IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI AUTOMATED CANDIDATE FILTERING

Abhay Tiwari
Student, Computer Science and Engineering Chandigarh University Mohali, Punjab, India

Monika
Assistant Professor
Chandigarh University

Subhajit Paul
Student, Computer Science and Engineering Chandigarh University Mohali, Punjab, India
Abhijeet Kumar Student, Chandigarh university

Shreyansh Shrey
Student, Computer Science and Engineering Chandigarh University Mohali, Punjab, India

Subham Kumar Jha Student, Computer Science and Engineering Chandigarh University Mohali, Punjab, India

Abstract: AI Automated Candidate Filtering is one of the most used technologies that many of us are familiar with. If we ever applied for job through resume, we are very much known to the fact that our resume goes through various filtering and processing before our profile gets selected for the applied job. Its primary function is to manage job applications and help companies filter and organize the large volume of resumes and job applications they receive. Applicant Tracking Systems (ATS) rely on a combination of software, databases, and algorithms to manage and process job applications efficiently. The specific tools and technologies used can vary depending on the ATS software vendor. The most common working technology is NLP (Natural Language Processing) and OCR(Optical Character Recognition). NLP and OCR are two widely used powerful AI tools whose usage can be seen in everyday life other than ATS (Applicant Tracking System). Natural language processing (NLP) and optical character recognition (OCR) are two technologies that can be used to improve the performance of ATSs. NLP can be used to extract more relevant and sensible words, sentences and paragraphs which resembles the keywords and match them according to the job description. To pick out these words, OCR comes into play where it an detect text from various formats of resume. For ex- ppt, pdf, docx, URL of candidate's profile etc. All these mechanism work in harmony to make a fair hiring process.

Keywords— CNN, SVM, DEEP LEARNING, ARTIFICIAL INTELLIGENCE, FEATURE EXTRACTION, ATS.

I. Introduction

ATSs are used by many companies to automate the process of screening job applications. ATSs typically use a variety of techniques to assess a candidate's qualifications, including keyword matching, resume parsing, and sentiment analysis. Keyword matching is the simplest technique used by ATSs. The recruiters fill in the details and needed requirements in the database of the ATS software. So while matching though candidates resume, it matches the keywords of the job description along with the resume. This helps filtering out the candidate automatically and ranks them accordingly in another database. This automation process makes the recruit process easy and time saving.

Need: First and the most primary cause of recruiter to hire candidates though LinkedIn is because of it's a large and active platform, with over 830 million members. This gives companies access to a large pool of potential candidates. Second, LinkedIn profiles typically contain a wealth of information about candidates, such as their skills, experience, and education. This information can be used by ATS systems to identify qualified candidates and filter out unqualified candidates. Relevant Contemporary Issue: There are also some challenges associated with using ATS software to source candidates through LinkedIn. First, it is important to note that LinkedIn prohibits the scraping of profiles without the user's consent. As a result, companies must use a LinkedIn-approved method to scrape profiles. Second, it is important to clean and normalize the data that is scraped from LinkedIn profiles. This is because LinkedIn data can be messy and inconsistent. with performing assignments like tokenization, grammatical form labeling, and named element acknowledgment without building all that without any preparation. Knowledge of these libraries empowers people to use the work done by others and expand upon it to tackle their NLP issues. Understanding AI calculations is likewise important to investigate the abilities and restrictions of NLP in text age and understanding. NLP is a field where AI calculations are widely utilized. AI calculations empower the PC to gain from models and make forecasts on new information. Profound learning calculations are a subset of AI calculations that have demonstrated to be especially successful in NLP errands like message age and feeling examination. Choice trees and Gullible Bayes are other famous AI calculations utilized in NLP errands like text grouping and point demonstrating. Information on these calculations empowers people to pick the right calculation for their NLP task and work on the presentation of their models.

II. Background information

ATS although uses advanced tools such as NLP and OCR, they are still in development as this technology needs lots of training to be perfect. Natural Language Processing (NLP) has made significant strides in recent years, yet it still faces various challenges and limitations. Identifying these issues is crucial for improving the capabilities of NLP systems. NLP struggles with disambiguating words or phrases with multiple meanings, especially in the absence of sufficient context. Resolving the ambiguity and selecting the correct meaning is a persistent challenge. NLP models often struggle to comprehend the context of a sentence or a document fully. NLP systems can potentially invade upon user privacy, especially when processing sensitive information. Maintenance of ethical standards and ensuring user data protection is a critical concern that requires robust solutions and stringent privacy policies. image files, turning into fully searchable documents with text content recognized by computers. Optical Character Recognition extracts the relevant information and automatically enters it into electronic database instead of the conventional way of manually retyping the text. Although the technology sounds quite advanced but needs lots of training from various dataset of individual character to be able to recognize the character. This is because "e" can sometimes look like "a" if the clarity of the image isn't clear. Similarly "I" can be confused with small "L" which looks like "l". Such things makes OCR a challenging technology to handle. There has been a growing interest in the use of ATS software to source candidates through LinkedIn. In a 2022 survey by CareerBuilder, 63% of employers said that they use LinkedIn to recruit

candidates. A number of studies have examined the effectiveness of using ATS software to source candidates through LinkedIn. For example, a 2021 study by LinkedIn found that companies that use LinkedIn to recruit candidates hire 50% more qualified candidates than companies that do not use LinkedIn. Another 2021 study by Indeed found that companies that use LinkedIn to recruit candidates have a
20% higher fill rate for open positions than companies that do not use LinkedIN we first identified the client's need or relevant contemporary issues. We define and differentiate the tasks required to identify, build, and test the solution to the client/ consultancy problem. This chapter also includes the task definition, broad problem statement, and the scope of the project. We maintained a specific timeline in the form
of a Gantt Chart to complete the tasks in the given amount of time. We conducted a Literature Review for our problem statement. We went through the old method of organizing tasks, and its merits and demerits were all considered, and with the help of the latest technologies we have arrived at an optimal solution for this old and absurd method of scheduling tasks. In this paper, the motive is to address the working principle of new age candidate filtering system through benefits of advancement in AI.

III. Literature review

Several existing solutions have made significant strides in the realm of ATS using NLP and OCR, enabling candidate filtering. While NLP and OCR has proven to be a powerful tool, it still encounters certain limitations. A general overview of NLP and OCR is briefly explained through following working mechanism of extracting candidate's information from LinkedIn profiles. OCR can be used to extract information such as the candidate's name, contact information, skills, experience, and education from their LinkedIn profile. This information can then be added to the candidate's profile in the ATS system, eliminating the need for the recruiter to manually enter it. NLP can be used to match candidates to open positions based on their skills and experience. This save the recruiter a lot of time and hassle to manually enter and check all details whether candidates are suitable. NLP can be used to pre-screen candidates by identifying those who meet the basic requirements for the position. This can help recruiters to save time by eliminating candidates who are not a good fit for the position..

NLP Libraries: NLP libraries give a scope of instruments and methods for handling and dissecting text information. NLTK is one of the most generally utilized NLP libraries, and gives a complete set-up of devices for text handling, including tokenization, stemming, and parsing. spaCy is one more well known NLP library that is upgraded for execution and gives a scope of devices to message handling and examination. PyTorch is an AI library that is generally utilized for NLP undertakings, and gives a scope of instruments to building and preparing profound learning models.

AI Calculations: AI calculations are a fundamental part of NLP, and are utilized for errands like message grouping, opinion examination, and machine interpretation. Profound learning calculations, for example, RNNs and CNNs have altered the field of NLP as of late, and are generally utilized for undertakings like machine interpretation and text outline. Choice trees and Credulous Bayes are other AI calculations that are generally utilized for text arrangement assignments.

Limitations of NLP and OCR in ATS-: Although NLP and OCR are powerful technologies that can improve the efficiency and accuracy of the recruitment process, there are some limitations to using them in ATS with LinkedIn URLs.

NLP: One limitation of NLP is that it can be inaccurate, especially when dealing with complex or nuanced language. This is because NLP models are trained on large datasets of text, but these datasets may not be representative of the language that is used in LinkedIn profiles. As a result, NLP models may misinterpret or misclassify information in LinkedIn profiles.

OCR: OCR is also limited in its accuracy. This is because OCR models are trained on images of text, but these images may be of varying quality. As a result, OCR models may not be able to accurately extract information from LinkedIn profiles, especially if the profiles are poorly formatted or contain images experiences in huge text datasets, making it more clear the basic information. All in all, this writing review gives an outline of the present status of exploration in the key specialized abilities expected for NLP, including Python programming abilities, information on NLP ideas and strategies, experience with NLP libraries, comprehension of AI calculations, and knowledge of information pre-handling and perception techniques. These abilities are fundamental for investigating the capacities and restrictions of NLP in text age and understanding. Continuous exploration there is probably going to drive further advancement in NLP, empowering us to communicate with computerized information in additional regular and natural ways.

IV. FEATURES

Key capabilities of Automatic Candidate Filtering includes-

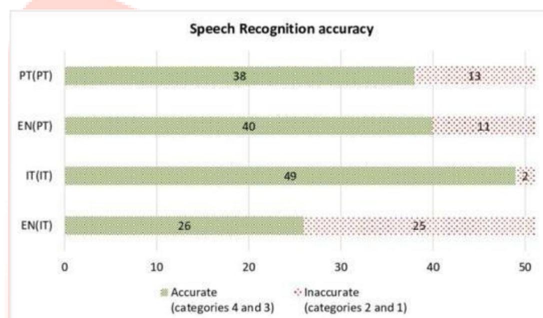
A. Resume Parsing: NLP in ATS can extract relevant information from resumes or LinkedIn profiles. This includes parsing details such as skills, work experience, education, and contact information. Semantic Analysis: NLP helps in understanding the context and meaning of the text. This allows the system to identify key skills, achievements, and qualifications, providing a more comprehensive view of a candidate's profile.

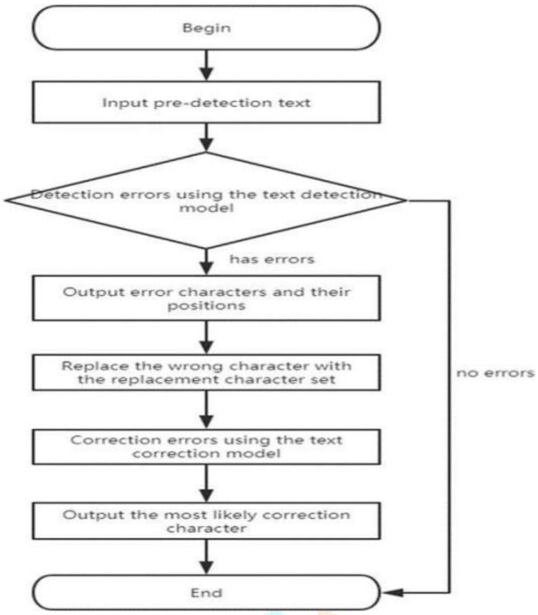
B. LinkedIn URL Handling: OCR may not be directly involved in processing LinkedIn URLs, as URLs are text-based and don't require optical character recognition. Instead, typical web scraping or API-based methods may be used to retrieve the content from the LinkedIn profile.

C. Handling Image-based Content: If a LinkedIn profile includes images with text (e.g.,screenshots or images containing job details), OCR may be employed to extract text from these images.

D. PDF Resumes: OCR can be applied to extract text from scanned or image-based PDF resumes, making the information accessible to the ATS.

E. Evaluation: Evaluate the performance of the speech-to-text system using appropriate evaluation metrics such as word error rate (WER). Identify areas for improvement and iterate on the system by adjusting the models, data, or algorithms as necessary.





Despite these capabilities, Automated Candidate Filtering software still faces several limitations and challenges:

A. Data Privacy: It's crucial for ATS systems to comply with data privacy regulations.

B. Being Ethical: NLP and OCR should be applied ethically and securely, ensuring that candidate information is handled appropriately.

C. Consent Management: Systems may need to manage candidate consent for accessing and processing data from LinkedIn profiles, considering privacy concerns.

D. Overemphasis on Keywords: If the NLP algorithm relies heavily on keyword matching, it may overlook candidates with relevant skills but different phrasing in their profiles.

E. Image Quality: OCR accuracy depends on the quality of the images. Poorly scanned or low-quality image causes confusion in OCR's running mechanism.

V. Result and analysis

In this section, we will present the results and analysis of our research paper on facial emotion recognition using CNN. We developed a CNN model using a modified VGG16 architecture to classify seven emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral.

Dataset:

For audio modeling:

LibriSpeech: a popular dataset for speech recognition and audio modeling, containing a large collection of English audiobooks across various speakers.

For text generation:

A. OpenAI's GPT-2/GPT-3 Datasets: These dataset were used to train huge AI models. For ex- GPT2/GPT3. With the help of mass dataset and reliable training data, though books, internet etc . this feat was achieved.

B. Model Training:

The training phase involves training a text generation model on the chosen dataset. You can use state-of-the-art models like GPT-3 or GPT-4, or even develop your own model using frameworks like TensorFlow or PyTorch. The training process typically involves techniques such as tokenization, language modeling, and fine-tuning.

C. Evaluation Metrics:

To assess the performance of our text generation model, we employed several evaluation metrics tailored to the task of audio-based text generation. Each metric provides unique insights into the model's capabilities and allows for a comprehensive evaluation. The following evaluation metrics were utilized:

(a).Perplexity: Perplexity is a widely adopted metric in language modeling tasks, including text generation. It quantifies the model's ability to predict the next word in a sequence. We calculated perplexity on our test dataset measure the model's overall performance. A lower perplexity score indicates a better ability to generate coherent and fluent text.

(b).BLEU (Bilingual Evaluation Understudy): BLEU is a popular metric used for machine translation evaluation. Although originally designed for translation tasks, it can be adapted for text generation. We employed BLEU to evaluate the overlap between the generated text and reference text. A higher BLEU score signifies better performance in generating text that closely aligns with the expected output.

D.Results: The emotional matrix graph provides a visual representation of the emotional dynamics present in famous person's speeches. The graph captures the emotional content expressed throughout the speech by mapping different emotions along with their intensity over time.

The x-axis represents the chronological progression of the speech, while the y-axis represents the intensity or magnitude of emotions expressed. Each emotion is represented by a distinct color or marker on the graph.

The graph reveals the emotional journey of the speech, showing how the intensity of different emotions fluctuates throughout the duration. It allows us to identify key moments or sections where specific emotions are more prevalent or pronounced. Peaks in the graph indicate high emotional intensity, while valleys represent moments of lower emotional expression.