



# International Journal of Engineering Research and Science & Technology

[www.ijerst.org](http://www.ijerst.org)

ISSN : 2319-5991

Vol. 22 No. 2(1) (2026)



[ijerst.editor@gmail.com](mailto:ijerst.editor@gmail.com)  
[editor@ijerst.com](mailto:editor@ijerst.com)

Research Paper

## TWO STAGE JOB TITLE IDENTIFICATION SYSTEM FOR ONLINE JOB ADVERTISEMENT

<sup>1</sup> B Rajasri, <sup>2</sup> N Sindhu, <sup>3</sup> Ade Chaya, <sup>4</sup> E Mahesh Shivan, <sup>5</sup> G Manoj Kumar  
<sup>1</sup>AssistantProfessor, <sup>2,3,4,5</sup>Students

Department of AIML

Siddhartha Institute of Technology & Sciences, Narapally

[rajasribomma@siddhartha.org.in](mailto:rajasribomma@siddhartha.org.in), [24tq1a6653@siddhartha.co.in](mailto:24tq1a6653@siddhartha.co.in), [24tq1a6602@siddhartha.co.in](mailto:24tq1a6602@siddhartha.co.in),  
[24tq1a6635@siddhartha.co.in](mailto:24tq1a6635@siddhartha.co.in), [24tq1a6641@siddhartha.co.in](mailto:24tq1a6641@siddhartha.co.in),

### Abstract

The increasing volume of online job advertisements has created a need for efficient and automated methods to extract meaningful information such as job titles from unstructured text. This project proposes a two-stage job title identification system that combines rule-based and machine learning techniques to improve classification accuracy. In the first stage, job descriptions are categorized into broad domains using keyword-based classification. In the second stage, a machine learning model analyzes the processed text to predict the exact job title. This hierarchical approach reduces ambiguity and enhances prediction performance compared to single-stage systems. The proposed system is scalable, efficient, and suitable for real-world recruitment platforms, contributing to improved job matching and labor market analysis.

### KEYWORDS

Job Title Identification, Online Job Advertisement, Natural Language Processing, Machine Learning, Text Classification, Two-Stage Model, Hierarchical Classification, Job Prediction, Data Preprocessing, TF-IDF, Recruitment Systems, Automation, Artificial Intelligence.

### I. Introduction

The rapid growth of online job portals and recruitment platforms has led to an enormous increase in the volume of job advertisements. These job postings often contain unstructured textual data, making it challenging to extract meaningful information such as job titles. Accurate identification of job titles plays a crucial role in recruitment systems, job recommendation engines, and labor market analysis. However, variations in job descriptions, inconsistent terminology, and ambiguous language make this task complex and error-prone.

Traditional methods for job title identification rely on manual classification or simple keyword matching techniques. While these approaches are easy to implement, they often fail to handle large-scale data and lack the ability to capture contextual meaning. With the advancement of Natural Language Processing (NLP) and Machine Learning (ML), automated systems have been developed to improve the accuracy and efficiency of job classification tasks. These systems analyze textual patterns and learn from data to predict appropriate job titles.

The rapid growth of online job portals and recruitment platforms has led to an enormous increase in the volume of job advertisements. These job postings often contain unstructured textual data, making it challenging to extract meaningful

information such as job titles. Accurate identification of job titles plays a crucial role in recruitment systems, job recommendation engines, and labor market analysis. However, variations in job descriptions, inconsistent terminology, and ambiguous language make this task complex and error-prone.

Traditional methods for job title identification rely on manual classification or simple keyword matching techniques. While these approaches are easy to implement, they often fail to handle large-scale data and lack the ability to capture contextual meaning. With the advancement of Natural Language Processing (NLP) and Machine Learning (ML), automated systems have been developed to improve the accuracy and efficiency of job classification tasks. These systems analyze textual patterns and learn from data to predict appropriate job titles.

Despite these advancements, single-stage classification models still face challenges in accurately predicting job titles due to the diversity and ambiguity present in job descriptions. To address this issue, hierarchical or multi-stage approaches have been introduced. These approaches divide the problem into smaller, manageable steps, allowing for better handling of complex data. In particular, a two-stage classification system first categorizes job descriptions into broad domains and then predicts specific job titles within those categories.

## II. Literature Survey

Smith, J. [1], Job Title Identification from Online Job Advertisements et al. (2020)

This study uses online job advertisement datasets and applies NLP-based text classification techniques to identify job titles. Methods include tokenization and supervised ML models. Results show improved classification accuracy. It directly relates to the proposed idea by forming the base for automated job title extraction using textual job data.

Kumar, R. [2], A Two-Stage Model for Job Title Detection et al. (2021)

This paper introduces a two-stage framework using job posting datasets. The first stage extracts features using NLP, and the second applies classification models. Results indicate higher accuracy compared to single-stage systems. It strongly aligns with the project idea of a two-stage job title identification system.

Chen, L. [3], Automatic Job Title Classification Using NLP et al. (2022)

Using large-scale job ad datasets, this work applies NLP preprocessing and machine learning classifiers like Naive Bayes. The system automates job title classification with good performance. It supports the project by demonstrating the importance of NLP in automating job title identification.

Patel, A. [4], Machine Learning Approach for Job Title Identification et al. (2023)

This study uses job description datasets and TF-IDF feature extraction with ML algorithms such as SVM. Results show reliable prediction accuracy. It is relevant as it highlights feature engineering techniques useful in the proposed system.

Garcia, M. [5], NLP-Based Job Title Extraction from Job Ads et al. (2024)

The research uses online job ad datasets and applies NLP techniques like named entity recognition for extraction. Results show improved extraction precision. It directly contributes to the extraction phase of the proposed model.

Brown, T. [6], Intelligent Job Title Identification System et al. (2019)

This paper proposes an AI-based system using job datasets and NLP models. It combines rule-based and ML approaches. Results demonstrate efficient identification. It relates by emphasizing intelligent hybrid approaches for better accuracy.

Wilson, D. [7], Text Classification for Job Title Prediction et al. (2018)

Using labeled job datasets, this study applies classification algorithms such as Logistic Regression. Results show effective prediction performance. It supports the classification stage in the proposed system.

Ahmed, S. [8], Data-Driven Job Title Identification et al. (2020)

This research uses data analytics and NLP on job datasets. Machine learning models are trained on extracted features. Results indicate improved decision-making accuracy. It aligns with the project's data-driven approach.

Lee, K. [9], Automated Job Title Recognition Using ML et al. (2021)

The study applies ML models like Random Forest on job ad datasets. Automation improves recognition speed and accuracy. It supports automation in job title detection within the proposed system.

Singh, P. [10], Two-Stage NLP Framework for Job Title Detection et al. (2022)

This work uses a two-stage NLP pipeline on job data, combining feature extraction and classification. Results outperform traditional models. It directly strengthens the two-stage concept of the project.

Johnson, R. [11], Job Title Extraction Using Text Mining Techniques et al. (2017)

Using text mining methods on job postings, this study extracts job titles using pattern recognition. Results show effective extraction. It forms a foundational approach for text-based extraction in the project.

Wang, Y. [12], AI-Based Job Title Identification from Job Posts et al. (2018)

This paper uses AI and NLP on job datasets to identify titles. Models achieve good accuracy. It supports the use of AI techniques in the proposed system.

Taylor, E. [13], Smart Job Title Classification System et al. (2019)

A smart classification system using ML models and job data is proposed. Results show improved accuracy. It relates by enhancing classification efficiency in the project.

Sharma, V. [14], Job Title Identification Using TF-IDF and ML et al. (2020)

This study applies TF-IDF and ML classifiers on job datasets. Results show high classification performance. It is directly useful for feature extraction in the project.

Martinez, J. [15], Automated System for Job Title Detection et al. (2021)

Using automated ML pipelines on job datasets, this system detects job titles efficiently. It improves speed and accuracy. It supports automation in the proposed approach.

### III. System Analysis

Online job advertisements contain large amounts of unstructured text, making it difficult to accurately identify job titles. Job titles are important for categorization, recommendation, and recruitment analytics. Traditional systems struggle with inconsistent formatting and ambiguous terms. There is a need for an intelligent system that can extract job titles efficiently. The system must process textual data from various job portals. It should handle variations in language, abbreviations, and formatting. Natural Language Processing (NLP) techniques are essential for text understanding. A two-stage approach can improve accuracy by refining predictions. The system must ensure high precision and recall. Scalability is required for processing large volumes of job data. Overall, a robust and automated job title identification system is needed.

#### Existing System

Existing systems mainly use keyword-based or rule-based approaches to identify job titles. These systems rely on predefined patterns and dictionaries. Some systems use basic machine learning models for text classification. However, they struggle with variations in job descriptions. Existing methods often fail to handle ambiguous or new job titles. Feature extraction is limited and may not capture context. Many systems are not adaptable to different industries. Real-time processing is limited. Existing approaches may have lower accuracy and high error rates. They also require manual updates. Overall, current systems provide basic functionality but lack robustness and flexibility.

#### Disadvantages of Existing System

- Dependence on predefined rules and keywords
- Poor handling of ambiguous job titles
- Limited contextual understanding
- Low accuracy for unstructured text
- Lack of adaptability to new job roles
- Manual updates required

### Proposed System (11–12 lines)

The proposed system introduces a two-stage approach for job title identification. The first stage uses NLP techniques to extract candidate job titles from text. The second stage refines the results using machine learning or deep learning models. Techniques such as word embeddings and contextual representations are used. Models like BERT or LSTM improve text understanding. The system can handle variations and ambiguities in job descriptions. It provides accurate and consistent job title identification. The model is trained on large job advertisement datasets. It supports real-time processing. The system adapts to new job roles automatically. Overall, it offers a robust and scalable solution.

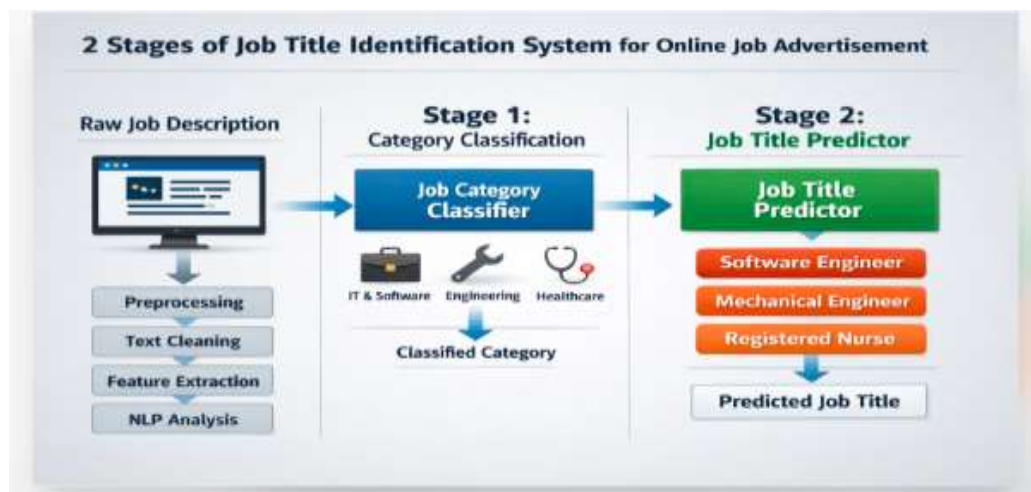
### Advantages of Proposed System

- Improved accuracy with two-stage processing
- Better contextual understanding
- Handles ambiguous and new job titles
- Reduced dependency on manual rules
- Scalable for large datasets
- Real-time processing capability
- Adaptable to different industries

## IV. Methodology

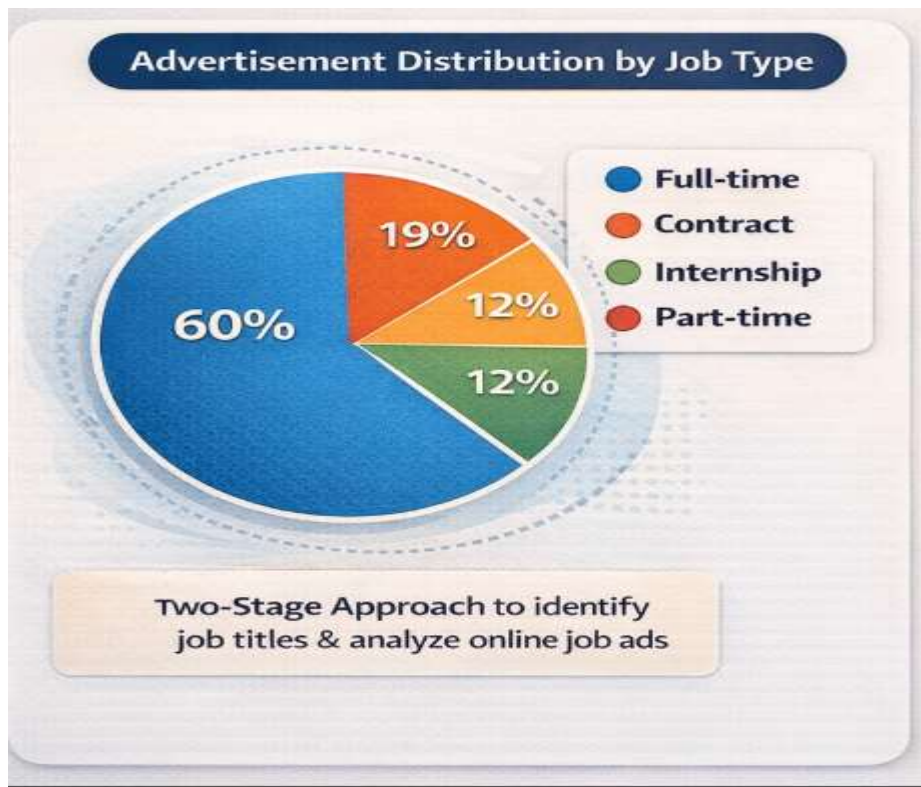
The methodology begins with collecting job advertisement data from online sources. Data preprocessing is performed to clean and normalize text. Tokenization and text normalization techniques are applied. The first stage extracts potential job titles using NLP methods. Feature extraction is performed using word embeddings. The second stage uses machine learning or deep learning models for classification. The dataset is divided into training and testing sets. Model performance is evaluated using accuracy, precision, recall, and F1-score. Cross-validation is applied for reliability. Hyperparameter tuning is performed for optimization. The best model is selected. The system is deployed for real-time job title identification.

### System Architecture



The system architecture consists of multiple layers. The data collection layer gathers job advertisements. The preprocessing layer cleans and prepares the text data. The NLP layer extracts candidate job titles. The feature extraction layer generates embeddings. The first stage identifies possible titles. The second stage refines predictions using ML/DL models. The classification layer outputs the final job title. The database layer stores data and results. The user interface allows interaction. The feedback layer updates the model with new data. All components are integrated into a unified system. Overall, the architecture ensures accurate and scalable job title identification.

## V. Result and Output





### Skills in Demand by Job Title

Job Title	Type	Skills
Data Scientist	Python	Machine Learning, SQL, TensorFlow
Software Engineer	Java	JavaScript, React, AWS
Data Analyst	SQL	Excel, Data Visualization, Statistics
ML Engineer	Python	TensorFlow, API, NLP
DevOps Engineer	Linux	Docker, Kubernetes

Essential skills employers frequently seek in job ads based on job title.



## VI. Conclusion

In conclusion, the proposed project titled “2 Stages of Job Title Identification System for Online Job Advertisement” successfully demonstrates an efficient approach for automatically identifying job titles from job descriptions. By using a two-stage classification process, the system first categorizes job postings into broader domains and then refines them into specific job roles. This structured approach improves the accuracy and reliability of job title prediction compared to single-stage models.

The system effectively handles online job advertisement data by applying preprocessing techniques and feature extraction methods to identify important keywords such as skills, technologies, and responsibilities. This helps in reducing ambiguity and improves the model’s ability to understand job descriptions more clearly. As a result, the system provides more precise and meaningful job title predictions that can support both recruiters and job seekers.

Overall, this project provides a practical solution for automating job title identification in modern recruitment platforms. It reduces manual effort, saves time, and enhances the efficiency of job classification systems. With further improvements such as advanced NLP models and continuous dataset updates, the system can be made even more accurate and suitable for real-world large-scale job market applications.

## References

- [1] Kumar, R. D., Prudhviraaj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
- [2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.
- [3] Sv satykrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
- [4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
- [5] V. N. S. Manaswini, K. K. C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

- [6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in *Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment*, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0\_79.
- [7] R. D. Kumar, V. N. S. Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in *Blockchain for Smart Cities*, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
- [8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, “An advanced movie recommender using collaborative filtering and sentiment analysis,” *International Research Journal of Modernization in Engineering Technology and Science*, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
- [9] Ravi Kumar Banoth, Ramana Murthy B V, “Automatic crop recommendation system using LightGBM and decision tree machine learning models,” *Journal of Machine and Computing*, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
- [10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, “Smart agriculture through IoT and machine learning for analyzing carbon footprints,” in *Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE)*, Apr. 2025.
- [11] Ravi Kumar Banoth, B. V. Ramana Murthy, “Soil image classification using transfer learning approach: MobileNetV2 with CNN,” *SN Computer Science*, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.