female participation rates matched those of men - and there is evidence that low participation rates are partly caused by discrimination and other hiring barriers. India's female labor force participation rate (FLPR) is low, at 24% overall and 16% in urban areas, placing it 120th out of 131 nations [6]. Even Indian women who have opted to work face difficulties in finding jobs. Involuntary unemployment has a 4.7% point gender disparity (8.7% of working-age women are unemployed, but only 4.0% of comparable men are) [7].

This disparity is even greater among highly educated urban dwellers, with both groups losing 8.8% points, as shown in Figure 1. The state with the least gender inequality is ranked one, and the state with the highest gender inequality is ranked 15th. Challenges in finding suitable employment partly explain the low labor rate. Women have more difficulty finding jobs than men [8]. While the internet and online recruiting techniques enhance access to work chances for female applicant groups, an increase in the number of applicants might encourage managers to depend more heavily on demographic-based preconceptions, whether purposely or accidentally. This type of prejudice is a global problem, and new and unique remedies to eliminate the bias are required to counteract these tendencies.
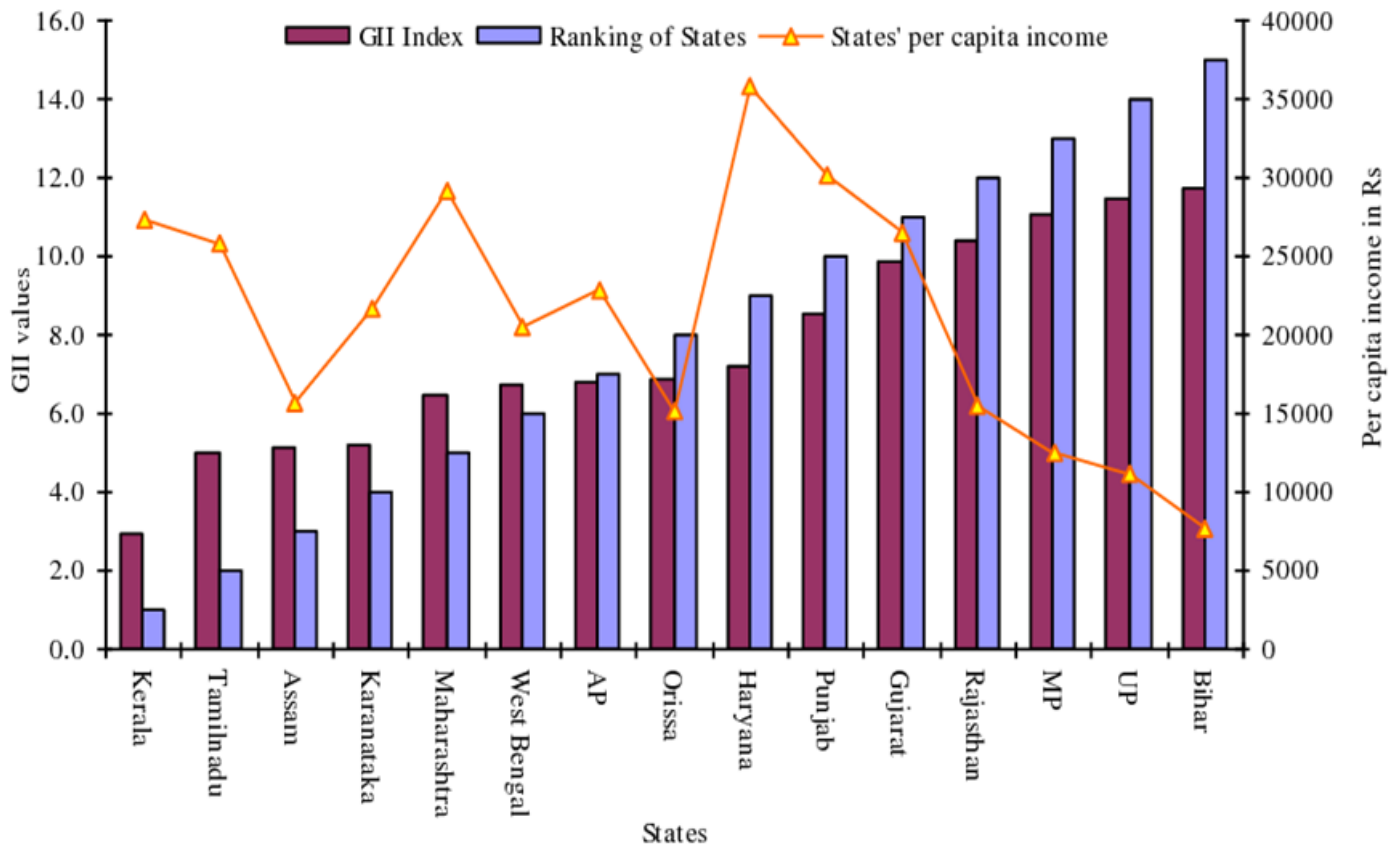

Figure 1: Gender inequality in highly educated urban dwellers

Additionally, a study by LeanIn found that women are often passed over for promotions and are less likely to receive critical stretch assignments that can help them advance in their careers [9, 10]. The job search process can be particularly challenging for women, who may face discrimination and bias in the workforce [11]. According to a study by the National Women's Law Center, women are often subject to gender bias in the hiring process and are less likely to be offered job interviews or receive job offers than men [12]. Additionally, a study by the Center for Talent Innovation found that women are often judged on their perceived potential rather than their qualifications and experience [13]. Such judgments make it difficult for women to advance in their careers and lead to a lack of representation in many industries, thus emphasizing the need for a technique to ease the job application process. Text extraction and summarization techniques are commonly used in natural language processing to extract relevant information from the text and present it in a condensed format. These techniques can extract information from job postings and summarize it concisely, making it easier for job seekers to identify job opportunities that align with their skills and qualifications. Common text extraction and summarization techniques include keyword extraction, named entity recognition, and sentence compression [14–16]. Text classification algorithms are commonly used in natural language processing (NLP) to classify text into different categories or labels. These algorithms can detect discriminatory language in job postings and filter them out [17]. Common text classification algorithms include support vector machines, decision trees, and naive Bayes [18]. Figure 2 represents the block diagram of the discussed text extraction technique, while Figure 3 represents the block diagram of the text summarization technique.

The proposed NLP-based job portal, SASHAKT, extracts relevant information from job postings and summarizes it, allowing women to quickly identify job opportunities that align with their skills and qualifications. Additionally, the portal utilizes text classification algorithms to identify and filter out job postings that may be discriminatory or biased toward women. This approach addresses women's need for a more inclusive job search experience by utilizing natural language processing techniques. The development of SASHAKT involved thoroughly analyzing women's current job search landscape and challenges. The existing job portals were reviewed, and key features that could be improved to better serve women's needs were identified. Text extraction and summarization techniques were then implemented to extract relevant information from job postings and present it concisely. Additionally, text classification algorithms were implemented to detect the discriminatory language in job postings and filter them out.

Figure 2: Text extraction block diagram

Figure 3: Text summarization block diagram

2 Method

2.1 Development of SASHAKT

SASHAKT is developed to cater to both ends of the recruitment process - recruiters and job seekers. It utilizes a React-based frontend that communicates with the Django backend via REST APIs written using Django Rest Framework (DRF). The data is stored in a SQL/Postgres database hosted on Heroku, while the frontend is hosted on Netlify. Figure 4 shows the hiring mechanism proposed using SASHAKT to create job postings, connect job seekers, and organize them in a single place while streamlining the complete recruitment process using text extraction, summarization, and classification techniques of natural language processing. Figure 5 represents the client-side of the job portal. SASHAKT simplifies data management for registration details such as personal information, contact information, and login credentials through safe storage in a database. The portal operates like any other web application, utilizing a database and REST APIs on the backend server to seamlessly communicate with the program logic and present the desired results to the user, as illustrated in Figure 6.
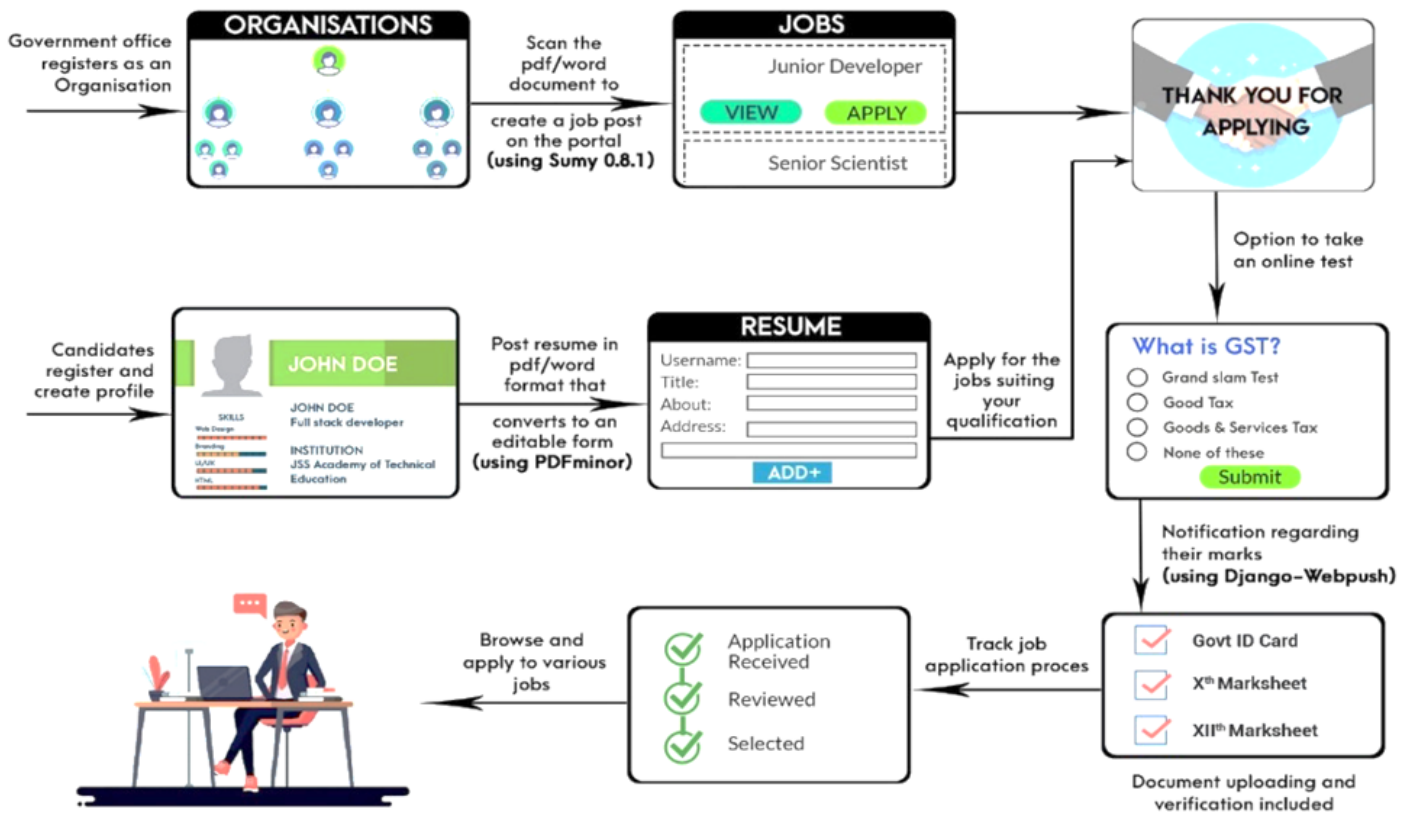
16

Figure 4: Hiring mechanism through online job portals
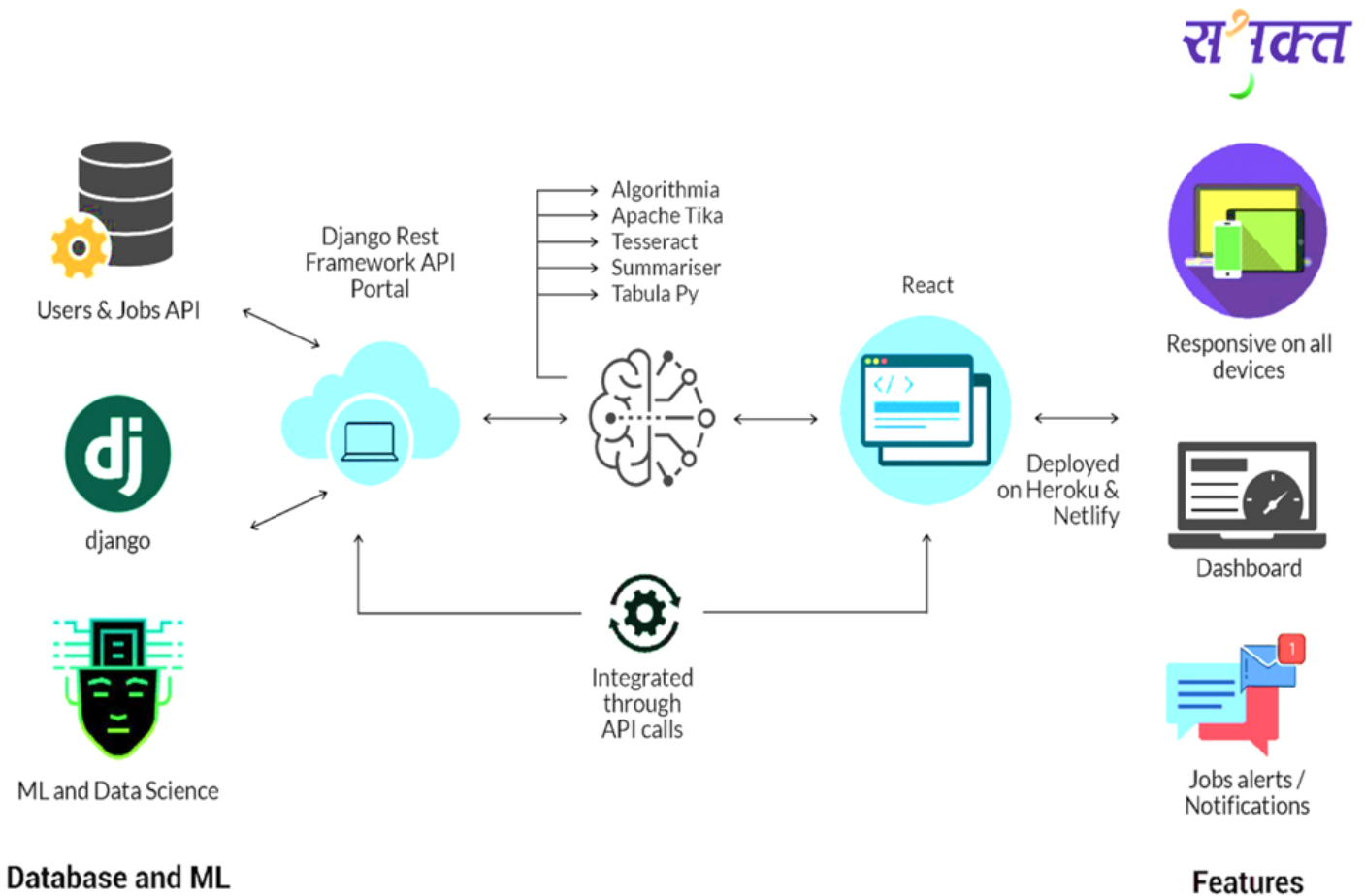

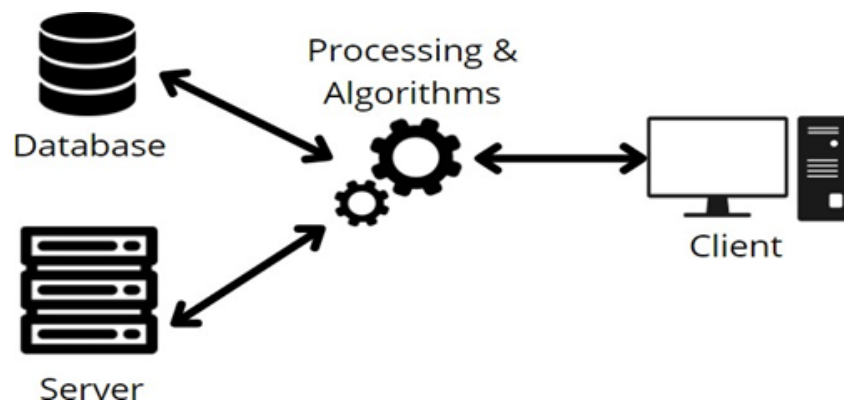Figure 5: Technology stack for the online job portal

Figure 6: Architecture diagram

The development of SASHAKT involved thoroughly analyzing women's current job search landscape and challenges. The working of the web application can be explained in the following four steps:

1. Step 1: The user uses a web browser or a mobile application to access the web application, which sends a request to the web server via the internet. Security mechanisms such as firewalls or cloud access security brokers and load balancers may be in place.

2. Step 2: The request is forwarded to the web application server via the web server. The web application server completes the required operation, such as accessing the database or processing data, and generates the requested data results.

3. Step 3: The results are returned to the web server by the web application server.

4. Step 4: The web server sends the requested data to the client (desktop, mobile device, and tablet), and it is displayed on the user's screen.

2.2 Text Extraction and Summarization Techniques

The existing job portals were reviewed, and key features that could be improved to better serve women's needs were identified. Text extraction and summarization techniques were then implemented to extract relevant information from job postings and present it in a concise format. The text extraction and summarization techniques used in SASHAKT included keyword extraction and sentence compression. Keyword extraction was used to identify the most important terms and phrases in a job posting, while sentence compression condensed the text into a more concise format.

2.3 Text Classification Algorithm

Additionally, text classification algorithms were implemented to detect discriminatory language in job postings and filter them out. The text classification algorithms implemented in SASHAKT include support vector machines and naive Bayes. These algorithms were trained on a dataset of job postings and were used to detect the discriminatory language in job postings and filter them out.

2.4 User Testing and Evaluations

The effectiveness of SASHAKT was evaluated through user testing and evaluations. Participants were asked to search for job opportunities using SASHAKT and provide feedback on their experience. The evaluations included questions about the usability of the interface, the effectiveness of the text summarization feature, and the accuracy of the text classification algorithms.

3 Results and Discussion

Figure 7 represents a sample Portable Document Format (PDF) advertising the job call for an apprentice post in the State Bank of India, which was added to the system to extract information. The goal was to obtain basic information about the job and list it on the portal. The main objective was to test the functionality of the proposed text extraction algorithm-based job portal and see how the job details are displayed in a summarized form. Figure 8 represents the result obtained for the uploaded PDF. The basic information from the PDF is extracted from the uploaded form, as shown in Figure 8, and it has automatically updated the job creation form. This listing can be posted like any other job on the portal; every eligible candidate can see it and apply for it if interested. However, during the development process, some drawbacks were identified. The time taken for the extraction and summarization process was determined to be dependent on the algorithm used by the library. In some cases, not all data from the PDF were extracted properly, so manual input was required.

SBI

Central Recruitment & Promotion Department, Corporate Centre, Mumbai
Phone: 022-22820427; Fax: 022-22820411; Email ID – crpd@sbi.co.in
ENGAGEMENT OF APPRENTICES UNDER THE APPRENTICES ACT, 1961
ADVERTISEMENT NO: CRPD/APPR/2021-22/10
(ONLINE REGISTRATION OF APPLICATION: FROM 06.07.2021 TO 26.07.2021)
ONLINE EXAMINATION: TENTATIVELY IN THE MONTH OF AUGUST 2021

Applications are invited from Indian Citizens for engagement as apprentices under the Apprentice Act 1961 (as amended from time to time).

1. Candidates can apply for engagement in one state only. Candidates can appear for the examination only once under this engagement project.
2. The process of registration is complete only when fee is deposited with the Bank through online mode on or before the last date for payment of fees.
3. Candidates are requested to apply Online through the link given on the following website https://nsdcindia.org/apprenticeship or https://apprenticeshipindia.org or http://bfsissc.com or https://bank.sbi/careers or https://www.sbi.co.in/ careers.
4. All revisions/corrigendum (if any) will be hosted on the Bank's website only.
5. State wise seats of apprentices, State wise Local languages, District wise training seats, Age, Education qualification, Duration of training, Training, Stipend, Selection Process, Examination Centres, Fees and other related parameters are mentioned below:

A. State/UT, Local Language and Tentative No. Of Training Seats:

| State / UT | Language* | SC | ST | OBC | EWS | UR | Total** | PWD | | | |
								LD	VI	HI	D&E
Gujarat	Gujarati	56	120	216	80	328	800	8	8	8	8
Andhra Pradesh	Telugu/Urdu	16	7	27	10	40	100	1	1	1	1
Karnataka	Kannada	32	14	54	20	80	200	2	2	2	2
Madhya Pradesh	Hindi	11	15	11	7	31	75	1	1	1	0
Chhattisgarh	Hindi	9	24	4	7	31	75	1	1	1	0
West Bengal	Bengali/Nepali	164	35	157	71	288	715	8	7	7	7
Andaman & Nicobar Islands	Hindi/English	0	0	2	1	7	10	1	--	--	--
Sikkim	Nepali/English	1	5	6	2	11	25	1	--	--	--
Odisha	Odia	64	88	48	40	160	400	4	4	4	4
Himachal Pradesh	Hindi	50	8	40	20	82	200	2	2	2	2

Figure 7: Sample PDF uploaded to the system for the job posting of Apprentice at the State Bank of India

Figure 8: Result obtained for the uploaded PDF

This issue can be improved by feeding the dataset used by the library with more information. Nevertheless, SASHAKT is a potential online employment platform for rural women, linking them with companies searching for intelligent and competent women. The platform promises various features to make the process easier for rural women and employers. Machine learning algorithms, if implemented and integrated with the backend, which can be called through the API endpoints, can provide a smoother and faster-personalized user experience on the platform.

4 Conclusion

Automatic text identification and extraction from images is a crucial study area in content-based information retrieval and text-based picture indexing. Text extraction applications include mobile robot navigation, automobile license detection, recognition, item identification, document retrieval, page segmentation, etc. Based on the data gathered from various methodologies, it has been discovered that morphological and edge-based algorithms may efficiently and effectively locate and extract text from pictures. SASHAKT is an innovative solution that utilizes text extraction summarization and classification techniques of natural language processing techniques to address the need for a more inclusive job search experience for women. It provides an efficient and effective way for women to find job opportunities that align with their skills and qualifications while also addressing the issue of discrimination and bias in the workforce. This study provides valuable insights into the potential of text-based approaches to improve the job search experience for underrepresented groups and paves the way for future research in this area.

Declaration of Competing Interests

Funding Declaration

Author Contribution

Jaspreet Kaur: Supervision, Conceptua;ization and Methodology, Writing- Reviewing; **Pragati Verma**: Conceptualization, Visualization, Investigation, Methodology, Data curation, Writing- Reviewing; **Sanyuktaa Bajoria**: Conceptualization, Visualization, Investigation, Methodology, Data curation, Writing- Original draft preparation, Writing- Reviewing.

References

[1] M. L. Johns, "Breaking the glass ceiling: structural, cultural, and organizational barriers preventing women from achieving senior and executive positions," *Perspectives in health information management*, 10 (Winter), 2013.

[2] K. Yanosek, D. Abramson, and S. Ahmad, "How women can help fill the oil and gas industry's talent gap," *McKinsey & Company*, pp. 1–11, 2019.

[3] D. Barton, J. Woetzel, J. Seong, and Q. Tian, "Artificial intelligence: implications for China," *McKinsey Global Institute*, Discussion (April), p. 20, 2017.

[4] N. Waller, "How men and women see the workplace differently," *The Wall Street Journal*, pp. 1–2, 2016.

[5] R. Lahoti and H. Swaminathan, "Economic development and women's labor force participation in India," *Feminist Economics*, 22 (2), pp. 168–195, 2016.

[6] L. Narayan, "Women's labour force participation in Haryana: a disaggregated analysis," *Imperial Journal of Interdisciplinary Research*, 2 (11), 2016.

[7] A. Ghazala, M. Guell, and A. Manning, "Gender Gaps in Unemployment Rates in OECD countries," *Journal of Labor Economics*, 24 (1), pp. 1–37, 2006.

[8] E. Fletcher, R. Pande, and C. M. T. Moore, "Women and work in India: descriptive evidence and a review of potential policies," *SSRN Electronic Journal*, 2017.

[9] A. L. Beckwith PhD, D. R. Carter DM, and T. Peters PhD, "The Underrepresentation of African American women in executive leadership: what's getting in the way?," *Journal of Business Studies Quarterly*, 7 (4), pp. 115–134, 2016.

[10] J. R. Cohen, D. W. Dalton, L. L. Holder-Webb, and J. J. McMillan, "An analysis of glass ceiling perceptions in the accounting profession," *Journal of Business Ethics*, 164 (1), pp. 17–38, 2020.

[11] M. M. Henley, "Women's success in academic science: challenges to breaking through the ivory ceiling," *Sociology Compass*, 9 (8), pp. 668–680, 2015.

[12] J. K. Pinto, P. Patanakul, and M. B. Pinto, "The aura of capability": Gender bias in selection for a project manager job, *International Journal of Project Management*, 35 (3), pp. 420–431, 2017.

[13] C. Ashcraft, B. Mclain, and E. Eger, "Women in Tech: The Facts," 2016.

[14] N. Kumari and P. Singh, "Text summarization and its types," 2021, pp. 368–378.

[15] A. Nenkova and K. McKeown, "A survey of text summarization techniques," in *Mining Text Data*, Boston, MA: Springer US, 2012, pp. 43–76.

[16] R. Parimoo, R. Sharma, N. Gaur, N. Jain, and S. Bansal, "A review on text summarization techniques," *International Journal for Research in Applied Science and Engineering Technology*, 10 (5), pp. 871–873, 2022.

[17] F. Zhang, H. Fleyeh, X. Wang, and M. Lu, "Construction site accident analysis using text mining and natural language processing techniques," *Automation in Construction*, 99, pp. 238–248, 2019.

[18] T. Pranckevičius and V. Marcinkevičius, "Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification," *Baltic Journal of Modern Computing*, 5 (2), 2017.