Resume Checker: A Dynamic NLP-based Evaluation System

Mahika Shah I038 | Nidhi Lavand I041 | Preeyaj Safary I043

Abstract

This paper presents a dynamic system for automated resume evaluation using Natural Language Processing (NLP) techniques. By analyzing resumes in various formats such as `.docx`, `.pdf`, and `.txt`, the system scores resumes based on accuracy and relevance to specific job descriptions. The system incorporates both front-end and back-end processing to assess skills, qualifications, and overall resume quality, providing actionable insights into areas for improvement. Through a combination of scoring algorithms and NLP-driven analysis, this system contributes to the field of recruitment by offering a robust and objective method for measuring resume effectiveness and job-fit potential.

Index Terms—Resume Evaluation, NLP, Natural Language Processing, Scoring System, Automated Assessment, Resume Parsing, Job Matching.

1 Introduction

The rising demand for automated solutions in recruitment has driven the need for reliable systems capable of efficiently evaluating resumes. NLP-based resume evaluation systems offer a solution by analyzing resumes for job-relevant content and scoring them based on various criteria, thus streamlining the recruitment process. Traditional resume evaluation methods are often subjective and time-consuming, creating a gap that automated systems aim to fill. This study presents an NLP-powered system that processes resumes in multiple formats, providing an accuracy score that aligns with the relevance and quality of resume content. By integrating both front-end and back-end technologies, our system offers a user-friendly and powerful tool for recruiters.

The goal of this study is to develop a model that evaluates resumes based on alignment with job requirements, assessing factors such as skill match, experience relevance, and content clarity. This work aims to provide valuable insights into resume quality, allowing both candidates and recruiters to understand key areas for improvement.

2 Related Work

Automated resume evaluation has been an area of interest in recent years due to its potential to enhance efficiency and objectivity in recruitment. Several studies have explored the use of NLP and machine learning techniques to assess resume quality. For instance, [?] developed a system that parses resumes and evaluates the alignment of skills with job descriptions, highlighting the importance of job-specific keyword matching. Similarly, [?] proposed a model that applies machine learning algorithms to predict job-fit scores based on resume content, with a focus on enhancing candidate-job matching.

Existing systems often face challenges in handling diverse resume formats and maintaining accuracy in skill extraction and job alignment. Our approach aims to address these issues by utilizing advanced NLP techniques, allowing for a more accurate and adaptable resume evaluation framework. Additionally, our system provides a user-friendly interface, facilitating seamless interaction for both candidates and recruiters.

3 Dataset

To evaluate resume quality for specific job roles, we compiled a dataset consisting of various documents in different formats, including `.docx`, `.pdf`, and `.txt`.

The dataset consists of a collection of diverse resumes obtained from both synthetic sources and sample resume databases. Each resume reflects different levels of experience and various professional fields, allowing for comprehensive evaluation across multiple job sectors. These resumes were categorized by industry and analyzed based on job relevance, highlighting specific attributes such as skills, experience, and educational background. The dataset was carefully curated to ensure a wide representation of job types, including technology, business, and healthcare, providing a robust foundation for testing and validating our resume scoring model.

The dataset includes:

- **Document Types**: Includes Word documents (`.docx`), PDF files (`.pdf`), and plain text files (`.txt`).
- **Number of Resumes**: A total of 5 best resumes are selected with relevant score value and accuracy, spanning multiple job roles and industries.
- **Key Attributes**: Each resume was annotated with attributes like skills, work experience, education, and certifications.
- **Job Descriptions**: A set of 20 job descriptions was used to match and score resumes for job relevance.

This dataset was essential for training and testing our NLP-based resume evaluation system, as it allowed us to measure accuracy and relevance scores effectively for each job role.

4 Methodology & Implementation

This section describes the methodology and implementation steps used in developing the resume evaluation system. The methodology is divided into several key stages.

4.1 Data Preprocessing

To prepare the resumes for analysis, we performed a series of preprocessing steps on the text data to ensure consistency and improve the effectiveness of our NLP-based evaluation system. The following steps were applied:

- **Text Extraction**: Text was extracted from documents in different formats, including `.docx`, `.pdf`, and `.txt`, using appropriate libraries for each file type.
- **Tokenization**: The extracted text was split into individual words or tokens. This step enables detailed analysis of each word in the resume for attributes such as skills and experience.
- **Lowercasing**: All text was converted to lowercase to maintain uniformity and avoid case-sensitive mismatches during analysis.
- **Removal of Stop Words**: Common words that do not add significant meaning, such as "and," "the," and "is," were removed using stop word lists. This reduces noise and helps focus on more informative words.
- **Removal of Punctuation**: Punctuation marks and other special characters were stripped from the text to prevent them from interfering with keyword matching and other analytical steps.
- **Lemmatization**: Words were reduced to their base or root forms (e.g., "running" to "run"), which helps in matching terms that may appear in different grammatical forms within the resumes.
- **Extracting Name and Experience using Regular Expressions (Regex)**: Regular expressions were used to identify and extract key information, such as the candidate's name and years of experience. This step is crucial for structured data extraction and helps categorize resumes based on experience level.

4.2 Feature Extraction

Purpose: Identify important features from resumes that relate to job relevance.
Techniques:

- **Text Parsing:** Extract key sections like "Education," "Experience," "Skills," etc.

- **Keyword Matching:** Extract keywords based on the job description, using lists of terms under categories like "skills," "certifications," or specific domain knowledge.
- **NER (Named Entity Recognition):** Use NLP models to tag entities like organization names, job titles, and qualifications.
- **Vectorization:** Convert resume text into embeddings using NLP techniques (e.g., TF-IDF, Word2Vec, or pre-trained models like BERT) to understand semantic relevance.

Libraries/Tools: spaCy for NER, Scikit-Learn for vectorization, and possibly Hugging Face for pre-trained embeddings.

4.3 Scoring Mechanism

Purpose: Rank resumes based on their relevance to a specific job description.
Rule-based Scoring:

- **Keyword Matching Score:** Count the number of relevant keywords from the job description found in the resume.
- **Experience Level:** Score based on years of experience and relevance to the required domain.
- **Skill Match Score:** Score based on the match rate of required skills against extracted resume skills.

4.4 Experimentation

- **Rule-based Model:** Test different keyword lists and weight combinations for sections (e.g., more weight to technical skills over soft skills).
- **ML/DL Model (if used):** Experiment with different vectorization techniques and models (e.g., BERT embeddings + logistic regression) for relevance prediction.
- **Hyperparameter Tuning:** For both rule-based and ML models, tune parameters like keyword importance weights, experience multipliers, and model-specific parameters.
- **ML/DL Model Scoring (if applicable):**
  - Use supervised ML models like SVM or deep learning models if you have a labeled dataset with resumes and job descriptions with relevance labels. Train the model to classify the relevance based on resume content.

4.5 Results and Evaluation

Output Presentation:

- For each resume, show the calculated scores (e.g., Skill Score, Experience Score, Overall Relevance Score).
- Rank resumes by their scores and display top-scoring resumes as the most relevant.

Evaluation Metrics:

- Precision, Recall, and F1-score: Evaluate the accuracy of resume relevance classification if using ML/DL models.
- Ranking Correlation: Compare the ranking output against a human-ranked list for validation.

4.6 Documentation in Notebook

Analysis: After scoring, provide an analysis of why certain resumes scored high and potential areas of improvement.

Each of these preprocessing steps ensures that the text data is standardized and ready for subsequent stages of the resume evaluation process, facilitating accurate scoring and job-fit assessment.

5 Future Scope of Resume Screeners

As the job market continues to evolve, the role of AI and NLP in recruitment is becoming more critical. The current resume screener leverages simple keyword matching and experience-based scoring. However, future enhancements can make it more sophisticated by incorporating the following:

1. **Advanced Natural Language Understanding (NLU):** Moving beyond keyword-based matching, deep learning models like BERT and GPT can be used to understand the context of the resume, allowing the system to assess the quality of experience, skills, and education more accurately.

2. **Machine Learning Models for Recommendation:** By using machine learning models, the system can recommend career paths, job roles, or further education options based on the candidate's profile.

3. **Diversity and Bias Reduction:** Implementing fairness and bias detection mechanisms in the scoring model to ensure that the system does not favor any specific demographic over others.

4. **Real-Time Resume Evaluation:** Enabling real-time resume evaluation during the job application process to give immediate feedback to applicants, enhancing the user experience.

6 Conclusion

This project demonstrates a practical application of Natural Language Processing (NLP) in automating the resume screening process. By analyzing resumes for key elements such as experience and skill keywords, the system can generate an objective score, aiding HR professionals in making quicker, data-driven decisions. While the current approach focuses on basic text processing and keyword matching, the future holds many opportunities to incorporate more complex models that can improve accuracy and fairness in the recruitment process.

The resume screener is an important step in reducing the manual effort involved in recruitment while enhancing the quality of candidate selection. By adopting more advanced techniques in NLP and machine learning, the screening process can be further optimized to support diverse hiring needs and ensure a better fit for job roles.

7 Takeaways

1. **Practical Application of NLP:** This project showcases how NLP techniques like tokenization, stopword removal, and lemmatization can be applied to real-world problems, such as resume screening.

2. **Importance of Data Quality:** The accuracy of the resume screener is highly dependent on the quality and quantity of data used to train and fine-tune the model. Proper preprocessing of resumes plays a critical role in ensuring better results.

3. **Scalability and Flexibility:** The system is designed to be scalable and flexible, allowing easy integration of new keywords and improvement of the algorithm to handle more complex resume formats and unstructured data.

4. **Ethical Considerations:** As automation in hiring grows, it is important to ensure that the models are transparent, fair, and free from bias. Monitoring and continuously improving the system will help maintain ethical standards in recruitment.

8 References

1. HaasS. Tools for Natural Language Processing. 2011. *http://ils.unc.edu/wstephani/nlpsp08/resources.html#tools* (accessed 1 Jun 2011).

2. Bhowmik, R., Garg, N., & Gupta, A. (2021). "Resume Screening Using Semantic Similarity and Clustering Algorithms". In Proceedings of the 2021 3rd International Conference on Communication, Devices and Computing.

3. Ma, Z., Wang, Y., & Zhao, Y. (2021). "Automated Resume Screening with Semantic Similarity and Gradient Boosting". In Proceedings of the 2021 3rd International Conference on Cybernetics, Robotics and Control.

4. Resume Screening with NLP: *https://www.jetir.org/papers/JETIR2303510.pdf*