

## Efficient Resume Screening Tool

Manjula Patil<sup>1</sup>, Md Ruhaan<sup>2</sup>, Syed Mohammed Taqhi<sup>3</sup>, K Naveed Ahmed<sup>4</sup>, Shailender Babbar<sup>5</sup>

*Department of Information Science and Engineering, Rao Bahadur Y Mahabaleshwarappa Engineering College*

\*\*\*

**Abstract** - In the evolving landscape of recruitment, traditional manual resume screening methods are increasingly inadequate due to the sheer volume and diversity of applications. This paper introduces an automated resume screening tool that leverages Machine Learning (ML) and Natural Language Processing (NLP) to enhance the efficiency and accuracy of candidate evaluation. The system employs NLP techniques such as named entity recognition and part-of-speech tagging to extract pertinent information from resumes. Subsequently, ML classifiers, including K-Nearest Neighbors and Support Vector Machines, are utilized to assess and rank candidates based on their alignment with job descriptions. By automating the extraction and interpretation of resume data, the proposed tool significantly reduces the time and effort required in the initial stages of recruitment, while also mitigating human biases inherent in manual screening processes. Experimental results demonstrate the tool's effectiveness in accurately identifying suitable candidates, thereby streamlining the hiring process and contributing to more objective and data-driven recruitment decisions.

### 1. INTRODUCTION

In today's dynamic job market, organizations are inundated with a vast number of applications for each open position. Traditional manual resume screening methods are increasingly inadequate, as they are time-consuming and prone to human biases. Recruiters often face challenges in efficiently identifying suitable candidates from large applicant pools, leading to potential oversight of qualified individuals.

To address these challenges, the integration of Machine Learning (ML) and Natural Language Processing (NLP) techniques has emerged as a promising solution for automating the resume screening process. ML algorithms can analyze vast amounts of data to identify patterns and make predictions, while NLP techniques enable the extraction and interpretation of relevant information from unstructured resume texts. By leveraging these technologies, automated systems can evaluate and rank candidates based on their alignment with job requirements, thereby enhancing the efficiency and objectivity of the recruitment process.

This paper explores the development and implementation of an automated resume screening tool that utilizes ML and NLP methodologies. The proposed system aims to streamline the initial stages of recruitment by accurately assessing candidate qualifications and reducing the reliance on manual screening. Through this approach, organizations can achieve a more efficient, fair, and data-driven hiring process.

### 2. Body of Paper

The recruitment landscape has undergone significant transformation with the advent of automated resume screening tools, primarily driven by advancements in Natural Language Processing (NLP) and Machine Learning (ML). These technologies aim to enhance the efficiency, accuracy, and fairness of the hiring process by automating the initial screening of candidate resumes.

#### Early Developments in Automated Screening

Initial efforts in automating resume screening focused on keyword-based matching systems. These systems relied on predefined rules to identify relevant information within resumes. However, they often lacked the contextual understanding necessary to accurately assess candidate suitability, leading to potential misclassifications.

#### Integration of NLP and ML Techniques

To address the limitations of rule-based systems, researchers began integrating NLP and ML techniques into resume screening tools. For instance, Saatçi et al. (2024) developed an NLP-based system that significantly reduced the time and effort required for manual evaluations, providing a more effective and unbiased candidate assessment process. Similarly, Nandhini et al. (2021) proposed an automated resume screening system that extracts data from resumes using NLP techniques and ranks them based on their match with job descriptions.

#### Advancements with Large Language Models

Recent advancements have seen the incorporation of Large Language Models (LLMs) into resume screening processes. Lo et al. (2025) introduced a context-aware and explainable multi-agent framework utilizing LLMs for resume screening. This framework consists of agents responsible for extracting, evaluating, summarizing, and formatting resume data, enhancing the contextual relevance of candidate assessments.

#### Semantic Analysis and Topic Modeling

Beyond keyword extraction, some studies have explored semantic analysis and topic modeling to improve resume screening. Jagwani et al. (2023) proposed a method using Latent Dirichlet Allocation

(LDA) and entity detection to evaluate resumes. Their approach aimed to provide a more content-driven assessment by analyzing the semantic structure of resumes, achieving notable accuracy in candidate selection.

## Challenges and Ethical Considerations

Despite technological advancements, challenges persist in automated resume screening. Bias in AI algorithms remains a significant concern, as highlighted by studies indicating that automated systems can inadvertently perpetuate existing biases in hiring practices. Ensuring transparency, fairness, and accountability in these systems is crucial for ethical recruitment.

## Research Gap

While significant progress has been made in sentiment analysis on platforms like Twitter, relatively fewer studies focus on YouTube comment sentiment analysis in a real-time web-based environment. Most existing solutions are either limited in their UI integration or require extensive computational resources. Furthermore, they rarely offer interactive visualizations or are difficult to use for non-technical stakeholders.

## Contribution of This Study

The *Efficient Resume Screening Tool* bridges this gap by:

- Offering a lightweight, real-time sentiment analysis web app
- Leveraging accessible machine learning models for effective classification
- Visualizing insights through graphs and word clouds
- Focusing on ease-of-use, accuracy, and extensibility.

## 3. Methodology

The development of *Efficient Resume Screening Tool* follows a systematic approach involving data collection, preprocessing, sentiment classification, and result visualization. The primary objective is to analyze the sentiments expressed in YouTube comments using machine learning and natural language processing (NLP) techniques and provide an interactive web-based interface for end users.

### 3.1 System Architecture

The system is designed as a full-stack application consisting of:

- **Frontend:** HTML, CSS, and JavaScript
- **Backend:** Python (Flask framework)
- **Libraries:** scikit-learn, TextBlob, NLTK, Matplotlib, WordCloud
- **Deployment:** Localhost or web server

The architecture follows a pipeline structure:

1. Input YouTube video URL
2. Extract comments via the YouTube Data API
3. Preprocess text data
4. Perform sentiment classification
5. Visualize and display results

### 3.2 Data Collection

- **Tool Used:** YouTube Data API v3
- **Method:** A video ID is extracted from the provided URL and used to fetch a maximum of 100-500 comments (configurable) per video.

- **Data Format:** JSON, containing comment text, author, likes, and timestamp (only comment text is used for sentiment analysis).

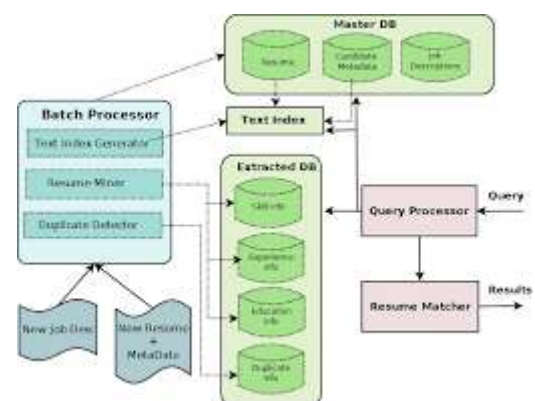
### 3.3 Data Preprocessing

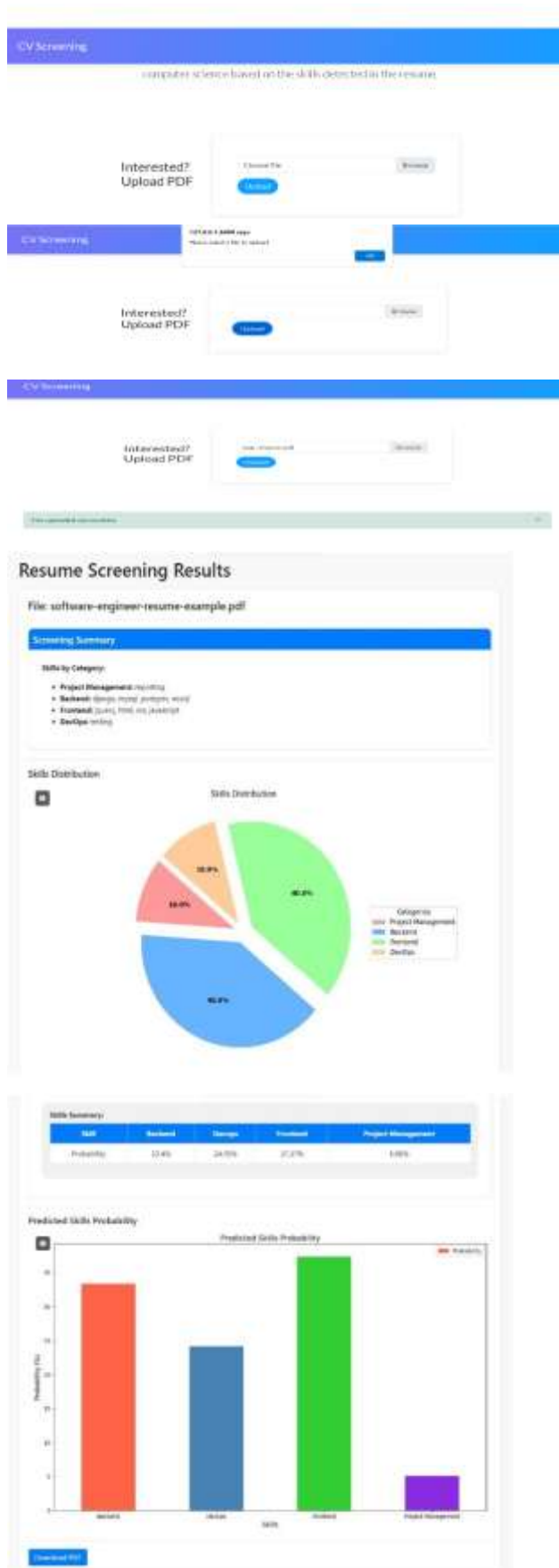
Text preprocessing is critical for effective sentiment classification. The following steps are performed using Python's nltk and re modules:

- **Lowercasing:** Converts all text to lowercase for uniformity.
- **Punctuation Removal:** Eliminates punctuation, emojis, and special characters.
- **Stopword Removal:** Removes commonly used words (e.g., "is", "the", "and") using nltk.corpus.stopwords.
- **Tokenization:** Breaks text into individual words or tokens.
- **Lemmatization:** Converts words to their base or root form.

### 3.4 Sentiment Classification

To enhance user interaction and understanding, results are visualized using: Automated resume screening systems utilize classification techniques to categorize resumes based on their relevance to specific job roles. By employing various machine learning (ML) algorithms and natural language processing (NLP) methods, these systems aim to streamline the recruitment process by efficiently identifying suitable candidates.





## CONCLUSION

The integration of Artificial Intelligence (AI) and Natural Language Processing (NLP) into resume screening has revolutionized the recruitment landscape, offering enhanced

efficiency and objectivity in candidate evaluation. Automated systems employing machine learning algorithms, such as Support Vector Machines, Random Forests, and Large Language Models (LLMs), have demonstrated significant improvements in processing speed and accuracy, with some frameworks achieving up to 11 times faster screening compared to traditional methods. These advancements have not only streamlined the hiring process but also hold the potential to mitigate human biases inherent in manual screening. For instance, studies have shown that AI-driven tools can reduce gender disparities in employment opportunities. However, the deployment of such technologies is not without challenges. Concerns regarding algorithmic transparency, data privacy, and the perpetuation of existing biases have been raised, necessitating ongoing scrutiny and ethical considerations.

## ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our Head of the Department, **Dr. B Sreepathi** and guide, **Prof. Manjula Patil**, for their continuous support, valuable guidance, and encouragement throughout the development of this research project. Their insights and constructive feedback played a vital role in enhancing the quality of this work.

We are also thankful to the Department of Information Science and Engineering, Rao Bahadur Y Mahabaleswarappa Engineering College, for providing the necessary resources and a collaborative environment that helped us explore and apply advanced concepts in Natural Language Processing and machine learning.

A special thanks to our peers and friends who participated in the testing phase of the *Efficient Resume Screening Tool* application and provided useful feedback to improve its usability. Finally, we thank our families for their unwavering support and motivation throughout this journey.

## REFERENCES

1. Saatçı, M., Kaya, R., & Ünlü, R. (2024). *Resume Screening with Natural Language Processing (NLP)*. *Alphanumeric Journal*, 12(2), 121–140.
2. □ Tayal, S., et al. (2024). *Resume Screening using Machine Learning*. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 10(2), 602–606.
3. □ Jagwani, V., Meghani, S., Pai, K., & Dhage, S. (2023). *Resume Evaluation through Latent Dirichlet Allocation and Natural Language Processing for Effective Candidate Selection*. *arXiv preprint arXiv:2307.15752*.
4. □ Li, C., Fisher, E., Thomas, R., Pittard, S., Hertzberg, V., & Choi, J. D. (2020). *Competence-Level Prediction and Resume & Job Description Matching Using Context-Aware Transformer Models*. *arXiv preprint arXiv:2011.02998*.

## BIOGRAPHIES



**Mrs. MANJULA PATIL**

**Assistant Professor, Dept of ISE, RYMEC, Ballari.**

**MD RUHAAN,**

**Final Year student of Information  
Science and Engineering (ISE)  
Department, RYMEC, Ballari.**

**K NAVEED AHMED,**

**Final Year student of Information  
Science and Engineering (ISE)  
Department, RYMEC, Ballari.**

**SYED MOHAMMED TAQHI,**

**Final Year student of Information  
Science and Engineering (ISE)  
Department, RYMEC, Ballari.**

**SHAILENDER BABBAR,**

**Final Year student of Information  
Science and Engineering (ISE)  
Department, RYMEC, Ballari.**