AI-Powered Application Tracking System with NLP-Based Resume Scoring

Jyotsna Nanajkar[1], Aniket Sable[2], Aniket Machrekar[3], Sakshi Wagh[4], Pooja Udbatte[5]
[1]Assistant Professor, Department of Information Technology, Zeal College of Engineering and Research, Pune, Maharashtra, India
[2,3,4,5]BE Students, Department of Information Technology, Zeal College of Engineering and Research, Pune, Maharashtra, India

Abstract:
The proposed AI-powered ATS assists in recruitment efficiency by automating the process of evaluating and scoring resumes. With Natural Language Processing (NLP) and machine learning, the system analyses applications, awarding scores that reflect how closely each resume matches job descriptions as well as skill relevance and experience. The ATS provides a percentage-based score indicating how closely the candidate matches the job and ranks them for easy hiring using models like Google Gemini Flash 2.0. We use PHP and MySQL for the backend to ensure safe storage of resumes as well as speedy access. A configurable scoring method that enables recruiters to modify attribute weightages according to positional requirements can help shortlist the right candidates for each job. By reducing the tedious effort and facilitating better hiring, this AI-based method streamlines the recruitment process and allows for decisions based on data rather than instinct.

Keywords: AI-powered Application Tracking System (ATS), Resume scoring, Natural Language Processing (NLP), Candidate ranking, Recruitment automation.

I. INTRODUCTION

While in the 21st century [1] at an accelerated pace of hiring industry, today recruiting employers are escapably worthy to screen thousands and thousands of resumes come for a particular job role.
Conventional manual screening [2] processes are time consuming, prone to mistakes and are low in scale which brings inefficiencies into candidate selection and potential mismatch.

In order to solve these issues, various organizations are implementing [3] AI-enabled Application Tracking Systems (ATS) which utilize artificial intelligence in recruiting processes. AI-driven ATS solutions have surged in popularity with modern recruitment, since they boost hiring effectiveness through resume evaluation automation and data-based decision- making support.

An AI-based ATS employs algorithms such as Natural Language Processing (NLP) and develops skills for identifying suitable resumes with precision and speed [4], which is impossible by a human being. Such systems evaluate resumes using a number of parameters – keyword relevance, experience alignment, skill match etc and provide a score reflecting how well each candidate matches the job [5].

This scoring method, known as resume matching/ranking, helps recruiters to get top candidates without the need to go through each resume individually [6]. These systems cut the time-to- hire and also assist to make sure hiring decisions based on equal meritocratic performances of candidates.

Moreover, along with AI resume screening software based on advanced algorithms such as that of Google Gemini Flash 2.0 [7,14], which offer context Elucidation. Unlike traditional keyword-matching systems, these algorithms take it one step further by using the semantic meaning of words and phrases to analyze skills and experience more closely matched with job requirements. Also, technologies such as PHP and MySQL will support the ATS backend by keeping secure data storage and enabling instant access of huge bulk resumes information for real time processing and ranking purposes.

II. MOTIVATION, AIM AND OBJECTIVE

Due to the impact of digitalization and technology on recruitment processes, companies are inundated with hundreds or thousands of resumes for any jobs they post [8]. Such an avalanche of applications leads to longer screening times, human bias, and a high likelihood of missing out on quality candidates.

We will examine how this project aims to build a more efficient AI-powered solution that can help automate these challenges with resume screening and candidate evaluation at scale [9]. We are trying to change the process of talent acquisition for the companies by using AI powered scoring, that helps make it faster, fairer and better.

Ensure, this ATS will use Natural Language Processing (NLP) and machine learning to provide the most accurate assessment of candidates based on skill, experience and keyword matching. In summary, using a customizable scoring model for training & placing the candidates on various roles in an organization; secure data storage and rapid scannability using PHP & MySQL and to improve hiring speed with accuracy.

Now with this AI-driven ATS, we hope to automate the initial phase of screening and in turn drastically decrease the time and cost of recruitment process [10]. Due to human bias in candidate selection, the system will minimize human error and provide better accuracy. ATS will help recruiteres to move their focus to more strategic work, improving overall quality of hire (and possibly density of the resume content by mass, as higher quality hires are more motivated and less reliant on word vomit), while offering candidates a more equitable and transparent selection process simply be eliminating bias from that stage all together.

III. LITERATURE REVIEW

3.1. " Information Extraction from Free- Form CV Documents in Multiple Languages" (2021)

Information extraction from free-form CV documents in multiple languages involves using natural language processing (NLP) and machine learning techniques to analyze unstructured resumes. Unlike standardized CV formats, free-form CVs vary in structure, wording, and language, making extraction challenging.

Advanced NLP models, often powered by deep learning, identify key details such as names, contact information, skills, work experience, and education. Integration with an Applicant Tracking System (ATS) ensures structured data is used for ranking and matching candidates efficiently. Continuous improvements through reinforcement learning and fine-tuning on diverse datasets further enhance the system's adaptability to new languages and CV formats.

Once extracted, the structured data is integrated into an Applicant Tracking System (ATS) for candidate ranking and job matching. The ATS applies scoring algorithms that compare extracted skills and experiences against job descriptions, ensuring better candidate-job fit. Machine learning models further refine candidate recommendations by learning from historical hiring decisions. Handling multiple languages requires multilingual embeddings or translation-based approaches to ensure cross-linguistic compatibility. Challenges include variations in date formats, abbreviations, and domain-specific terminology.

3.2. "Resume Ranking using NLP and Machine Learning" (2016)

Here we work on the project of Automated Ranking of resume using NLP(Machine Learning) methods. In addition to analyzing resumes in accordance with the unique constraints that the hiring company has identified in their hiring processes, the proposed system also extracts relevant information from social profiles (e.g. LinkedIn profile or GitHub) and identifies top match based on extracted features values.

The idea is to shorten the hiring time by sorting resumes based on pre-set standards and make the process of discovering talent efficient and reliable. It automatically scans and rates the resumes based upon various features such as keyword extraction and matching, all with the aim of making the screening process more objective, less reliant on manual effort.

3.3. "Resume Evaluation through Latent Dirichlet Allocation and Natural Language Processing for Effective Candidate Selection" (2014)

This study proposes a new approach for comparing resumes automatically based on Latent Dirichlet Allocation (LDA) and Named Entity Recognition (NER) using the SpaCy library. It parses out various entities within resumes like education, skills and experience and utilizes LDA to identify the most relevant topics or phrases of interest found in this text.

This means it rates resumes on criteria based on the content and not only based on keyword matching as done in a traditional ATS model and can achieve up to 82% accuracy across different assets. Abstract— This paper demonstrates how topic modeling and entity recognition can be combined to bring about a significant increase in the level of objectivity derived from resume assessments that is particularly useful for high-volume hiring situations.

3.4. "Data Analysis From The Applicants tracking System" (2022)

This study examines the role of Applicant Tracking Systems (ATS) for organizations to streamline recruitment, automate resume screening and enable data-driven decisions. It explains ATS as an online software package that enables electronic handling of recruitment needs– job postings, applicant tracking and interview scheduling etc., which frees up time to do other human work for recruiters.

The paper uses examples like BambooHR, the ATS tool that offers in-depth reports on various metrics including candidate source, yield ratios, offer acceptance rates and time-to-hire that help to analyse recruitment process and make necessary changes in strategy.

3.5. "A descriptive study on Applicant Tracking System: Automation software for Recruitment and Selection" (2019)

Introduces a natural language processing (NLP) technique to parse multilingual, unstructured CVs. They suggest transformer and BERT based models that extract and classify information including personal details, education, skills. The model produces data in two levels; section (e.g., personal info) and item (e.g., name, address), running across five vernaculars.

With a system that achieved high accuracy using end-to-end training and attention-based interpretability (and even teaching people how to classify competence in skills listed on CVs), it also includes the ability for job-seekers to self-assess their skills.

3.6. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" (2019)

Released BERT (Bidirectional Encoder Representations from Transformers), a new model for NLP tasks that was, at the time, state-of-the-art. BERT takes the advantage of bidirectional training using a masked language model that conditions on both left and right context for making predictions during pre-training.

This model achieved state-of-the-art results on a wide range of NLP tasks, including question answering and language inference. With a flexible architecture for fine-tuning time-dependent downstream tasks, BERT is leaps forward from the previous unidirectional models in language representation learning.

3.7. "Making Hiring Process Effective through Application Tracking System" (2023)

It explains the functionalities of ATS software and how they automate sourcing, screening and managing applicants. ATS software makes it faster and easier for HR professionals to sift through a high volume of applications while ensuring accuracy. reducing the manual tasking associated with screening, as well as streamlining identifying top candidates.

It helps in improving the quality of candidates and speed of recruitment by filtering out the best ones from the rest using specific keywords matching with criteria; whereas, ATS gives an improved experience to job seekers through timely notifications and feedback.

In addition, by centralizing and securing candidate information these tools help organizations save on recruitment cost and allow improved compliance with data protection regulations.

3.8. "Applicant Helper System for Resume Using Python and NLP" (2023)

An NLP-based Hybrid System, AHS, to stand out candidates resume preparation. While a traditional ATS filters resumes against the requirements for a certain job, AHS helps users to make their resume match with the job description as much as possible so that they will have higher chances of being selected.

AHS employs natural language processing to extract and analyze keywords, enabling applicants to tailor their resumes for particular job positions. This open-source AHS guides users on the strengths that are needed in their resumes while also increasing interview chances.

This analysis surpasses the potential of AHS to diminish resume rejection rates, as it has been stated that future development can be performed on mobile platforms. AHS ensures resumes are optimized for Applicant Tracking Systems (ATS) by structuring content effectively, improving parsing accuracy. It adapts to different industries by recommending relevant skills, job roles, and achievements based on sector-specific requirements.

Reference No.	Methodology	Technology/Algorithm Used	Limitations / Research Gap
1.	Extracting key details from unstructured CVs in different languages using NLP and machine learning.	Deep learning models, reinforcement learning, ATS integration.	Struggles with date formats, abbreviations, and industry-specific terms.
2.	Ranking resumes by analyzing extracted skills and experience for better hiring decisions	NLP, keyword extraction, LinkedIn/GitHub profile analysis.	May miss strong candidates due to feature extraction limitation.
3.	Resume assessment using LDA and NER to categorize relevant topics and entities.	LDA modeling, SpaCy NER, entity recognition.	Keyword-based ATS models may not capture deeper context.
4.	Using ATS data analytics to enhance recruitment strategies and decision-making.	ATS software, data tracking, recruitment metrics.	Effectiveness relies on high-quality data; possible biases in automation.
5.	Automating multilingual resume parsing with transformer models.	BERT-based NLP, attention mechanisms, end-to-end learning.	Requires large datasets and high computational resources.
6.	BERT model for deep bidirectional text understanding in hiring processes.	Transformer architecture, masked language modeling, transfer learning.	Needs extensive fine-tuning for domain-specific applications.
7.	Streamlining hiring via ATS automation for screening and candidate selection.	Keyword-based ranking, automated notifications, applicant filtering.	Can exclude strong candidates if resumes lack exact keyword matches.
8.	Enhancing resume content to align with job descriptions for better selection chances.	NLP, keyword extraction.	Limited accuracy in highly specialized or creative job fields.

IV. SYSTEM DESIGN:

IV.1 System Architecture:

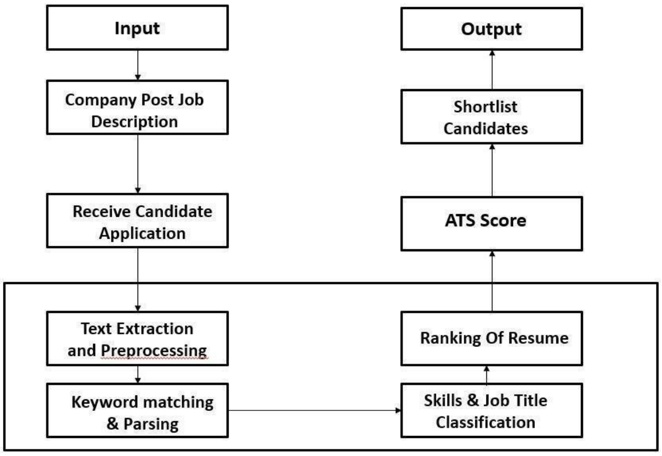

Fig.1.: System Architecture

The system consists of the following components:

1. **Company Post Job Description:**

It begins with the company posting a job description outlining the requirements and skills needed for the qualifications.

2. **Receiving an Application from a Candidate:**

The candidates who apply for the job add up to the ATS,Sourcing recruitment basics[11] | full course. Such applications often consist of resumes and, occasionally, supplementary documents.
The first step in the process is text extraction and preprocessing.
Text extraction — This step consists in obtaining the textual information from the resumes that were submitted to you. Data preprocessing (such as removing noise, text normalization and formatting and other process to prepare the data for analysis & processing) This step may consist of things like eliminating unnecessary characters, such as applying lower case conversion to texts, etc.

3. **Keyword Matching & Parsing:**

In this module, the text extracted from resumes is parsed and keyword matching will be done to check the presence of keywords related to skills, qualifications and any terms specific to jobs.

4. **Classification of job title and various skills:**

This step processes the resume and classifies certain skills and job titles found in it. This helps the ATS assess how closely a candidate's experience aligns with the job requirements.

5. **Ranking of Resume:**

Resumes are scored based on keyword matching and classified in this way. It can classify based on Different types like skill relevance, years of experience, and matching the title.

6. **ATS Score:**

The ATS rates each resume in accordance with the job description. Score — indicates a resumé alignment with job requirements of the candidate.
Shortlist Candidates:
After that, interview call for next rounds.

IV.2 Methodologies:

Powered by Google's Gemini Flash 2.0 model, the Al Application Tracking System analyzes resumes and job descriptions and uses capability matching to improve recruitment efficiency.

Step one is collection, cleaning and standardizing resumes and job descriptions for subsequent analyses. NLP techniques are used to extract key features such as skills, experience and education to provide a structured dataset.

Gemini Flash 2.0 uses semantic analysis to analyze resumes and job descriptions on context rather than just keyword matching

It understands these relationships, not only the profile of a candidate vs job requirements level but also their agreement and difference with regards to specific terms.

Candidate rank is determined through a proprietary ATS scoring algorithm that merges keyword and semantic analysis results [12], prioritizing skills fit, years of experience, and education level.

We save the scores in a MySQL backend managed by PHP scripts such that recruiters can easily get and give rankings.

The user-facing frontend built with HTML, CSS and JavaScript allows the recruiters to see and filter out the ranked resumes in real-time.

This simple, Al-driven system improves candidate selection and helps hiring professionals get there much faster.

V. CONCLUSION

To sum up, the benefits of Al-powered Applicant Tracking Systems (ATS) is that it simplifies recruitment process while allowing data-driven candidate screening and reduce hiring team's workload [13].

Al-based ATS automate resume parsing, keyword matching and candidate ranking to work through thousands of applications and find the most relevant profiles.

That said, they do have drawbacks such as bias features, inability to process some resume formats and grades metrics like creativity or cultural fit[14].

This is a terrific start, but as we move into the future, further developing these systems to ensure they are free of bias and provide an accurate assessment will be key.

An Al-powered ATS will never be perfect alone and it should always be complemented with human supervision to help ensure that the hiring process is as fair, effective, & efficient as possible[15].

Improvements in natural language processing and techniques for reducing bias will further improve the performance of ATS tools as Al continues to advance. The use of Al needs to be combined with human judgement and continuous improvements in the algorithm behind it will need to avoid overfitting, that is what a true organization defining an efficient hiring tool should aim for it [16].

VI. REFERENCES

[1] Davor Vukadin, Adrian Satja Kurdija, Goran Delac, Marin Silic, ‘‘Information Extraction From Free-Form CV Documents in Multiple Languages’’, Jun,2021, vol.9, pp.129-139.

[2] T.Khan,A. Juneja, Adnan Ayyas Shaheen Momin, “Resume Ranking using NLP and Machine Learning”, vol.20, pp. 1735–1780,May 2016.

[3] Vidhita Jagwani, Smit Meghani, Krishna Pai, ‘‘Resume Evaluation through LDA and NLP for Effective Candidate Selection’’, pp. 1–9, Dec. 2014.

[4] Miroslava Peicheva, ‘‘Data Analysis From Applicant Tracking System’’, pp.6-15, Dec, 2022.

[5] Suraj M, Aruna Kumari K, Binila B Chandran “A descriptive study on Applicant Tracking System: Automation software for Recruitment and Selection”, vol. 9, pp. 1735–1780, Feb 2019.

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, pp. 143-152, May, 2019.

[7] Ankitha K, Balu.L, “Making Hiring Process Effective through Application Tracking System”, pp. 499–506, Dec 2023.

[8] Lynsha Helena Pratheeba, Prajwal S, Pawan Kumar,‘‘Application Helper System for Resume Using Python and NLP’’ pp. 156–178, April 2023.

[9] C. H. Ayishathahira, C. Sreejith, and C. Raseek, ‘‘Combination of neural networks and conditional random fields for efficient resume parsing,’’ in Proc. Int. CET Conf. Control, Commun., Comput. (IC4), Jul. 2018, pp. 388–393.

[10] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, ‘‘Conditional random fields: Probabilistic models for segmenting and labeling sequence data,’’ in Proc. 8th Int. Conf. Mach. Learn. (ICML), San Francisco, CA, USA: Morgan Kaufmann Publishers, 2001, pp. 282–289.

[11] J. Pennington, R. Socher, and C. D. Manning, ‘‘Glove: Global vectors for word representation,’’ in Proc. Empirical Methods Natural Lang. Process. (EMNLP), 2014, pp. 1532–1543.

[12] M. Tosik, C. Lygteskov Hansen, G. Goossen, and M. Rotaru, ‘‘Word embeddings vs word types for sequence labeling: The curious case of CV parsing,’’ in Proc. 1st Workshop Vector Space Modeling Natural Lang. Process., Denver, CO, USA, 2015, pp. 123–128.

[13] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, ‘‘Distributed representations of words and phrases and their compositionality,’’ in Proc. Adv. Neural Inf. Process. Syst., 2013, pp. 3111–3119.

[14] P.Shivratri, P.Kshirsagar, R.Mishra, R.J. Damania, N. Prabhu, ‘‘Resume parsing and standardization”, Int.J.Comput, Sci.Eng., vol.3, pp.129-132, 2015.

[15] D. Chandola, A. Garg, A. Maurya, A. Kushwaha,‘‘Online resume parsing system using text analytics,’’ J. Multi-Disciplinary Eng. Technol., vol 9, no.1, pp. 1-5,2015.

[16] A. Cernian, D. Carstoiu, B.Martin, “Semi-automatic tool for parsing CVS and identifying candidates abilities and competencies”, pp.1-5, Mar.2017.