



(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 6, June 2025

⊕ www.ijirset.com ⊠ ijirset@gmail.com 🖄 +91-9940572462 🕓 +91 63819 07438





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406173|

Resume Classification using Machine Learning

Abhijith K S, Anita Babu

MCA Student, Department of Computer Science and Engineering, SCMS School of Engineering and Technology,

Ernakulam, India

Assistant Professor, Department of Computer Science and Engineering, SCMS School of Engineering and Technology,

Ernakulam, India

ABSTRACT: The increasing volume of job applications necessitates an automated and efficient approach to resume screening. This project presents a Resume Classification Systemthat categorizes resumes based on predefined job roles using machine learning techniques. Thesystem processes resumes using TF-IDF vectorization for feature extraction and employsALogistic Regression to classify resumes into relevant categories. The dataset comprisesmultiple resume datasets, including UpdatedResumeDataSet.csv, clean_resume_data.csv, andResume.csv, which contain resumes labeled according to different job roles. The model istrained and evaluated using Scikit-learn, with performance analysis conducted throughvisualization techniques using Matplotlib and Seaborn. A Flask-based web application is developed to allow users to upload resumes for realtime classification. The project is executed in a JupyterHub environment within AnacondaNavigator, ensuring an interactive and flexible development workflow.This resumeclassification system enhances recruitment efficiency by automating resume screening, reducing manual efforts, and improving the hiring process for organizations.

KEYWORDS: Resume Classification, Machine Learning, TF-IDF, Logistic Regression

I. INTRODUCTION

With the current competitive job market, organizations are receiving a huge number of resumes for various job posts, and it is a time-consuming and inefficient process to filter themmanually. To overcome this, automated resume categorization uses machine learningprocedures to categorize resumes into their content-based categories. Using TF-IDFvectorization in text processing and Logistic Regression for categorization, the systemprocesses resumes easily and categorizes them into proper job categories. The resume-taggeddataset aids in pattern detection and enhancing accuracy by the model. This project set aFlask-based web application to support an easy user interface for uploading resumes andrealtime classification from a pretrained model. The automated resume screening process reduces recruitment inefficiency, manual workload, and allows organizations to find suitablecandidates sooner.

The primary objectives of this project focus on enhancing the efficiency and accuracy of resume screening through automation. The project is designed to create a machine learning model that can classify resumes into specified job categories, largely eliminating the need for manual work in the hiring process. It uses TF-IDF vectorization for feature extraction and Logistic Regression for classification to provide strong and efficient categorization. To further enhance usability, a web application based on Flask is built, which enables users to upload resumes for real-time classification. The system also integrates a pretrained machine learning model, enabling efficient and scalable deployment for consistent performance across various use cases.

The project scope includes creating an automated resume categorization systemthat appliesmachine learning algorithms to classify resumes into job roles. The systemreads resumedatathrough TF-IDF vectorization and classifies them through a Logistic Regressionfor efficient and accurate prediction. A Flask web application is employed to develop an easy-to-useinterface whereby users can upload resumes and receive real-time feedbackforclassification. This is a project made a scale, as it deals with large data sets. Organizationswith a huge recruitment need will find this very scalable. Matplotlib and Seaborn tools arealso utilized for the analysis of data trends and performance in the model. The systemcanbeenhanced by adding deep learning models, feature engineering, or multi-language support, thus being adaptable to all industries and needs of recruitment.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406173|

II. LITERATURE SURVEY

An automated resume screening system uses ML and NLP (TF-IDF, Random Forest) to parse resumes and enhance candidate filtering, improving hiring efficiency [1]. A structured study details resume screening methodology, data analysis, and discussion, aiding logical flow and practical application [2]. ResumeAtlas uses big data (13,389 resumes) and sophisticated models such as BERT, with 92% top-1 accuracy in classification tasks [3]. Semantic similarity models (all-roberta-large-v1, etc.) and deep networks (LSTM, CNN) are employed for resume clustering and ranking against job descriptions [4]. Gradient Boosting improves resume categorization by boosting precision and efficiency in talent sourcing, promoting transparent recruitment processes [5]. An NLP system employing Jaccard Similarity optimizes candidate-job matching and includes LinkedIn functionality for enhanced screening [6]. Naïve Bayes, Random Forest, and SVM models are compared for resume categorization, with Random Forest achieving 70% accuracy; future work suggests video and voice analysis [7]. A separate study explores particle diffusion on percolation clusters, analyzing unique diffusion behaviors and ergodicity, relevant in physics and materials science [8]. CareerBERT predicts career paths using unstructured resume text with ESCO labels, improving job recommendations and internal mobility policies [9]. Improved English resume corpus makes AI-powered screening more accurate through models such as BERT, RoBERTa, and T5 and demonstrates improved classification with increased datasets [10].

III. METHODOLOGY

The Resume Classification System automates the screening process by applying machine learning techniques to categorize resumes into predefined job roles. The system starts by collecting resumes from multiple datasets such as UpdatedResumeDataSet.csv, clean_resume_data.csv, and Resume.csv. The content from these CVs is preprocessed by stopword removal, tokenization, and TF-IDF vectorization for representing the unstructured text as useful numerical features. To enhance classification accuracy, the dataset is split into training (80%) and testing (20%) sets, ensuring proper model evaluation.

Once preprocessing is completed, a Logistic Regression model is trained on the transformed resume data. Logistic Regression is selected due to its effectiveness in binary and multi-class classification tasks. The model acquires the patterns and key features of different job postings. During training, the algorithm optimizes the decision boundary by minimizing classification errors. After training, the model's performance is measured using accuracy, precision, recall, and F1-score along with confusion matrix to compare correct and incorrect predictions. Additionally, other models like Random Forest, K-Nearest Neighbors (KNN), and Support Vector Classifier (SVC) are experimented with for comparison purposes.

The trained model is integrated into a Flask web application for real-time resume categorization. When a user provides a resume, the application retrieves the text, performs preprocessing, vectorizes it using the TF-IDF vectorizer, and feeds it to the pretrained model for prediction. The system then displays the predicted job category instantly on the web interface. Visualization tools such as Matplotlib and Seaborn are employed to present data trends, model accuracy comparisons, and classification insights. The architecture is scalable, allowing future enhancements like deep learning integration, multilingual support, and expanded job role categories to further improve the system's performance and flexibility.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|



Volume 14, Issue 6, June 2025

Fig. 1. Shows the system architecture of resume categorization using machine learning

IV. EXPERIMENTAL RESULTS

The Resume Classification System was evaluated using multiple machine learning algorithms, including Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), and Support Vector Classifier (SVC). After preprocessing the dataset using TF-IDF vectorization and SMOTE balancing, the dataset was split into 80% train and 20% test sets. The models were trained and tested, and the results showed that Logistic Regression, Random Forest, and SVC each achieved a high accuracy of 99.48%, while KNN achieved a slightly lower but still impressive accuracy of 98.44%. Logistic Regression was utilized for web application deployment because it is simple, efficient, and highly interpretable. Besides accuracy, other parameters such as precision, recall, and F1-score too reaffirmed the model's high performance for various categories of jobs. Comparison between the existing manual system and the suggested automated system also highlighted significant advances in time saving, scalability, consistency, cost reduction, and candidate experience. Data visualizations created using Matplotlib and Seaborn also supported the findings, providing insights into dataset distributions and classification performance, validating the effectiveness and reliability of the proposed system.







www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406173|

Reg	ister	
sername		
Enter your username	e	
assword		
Enter your password	1	
onfirm Password		
Confirm your passw	ord	
Reg	ister	
Already have an	n account? <mark>Logi</mark> n	n

(b)

Home	About	Contact	logout
			Resume Classification
			Upload your resume (PDF or text file): Choose File No file chosen
			Predicted Category: Python Developer
			Resume Text:
			AdHIJITH K S abhijith.ks.shajidg mall. com +91891552360 PROFILE

(c)

Fig. 2. Shows the result of resume categorization using machine learning (a)Login page, (b)Register page, (c)Job prediction page

Performance Comparison,

Models	Accuracy
K-Neighbors Classifier (KNN)	0.9844559585492227%
Logistic Regression	0.9948186528497409%
Random Forest	0.9948186528497409%
Support Vector Classifier (SVC)	0.9948186528497409%

V. CONCLUSION

The Resume Classification System successfully automates the process of resumescreening using machine learning techniques, reducing the time and effort requiredformanual evaluation. The system efficiently and accurately classifies resumes into predefinedjob roles by using Logistic Regression for classification and TF-IDF vectorization for featureextraction. The system's intuitive real-time interface resume classification in The system'suser-friendly interface for a Flask-based web application allows recruiters and job seekerstoaccess it in real time. Data visualization analyzing help the dataset distributions andmodel performance, and a Flask-based web application makes it available to recruiters and jobseekers using Matplotlib and Seaborn. This project increases the effectiveness of hiring. Thisproject enhances





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406173|

recruitment efficiency and, ultimately, the hiring process overall by enablingfaster and more accurate candidate selection.

REFERENCES

[1] Daryani, Chirag, et al. "An automated resume screening system using natural languageprocessing and similarity." ETHICS AND INFORMATION TECHNOLOGY [Internet]. VOLKSON PRESS (2020): 99-103.

[2] Padmaja, D. Lakshmi, et al. "Automated resume screening using natural languageprocessing." Journal of Emerging Technologies and Innovative Research (JETIR) 10.3(2023): 100-104.

[3] Heakl, Ahmed, et al. "ResuméAtlas: Revisiting Resume Classification with Large-ScaleDatasets and Large Language Models." Procedia Computer Science 244 (2024): 158-165.

[4] Roy, Pradeep Kumar, Sarabjeet Singh Chowdhary, and Rocky Bhatia. "AMachineLearning approach for automation of Resume Recommendation system." Procedia ComputerScience 167 (2020): 2318-2327.

[5] Venkatakrishnan, R., et al. "Revolutionizing Talent Acquisition: A Comparative StudyofLarge Language Models in Resume Classification." 2024 5th International ConferenceonInnovative Trends in Information Technology (ICITIIT). IEEE, 2024.

[6] Saatci, Mehtap, Rukiye Kaya, and Ramazan Ünlü. "Resume Screening withNatural Language Processing (NLP)." Alphanumeric Journal 12.2 (2024): 121-140.

[7] Pal, Riya, et al. "Resume classification using various machine learning algorithms." ITMweb of conferences. Vol. 44. EDP Sciences, 2022.

[8] Pacheco-Pozo, Adrian, and Igor M. Sokolov. "Universal fluctuations and ergodicityofgeneralized diffusivity on critical percolation clusters." Journal of Physics A: Mathematical and Theoretical 55.34 (2022): 345001. 22

[9] Decorte, Jens-Joris, et al. "Career path prediction using resume representation learningand skill-based matching." arXiv preprint arXiv:2310.15636 (2023).

[10] Gan, Chengguang, and Tatsunori Mori. "Construction of English resume corpus andtest with pre-trained language models." arXiv preprint arXiv:2208.03219 (2022).









INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com