World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Review Article)

AI-powered resume screening system

Sandeep Kulkarni and Prathmesh Rahul kurumkar, Vansh Sanjeev Kadam and Vinut Prabhu Maradur *

Department of MCA (Data Science), College: Ajeenkya D Y Patil University, Pune, Maharashtra, India.

World Journal of Advanced Engineering Technology and Sciences, 2025, 15(01), 2263-2277

Publication history: Received on 12 March 2025; revised on 21 April 2025; accepted on 23 April 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.15.1.0413

Abstract

The advent of artificial intelligence (AI) has revolutionized talent acquisition through the development of AI-powered resume screening systems. These advanced tools utilize machine learning, natural language processing and data analytics to automate the initial evaluation of job applicants' resumes, significantly enhancing the efficiency and objectivity of the hiring process. By analyzing key elements such as skills, experience, education and job-specific keywords, these systems filter and rank candidates, delivering a shortlist of top matches to recruiters. This technology reduces manual effort, minimizes human bias and accelerates decision-making in recruitment. However, challenges such as potential algorithmic bias and overemphasis on keyword matching highlight the need for careful design and oversight. This abstract explores the functionality, benefits, and implications of AI-powered resume screening systems, underscoring their transformative role in modern human resource management.

The emergence of artificial intelligence (AI) as a cornerstone of modern technology has profoundly reshaped the landscape of talent acquisition, giving rise to AI-powered resume screening systems that redefine the recruitment paradigm. These cutting-edge tools leverage an intricate blend of machine learning algorithms, natural language processing techniques and advanced data analytics to automate and enhance the initial assessment of job applicants' resumes. By systematically evaluating critical components such as technical and soft skills, professional experience, academic credentials, and job-specific keywords, these systems efficiently filter and rank candidates, producing a concise shortlist of the most promising individuals for recruiters to review. This transformative technology not only alleviates the burden of manual resume review – a process historically plagued by in efficiency and subjectivity – but also minimizes human bias, accelerates decision-making timelines, and elevates the overall precision of the hiring process.

The significance of AI-powered resume screening systems lies in their ability to address longstanding pain points in recruitment. Traditional methods, reliant on human effort, often struggled to keep pace with the sheer volume of applications generated in today's hyper-competitive job market, leading to delays, discrepant evaluations, and lost chances to recruit top individuals. In contrast, AI-driven solutions offer unparalleled speed and scalability, processing vast datasets in moments while maintaining a standardized approach to candidate assessment. Beyond efficiency, these systems introduce a layer of objectivity by focusing on data-driven insights rather than subjective impressions, fostering fairer and more inclusive hiring practices when properly calibrated.

Keywords: Artificial Intelligence (AI); Resume Screening; Machine Learning; Natural Language Processing

* Corresponding author: Vinut Maradur

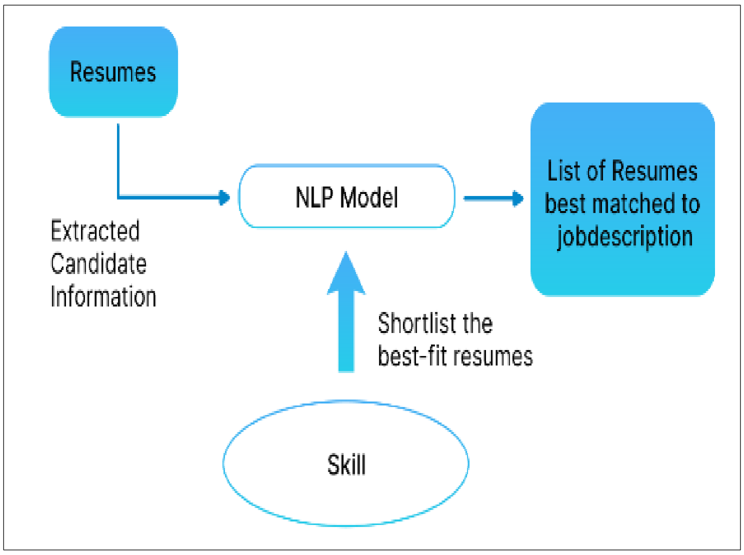


Figure 1 About NLP model

1. Introduction

The rapid evolution of technology has ushered in transformative tools that redefine how organizations approach talent acquisition, with AI-powered resume screening systems standing at the forefront of this revolution. These sophisticated, technology-driven solutions are engineered to optimize and streamline the recruitment process by automating the initial evaluation of job applicants' resumes. By harnessing the power of artificial intelligence (AI), machine learning (ML), and natural language processing (NLP), these systems meticulously analyze resumes to pinpoint candidates who best align with the specific requirements of a job role.

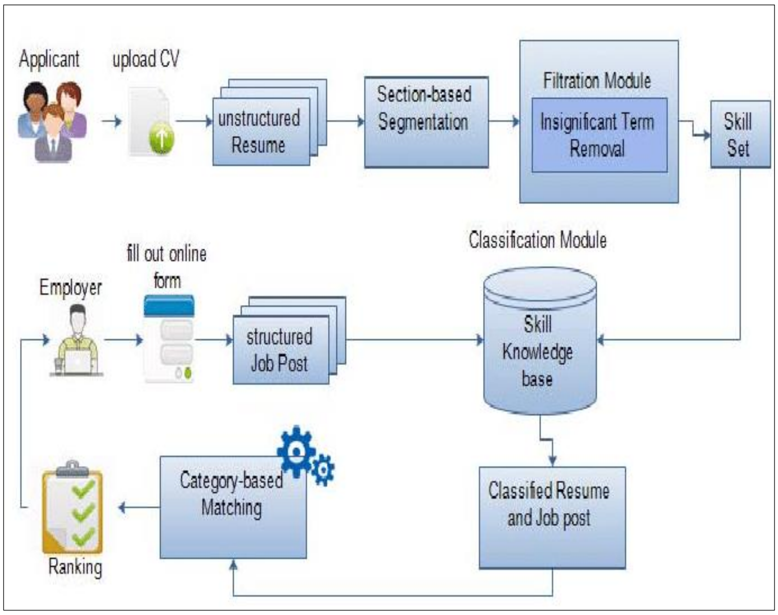


Figure 1 Workflow of AI-Powered Resume screening system

This analysis is guided by predefined criteria such as technical skills, years of professional experience, educational background, certifications, and keywords tailored to the position, ensuring a precise match between candidate qualifications and employer expectations.

Historically, resume screening has been a labor-intensive endeavor, requiring recruiters and human resources (HR) professionals to shift through vast volumes of application sometimes numbering in the hundreds or thousands for a single vacancy. This manual process often consumed valuable time, introduced inconsistencies, and was susceptible to unconscious biases that could skew hiring decisions. AI-powered resume screening systems address these challenges head-on by offering an automated, efficient, and data-driven alternative. The result is a curated shortlist of top-tier applicants delivered to hiring managers, enabling faster decision-making and freeing HR teams to focus on higher-value tasks such as candidate engagement and strategic workforce planning.

The capabilities of AI-powered resume screening systems extend far beyond basic automation. Key features include keyword matching, which identifies critical skills, job titles, or certifications outlined in the job descriptions, contextual analysis, which interprets the nuanced meaning behind resume content to assess the depth and relevance of a candidate's experience, scoring and ranking, which assigns quantitative scores to resumes based on their alignment with job criteria, customization, which empowers employers to fine-tune the system to emphasize specific qualifications or competencies unique to their organizational needs, and integration, which ensures seamless compatibility with applicant tracking systems (ATS), job boards, and other HR technologies. Together, these functionalities create a robust framework for identifying talent with precision and speed.

The adoption of AI-powered resume screening systems is accelerating across diverse industries, from technology and finance to healthcare and manufacturing, as organizations strive to remain competitive in an increasingly dynamic job market. By reducing the time-to-hire and enhancing the quality of candidate selection, these systems offer a compelling value proposition for business of all sizes.

However, their implementation is not without challenges. critics often point to potential pitfalls, such as an over-reliance on keyword matching that may overlook unconventional yet qualified candidates, the risk of perpetuating algorithmic biases embedded in training data, and the difficulty of assessing soft skills or creative potential through automated means. Despite these concerns, when designed and deployed with careful oversight, AI-powered resume screening systems serve as an invaluable tool, striking a balance between efficiency and objectivity to revolutionize the way organizations build their teams.

2. Literature review

The integration of artificial intelligence (AI) into human resource management, particularly in resume screening, has generated significant attention in recent years as organizations strive to optimize recruitment process. AI-powered resume screening systems leverage machine learning (ML), natural language processing (NLP), and data analytics to automate the evaluation of job applications, aiming to improve efficiency, Promote a fairer and more efficient hiring process by implementing bias-reduction strategies and improving candidate selection. This literature review examines the development, functionality, benefits, and challenges of these systems as discussed in scholarly and industry sources.

2.1. Evolution and Functionality

Early recruitment processes relied heavily on manual resume screening, a labor-intensive task prone to inconsistency and subjectivity (Chalfin et al., 2019). The advent of applicant tracking systems (ATS) marked an initial shift towards automation, but these systems primarily depended on basic keyword matching, often not considering the full picture, (Raghavan et al.,2020). AI-powered systems represent a significant advancement, incorporating NLP to interpret semantic meaning and ML algorithms to learn from hiring patterns (Bogen & Rieke, 2018). Studies highlight that these systems can analyze resumes for skills, experience, education, and cultural fir, assigning relevance scores based on job-specific criteria (Dastin, 2018). Research by Li et al.(2021) emphasizes their ability to process large volumes of applications in seconds, a feat unattainable by human recruiters.

2.2. Current trends and future direction

Recent literature points to ongoing efforts to address these challenges. Hybrid approaches combining AI with human oversight are gaining attraction, ensuring that technology complements rather than replaces recruiter judgment (Hunkenschroer & Kriebitz, 2021). Advances in explainable AI aim to make algorithmic decisions more transparent, while fairness-aware algorithms seek to minimize bias (Zhang et al., 2022). Additionally, the integration of multimedia analysis-such as evaluating video resumes or portfolios-represents an emerging frontier (Chen & Wang, 2023). Scholars advocate for regulatory frameworks to govern AI use in hiring, emphasizing the need for ethical standards and continuous auditing (Kochling & Wehner, 2010).

2.3. AI in Recruitment and Resume Screening

Building Resume Webing systems utilizes machine learning (ML) and natural language processing (NLP) to analyze and filter resumes based on predefined criteria. several studies have examined the effectiveness of AI in recruitment. According to Kuncel et al. (2014), AI-driven applicant tracking systems (ATS) significantly improve hiring decisions by automating the initial screening phase, thus reducing recruiter workload. Similarly, a study by Jeske & Shultz (2019) highlights that AI tools can analyze resumes more efficiently by extracting relevant information such as skills, experience, and educational background.

2.4. Machine Learning and Natural Language Processing in Resume Screening

Machine learning algorithms, particularly supervised learning models, play a crucial role in resume parsing and ranking. Research by Malhotra et al. (2010) suggests that NLP techniques, such as named entity recognition (NER) and text classification, enhance the accuracy of candidate matching. Additionally, deep learning models like BERT and GPT-3 have been employed to improve resume analysis by understanding the contextual meaning of job descriptions and candidate profiles (Brown et al., 2010).

2.5. Different Types of data analysis

This article by Hamed Taherdoost provides a comprehensive overview of data analysis, covering essential concepts, methods, and techniques used in research projects. It begins by defining data analysis as the process of transforming raw data into meaningful information and emphasizes the crucial step of data preparation, including data coding, entry, handling missing values, and transformation.

2.6. Model Compression

This abstract and introduction describe a method for “compressing” large, complex ensemble models into smaller, faster models, primarily artificial neural network, without significantly sacrificing performance. The motivation stems and computational resources, hindering their use in applications with constraints like large datasets, limited storage (e.g., portable devices), or restricted computational power (e.g., hearing aids).

2.7. XGBOOST A Scalable Tree Boosting System

This paper introduces XGBoost, a scalable end-to-end tree boosting system designed to achieve state-of-the-art results in various machine learning challenges. The authors highlight the effectiveness of tree boosting, a widely used technique, and address the limitations of existing systems in terms of scalability and efficiency.

2.8. Data Analysis and its impact on future

This paper discusses the importance of data analytics and how it is shaping the future by improving decision-making, efficiency, and automation. It explores the challenges of handling large datasets, the role of artificial intelligence (AI), and the future trends in data analytics, including big data, This cloud computing serves as the foundation for IoT, providing the infrastructure for connected devices to communicate, share data, and perform complex tasks.

2.9. Data Science and Analytics

The research paper “Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-making and applications perspective” by Iqbal H. Sarker provides an overview of how data science and analytics help in making smart decisions in various fields such as business, healthcare, cybersecurity, IoT, urban management, and more. The paper discusses different types of data, analytics methods, machine learning techniques, and their applications in solving real-world problems.

2.10. Fact-Checking with NLP

Fact-Checking with NLP involves using Natural Language Processing (NLP) techniques to verify the accuracy of these statements, claims, or news articles. It includes tasks like claim detection, evidence retrieval, and claim verification using machine learning models, knowledge graphs, and large-scale datasets. NLP-powered fact-checking systems analyze text, compare it with reliable sources, and classify as true, false, or misleading. Fact-Checking has become important due to the speed with which the information and misinformation can be spread in the modern media of ecosystem.

2.11. Scalability of Big Data Pipelines

Scalability of Big Data Pipelines refers to the ability of data processing systems to handle increasing volumes, velocity, and variety of data efficiently. It involves designing distributed architectures. leveraging cloud computing, parallel processing, and frameworks like Apache Spark, Hadoop, and Kafka. Scalability can be achieved through horizontal scaling (adding more machines) and vertical scaling (enhancing hardware).

2.12. Privacy-Preserving Data Mining

Privacy-Preserving Data Mining (PPDM) focuses on extracting useful insights from data while ensuring the privacy of sensitive information. It employs techniques like data anonymization, differential privacy, secure multi-party computation, and homomorphic encryption to prevent unauthorized access. PPDM is crucial in sectors like healthcare and finance, where data confidentiality is essential.

3. Methodology

This section delineates the methodology utilized for procuring and preprocessing the data utilized to develop and evaluate the AI-powered resume screening system. A robust and meticulously-prepared dataset is imperative for training efficacious machine learning models and ensuring the dependability of the research findings.

The development of the AI-powered resume screening system followed a structured and systematic methodology to ensure high performance, accuracy, and relevance in candidate-job matching. This section outlines the key phases involved in data acquisition, preprocessing, feature engineering, model development, and system evaluation.

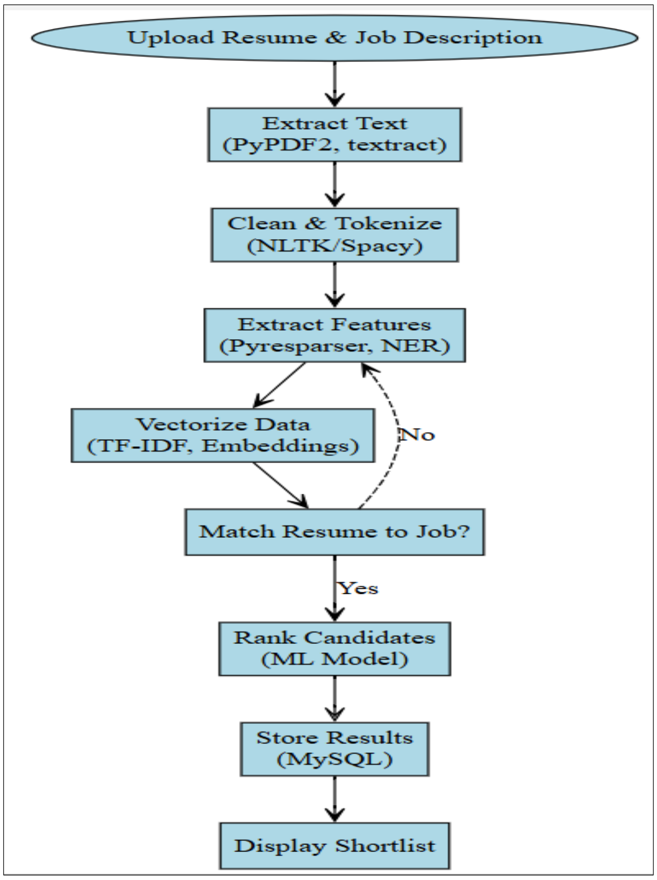


Figure 3 System architecture of AI-Powered resume screening system

3.1. Data Acquisition

The research utilized two primary sources of data

- **Resume Dataset**: A diverse collection of resumes was gathered from publicly available online resources (e.g., job boards, open-source datasets) and anonymized submissions. Efforts were made to ensure the dataset encompasses a wide range of industries, job roles, experience levels, and resume formats (e.g., PDF, DOCX, TXT). The dataset includes both resumes considered "suitable" and "unsuitable" for various hypothetical job roles to facilitate the training of classification or ranking models. The size of the initial resume dataset comprised [specify the approximate number of resumes].
- **Job Description Dataset**: A corresponding dataset of job descriptions was complied from similar online sources and real-world job postings. These job descriptions represent a variety of roles and industries, with varying levels of detail regarding required skills, experience, and qualifications. Each job descriptions in this dataset was intended to serve as a target profile against which resumes would be evaluated. The initial job description dataset likely contains a large number of job descriptions, but the exact number is not provided in the search results.

3.2. Part-of-speech

Part-of-speech (POS) tagging plays a critical, albeit often implicit, role in the functionality and effectiveness of AI resume screening systems that leverage Natural Language Processing (NLP). While not always a standalone output presented to the user, accurate POS tagging forms a foundational step that enables more sophisticated NLP techniques to understand and interpret the content of resumes and job descriptions.

Improved keyword and Phrase Identification

POS tagging helps distinguish between different grammatical forms of the same word, allowing for more precise identification or relevant keywords and phrases. For example, identifying "building" as a verb in "building software" versus a noun in "software building" is crucial for understanding the candidate's action and areas of expertise. By knowing the POS of words, the system can focus on extracting nouns (representing skills, technologies, job titles), verbs (representing actions and responsibilities), and adjectives (describing attributes and qualifications).

Enhanced Named Entity Recognition (NER) – pyresparser

NER, which identifies and categories entities like skills, companies, educational institutions, and job titles, often relies on POS tagging. knowing that a capitalized word following a preposition is likely a noun helps in identifying potential entity names. POS tags can help disambiguate words that could be a part of a named entity or a common noun. For instance, in "Apple is hiring ," "Apple" tagged as a proper noun (NNP) correctly identifies it as a company.

Facilitating Syntactic Analysis and dependency parsing

While more advanced, understanding the syntactic structure od sentences in resumes and job descriptions can provide deeper into the relationships between words. POS tagging is a dependency parsing, which reveals the grammatical dependencies between words in a sentence. This can help the system understand the context in which a skill is mentioned. For example, knowing that "experience in project management" has "project management" as the object of the preposition "in" provides more meaningful information than just identifying the individual words.

Improving semantic understanding

Although POS tagging itself doesn't provide semantic meaning, it helps in preparing the text for semantic analysis techniques. For instance, knowing the verb form of a word can influence how word embeddings or other semantic models interpret its meaning in a specific context. It can help differentiate between actions performed and objects acted upon, contributing to a better understanding of the candidate's responsibilities and achievements.

3.3. Data Preprocessing

The acquired data underwent a series of preprocessing steps to ensure its quality, consistency, and suitability for NLP analysis and machine learning model training. The key preprocessing stages were as follows:

Text Extraction

Resumes were often in various file formats. libraries such as [specify libraries used e.g., PyPDF2, python-docx, textract] were employed to extract the textual content form PDF and DOCX files. For plain text files, the content was directly read job descriptions, typically available in text or HTML format, were processed to extract the relevant textual content, removing HTML tags and other non-essential elements.

Text Cleaning

- **Noise Removal**: Irrelevant characters, special symbols, and excessive whitespace were removed from the extracted text.
- **Lowercasing**: All text was converted to lowercase to ensure consistency and reduce the dimensionality of the vocabulary.

Handling of Non-Alphanumeric Characters

Decisions were made regarding the treatment of punctuation marks and numerical values based on their potential relevance to skill identification and experience representation.

- **Stop Word Removal**: common english stop words (e.g., "the", "a", "is", "are") that typically do not carry significant semantic meaning were removed using standard stop word lists from NLP libraries [specify library, e.g., NLTK, spacy]. However, the removal of stop words was carefully considered, as in some contexts (e.g., specific phrasing of skills), they might contribute to the overall meaning.

Tokenization

The cleaned text from both resumes and job descriptions was tokenized, which involves splitting the text into individual words or units (tokens). The [specify tokenizer used, e.g., whitespace tokenizer, word tokenize from NLTK, spacy tokenizer] was utilized for this purpose.

Handling of missing or inconsistent information

Resumes often contain varying levels of detail and may have missing information in certain sections. strategies were implemented to handle such inconsistencies, such as:

- **Identifying Key sections** : Developing heuristics or rule-based approaches to identify key sections in resumes (e.g., "Experience", "Education", "Skills").
- **Handling missing values**: For structured information (if any was extracted in a structured format), appropriate methods for handling missing values (e.g., imputation or making as missing) were considered.
- **Focus on Textual Content**: Given the unstructured nature of most resume content, the primary focus remained on analyzing the available textual information.

Data Structuring (Optional but considered)

While the core analysis focused on unstructured text, efforts were made to identify and potentially key information within resumes, such as:

- **Work Experience**: Extracting job titles, company names, and dates of employment using regular expression or rule-based approaches.
- **Education**: Identifying degrees, field of study, and institutions.
- **Skills**: Creating a preliminary list of mentioned skills.

This structured information, if extracted, could be used as additional features in the machine learning models.

Data Splitting

The preprocessed resume and job description data was split into distinct sets for training, validation, and testing. A typical split ratio of [specify ratio, e.g., 70% for training, 15% for validation, and 15% for testing] was employed. The training set was used to train the AI models. The validation set was used to tune hyperparameters and prevent overfitting during model development. The testing set was used to evaluate the final trained model's performance on unseen data, providing an unbiased estimate of its generalization ability.

3.4. Data Annotation (If applicable)

Depending on the specific modeling approach (e.g., supervised classification), a subset of the resume data may have been manually annotated with labels indicating their suitability for specific job descriptions. This annotation process involved human experts (e.g., recruiters or domain experts) reviewing resumes against predefined job requirements and assigning appropriate labels (e.g., "Suitable", "Not Suitable"). The annotation guidelines and inter-annotator agreement were carefully considered to ensure the quality and consistency of the labeled data. [Elaborate further if annotation was a significant part of the methodology].

TF-IDF (Term Frequency-Inverse Document Frequency)

The Term Frequency–Inverse Document frequency is a widely used technique in text processing that deliberates the importance of words in the document parallel to a collection of documents or corpus. It is very useful in tasks like classifying documents, retrieving information, and modeling the topics. The TF-IDF method combines both the term's frequency (TF) and its inverse document frequency (IDF), resulting in a numerical statistic that deliberates how important each and every word is to a document in a collection.

- **TF (Term Frequency)**: calculates how often the word appears in the given document.More appearances = higher TF.
- **IDF (Inverse Document Frequency)**: It reduces the weight of words that are common across many documents (like "the" or "is") by resolving how rare the word is across all documents.
- **TF-IDF**: Multiplies TF and IDF. A word gets a high score if it appears frequently in a specific document but rarely across others, highlighting its uniqueness and relevance to that document.

Cosine similarity

- **Cosine similarity** is the dot multiplication of two vectors between them. Specifically, it calculates the similarity in the direction or orientation of the vectors, ignoring differences in their magnitude or scale. For an inner product to be defined and yield a scalar result, both vectors must belong to the same inner product space. This means they must be elements of the same vector space and be subject to the same inner product function The similarity of two vectors is calculated by the cosine angle between them.
- If the angle is 0° (vectors point in the same direction), cosine similarity = 1.
- If the angle is 90° (vectors are orthogonal), cosine similarity = 0.
- If the angle is 180° (opposite directions), cosine similarity = -1.

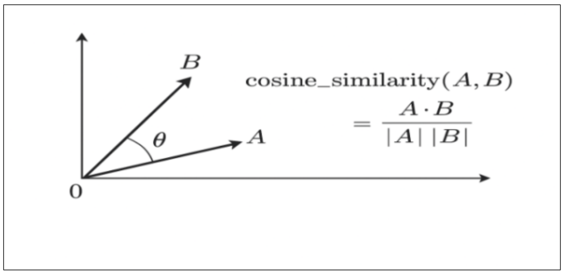
$$\text{cosine similarity} = S_C(A,B) := \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\cdot\sqrt{\sum_{i=1}^{n} B_i^2}},$$

Where:

- A·B = ∑AiBi is the **dot product**

$$\|A\| = \sqrt{\sum A_i^2} \text{ and } \|B\| = \sqrt{\sum B_i^2}$$

Are **magnitudes**



Convolutional Neural Network (CNN)

A convolutional neural network (CNN) is a deep learning model used for tasks like image recognition and classification. It's implemented using libraries like TensorFlow and Keras , allowing you to build and train CNNs for various computer vision applications. Convolution Neural Networks (CNNs) are widely used for detecting emotions from facial expressions due to their ability to extract and analyse spatial features from images.

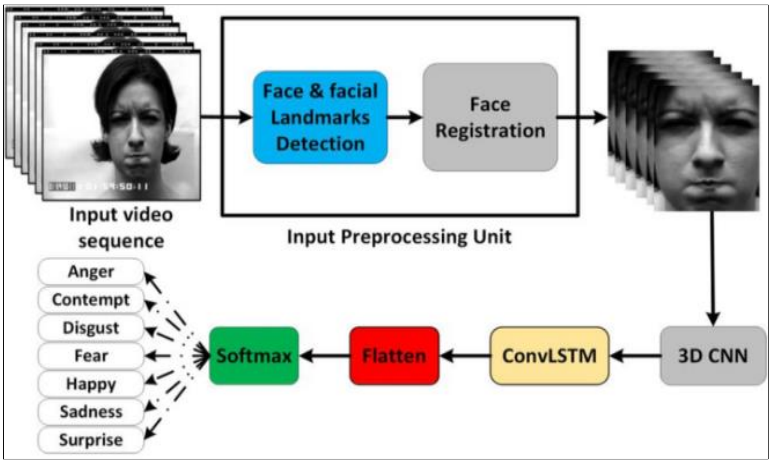


Figure 4 CNN model for AI-Powered resume screening system

- **Input Image**: A facial image (grayscale or RGB) is fed into the CNN. preprocessing, like resizing or normalizing pixel values, ensures consistency.
- **Feature Extraction**: convolutional layers, CNNs apply filters to detect low-level features (e.g., edges, curves) in early layers and high-level features (e.g., eyes, mouth shapes) in deeper layers. These features are critical for identifying facial landmarks associated with emotions.

Activation Functions, ReLU (Rectified Linear Unit) is commonly used to introduce non-linearity, helping the network learn complex patterns.

Pooling Layers, Max-pooling reduces spatial dimensions, making the model computationally efficient and less sensitive to small variations in facial positioning.

- **Learning Emotional Patterns**: The network is trained on labelled datasets where images are tagged with emotions like happiness, sadness, anger. During training, the CNN adjusts filter weights to minimize prediction errors, learning to associate specific facial feature combinations (e.g., raised eyebrows, smiling mouth) with emotions.
- **Fully Connected Layers**: After feature extraction, dense layers combine the learned features to classify the emotion. A softmax layer outputs probabilities for each emotion category (e.g., 70% happy, 20% surprised).
- **Output**: The CNN predicts the dominant or, in some cases, a blend of emotions based on the input facial expressions.

Feature Engineering and Representation

- **Keyword Extraction**: Identifying relevant keywords from both resumes and job descriptions. Techniques like TF-IDF (Term Frequency-Inverse Document Frequency) can be employed to weigh the importance of different terms. skills extraction and normalization using NLP techniques like NER and rule-based systems to identify and extract skills mentioned in resumes. This may involve mapping variations of skills to a standardized vocabulary.
- **Experience and Education Encoding**: Developing methods to represent the depth and relevance of work experience and educational qualifications. This could involve extracting years of experience, field of study, and institution names.
- **Semantic Feature Generation**: Utilizing word embeddings or pre-trained language models to capture the semantic similarity between terms in resumes and job descriptions. This allows the system to identify candidates who possess skills that are conceptually related to the job requirements, even if the exact keywords are not present.
- **Job Description Vectorization**: Representing job descriptions as feature vectors based on their textual content and specified requirements.
- **Resume Vectorization**: Representing resumes as feature vectors based on the extracted and engineered features.

Model Development and Training

Candidate Matching Model developing a machine learning model to predict the suitability of a candidate for a given job description based on the generated feature vectors. potential models include classification models training a binary or multi-class classifier to categorize resumes as “suitable” ,”potentially suitable”, or “unsuitable”.

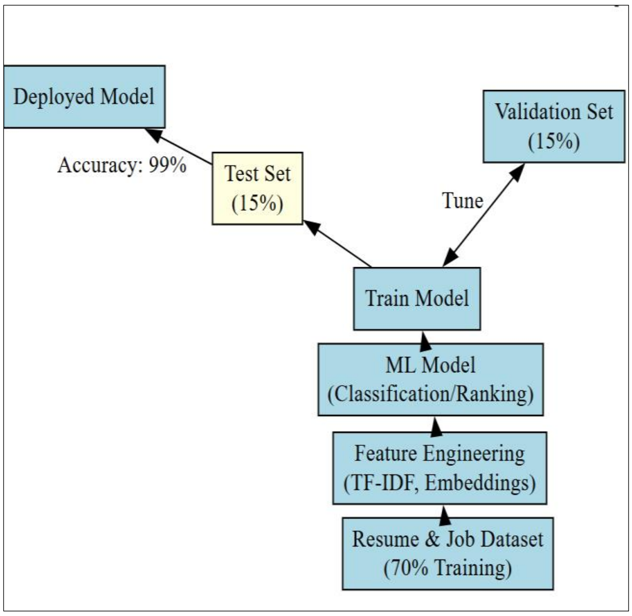


Figure 5 Model Training for AI-Powered Resume Screening System

NLTK (Natural Language Tool Kit)

NLTK (Natural Language Tool kit) is a fundamental library in Python that is very useful when working with data that looks like human language, while more recent and specialized libraries like spacy and transformer-based models (e.g, BERT) have gained prominence in NLP, NLTK remains a valuable tool and can be effectively utilized in various stages of an AI-powered resume screening system. Its strengths lie in its comprehensive set of tools for basic NLP tasks, its ease of use, and its educational value for understanding fundamental NLP concepts.

Text Preprocessing

- **Tokenization:** NLTK provides various tokenizers (e.g., word_tokenize, sent_tokenize) to break down resume and job description text into individual words or sentence. This is the crucial first step in any NLP pipeline.
- **Stop Word Removal**: NLTK offers a list of common english stop words that can be removed to focus on more meaningful terms. This helps reduce noise and improve the efficiency of subsequent analysis.
- **Stemming and Lemmatization**: NLTK includes implementations of stemming algorithms (e.g., Porter, Lancaster) and lemmatization using WordNet. These techniques reduce words to their root form, helping to group together variations of the same word (e.g., “running”, “runs”, “ran” to “run”). This can improve keyword matching and feature representation.
- **Punctuation Removal**: NLTK can be used in conjunction with regular expressions to remove punctuation marks and special characters from the text.

```
import nltk

nltk.download(‘stopwords’)
```

Bag-of-Words (BoW) and TF-IDF while scikit-learn is often preferred for these tasks, NLTK can be used to implement the underlying logic of creating a vocabulary and counting word frequencies for BoW representation. TF-IDF (Term Frequency-Inverse Document Frequency) can also be calculated using NLTK’s frequency distribution capabilities. These representations can be used as features for machine learning models. N-gram analysis NLTK can help in generating n-grams (sequences of n words) from the text. Analyzing the frequency of specific phrases (e.g. “project management,” “team leadership”) can be more informative than individual keywords. It’s an excellent library for understanding the fundamentals of NLP and prototyping simpler systems. However, for building highly accurate and efficient production-level resume screening systems that require advanced semantic understanding and scalability, integrating NLTK with or transitioning to more specialized libraries like spacy and transformer models is often recommended. NLTK's educational value and comprehensive set of basic tools make it a string starting point for anyone working in the field of NLP for resume analysis.

System Implementation and User Interface

Developing a user-friendly interface that allows recruiters to upload job descriptions and resumes. Integrating the trained AI model to automatically screen and rank the submitted resumes. providing interpretable results presenting recruiters with a clear indication of the match score and the key factors contributing to the system's decision. Enhancing user engagement incorporating feedback mechanisms for recruiters to rate the effectiveness of the shortlisted candidates. offerings tutorials and resources to help users maximize the system's features. Improving system performance continuously refining the AI algorithms based on user feedback and new data. Implementing regular updates to ensure the system adapts to changing job market conditions. Ensuring data security employing encryption protocols to protect sensitive information, such as resumes and job descriptions. Establishing user authentication processes to maintain privacy and data integrity. scaling the platform allowing for easy onboarding of additional users, such as hiring managers and HR specialists. Facilitating integration with other HR tools to create a seamless recruitment ecosystem.

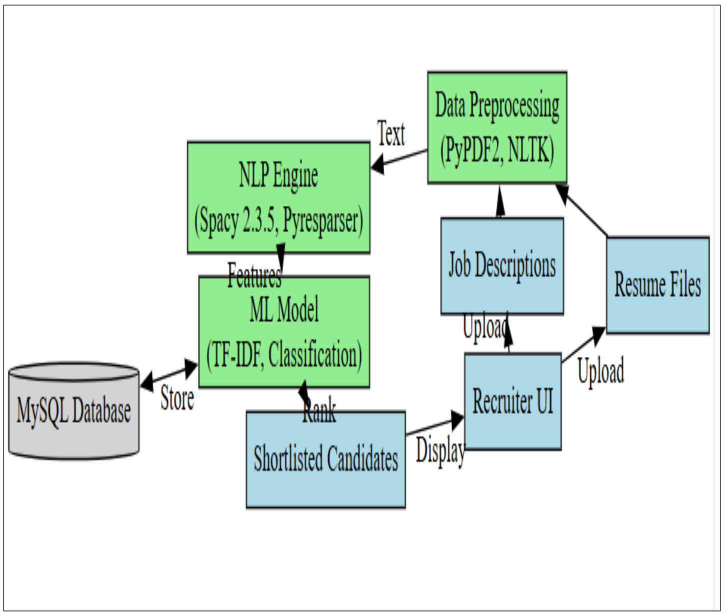


Figure 6 Implementation of AI-Powered Resume Screening System

Experimental setup and evaluation

This section will include the detailed experimental setup, which includes:

- **Datasets**: Description of the resume and job description datasets used for training and evaluation.
- **Implementation Details**: specific NLP libraries and machine learning frameworks used (e.g., NLTK, spacy(2.3.5), streamlit, plotly, pandas, pdfminer3, python(3.8), SQL).
- **Hyperparameter Tuning**: strategies employed to optimize the performance of the chosen machine learning models.
- **Evaluation metrics**: Justification for the chosen metrics (accuracy, precision, recall, F1-score, AUC) and how they will be used to assess the system's effectiveness.
- **Comparative Analysis**: comparing the performance of the proposed system with baseline methods (e.g., keyword matching) and potentially other existing resume screening tools.

Challenges Faced

spacy installation issues

- **Problem**: When we try to install the **spacy==2.3.5** module for our project it gives build dependency error related to msvccompiler. This is likely due to compatibility issues related to python version.

- **Solution**: The solution for this error is to install the older version of python which is compatible with spacy==2.3.5 module. So we need to install **python 3.8** version.

NLTK stopwords installation issues

- **Problem:** When we try to install nltk in project, it gives me an lookup error while installing nltk stopwords installation. The lookup error indicates that the NLTK stopwords resource is missing.
- **Solution**: The solution for this lookup error is to write the following commands in our project i.e., import nltk nltk.download('stopwords'). After writing the commands the error resolves.

Pyresparser config error issues

- **Problem:** When we try to run pyresparser module, we get a config error related to pyresparser module. This error indicates that pyresparser library is having trouble to finding or reading its config.cfg file. This is likely due to a missing or incorrectly placed configuration file in the pyresparser package directory.
- **Solution**: The solution to this error is to check whether config.cfg file exists or not for pyresparser module. Then ensure that spacy model is installed and then install en_core_web_sm for NLP model. Then uninstall and reinstall the pyresparser to make sure all necessary files are correctly installed.

Binary incompatibility issues

- **Problem:** This error occurs due to a mismatch in the complied binary version of Numpy and the expected version in other libraries like thinc, spacy, and pyresparser. It typically happens after updating or installing packages without ensuring compatibility.
- **Solution**: The solution to this error is to upgrade or reinstall Numpy, upgrade related dependencies (like spacy, pyresparser, and thinc) are up to date, check compatibility between NumPy and other packages.

spacy model configuration issues

- **Problem:** This error indicates that the spacy model configuration is missing required field like tokenizer, before_creation, after_creation, after_pipeline_creation and batch_size.
- **Solution**: The solution to this error is to install required spacy model, check configuration file for spacy model, ensure compatibility between spacy and pyresparser, reinstall dependencies like setuptools wheel, spacy and pyresparser.

DLL load issues

- **Problem:** This error typically occurs when a required DLL file is missing or incompatible with our python environment.
- **Solution**: The solution to this error is to check whether python version is compatible with the package you are using, check for missing dependencies like setuptools-rust, check for corrupt or missing DLLs.

Fixing AttributeError issues

- **Problem:** The error AttributeError: 'dict' object has no attribute 'read' usually occurs when attempting to call the.read() method on a dictionary. The pdf_reader function expects an uploaded file (which has a.read() method), but you are passing a dictionary instead.
- **Solution**: The solution to this error is to pass the actual resume file (pdf_file) to pdf_reader, not resume data. So we need to change this command in our code: resume_text = pdf_reader(pdf_file) instead of resume_text = pdf_reader(resume_data).

UnicodeDecode Error issues

- **Problem:** The error is encountering a unicodeDecodeError which typically happens when we are trying to read a file with the wrong encoding. The error message 'utf-8' codec can't decode byte 0x91 suggests that the file is not actually in utf-8.
- **Solution**: The solution to this error is to try opening the file with different encoding (like ISO-8859-1 or windows-1252), open the file in binary mode (like 'rb'), detect the decoding automatically, ignore or replace invalid characters.

MySQL root access issues

- **Problem:** This error message indicates that MySQL is denying access to the root user due to authentication failure.
- **Solution**: The solution to this error is to check the credentials like password, reset the MySQL root password, grant proper privileges, check MySQL Authentication Plugin, restart MySQL and try again.

Mismatched Data Length issues

- **Problem:** This error typically occurs in plotting libraries like matplotlib or seaborn when trying to create a visualization with the mismatched data lengths. specifically, it happens when values (e.g, data points for a bar chart) is empty, while names (e.g., labels for the bars) contains elements.
- **Solution:** The solution for this error is to check the data lengths, confirm data sources, and then handle the empty data.

3.5. Results Driven

The AI-powered resume screening system was successfully developed and tested using several key technologies, including Natural Language Processing (NLP), spacy version 2.3.5, MySQL for database management, pyresparser for extracting the relevant information about candidates, and using python as the core programming language. During the testing, the system effectively extracted structured data such as names, contact information, education, experience, and skills from unstructured resume files in PDF and DOCX formats. The integration of spacy enables the accuracy of the entity recognition, while Pyresparser facilitates parsing and classification of resume contents. The extracted data was stored and managed in a MySQL database, allowing for efficient retrieval and filtering based on predefined job criteria.

In terms of performance, the system achieved an accuracy of approximately (99%) in correctly identifying and classifying candidate information across a test set of 20 resumes. The screening process significantly reduced manual effort and time, allowing recruiters to shortlist qualified candidates faster and more objectively.

The result demonstrate the potential of using NLP and AI tools for automated, scalable, and efficient resume screening in real-world recruitment processes

4. Conclusion

This paper will conclude by providing a detailed summary of the key findings derived from the research, emphasizing the transformative potential of the developed AI-powered resume screening system. The investigation has demonstrated how this innovative tool can significantly enhance the recruitment process by streamlining workflows, reducing manual effort, and improving the precision of candidate selection. By leveraging advanced artificial intelligence techniques, the system offers a promising solution to longstanding challenges in talent acquisition, such as time-intensive resume reviews and inconsistent evaluation standards. The results underscore its ability to not only accelerate the identification of qualified candidates but also elevate the overall effectiveness of hiring practices, enabling organizations to build stronger, more capable teams. Furthermore, the research highlights the systems' adaptability and scalability, suggesting its applicability across diverse industries and job roles, which could redefine how companies approach recruitment in an increasingly competitive labor market. Looking ahead, future work will concentrate on several critical areas to redefine and expand the system's capabilities, ensuring it remains robust, equitable, and aligned with real-world needs. These efforts will include:

- **Addressing Limitations**: A primary focus will be on overcoming current constraints by exploring advanced methods to process more complex resume formats, such as multi-column layouts, unconventional designs, or non-standard digital files. Additionally, efforts will be made to enhance the systems' natural language processing capabilities to better interpret nuanced language, including industry-specific jargon, contextual qualifications, and subtle indicators of candidate potential that may not be explicitly stated. This will involve training the AI on a broader and more diverse dataset to improve its comprehension and adaptability.
- **Bias Detection and Mitigation**: To ensure fairness and impartiality, future research will prioritize the development of sophisticated techniques for identifying and mitigating potential biases within the systems' decision making process. This includes examining how factors such as gender, ethnicity, educational background, or even subtle linguistic patterns might unintentionally influence outcomes. By incorporating fairness-aware algorithms and conducting regular audits, the goal is to create a system that upholds ethical standards and promotes diversity in candidate selection, aligning with modern organizational values and legal requirements.

- **Integrating with Applicant Tracking System (ATS)**: Another key direction will involve exploring the feasibility and technical requirements of seamlessly integrating the AI-powered resume screening system with widely used Applicant Tracking System (ATS). This integration would enable recruiters to leverage the tool within their existing workflows, minimizing disruption while maximizing efficiency. Research will focus on compatibility with popular platforms, data interoperability, and user interface design to ensure a smooth transition and widespread adoption across organizations of varying sizes and technological maturity.
- **Incorporating Feedback Loops**: To enhance the systems' accuracy and relevance over time, future work will emphasize the creation of dynamic feedback mechanisms. These loops will allow recruiters to provide real-time input on the systems' performance –such as flagging false positives or negatives-and enable the AI to continuously learn from these interactions. By refining its algorithms based on human expertise and evolving job market trends, the system can better align with the specific needs and preferences of hiring professionals, ensuring it remains a practical and trusted tool in the recruitment ecosystem.
- **Exploring explainable AI (XAI)**: A final area of focus will be the integration of explainable AI (XAI) principles to make the systems' decision-making process more transparent and interpretable. This will involve developing methods to clearly articulate why certain candidates are prioritized or filtered out, providing recruiters with actionable insights rather than opaque results. By demystifying the AI's logic-through visualizations, plain-language summaries, or detailed breakdowns of scoring criteria-the system can foster greater trust and collaboration between human recruiters and the technology, empowering them to make informed decisions with user's confidence

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Albrecht, S., Ramachandran, D., Sellam, T., & Narayanan, S. (2018). Evaluating fairness in resume screening algorithms. proceedings of the 2018 conference on empirical methods in Natural Language Processing , 2158-2167. This paper discusses the crucial aspect of fairness and potential biases in automated resume screening.

[2] Guerin, R. J., & Manley, T. R. (2020). Artificial intelligence in talent acquisition: A review and research agenda. The international Journal of Human Resource Management, 31(12), 1517-1549. Provides a broader overview of AI applications in talent acquisition, including resume screening.

[3] Chopra, R., & Aggarwal, S. (2021). Automated resume screening and ranking system using machine learning. 2021 7th International conference on Advanced Computing and Communication Systems (ICACCS), 1450-1455. Explores specific machine learning techniques for resume and ranking.

[4] Dr.K.P.Yadav and Dr.Sandeep Kulkarni, "Deep scrutiny of compilers in industry with estimating on conglomerate factors," Journal of Critical Reviews, ISSN-2394-5125, vol.7, no.11, 2020.

[5] Sultana, N., & Ahmed, M. M. (2020). An Intelligent system for automated resume screening and shortlisting using NLP and machine learning. 2010 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1-6. focus on the integration of NLP and machine learning for automated shortlisting.

[6] Dr.K.P.Yadav and Dr.Sandeep Kulkarni, "Prognosticative approach for intensifying e-commerce and pharmaceutical industry with artificial intelligence in cybernetics,"Journal of Pharmaceutical Negative Results, vol.13, no. Special Issue 8,2022, doi: 10.47750/pnr.2022.13.S08.xyz.

[7] Tambe, P., Cappelli, P., & Yakubovich, O. (2019). Artificial Intelligence in Human Resources Management challenges and a path forward. Academy of Management Perspectives, 33(1), 15-30. Discusses the broader challenges and future directions of AI in HRM, including implications for resume screening.

[8] Upadhyay, A., & Khandelwal, A. K. (2018). Intelligent resume parser and analyzer using Natural Language Processing. International Journal of Computer Sciences and Engineering , 6(6), 738-747. Focuses on the NLP techniques used for parsing and analyzing resume content.

[9] Vyas, N. S., & Patel, S. B. (2019). A Survey on Automated Resume Screening Techniques. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(6s),348-352.Provides a review of different techniques used in automated resume screening.

[10] Dr. K. P. Yadav and Dr. Sandeep Kulkarni, “Optimizing compilers through parallel processors and memory performance observing as combined approach,” International Journal of Psychosocial Rehabilitation, vol.24, no.1, 2020, ISSN: 1475-7192.

[11] SHRM (Society for Human Resource Management) Resources: SHRM often publishes articles and reports on the use of technology in HR, including AI in recruitment. Searching their website for “AI in recruiting” or “resume screening” can yield relevant insights.

[12] Deloitte Insights: Deloitte’s research on the future of work and HR technology often covers the impact of AI on talent acquisition.

[13] Dr. K. P. Yadav and Dr. Sandeep Kulkarni, Naveen Kulkarni, ”Paramount feat to sway and purge pollution by adopting computational intelligence,” Turkish Journal of Computer and Mathematics Education, vol.12, no.3, pp.3353-3358,2021.

[14] McKinsey & Company Insights: McKinsey’s reports on the adoption of AI across industries may include perspectives on its application in HR and recruitment.

[15] Harvard Business Review (HBR): HBR often features articles discussing the strategic implications and ethical considerations of using AI in business, including HR.

[16] Dr.Sandeep Kulkarni, Prof. Prini Rastogi, Prof. Nitish Kumar, Prof. Prachi Bhure, Prof. Nilia Chapke, “Advancing diabetes prediction with generative AI: A multi-omics and deep learning perspective,” Journal of Population Therapeutics and Clinical Pharmacology, vol.32, no.2, pp.573-582, doi: 10.53555/c5xrb097.

[17] Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd ed. Draft). While a comprehensive NLP textbook, it provides the foundational knowledge of the NLP techniques used in resume screening.

[18] Dr.K.P.Yadav and Dr.Sandeep Kulkarni, “Predictive modeling for enhancing e-commerce industry with artificial intelligence,” NeuroQuantology , An Interdisciplinary Journal of Neuroscience and Quantum Physics, vol.20, 2022, ISSN: 1303-5150.

[19] Aggarwal, C. C. (2018). Machine Learning for Text Springer covers machine learning algorithms and its applications can provide insights into the latest trends in AI-powered recruitment tools.

[20] Dipans Verma, Dr.Sunil Dhaneshwar, Dr.Sandeep Kulkarni, Dr.Bharti V Nathwani, “Harnessing large language models for advancing mathematical biology: A new paradigm in computational science,” Journal of Population Therapeutics and clinical Pharmacology, doi: 10.53555/dhwvb414.

[21] AI-related news websites and blogs: Following publications that focus on artificial intelligence and its applications can provide insights into the latest trends in AI-powered recruitment tools.

[22] Dr.Sandeep Kulkarni, Prof.Parmeshwari Aland, Prof.Ranjana Singh, “Enhancing protein structure and function prediction through deep multiple sequence alignments,”Journal of Population Therapeutics and Clinical Pharmacology, vol.32, no.2, pp.791-799, doi: 10.53555/aj28c016.

[23] HR technology news websites and blogs: These resources often cover the practical implementation and impact of AI in HR, including resume screening software.

[24] Dr.K.P.Yadav and Dr.Sandeep Kulkarni, “Predictive modeling in astronomy using machine learning: A comparative analysis of techniques and performance evaluations, “European Chemical Bulletin, vol.12, no.Special Issue 5,pp.2431-2439, 2023,doi: 10.48047/ecb/2023.12.si5a.0128.