International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Resume Authenticity Assessment and Credibility Scoring Using Machine Learning and Linguistic Pattern Analysis

Dr. Sankati Ramakrishna[1], R. Ponna krishna vamsi[2], V. Shanmukha Rohit[3], P. Kushal[4], N. Chandu[5]
[1]Assistant Professor, Dept. of CSE–Data Science
KKR & KSR Institute of Technology And Sciences, Guntur
Email: rk4uin2010@gmail.com[1]
[2,3,4,5]B.Tech Student, Dept. of CSE–Data Science
KKR & KSR Institute of Technology And Sciences, Guntur
Emails: 22jr1a44b6@gmail.com, 22jr1a44c7@gmal.com, 22jr1a44b3.csd2@gmail.com, 22jr1a44a5@gmail.com

***Abstract*—In the era of digital recruitment resume screening is the crucial step if it is not done properly will lead to Fake extracted resumes and HR managers uses the key word based application tracking system and rule based filtering tools to overcome the challenges and speedup and reducing the cost effective the recruitment process and accurately and detecting important information by using natural language processing that fits for job description and identifying the fraud information in the resume by providing the credibility scoring this project uses the uses the combination of the machine learning and natural language processing framework for Resume Authenticity Assessment and Credibility Scoring our system integrate the BERT transformer architecture and word embedding techniques and AI-generation detection and email automation.and scores in ascending order Federated training for privacy and secure resume and to avoid the human error to train models using resumes data set this can be processed through ensemble classifier SHAP Explainability and this output gives the credibility score and a gives the specific skill mismatch under 10 to 15 seconds fastly scalable integration into ATS for cost-effective, equitable hiring.**

***Index Terms*—Credibilityscoring, ApplicationTrackingSystem, Bert, NaturalLanguageProcessing, Recruitmentautomation, Ma- chine learning**

I. Introduction

In the today's world the technology improved day by day and number of job applications also increased and in the earliest days the companies changing the hiring process and shifted to the online screening process due to the job portals and the Hr and company relay on the digital recruitment but there are some drawbacks with the digital recruitment such as keyword based filtering means the if words are matches the exact keyword with the job description and receiving the number. The resume screening is evaluated based on the application tracking system and they don't check semantics in the resume; it leads to poor hiring quality and misleading information and increasing overall cost and getting the fake resumes.

Traditional resume screening are completely depends on the manually and Keyword based Application tracking Systems and manual screening depends on the human declaration and it also more time consuming process and leads to high chances of the human errors and other side the key word based Application tracking Systems accurately screening process by keyword against the job description but it improves the time and cost but the high chances of the getting the accurate result and less understanding of semantics.candidates who strate- gically manipulate keywords may pass through automated filters despite having insufficient qualifications while genuine candidates less chances of getting the resume that fits for the job description.

The recent enhancement and technologies such as the Machine learning and the Natural language processing will give the right solution to the recruitment process the modern and new techniques in NLP capture semantics and linguistic patterns in the resume text data transformer based archi- tecture like the BERT(Bidirectional encoders representation from transformers) [5] it performs the contextual analysis and embeddings and it also understand what are the skills and jobs projects that matches the another resume and the secure and resumes are having the genuine and personal information it is more secure in the resume evaluation process to secure these make the centralized storage so it give the more au- thenticity scoring and securely. Where ensemble classification techniques are used to maintain predictive accuracy.

The development of the system will give more scalable and high accurate solution for the automatic resume screening and authenticity assessment ,its primary way to contribute that depends on the generating credibility scoring for the each resume and hrs professional identify the risks quickly the a main key features includes the resume ats given under the 10 to 15 seconds and suitable for the small to large amount of enterprises reduces the human error and mitigate the bias in screening process and improves the efficiency and effectiveness,reliably of trusted resumes in the recruitment process and automated screening time

II. Literature Review

Applying Google-BERT-Based NLP for Automated Resume Screening and Candidate Ranking [1] According to the survey it introduces the google bert architecture in this they uses the datasets of the resumes and job related information and these are analyzed and gives the we find that the Natural language processing techniques will increase resume filtering and accuracy over existing Application tracking systems. The study points to the advantages of only contextual understanding to the candidate credibility task

A Machine Learning for Automation of Resume Recommendation System (2019) [2] This study majorly we focus on resume recommendation system using some ml models and deep learning models and this study describes the select the jobseekers according to job and this Research results the and suggest on the performance based shortlisting candidate using deep learning techniques and machine learning techniques were better than old rule based Application Tracking Systems [3].

Design and Development of Machine Learning-Based Resume Ranking System (2021) [3] The study mainly focuses on the resume ranking system based on the different factors such as experience and skills that are required to job description skills and Achievements projects and normalizes the resumes and plans the resumes with a ranking system. And reducing cost for the recruitment process [4]

Application of LLM Agents in Recruitment: A Novel Framework for Resume Screening [4] Chengguang Gan, Qinghao Zhang, Tatsunori Mori (2024) In this paper the resume screening done through the large Language models and agents are used ,LLM-systems completely depends on the recruitment efficiency through automatically analyzing and summarizing, evaluating and access the ranking candidates better than normal ATS models The literature survey shows the LLMs are typically used the keyword based filtering and gives resume ranking It finds that transformer-based natural language models like BERT [5] have a high impact on recruitment process and efficiency and accurate by boosting recruitment process effectiveness via higher resume-job description fit, less human error and trust candidates that were shortlisted. Literature surveys and identify the research gaps is manual resume screening is consume the more amount and also include high cost process and traditional resume screening process such keyword based ats shortlisted and ranks based on the some keyword match with the job and the existing systems focus on the recruitment side not on the job seeker side for feedback mechanism and only gives the one resume not prioritize the resumes high to low order. It works with the multiple formats it leads to analyze the resumes in a difficult manner. And hrs spends lots of time screening and performance issues ,high chances of the human error .Some acts and systems ignore main meaning and sometimes reject the qualified candidates.

III. Problem Statement

In modern era the resume screening is difficult task for the recruiters.The company receives large no of resumes for precise job role. It is very complex task to review each and every resume. It is time consuming process and based on human judgement which lead to higher bias and error. To handle this problem, the companies mostly rely on keyword based applicaton tracking system.

IV. Proposed System

The proposed system that resumes authenticity assessment and credibility scoring using the Nlp and Machine learning approaches that used transformer based architecture that understand the contextual meaning and semantics to enhance the resume screening process and also send email to job aspirants quickly and give the feedback to the user. By adding authentication and authorization, it classifies the resume in high to low order. The system handles a large amount of massive training data.

V. Methodology

This part is mainly focus on the development resume authenticity and credibility scoring tools We used the step by step mechanism to develop the project and also project is divided into different phases and key functions using the resume drag and drop and evaluating resume screening planning and setup,core development ,testing and evaluation ,deployment,this ensure the scalability and efficiency the entire process is developed using the python and javascript And some python libraries to train and evaluate the models like the Hugging Face and Transformer ,flask for the backend ,scikit learn for evaluation metrics. And making a beautiful dashboard.

A. Planning & Setup

In this Phase they define all requirements that are required to develop the project. Such as frontend for react js and backend for flask and bert model for semantic checking and feedback system. And folder setup.python and javascript setup. Database setup and email authentication and authorization

B. Dataset Collection

Collecting the data of resumes dataset and job pair datasets from the the various sources like kaggle resume dataset and linkedin and indeed dataset and websites that include the both fresher and experienced people datasets we are using the labelled dataset that is fit or not fit data and preprocess the data such as lemmatization stopwords removal representing the numerical data and stored in the json or csv format to easy to access.

C. BERT Model & Semantic Understanding

Collect the pretrained Bert based and uncased model from hugging face and and fine tune with the data to read the full context and meaning and train with data set for binary classification (fit/Not) runs on the gpu for 10 to 20 epochs.and monitor the loss and epochs add semantic features and integrate the embeddings techniques for deep match and understanding Integrate with the backend by using flask/fastapi to monitor the api calls from frontend to backend and backend to frontend.

D. Backend development

Building the backend architecture using the flask to create backend functions that uploads the resume and match with bert scoring.adding the nlp preparation such skills and experience using spacy to handle large data uses the batch processing and integrate the bert model.call the model and pairs of output,labels.

E. Database Design

Design the database that user can logging details and resume information safely To establish a connection with the database using the SQLAlchemy add index fast quires it stores the processed resume.

F. Frontend

For Frontend using the react js that uses the component for usability and tailwind css for the coloring ui and interactive dashboard and shows the top 10 resumes. Works with different job descriptions and connects to the backend flask. Adding authentication and authorization.

G. Resume Ranking

Adding the backend logic to sort the bert score and adding different factors to implement this add the backend reuters the json format and frontend returns the table give resume ranking with time less than 1 second.

H. Authentication & Authorization

Implement the authentication and authorization for secure access to the tool and also from the recruiter side and store the logging information through the JWT(json web token) admin can access the data.

I. Email Automation & Feedback

By integration of smtp protocol automatic mail is sent to the candidate by the HR and SHAP will help to explain why the resume is not fit using the rule based summaries.

Validation and testing:validate the model using the different metrics such as ROC and F1 and precision confusion matrix score tested the workflow through the via postman and classification metrics to validate model accuracy by whether the model underfitted or overfitted and these key metrics are used to evaluate the accuracy of ranking.

VI. Flow Diagram

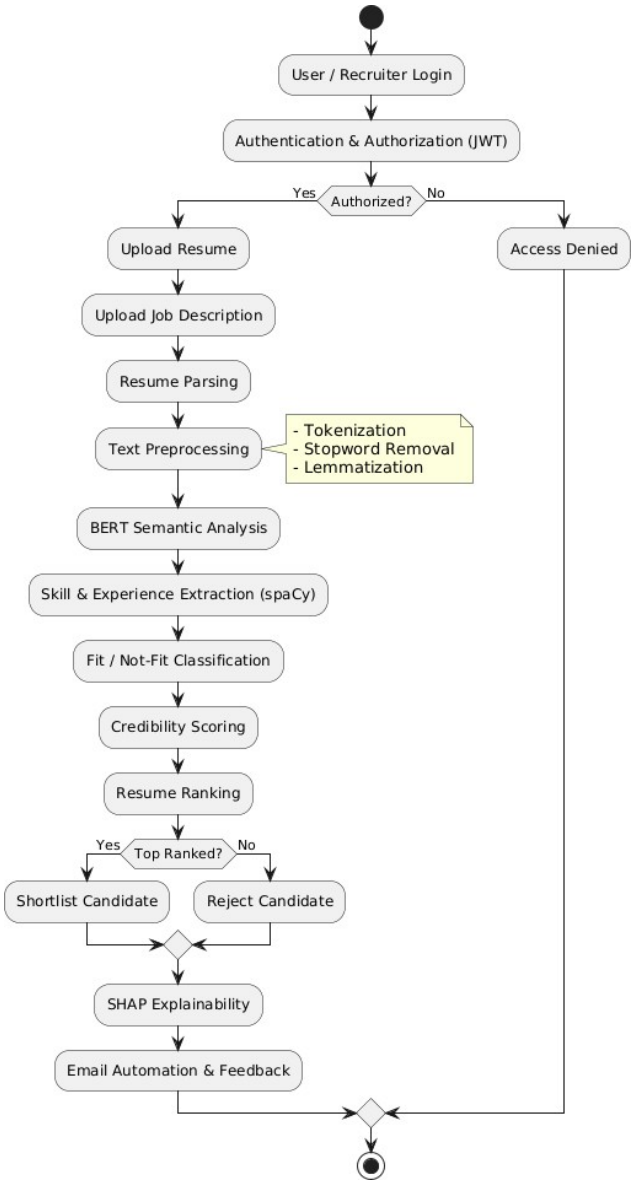

Fig. 1. System architecture flow diagram for resume authenticity assessment

VII. Results and Discussions

The systems that are tested using various resume and job descriptions and various factors and AI resume tools accuracy 27% higher than the normal keyword based tools accuracy and hr screen time reduced and efficient recruitment process and for feed back given by the tool user can understand the test and they align with job description.the system works with multiple kinds format data with the resultant. And also increase performance.

There are different metrics are used to evaluate the models and such classification metrics gives good amount of accuracy which is f1 score and precision and recall.

TABLE I
Model Evaluation Metrics

Metric	Value
F1 Score	92.5%
Accuracy	94.6%
Precision	99.9
Recall	96.09%

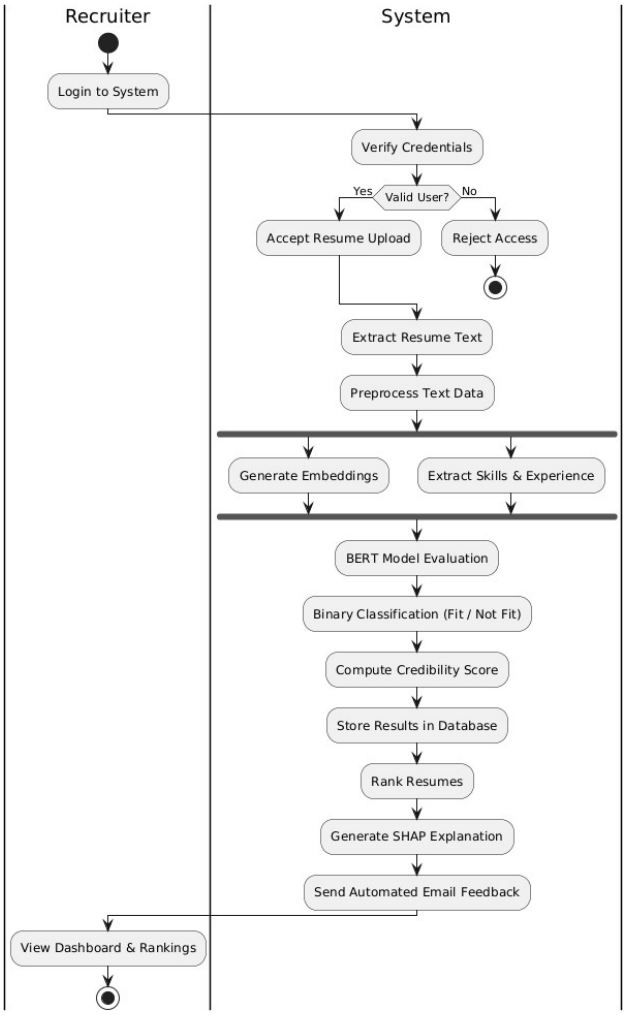

Fig. 2. Data processing pipeline for resume credibility scoring

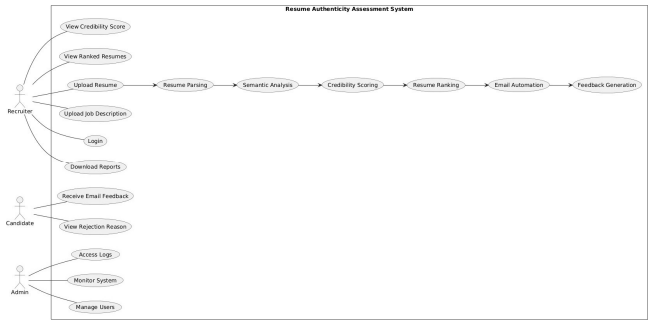

Fig. 3. Model training and evaluation workflow

VIII. Discussions

Resume authenticity and credibility scoring using the ml and bert transformers are make the unique in the recruitment process and semantically matching candidates with job descriptions and achieving the accuracy of 94 and reducing the screen time to the 75 percentage by integration of react and flask frontend and backend with machine learning models and secure data storage sql and framework for automation emails scalable in hiring process and evaluate the resume with various aspects such experience and freshers projects and semantic checking and authentication and authorizations

	Predicted: Suitable	Predicted: Not Suitable
Actual: Suitable	341	14
Actual: Not Suitable	0	396

Fig. 4. Credibility scoring distribution across different resume categories

Resume Type	No. of Fields	Correct Extraction	Accuracy (%)	Notes
Plain Resume	Text	56	96.7%	Minimal formatting errors
Standard Template Resume	60	54	90.0%	PDF structure preserved well
Modern/Graphic Resume	60	48	80.0%	Text in columns/images reduces extraction accuracy
Infographic Resume	60	42	70.0%	Icons, shapes, and charts difficult for parser
Infographic Resume	60	42	70.0%	Icons, shapes, and charts difficult for parser
Image-Converted Resume	60	32	53.3%	Only text-based OCR content extracted

Fig. 5. System performance comparison with traditional ATS

IX. Conclusion

The resume screening process will improve when compared to the keyword based ats tools through the fine tuned bert based transformer architecture and it will not only improve and it will improve the user experience and our system demonstrate the operational success and hiring process by decreasing the recruiter workflow, it works with different format of resumes very efficiently and effectively by bert model and ensure high authanticity explainable ai will help whether the person why rejected to user through the email automation. And top ten displayed for the job description and enhancement of security further.,the future enhancement will understand the

deeper understating of context and human bias reduced greatly the system performance will be improved through real time analytics and dynamically.

REFERENCES

[1] Jayanti, L. P. S. D., and Meditya Wasesa. ”Application of Predictive Analytics To Improve The Hiring Process In A Telecommunications Company.” Journal CoreIT 8, no. 1 (2022): 32-39.
[2] Baykal, Elif. ”Digital era and new methods for employee recruitment.” In Handbook of research on strategic fit and design in business ecosystems, pp. 412-430. IGI Global, 2020.
[3] Alrakhawi, Hazem AS, Randa Elqassas, Mohammed M. Elsobeihi, Basel Habil, Basem S. Abunasser, and Samy S. AbuNaser. ”Transforming Human Resource Management: The Impact of Artificial Intelligence on Recruitment and Beyond.” (2024).
[4] Rhemananda, Happy, Dima Roulina Simbolon, and Olivia Fachrunnisa. ”Blockchain technology to support employee recruitment and selection in industrial revolution 4.0.” In Proceedings of International Conference on Smart Computing and Cyber Security: Strategic Foresight, Security Challenges and Innovation (SMARTCYBER 2020), pp. 305-311. Springer Singapore, 2021.
[5] Allal-Che´rif, Oihab, Alba Yela Ara´nega, and Rafael Castan˜o Sa´nchez. ”Intelligent recruitment: How to identify, select, and retain talents from around the world using artificial intelligence.” Technological Forecasting and Social Change 169 (2021): 120822.
[6] Trinh, Quynh, and Thanh-Tuan Dang. ”Automatic process resume in talent pool by applying natural language processing.” In Proceedings of international conference on logistics and industrial engineering 2021, pp. 234-240. 2021.
[7] Thakkar, Darshit. ”Leveraging AI to transform talent acquisition.” International Journal of Artificial Intelligence and Machine Learning 3, no. 3 (2021): 7.
[8] Laurim, Vanessa, Selin Arpaci, Barbara Prommegger, and Helmut Krcmar. ”Computer, whom should i hire?–acceptance criteria for artificial intelligence in the recruitment process.” (2021).

