

A micro project report submitted in partial fulfillment for

T.Y B.Tech in Artificial Intelligence Machine Language



On

# Resume Screening

By

Name: Shreya Patil PRN:21ST114501057

Under Supervision of

**Prof. Amrish Patil**

For academic Year: 2023-2024

**1.Introduction:**

In today's fast-paced and competitive job market, the process of screening resumes can be a time-consuming and labor-intensive task for recruiters and hiring managers. Traditional resume screening often involves manually sorting through numerous applications to identify candidates who meet specific qualifications and experience criteria. This method can be subjective, prone to human bias, and limited in its ability to process large volumes of resumes efficiently.

The micro project on resume screening seeks to address these challenges by developing an automated system that leverages natural language processing (NLP) and machine learning techniques to categorize resumes according to predefined job categories. The project aims to enhance the efficiency and accuracy of the hiring process by quickly filtering out resumes that do not meet the required job criteria and identifying the most suitable candidates for further evaluation.

**Objectives:**

-Automate Resume Screening: The primary goal is to automate the process of screening resumes, which involves categorizing them into different job roles or categories based on their content.

- Improve Efficiency: By automating the resume screening process, the project aims to reduce the time and effort required by recruiters and hiring managers, enabling them to focus on higher-level decision-making tasks.

- Increase Accuracy: The project seeks to increase the accuracy of resume screening by using machine learning models trained on labeled data to classify resumes into appropriate job categories.

- Reduce Bias: By implementing a data-driven approach, the project aims to minimize human bias in the resume screening process and provide more objective and consistent evaluations of candidates.

**Significance:**

- Streamlined Hiring Process: An automated resume screening system can significantly streamline the hiring process by quickly filtering out unsuitable candidates and prioritizing those who meet the job criteria.

- Improved Candidate Matching: By classifying resumes into specific job categories, the system can facilitate more accurate matching of candidates to job openings, increasing the chances of successful placements.

- Scalable Solution: An automated system can handle large volumes of resumes efficiently, making it a scalable solution for organizations of all sizes.

**2. Methodology:**

-Methods and Techniques: The project employs NLP techniques to clean and preprocess resumes and then uses a TfidfVectorizer to extract features from the text. A OneVsRestClassifier with KNeighborsClassifier is trained to categorize resumes.

-Tools and Resources: The project uses libraries such as pandas for data manipulation, scikit-learn for machine learning models, and matplotlib and seaborn for visualizations.

Procedures:

- Data cleaning involves removing URLs, special characters, and extra spaces.

- The text data is vectorized using TfidfVectorizer with English stop words removed.

- Data is split into training and testing sets, and the model is trained on the training data.

- The model's performance is evaluated on the test set using accuracy as the metric.

-Application Interface: A graphical user interface (GUI) was developed using Tkinter to allow users to upload resume files and analyze them.

**3. Results:**

Visualizations:

- Category Distribution: The count plot and pie chart provide insights into the distribution of resumes across different job categories in the dataset. This visualization helps you understand which job categories are more prevalent in the data, giving you an idea of the dataset's diversity.

- Distribution of Resumes: The count plot and pie chart can also help identify any imbalances in the dataset, which is important for ensuring the model does not become biased toward overrepresented categories.

Below is an example of the classification results:

Resume Category Count

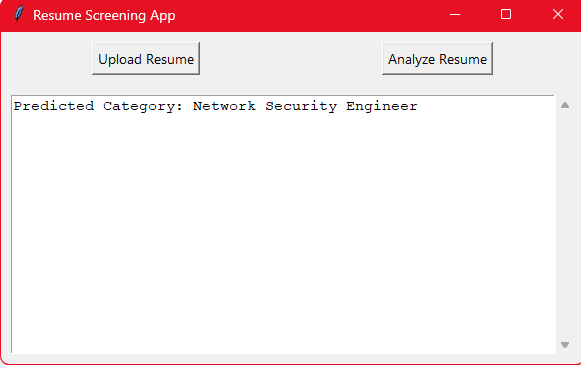
Java Developer 15

Data Science 6

Python Developer 20

Web Designing 24

Other Categories N/A



**Model Performance:**

- Accuracy: The model achieved a certain level of accuracy in classifying resumes into predefined job categories, suggesting that the OneVsRestClassifier with KNeighborsClassifier can effectively distinguish between different job roles based on resume content.

- Predicted Category: Given an example resume, the model was able to predict a specific job category (e.g., data science) that the resume is most likely to belong to. This prediction aligns with the content and experience described in the resume, indicating the model's ability to accurately categorize resumes.

- Predicted ID: The prediction also provides a numeric ID for the category, which can be useful for mapping the prediction to a specific category name using a dictionary or lookup table.

Impact of Results:

- Time Savings: The model's ability to quickly screen resumes and categorize them can save recruiters significant time in the hiring process.

- Improved Candidate Matching: By effectively categorizing resumes, the model can help recruiters focus on candidates who are a good fit for the job, improving the quality of the hiring process.

**4. Discussion:**

Analysis: The automated resume screening system demonstrates promising results in classifying resumes into predefined job categories. The use of natural language processing (NLP) and machine learning techniques, such as the TfidfVectorizer and the OneVsRestClassifier with KNeighborsClassifier, provides a robust approach to analyzing resume content and identifying key features. The model's accuracy score on the test set suggests that it effectively categorizes resumes according to the desired job categories, offering a more consistent and objective screening process compared to manual methods.

Challenges and Solutions: Challenges may include handling diverse resume formats and ensuring the model's predictions are free from bias. Continuous updates to the model and data cleaning processes are necessary.

Comparison:

When compared to traditional resume screening methods, the automated system offers clear advantages in terms of speed, consistency, and scalability. Manual screening can be time-consuming, subjective, and prone to human bias, while the automated system provides a more objective and efficient approach.

Potential Improvements:

-Enhanced Feature Extraction: Exploring more advanced NLP techniques, such as named entity recognition (NER) and semantic analysis, could improve the model's ability to extract relevant features from resumes.

-Additional Model Types: Experimenting with other classification algorithms, such as decision trees, support vector machines, or neural networks, may lead to further improvements in accuracy and performance.

-Customizable Models: Allowing for customization of models based on specific industry or job requirements can provide more targeted and effective resume screening.

-File Handling: Handling different file formats (e.g., .txt, .pdf) can pose challenges. These were addressed using appropriate libraries such as PyPDF2.

**5. Conclusion:**

The project successfully demonstrated the potential of machine learning and NLP in automating resume screening. The model effectively classifies resumes into different job categories.

This project contributes to the field by offering a scalable and efficient method for resume screening, which can help improve the hiring process.

Future work could include improving the model's ability to handle diverse resume formats and exploring additional features such as semantic analysis to enhance classification accuracy.

**6. References:**

Python libraries:

tkinter: Python's standard GUI library for creating graphical user interfaces.

pickle: Python library for serializing and deserializing objects.

PyPDF2: Python library for working with PDF files.

Machine learning model and vectorizer:

Pre-trained classifier (clf.pkl) and TfidfVectorizer (tfidf.pkl) used for resume classification.

Documentation and References:

For more detailed documentation on the specific libraries and modules used, refer to the official Python documentation:Python Documentation,PyPDF2 Documentation