

**Gain insightful perspectives**

# **Resume Parsing and Fake Job Listing Detection: Enhancing Recruitment Efficiency and Preventing Fraud**

**in**

**Partial Fulfillment of S.Y.B.Tech CSE (AI&ML) Course- Project Development - I**

**by**

**Group 9**

**PIYUSH AGARWAL(SYAIMLA01)**

**SOHAM NARSALE(SYAIMLA39)**

**SHREERAJ PATIL(SYAIMLA43)**

**PARTH SINGHAL(SYAIMLA53)**

**(A.Y. 2022-23 )**

**Under the guidance of**

**DR. SONAL GORE**

**Department of CSE (AI&ML)**

**PCET's Pimpri Chinchwad College of Engineering**

# Problem Statement and Introduction

- The process of reviewing resumes for job openings can be time-consuming and inefficient, particularly for large organizations with a high volume of applicants. Manually reviewing resumes also increases the risk of human error and bias, which can negatively impact the recruitment process.
- The process of reviewing resumes for job openings can be time-consuming and inefficient, particularly for large organizations with a high volume of applicants. Manually reviewing resumes also increases the risk of human error and bias, which can negatively impact the recruitment process.
- Resume parsing can streamline the recruitment process, reduce bias, improve data analytics, and protect job seekers from fraudulent recruitment practices.

# Motivation

There are several motivations for a project on *fake job posting detection* and *resume parsing*:

- Protecting job seekers:
- Protecting employers:
- Improving the job market:
- Advancing machine learning techniques:
- Streamlining recruitment processes:
- Improving candidate experience:
- Reducing bias:
- Improving data analytics:

Overall, a project on resume parsing can have significant practical, societal, and scientific value by streamlining the recruitment process, improving the candidate experience, reducing bias, and enhancing data analytics in recruitment.

# Objectives

The objectives for a project on resume parsing are:

- To develop an automated system for resume parsing to extract useful information from resumes and identify possible job opportunities..
- Use natural language processing and machine learning techniques to analyze resumes, identifying relevant information such as job requirements, candidate qualifications.
- To evaluate the performance of the system by testing it on a large dataset of resumes, comparing it to existing methods of manual screening and evaluation.
- Optimize the system to increase its accuracy extracting relevant information from resumes.
- To develop a user-friendly interface that allows recruiters and job seekers to easily interact with the system, providing feedback and suggestions for improvement.
- Conduct a cost-benefit analysis to determine the economic and societal impacts of the system, including its potential to reduce fraud, streamline the recruitment process, and improve data analytics in recruitment.

Overall, the objectives of the project are to develop an automated system that consists of resume parsing to improve the efficiency, accuracy, and fairness of the recruitment process.

# **Literature survey**

<b>Reference 1</b>	End-to-End Resume Parsing and Finding Candidates for a Job Description using BERT IIIT Delhi <a href="#">Vedant Bhatia</a> , <a href="#">Prateek Rawat</a> , <a href="#">Ajit Kumar</a> , <a href="#">Rajiv Ratn Shah</a>
<b>Objectives</b>	The objective of this research paper is to present an end-to-end solution for finding suitable candidates for a job along with resume parser using text classification technique (BERT.)
<b>Proposed solution</b>	The resume parser can extract information from LinkedIn resumes with 100% accuracy, and the ranking system uses BERT( Bidirectional Encoder Representations from transformers) to determine the degree of similarity between a job description and a candidate's past job experiences as mentioned in their resume.
<b>Results</b>	<ol style="list-style-type: none"><li>1. Differentiating between LinkedIn and non-LinkedIn resumes:</li><li>2. Building a BERT based model to convert resumes to a standard format.(LinkedIn format is converted to HTML format)</li><li>3. Data collection tool for employers in the form of a web application.</li><li>4. Ranking resumes as per their suitability to a job description using BERT sequence pair classification.</li></ol>

<b>Advantages</b>	<ol style="list-style-type: none"><li>1. BERT has ability to the context and meaning of words and phrases in a sentence.</li><li>2. BERT is a pre-trained deep learning model</li><li>3. High level of accuracy in ranking system for resume</li></ol>
<b>Limitation</b>	<ol style="list-style-type: none"><li>1. The system relies on the information provided in the resumes, which may not always be accurate or complete.</li><li>2. The system is currently limited to LinkedIn resumes and general non-LinkedIn formats.</li><li>3. We cannot train a network to rank resumes as per their suitability to the job description due to a lack of ground truth and information loss..</li><li>4. While the system provides an intuitive measure of similarity between the job description and candidate profiles, it does not take into account other factors such as soft skills or cultural fit.</li><li>5. While the system has achieved high levels of accuracy in both resume parsing and candidate ranking, there is still room for improvement in terms of accuracy and efficiency.</li></ol>



Reference 2	<p><b>Resume parser using NLP, University of Wolverhampton</b></p> <p><b>Prasuna Pokharel</b></p> <p><b><a href="https://www.researchgate.net/publication/361772014_RESUME_PARSER">https://www.researchgate.net/publication/361772014_RESUME_PARSER</a></b></p>
Objectives	<ol style="list-style-type: none"><li>1. To develop an automated system that can extract relevant information from resumes in a structured format.</li><li>2. To reduce the time and effort required to manually review resumes by HR professionals.</li><li>3. To improve the accuracy of candidate selection by providing a more comprehensive view of their skills and experiences.</li><li>4. To provide a transparent approach to project management, with tasks and progress constantly visible during each sprint.</li><li>5. To create a user-friendly system that is easy to learn and use, making it accessible for companies of all sizes.</li><li>6. To implement Natural Language Processing (NLP) techniques to extract minute data from resumes, such as education, experience, skills, and experience.</li></ol>
Proposed solution	<ol style="list-style-type: none"><li>1. The proposed solution is a system that uses NLP techniques to extract information from resumes.</li><li>2. The system consists of three modules: an administration management system, a file upload and parser system, and an information extraction system.</li><li>3. The system will use algorithms like BERT( Bidirectional Encoder Representations from transformers) Model, Neural Network, and NLP for information extraction.</li><li>4. BERT is a powerful algorithm that beats traditional models like RNN, LSTM, and Bi-LSTM in NLP tasks.</li><li>5. The administration management system and file upload and parser system details are not specified.</li><li>6. The information extraction system is a critical component of the solution, and it's not clear what kind of features will be extracted from resumes</li></ol>

<b>Results</b>	<p>The system is made up of three modules: an administration management system, a file upload and parser system, and an information extraction system. By parsing the resume, HR can select the best candidate for the job based on the company's needs. Although extracting information from resumes without losing candidate information is a challenging task, the NLP method is simple to adopt and provides meaningful results. Overall, this project presents a promising solution to make the recruiting process easier and more efficient for companies.</p>
<b>Advantages</b>	<ol style="list-style-type: none"><li>1. The system is made up of three modules that work together to provide a structured format for the extracted information.</li><li>2. The parsed information from the resume can be used by HR to select the best candidate for the job based on the company's needs.</li><li>3. The NLP method is simple to adopt and provides meaningful results compared to other methodologies described in research papers.</li><li>4. The system reduces the time taken to manually review resumes, allowing HR professionals to focus on other important tasks.</li><li>5. The system provides a transparent approach to project management, with tasks and progress constantly visible during each sprint.</li></ol>
<b>Limitation</b>	<ol style="list-style-type: none"><li>1. The system may not be able to extract all relevant information from resumes, especially if the resume is poorly formatted or contains unusual language.</li><li>2. The system may not be able to accurately identify certain skills or experiences that are not explicitly mentioned in the resume.</li><li>3. The system may require significant training data to improve its accuracy, which can be time-consuming.</li><li>4. The system may require regular updates and maintenance to keep up with changes in language and industry-specific terminology.</li></ol>

<b>Reference 3</b>	<b>Resume Parser Using Natural Language Processing Techniques International Journal of Research in Engineering and Science (IJRES) Volume 9 Issue 6    2021    PP. 01-06</b>  <b>Shubham Bhor<sup>1</sup> , Vivek Gupta<sup>2</sup> , Vishak Nair<sup>3</sup> , Harish Shinde<sup>4</sup> , Prof. Manasi S.Kulkarni<sup>5</sup></b>
<b>Objectives</b>	<ol style="list-style-type: none"> <li>1. Develop a job portal where employees and applicants can upload their resumes for a particular job.</li> <li>2. Use natural language processing techniques to parse resumes and extract necessary information.</li> <li>3. Rank resumes based on the skill set required by the company and the skills possessed by the employee.</li> <li>4. Enlarge the resume dataset and improve the performance of the proposed system.</li> <li>5. Parse resumes from different applications and websites like LinkedIn, GitHub, Naukri.com, etc.</li> <li>6. Add a wide range of psychometric tests to make the system more versatile.</li> <li>7. Extract data from social media platforms like LinkedIn to make recruitment processes easier and avoid unfair and discriminatory practices.</li> </ol>
<b>Proposed solution</b>	<ol style="list-style-type: none"> <li>1. It uses natural language processing techniques to parse resumes and extract relevant information.</li> <li>2. Resumes received will be parsed and ranked according to company requirements.</li> <li>3. The system will also extract data from social media platforms like LinkedIn to make recruitment processes easier and avoid unfair and discriminatory practices.</li> </ol> <p>Overall, the proposed solution aims to provide quality applicants for companies while dampening unfair and discriminatory practices in the recruitment process</p> <p>.</p>
<b>Results</b>	<ol style="list-style-type: none"> <li>1. Rank resumes based on company requirements and extract data from social media platforms like LinkedIn to make recruitment processes easier and avoid unfair and discriminatory practices.</li> <li>2. The future scope of the project includes enlarging the resume dataset, improving system performance, and adding a wide range of psychometric tests to make it more versatile.</li> </ol>

<b>Advantages</b>	<ol style="list-style-type: none"><li>1. Eases the recruitment process for companies and candidates.</li><li>2. Provides quality applicants for companies.</li><li>3. Avoids unfair and discriminatory practices in the recruitment process.</li><li>4. Uses natural language processing techniques to parse resumes and extract relevant information.</li><li>5. Ranks resumes based on company requirements.</li><li>6. Extracts data from social media platforms like LinkedIn to make recruitment processes easier.</li><li>7. Enlarges the resume dataset to improve system performance and accuracy.</li></ol>
<b>Limitation</b>	<ol style="list-style-type: none"><li>1. The accuracy of the system may depend on the quality and completeness of the resume data.</li><li>2. The system may not be able to handle :-<ol style="list-style-type: none"><li>a.non-standard resume formats or unconventional data.</li><li>b.resumes in languages other than English.</li><li>c.complex or ambiguous language structures.</li></ol></li><li>3. The system's ranking algorithm may not accurately reflect company preferences or requirements in all cases.</li><li>4. Extracting data from social media platforms like LinkedIn may raise privacy concerns for candidates.</li></ol>

<b>Reference 4</b>	<b>Fake Job Detection Using Machine Learning(IJRASET Volume 10 Issue IV Apr 2022)</b> Priya Khandagale <sup>1</sup> , Akshata Utekar <sup>2</sup> , Anushka Dhonde <sup>3</sup> , Prof. S. S. Karve <sup>4</sup> <sup>1, 2, 3</sup> Datta Meghe College of Engineering Airoli Navi Mumbai, Maharashtra, 400708 <sup>4</sup> Guide
<b>Objectives</b>	<ol style="list-style-type: none"> <li>1. The objective of this paper is to propose an automated solution based on machine learning-based classification approaches to prevent fraudulent job postings on the internet</li> <li>2. To distinguish fake job postings from a wider pool of job postings and notify the user in order to protect them from falling victim to online scams.</li> </ol>
<b>Proposed solution</b>	<ol style="list-style-type: none"> <li>1. An automated system that uses machine learning-based classification approaches to detect fraudulent job postings on the internet.</li> <li>2. A Kaggle dataset is used, the data is preprocessed and cleaned to make it prediction ready, then the data is provided to the classifier for predictions. The system trains the model to classify jobs as authentic or false based on previous data of bogus and legitimate job postings. Once a fake job posting is identified, it can be deleted so that job seekers can focus entirely on legitimate job openings.</li> </ol>
<b>Results</b>	<p>The result of this project is an automated system that uses machine learning-based classification approaches to detect fraudulent job postings on the internet.</p> <p>Overall, this paper aims to address the dangerous problem of fake job postings and protect users from falling victim to online scams.</p>

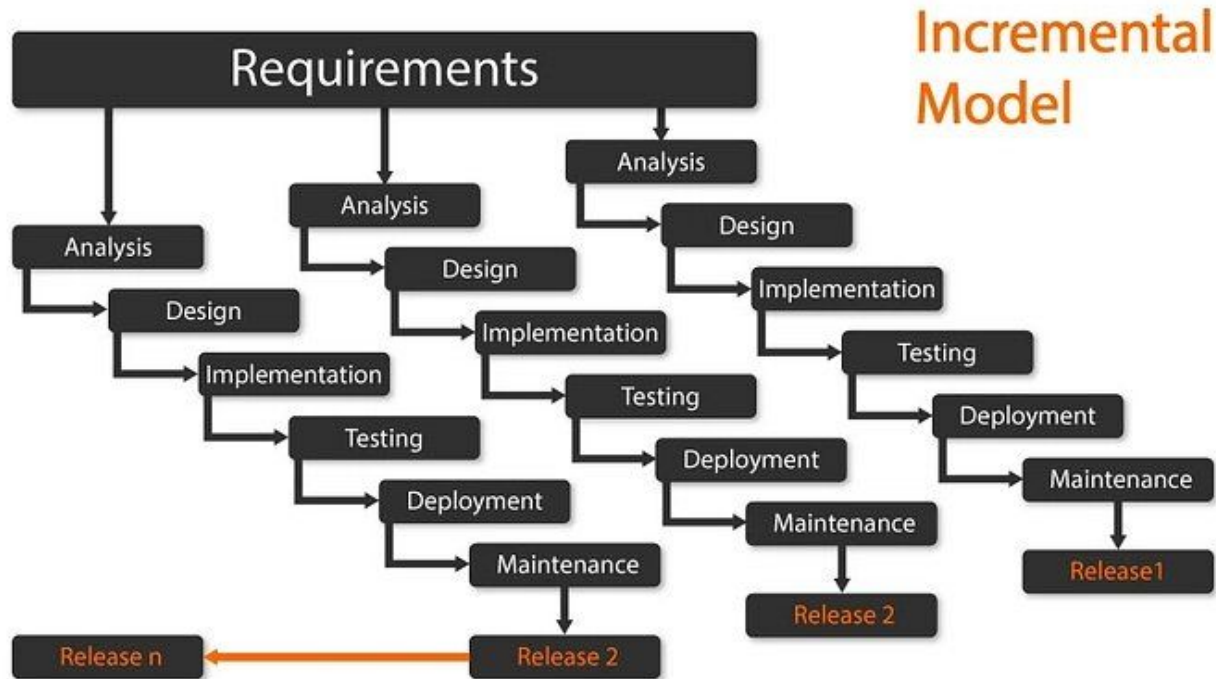
<b>Advantages</b>	<ul style="list-style-type: none"><li>● <u>Automated solution</u>: The proposed system is an automated solution that can detect fake job postings on the internet, which saves time and effort for job seekers.</li><li>● <u>Accuracy</u>: The system uses machine learning-based classification approaches to distinguish bogus job postings from a wider pool of job postings with high accuracy.</li><li>● <u>Protection from scams</u>: The proposed solution aims to address the dangerous problem of fake job postings and protect users from falling victim to online scams.</li><li>● <u>Easy to use</u>: The system is easy to use and does not require any technical expertise, making it accessible to a wide range of users.</li><li>● <u>Scalability</u>: The system can be scaled up easily to handle large volumes of data, making it suitable for use by organizations that receive a high volume of job applications.</li></ul>
<b>Limitation</b>	<ul style="list-style-type: none"><li>● <u>Limited dataset</u>: The system uses a Kaggle dataset that contains information on the job, which may not be representative of all job postings on the internet.</li><li>● <u>Language limitations</u>: The system is designed to work with English language job postings only, which may limit its usefulness in non-English speaking countries.</li><li>● <u>False positives</u>: The system may classify some legitimate job postings as fake, leading to false positives and potentially causing frustration for users.</li><li>● <u>Dependence on data quality</u>: The accuracy of the system depends on the quality of the data used to train it. If the data is incomplete or inaccurate, it may affect the performance of the system.</li><li>● <u>Lack of human oversight</u>: The system relies solely on machine learning-based classification approaches and does not involve human oversight, which may limit its ability to detect more sophisticated scams that require human intuition and judgment.</li></ul>

<b>Reference 5</b>	<p>Fake Job Post Prediction Using Different Data mining Techniques 2021 2nd International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST)</p> <p>Sultana Umme Habiba; Md. Khairul Islam; Farzana Tasnim</p>
<b>Objectives</b>	<p>The study aims to provide insights into the effectiveness of these techniques for detecting fake job postings. This paper proposed to use different data mining techniques and classification algorithm like KNN, decision tree, deep neural network etc. to predict a job post if it is real or fraudulent.</p>
<b>Proposed solution</b>	<p>The proposed solution for detecting fake job postings in this study involves using different data mining techniques and classification algorithms to analyze job post data and predict whether a job post is real or fraudulent. The study compares the performance of several techniques, including KNN, decision tree, support vector machine, naive bayes classifier, random forest classifier, and deep neural network. The authors also describe a three-step process for preparing the data for analysis: data pre-processing, feature selection, and fraud detection using a classifier.</p>
<b>Results</b>	<ol style="list-style-type: none"><li>1. Accuracy of predictions differed by each technique and classification algorithm.</li><li>2. Highest accuracy was achieved by<ol style="list-style-type: none"><li>a. Deep neural network technique, average of 97.7%.</li><li>b. Random forest classifier,average of 96.7.</li></ol></li></ol> <p>However, it is important to note that the accuracy of each technique may vary depending on factors such as the dataset used and other parameters chosen during training and testing.</p>

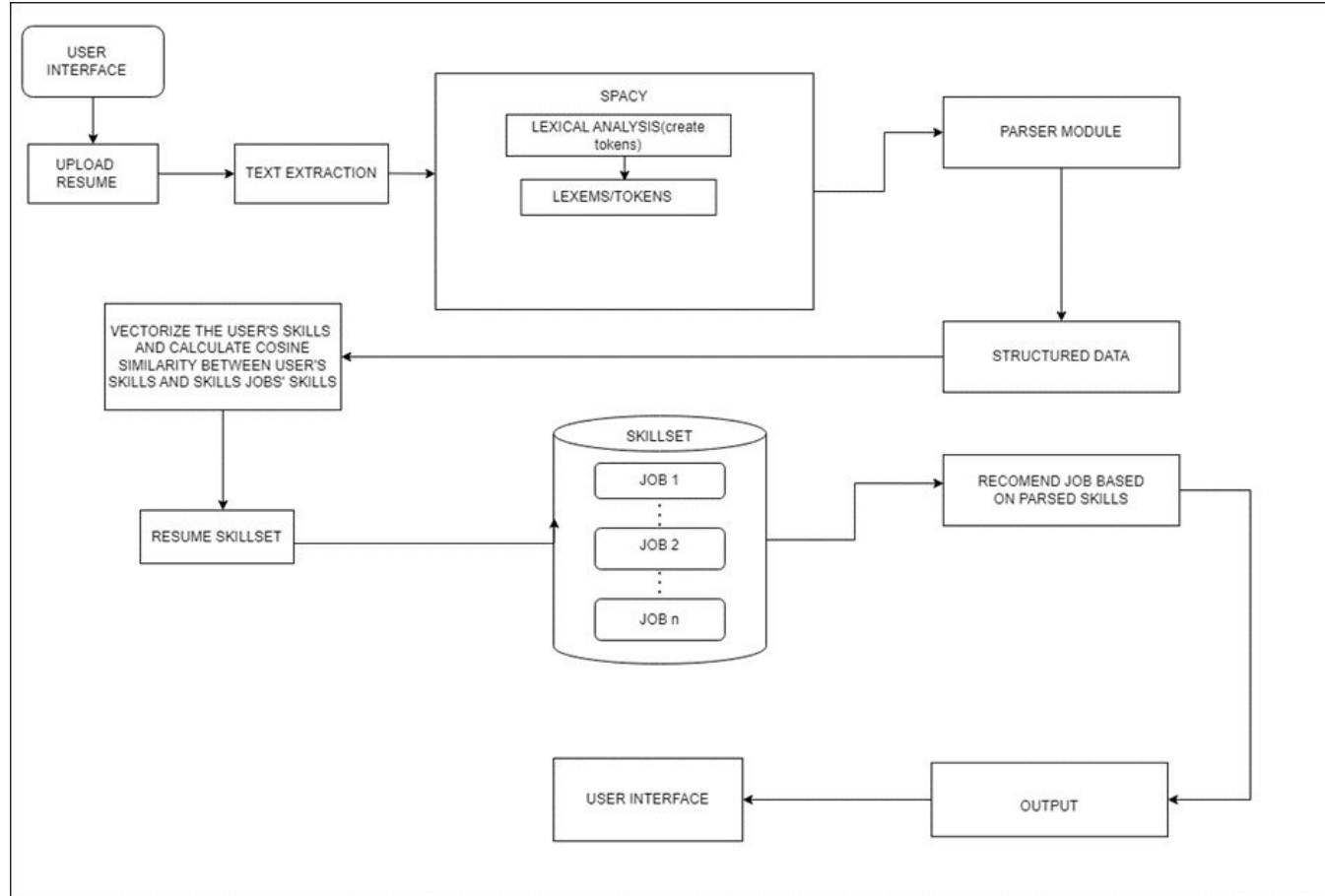
<b>Advantages</b>	<p>Here are some advantages of each technique mentioned in the study:</p> <ol style="list-style-type: none"><li>1. KNN: Simple and easy to implement, can handle both continuous and categorical data.</li><li>2. Decision tree: Easy to understand and interpret, can handle both continuous and categorical data.</li><li>3. Support vector machine: Effective in high-dimensional spaces, can handle both linear and non-linear data.</li><li>4. Naive Bayes classifier: Simple and fast, can handle large datasets with high dimensionality.</li><li>5. Random forest classifier: Can handle missing values and noisy data, provides an estimate of feature importance.</li><li>6. Deep neural network: Can learn complex non-linear relationships between input and output variables, can automatically extract features from raw data without manual feature engineering.</li></ol>
<b>Limitation</b>	<p>Here are some limitations of each technique mentioned in the study:</p> <ol style="list-style-type: none"><li>1. KNN: Sensitive to irrelevant features, requires a large amount of memory to store the entire dataset.</li><li>2. Decision tree: Prone to overfitting, can be biased towards features with many levels.</li><li>3. Support vector machine: Can be sensitive to the choice of kernel function and parameters, may not perform well on imbalanced datasets.</li><li>4. Naive Bayes classifier: Assumes that all features are independent, may not perform well on datasets with correlated features.</li><li>5. Random forest classifier: Can be slow to train on large datasets, may not perform well on datasets with high dimensionality.</li><li>6. Deep neural network: Requires a large amount of training data and computational resources, can be difficult to interpret and debug.</li></ol>



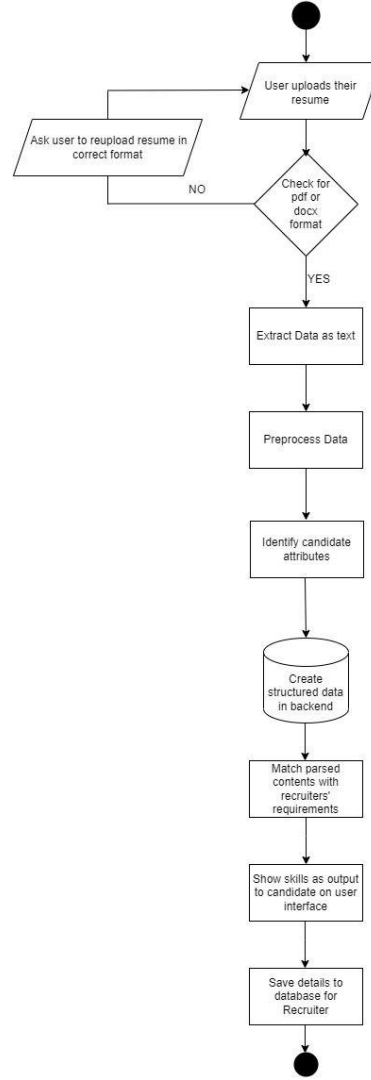
## Process Model followed by Project



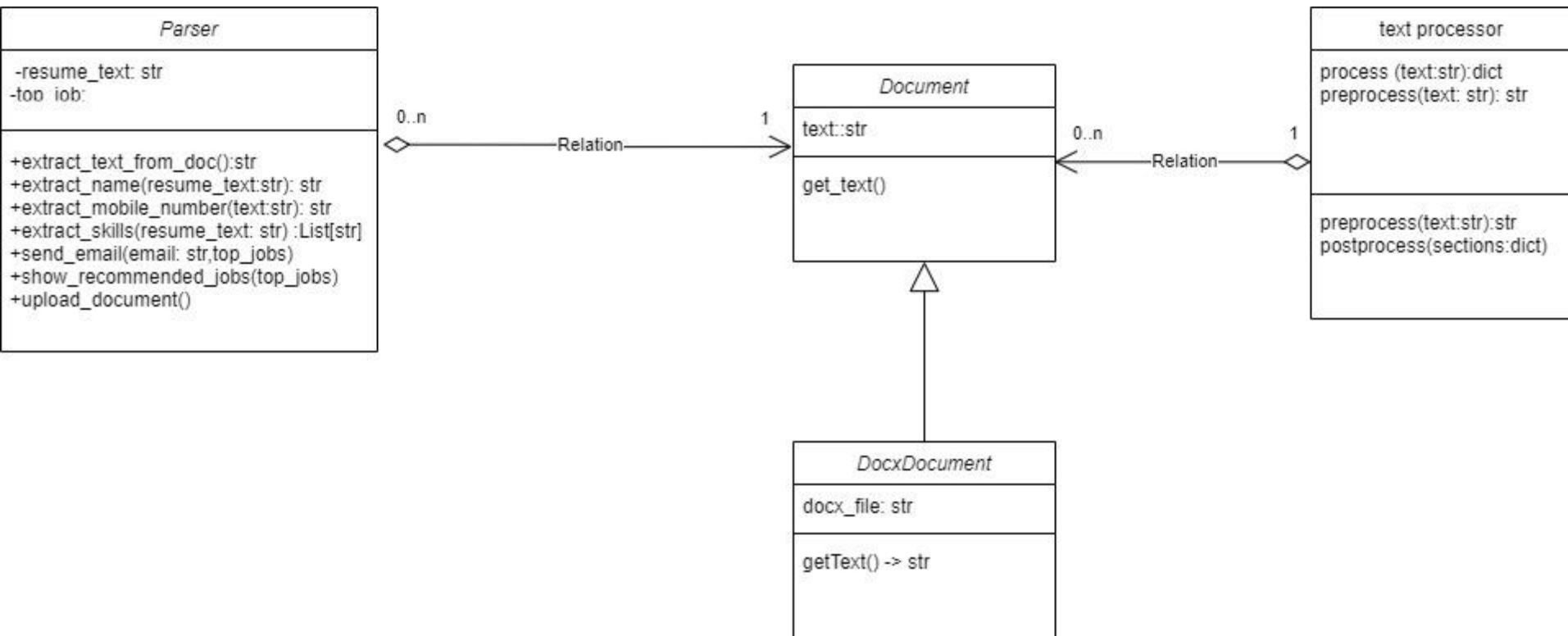
# ARCHITECTURE DIAGRAM



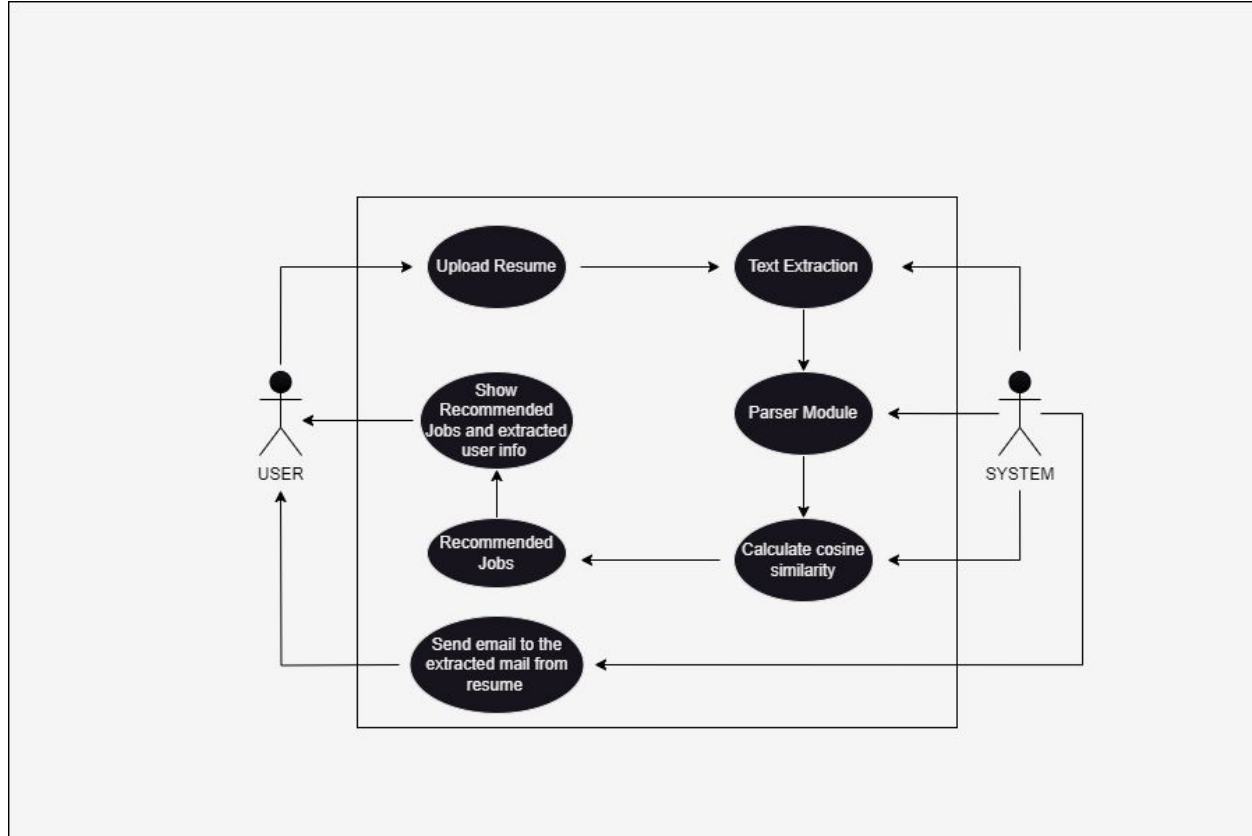
# ACTIVITY DIAGRAM



# CLASS DIAGRAM



## USE CASE DIAGRAM



# Planning for Paper Publication

1. IEEE Transactions on Knowledge and Data Engineering:

This journal focuses on the topics related to artificial intelligence, including NLP and machine learning,

2. ACM Transactions on Information Systems:

emphasis on information retrieval and NLP, making it a good fit for research on resume parsing and fake job detection.

3. IJRASET (**International Journal for Research in Applied Science and Engineering Technology**)

# Conclusion

- The system will be creating a completely automated hiring environment for both the job seekers as well as candidates.
- The machine learning model can analyze the candidate's resume and extract relevant information such as their work experience, education, skills, and achievements. This analysis can be used to match the candidate's profile with the job requirements and identify the best-suited candidates for the job. The deep learning and NLP models can also be used to extract more nuanced information, such as sentiment analysis of the resume and the candidate's personality traits.
- The deep learning, NLP, and machine learning models can work together to detect fraudulent job postings by analyzing the job description and company details, identifying patterns of fake job postings, and flagging them for further review by human experts. The machine learning model can also learn from past examples of fake job postings to improve its accuracy over time.
- The machine learning model can also be used to predict which candidates are more likely to be successful in the role based on past hiring data, improving the accuracy of the hiring process.
- In future the model can also be used to automate other aspects of the hiring process, such as scheduling interviews and sending automated responses to candidates.
- The model can also be used to personalize the candidate experience, such as recommending other job opportunities based on the candidate's skills and interests.

The implementation of these technologies has the potential to revolutionize the job market by providing a safer and more streamlined hiring process.

# References

- A Comparative Study on Fake Job Post Prediction Using Different Data mining Techniques. 2021 2nd International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST)
- End-to-End Resume Parsing and Finding Candidates for a Job Description using BERT IIIT DELHI 2019.
- Resume parser using NLP, University of Wolverhampton
- Resume Parser Using Natural Language Processing Techniques International Journal of Research in Engineering and Science (IJRES) Volume 9 Issue 6 | 2021 || PP. 01-06
- Fake Job Detection Using Machine Learning(IJRASET Volume 10 Issue IV Apr 2022)
- Fake Job Post Prediction Using Different Data mining Techniques 2021 2nd International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST)



**THANK YOU**