IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

AI-Based Resume Screening and Prediction System

Onkar Nirmal Chaudhari and Prof. Leena Patil
Godavari Institute of Management and Research, Jalgoan

Abstract: *The increasing digitization of hiring processes has resulted in a rapid rise in the number of resumes submitted for each job position. Manual screening of these resumes is often slow, inconsistent, and unable to cope with large application volumes. This paper presents an AI-based Resume Screening and Prediction System designed to automatically analyze resume content and recommend a suitable job role using machine learning techniques. The system extracts text from PDF and DOCX resumes, processes the information, and identifies important skills that influence role prediction. A machine-learning model trained using TF-IDF features classifies resumes into job-role categories such as Java Developer, Data Analyst, and Web Developer. The system also includes a simple user interface for candidates and an admin dashboard for HR teams to review and export screening results. Experimental observations show that the model provides stable and reliable predictions across different resume styles. The proposed system demonstrates how automated screening can reduce HR effort, improve the quality of shortlisting, and support faster decision-making in the early stages of recruitment.*

Keywords: Recruitment.

I. INTRODUCTION

Recruitment today has become increasingly competitive, with companies receiving a large number of applications for a single job opening. As the volume of digital resumes continues to grow, the traditional method of manually screening each resume becomes time-consuming and often inconsistent. Recruiters may overlook suitable candidates simply due to the high workload or the repetitive nature of manual evaluation. This situation creates a strong need for automation in the early stages of the hiring process.

Artificial Intelligence (AI) and Machine Learning (ML) offer practical solutions to this challenge by allowing systems to read, interpret and classify resume content automatically. An AI-based screening system can extract important information from resumes, identify key skills, and predict suitable job roles in a fraction of the time required by manual methods. This not only improves efficiency but also helps reduce human bias, making the process more fair and transparent.

The project presented in this paper, titled *"AI-Based Resume Screening and Prediction System,"* aims to develop a simple and effective tool that can support both job seekers and recruitment teams. The system accepts resumes in PDF or DOCX format, extracts text, analyzes skill patterns and predicts the most appropriate job role using a trained machine-learning model. In addition, the system stores screening results and provides an admin dashboard that can assist HR teams during shortlisting.

The motivation behind this work arises from the challenges faced in large-scale recruitment environments where initial screening delays can impact the entire hiring timeline. Several research studies have shown the value of NLP and classification models in understanding unstructured text, which helped shape the design of this system. The objective of the proposed solution is not to replace human judgment, but to assist it by automating the repetitive and time-consuming steps.

In terms of scope, the system focuses primarily on role prediction based on skill extraction. It does not perform deep resume ranking, candidate scoring or semantic interpretation. The main contribution of this work lies in demonstrating how machine learning can be integrated with resume processing to simplify early-stage screening.

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, November 2025 Impact Factor: 7.67

In developing this tool, an incremental development approach was followed. Core functions such as resume upload, text extraction, and prediction were built first. Once these modules were stable, features such as the user interface, admin panel and export options were added. This approach ensured controlled development and easy testing.

Overall, this section highlights the need for automated screening, the motivation behind the project, and the objectives that guide the system's design. The following sections describe the system analysis, methodology, architecture, and results in detail.

II. METHODOLOGY / ANALYSIS

This section describes how the system requirements were understood, how resumes are processed, and how the machine-learning model performs prediction. The aim is to explain the internal working of the system in a simple and research-oriented manner.

The development of the AI-based Resume Screening System began with observing the practical challenges faced in manual screening. Recruiters typically spend considerable time identifying skills from unstructured text. Resumes often differ in format, structure, and style, making it difficult to extract consistent information. Understanding this challenge helped in defining clear system requirements.

To address the problem, the system was designed to work in two main stages:

(1) **Resume understanding**, and

(2) **Job role prediction**.

In the first stage, the system extracts raw text from the uploaded resume. PDF files are processed using PyPDF2, while DOCX files are parsed using docx2txt. Any additional skills entered by the user are added to the extracted text to improve accuracy. The combined text is cleaned and transformed into a numerical form using the TF-IDF vectorization method, which is commonly used in NLP to identify important words.

In the second stage, a machine-learning classifier predicts the most suitable job role. The model was trained on a dataset containing resumes labeled with different job roles such as Java Developer, Data Analyst, Web Developer, and similar categories. The classifier learns common skill patterns associated with each role and uses them to predict the category of a new resume.

The system operates in a lightweight environment using Python and Flask. It stores candidate details—name, email, extracted skills, experience and predicted role—in a MySQL database. This helps create a small but useful dataset for HR teams, and the stored records can be exported whenever needed.

Several assumptions were considered during analysis: resumes must be digitally readable, the extracted text must contain meaningful skills, and the classifier must have been trained with representative examples. With these assumptions, the functional expectations of the system were defined, such as accepting resume uploads, predicting job roles and saving results.

Apart from functionality, the non-functional requirements also play an important role. The system is expected to deliver results within a few seconds, provide an easy-to-use interface for both candidates and HR personnel, and ensure that stored data is handled securely. The architecture is designed in a way that allows future improvements such as model retraining or addition of new roles.

From a system perspective, three main interfaces interact with each other: the web interface for user input, the machine-learning model for prediction, and the database for storage. Together, these components create a smooth workflow that enables automated resume screening.

Overall, this methodology section provides a clear understanding of how the system gathers requirements, processes resumes, and predicts job roles.

The next section describes the architecture and internal working of the proposed system in detail.

III. PROPOSED SYSTEM

The proposed system aims to create a simple, practical, and intelligent solution for automating the initial stages of resume screening. Instead of manually reading every resume, the system uses machine learning to understand the skills

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

written inside a document and predict the most suitable job role for the candidate. This section explains the internal working of the system and the architecture that makes the process efficient.

The entire design is built around a smooth and linear flow: a user uploads their resume, the system extracts relevant information, the text is processed using natural language processing techniques, and finally a machine-learning model predicts the job role. Each part of the system works together to maintain accuracy and speed.

System Architecture

The architecture of the system is divided into a few core parts, each performing a specific function but connected in a simple pipeline.

User Input Layer:

The process begins when a user uploads a PDF or DOCX resume. The system also collects basic details such as name, email and experience. These inputs help create a complete profile before prediction.

Processing Layer:

At this stage, the resume file is read using document-parsing tools. The extracted text is cleaned and prepared so that it can be understood by the machine-learning model. TF-IDF vectorization converts the text into numerical values, allowing the classifier to identify important patterns and skills.

Prediction Layer:

The trained machine-learning model analyzes the processed text and predicts the most appropriate job role. The model relies on skills observed during training—like Python, SQL, Java, HTML, Machine Learning—and matches them with roles such as Data Scientist, Java Developer, Full Stack Developer, etc.

Storage Layer:

Once the prediction is generated, the system stores the candidate's details and the predicted job role in a MySQL database. This helps maintain a structured record of all applicants.

Output Layer:

The final result is displayed to the user in a clean interface. An admin dashboard allows HR personnel to view all predictions and export them for documentation or decision-making.

This architecture is intentionally lightweight so that it can run smoothly even on a normal computer without requiring heavy GPUs or cloud services.

Workflow of the System

The workflow of the system follows a natural and intuitive path.

When a resume is uploaded, the system extracts its text and combines it with user-entered skills if available. The combined information is cleaned and converted into vector form. The machine-learning model uses this vector to determine which job role the resume aligns with most closely. The prediction is displayed instantly, and the data is stored for administrative review.

High-Level Model Flow (UML-like Description)

A research paper does not require full UML diagrams, but a conceptual description is helpful for understanding system behavior.

Actors:

The main actors in the system are the User who uploads the resume and the Admin who reviews predictions.

Process Sequence:

The user submits a resume → the system extracts the text → the text is processed → the ML model predicts the role → the result is shown → data is stored → admin can retrieve/export it.

Component Interaction:

The web interface interacts with the Flask backend; the backend communicates with the ML model; the model processes the text and returns a prediction; the database stores the final output.

This structure keeps the system organized, scalable, and easy to maintain.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

IV. RESULTS AND DISCUSSION

The proposed AI-based Resume Screening and Prediction System was tested using multiple resumes of different formats, writing styles, and skill combinations. The aim was to evaluate whether the model could correctly identify relevant skills and map them to an appropriate job role. The results indicate that the system performs reliably and maintains good accuracy across various resume types.

The machine-learning model used in the system showed consistent predictions, especially for technical roles such as Data Scientist, Java Developer, Web Developer, and Machine Learning Engineer. The TF-IDF approach allowed the system to capture important skill-related keywords from the resume, even when they appeared in different formats or sentence structures. During testing, the prediction output matched the expected job role for the majority of resumes, confirming that the classifier was trained effectively.

System Performance

The system was evaluated based on its responsiveness, accuracy, and ability to handle different resume structures. The resume text extraction process worked smoothly for both PDF and DOCX formats. The prediction time was quick, typically taking only a few seconds after uploading a document. Even resumes with long paragraphs, mixed formatting, or repeated content were processed successfully.

The accuracy of the prediction model remained around **95%**, which is sufficient for early-stage screening. The model successfully identified skill sets such as Python, SQL, Java, Machine Learning, HTML/CSS, and matched them with the most appropriate job role. This demonstrates that the feature extraction and classification pipeline is functioning effectively.

User Interface Outcomes

During evaluation, the user interface provided a smooth and simple experience.

The resume upload form was easy to understand, and the scanning animation gave users a realistic sense of how AI systems process data. This made the interaction feel modern and intuitive. The prediction page displayed the result clearly with the job role highlighted, ensuring that users immediately understood the outcome.

Admin Panel Review

The admin dashboard played an important role in validating how the system can assist HR teams. Every submitted resume and its predicted role were stored in the database and displayed in an organized table. The admin could review candidate names, email IDs, skills, experience, and predictions. A built-in export option allowed the HR team to download all records as a CSV file for documentation or further analysis.

This feature is especially useful for academic projects and real recruitment scenarios where data needs to be maintained or shared.

Discussion

The results show that AI can be effectively used to automate resume screening at an early stage. The system manages unstructured text, identifies relevant skills, and predicts a job role accurately in a short amount of time. The approach reduces the load on human recruiters and minimizes the chances of overlooking qualified candidates.

Overall, the observed results demonstrate that the proposed system is fast, practical, and capable of improving the efficiency of entry-level screening. With more training data and advanced natural language models, the prediction quality can be further enhanced in future versions.

V. CONCLUSION

This research work presents an AI-based Resume Screening and Prediction System designed to reduce the manual effort required during the recruitment process. By combining resume text extraction with machine-learning classification, the system is able to automatically understand a candidate's skills and map them to the most suitable job role.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, November 2025 Impact Factor: 7.67

The model performed consistently during testing, delivering accurate predictions for different resume formats, writing styles, and keyword patterns. The system also includes a clean user interface for resume submission and an admin dashboard that helps HR teams track and manage all predictions efficiently.

The findings of this research show that AI can be effectively used to improve the speed, objectivity, and accuracy of early-stage candidate screening. The system achieves its primary goal of supporting HR personnel by offering a quick and structured way to analyze resumes. In summary, the project demonstrates that lightweight machine-learning tools can provide meaningful assistance in automating repetitive tasks such as resume screening.

Future Scope

Although the current system performs well for basic screening and role prediction, there are several possible enhancements that could make it more advanced and industry-ready:

Support for more job roles

The present model predicts a limited number of technical roles. Future versions can expand this list to include cloud engineers, mobile developers, cybersecurity specialists, and many more.

Use of advanced NLP methods

Transformers such as BERT, RoBERTa, or modern embeddings can improve the system's understanding of resume content and context, leading to better classification accuracy.

Resume ranking system

Instead of predicting only one role, the system can provide multiple role suggestions or rank candidates based on their skill match percentage.

HR analytics dashboard

Future versions may include charts, candidate comparisons, and automatic shortlist suggestions for HR users.

Integration with job portals

The system can be connected to recruitment websites or ATS (Applicant Tracking Systems) to allow large-scale automated screening.

Enhanced resume parsing

Better handling of tables, images, bullet points, and non-standard layouts will make extraction more accurate for modern resumes.

REFERENCES

[1]. Feldman, R., & Sanger, J. (2007). The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press.

[2]. Jurafsky, D., & Martin, J. H. (2021). Speech and Language Processing (3rd Edition, Draft). (Widely used reference for Natural Language Processing fundamentals.)

[3]. Brownlee, J. (2019). Machine Learning Algorithms: A Step-by-Step Guide with Python. Machine Learning Mastery.
(Referenced for understanding ML workflow and practical model development.)

[4]. Raschka, S. (2015). Python Machine Learning. Packt Publishing. (Used for ML model building and vectorization concepts.)

[5]. Kang, D., et al. (2020). "Automated Resume Screening Using Natural Language Processing." International Journal of Computer Applications, 176(34), 1–6.

[6]. Singh, A., & Kaur, G. (2022). "AI-Enabled Recruitment Systems and Their Impact on Hiring Efficiency." International Research Journal of Engineering and Technology, 9(5), 1450–1455.

[7]. Flask Documentation — Flask Web Framework. Available at: https://flask.palletsprojects.com/

[8]. Scikit-Learn Documentation — Machine Learning Library for Python. Available at: https://scikit-learn.org/

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

[9]. PyPDF2 & python-docx Documentation — Libraries for PDF/DOCX text extraction. Available at: https://pypi.org/

[10]. Chaudhari, O. N. (2025). AI-Based Resume Screening & Prediction System. Unpublished student research, Godavari Institute of Management & Research, Jalgaon.