A New Automated System Approach to Detect Digital Forensics using Natural Language Processing to Recommend Jobs and Courses

Shahlaa Mashhadani[1] , Rajaa Mrayeh Mohammed[2], Nishtha Jatana[3], Charu Gupta[4], Oday Ali Hassen[5,*], Shweta Jindal[6]

[1]Department of Computer, College of Education for Pure Sciences Ibn Al-Haitham, University of Baghdad, Iraq

[2]Department of Computer Science, College of Science for Women, University of Baghdad. Iraq

[3]Maharaja Surajmal Institute of Technology, New Delhi, India

[4]Department of Computer Science and Engineering, Bhagwan Parshuram Institute of Technology, New Delhi, India

[5]Computer Department, College of Education for Pure Sciences, Wasit University, Iraq

[6]Indira Gandhi Delhi Technical University for Women, India

Email: shahlaa.t@ihcoedu.uobaghdad.edu.iq; rajaa.alnidway@uoitc.edu.iq; Nishtha.jatana@gmail.com; charu.wa1987@gmail.com; oday123456789.oa@gmail.com; oday123456789.oa@gmail.com; miss.shweta.singhal@gmail.com

Abstract

A resume is the first impression between you and a potential employer. Therefore, the importance of a resume can never be underestimated. Selecting the right candidates for a job within a company can be a daunting task for recruiters when they have to review hundreds of resumes. To reduce time and effort, we can use NLTK and Natural Language Processing (NLP) techniques to extract essential data from a resume. NLTK is a free, open source, community-driven project and the leading platform for building Python programs to work with human language data. To select the best resume according to the company's requirements, an algorithm such as KNN is used. To be selected from hundreds of resumes, your resume must be one of the best. Therefore, our work also focuses on creating an automated system that can recommend the right skills and courses to help the desired candidates by using Natural Language Processing to analyze writing style (linguistic fingerprints) and also used to measure style and analyze word frequency from the submitted resume. Through semantic search and relying on individual resumes, forensic experts can query the huge semantic datasets provided to companies and institutions and facilitate the work of government forensics by obtaining official institutional databases. With global cybercrime and the increase in applicants seeking work and leveraging their multilingual data, Natural Language Processing (NLP) is making it easier. Through the important relationship between Natural Language Processing (NLP) and digital forensics, NLP techniques are increasingly being used to enhance investigations involving digital evidence and leverage the support of NLP for open-source data by analyzing massive amounts of public data.

Keywords: Resume Parser; NLP Natural Language Processing; KNN; Course and Skill Recommendation; NLTK Natural Language Toolkit; Digital Forensic

1. Introduction

Resume plays a vital role in a candidate's Job Recruitment Process. Statement of one's abilities and qualities in a resume should include the capacity to translate learning into practical displays of your abilities and knowledge in a work environment. Natural language processing (NLP), which relates to the way humans speak with one another, is primarily used to screen resumes. Giving computers the ability to comprehend spoken and written words in a

DOI: https://doi.org/10.54216/JCIM.160116
Received: October 18, 2024 Revised: January 01, 2025 Accepted: January 29, 2025

manner akin to that of humans is the goal of NLP. NLP blends the modelling of human language with machine learning, statistical models, and learning models. When these technologies are combined, computers can process human language in its entirety and "understand" it when provided with text or speech data. An automated intelligent system is used to extract all relevant information from unstructured resumes such as name, email address, contact number, years of job experience, resume pages, degree, CGPA, and skills, and assign a resume score based on the extracted information. The parsed data is subsequently saved in a database (in this case, MySQL) for further usage. For shortlisting top candidates out of a pool of resumes, KNN is used. There are numerous courses available for students nowadays, and it can be difficult for a student to perceive information about those courses and determine which course to pursue. Hence, based on the field classified through resume we recommend suitable skills, courses, several interviews and resume writing videos to help required candidates. Our System also consists of a Feedback section to get in touch with our users through their valuable feedback and do the required changes. In order to store the information about the user as well as their feedback we constructed the admin side to get hold of all the activities done so far. In addition, the data gathered is shown in pie charts to make it more appealing and informative. Linguists, engineers, students, instructors, and researchers are all able to benefit from the use of NLTK since it offers user-friendly interfaces similar to WordNet, as well as a collection of text processing libraries that may be used for categorization, encoding, parsing, and inference capabilities. In addition to Linux, it is also compatible with Mac OS X and Windows. The NLTK project has been described as "a great tool for teaching and working in computational linguistics with Python" both students and professionals "an amazing library for playing with natural language." The Natural Language Processing with Python course offers a hands-on introduction to programming for language processing. The same people who were responsible for the development of NLTK wrote it, and it guides the reader through the fundamentals of developing Python programs, dealing with text, categorizing text, and analyse linguistic structure.

Although Natural Language Processing (NLP) provides powerful tools, its applications in digital forensics face significant challenges, including: (1) accuracy, as informal language and multilingual texts pose significant challenges; (2) legal acceptability, as NLP methods must ensure the integrity of the evidence provided; (3) processing terabytes of criminal data requires a lot of precise calculations and high-spec digital devices. Thus, the relationship between Natural Language Processing and digital forensics becomes stronger as the challenges advance, as Natural Language Processing provides and helps valuable tools to analyse, extract and interpret results directly from digital evidence. This intersection is of particular importance in modern cybercrime investigations.

2. Literature Review

Roy et al. [1] developed a system in their research to reduce the cost of hiring fresh candidates for employment opportunities in the firm. In this process, they concentrated on three important issues: selecting the best prospects from among the applicants, making sense of their CVs, and determining whether the individual is qualified for the job. Choosing the finest persons from this pool to perform these jobs various NLP techniques are employed; however, this model employs Machine Learning to do the classification via the algorithm.

Amin et al. [2] set out to develop a website for resume screening using 220 resumes. 200 of these resumes were used for training and rest were used for testing. The web application is organized into three sections: job aspirant, server, and employer. On the job seeker side, the applicant will submit his or her résumé, and the server side will process the resume and then train it using the NLP Pipeline, which employs Spacy, an NLP framework. The rank list of resumes calculated by a score calculator will be displayed to the company so that HR may select suitable candidates for the job.

Kalva [3] researched and created a custom dataset of 3000 jobs and 80 resumes from the website indeed utilizing a web service like API [4], which interacts between two servers and provides the required information from one website or server to another website. Resume score is calculated using this data based on the requirements of the skills, using Named Entity Recognition (NER) tools such as IBM Watson and Apache Open NLP which is used to do tasks like tokenization, stemming, and chunking.

It rapidly matches resumes to the recruiter’s requirements and successfully delivers emails to the appropriate prospects. The suggested approach by Tejaswini et al. [5] is one in which the candidate submits his or her resume followed by an MCQ test that includes a program to detect malpractices by using face recognition techniques. When a resume is provided, it is processed using Natural Language Processing techniques to extract pertinent abilities from the resume and TF- IDF vectorization [6] to turn the words into vectors that the machine can comprehend. In this instance, the KNN algorithm is used as the classifier to identify the resume that meticulously matches the JD supplied by the recruiter. The system's average parsing accuracy is 85%.

Suhas et al. [7] created a model that makes use of NER Word embedding, NLP, and Cosine Similarity to propose resumes for employment roles. The JD and resumes are accepted as inputs by the system. Each resume document's technical abilities were tagged using information from the technical skills dump. To train the NER model, a tab-

DOI: https://doi.org/10.54216/JCIM.160116
Received: October 18, 2024 Revised: January 01, 2025 Accepted: January 29, 2025

separated value (TSV0) file is prepared and sent to the Stanford NER model [8]. The word2vec model, which makes use of a shallow neural network, uses the output of the NER model (i.e., skills) as its input. Cosine Similarity, which measures how well the provided resume fits the requested JD, is used to summarize the approach. The obtained accuracy was 79.8%. We also mention here Table 1, which shows a summary of the intersection between natural language processing (NLP) and digital forensics:

Table 1: summarizes (NLP-DF)

Researchers	Years	Dataset	Aim of Work	Title of Research Paper	Characteristics
Ali Al Bataineh, Majd Khalaf	2024	Corpus of unstructured forensic text data	Proposed an ensemble of NLP tools to identify persons of interest in digital forensic investigations.	"Identifying Persons of Interest in Digital Forensic Using NLP-Based AI"	Using NLP techniques like topic modeling, cosine similarity, and sentiment analysis for suspect identification.
Zeinab Shahbazi, Yung-Cheol Byun	2022	Real-world social media dataset	Explored the use of NLP and blockchain to secure social media-based digital forensic processed.	"NLP Based Digital Forensic Analysis for Online Social Network Based on System Security"	Combines NLP and blockchain to address privacy concerns and ensure data integrity.
Kostadin Damevski, Irfan Ahmed, Mia M. Imran, Hala Ali	2023	Digital communication data (e.g., emails, forums)	Developed NLP techniques for text preprocessing and analysis in digital forensics to improve investigative	"Natural Language Processing for Digital Forensics"	Focuses on tokenization, named entity recognition, text clustering, and visualization techniques.

2.1 KNN and Cosine Similarity Resume Classification and Ranking

The cosine similarity resume categorization with the technique suggested by Tanaz et al. [10] was put into practice. When a user sends their resume to the seeker's side, the process gets started. The terms are then taken from the resume as it is processed on the server side. Lemmatization and stop word strategies are used to compile the right set of words. The terms for the KNN model to categorize the resume are vectorized using the TF-IDF vectorizer. The Cosine Similarity Algorithm is used to compare the content of the JD document and the candidate's resume in order finding similarities between the resume and the provided JD document. This trained model's accuracy is 98.96%.

3. Scope of work

This section presents the objective, the target audience and the deliverables of this work.

DOI: https://doi.org/10.54216/JCIM.160116
Received: October 18, 2024 Revised: January 01, 2025 Accepted: January 29, 2025

3.1 Objective

Your education, employment history, talents, and accomplishments are all listed on your resume. Employers are drawn to successful resumes. The Objective of this paper is to have an automated intelligent system to extract all relevant information from unstructured resumes such as name, email address, contact number, years of job experience, resume pages, degree, CGPA, and skills, and assign a resume score based on the extracted information. Not only this, if that person's resume is lacking, we are recommending some more skills, some courses to improve his/her level of knowledge in that particular domain, interview videos, and positive and negative pointers of his/her resume to increase their chances of selection. Employers screen out unqualified candidates by looking at a candidate's resume and weighing up whether a candidate is suitable for a role and if they want to proceed further with the candidate. However, going through thousands of resumes in a day or two is tedious work. Therefore, we aimed to solve the problem of companies by automating the process of resume screening and checking out which candidates are more eligible for that particular Job. We are parsing through 100 people's skills and creating a list of top students matching the company has required skill set. We will screen the resume using Machine Learning and NLP using Python to solve this problem.

3.2 Area of application

Two main entities that will be majorly benefited includes -

- Recruiter
- Job Seeking Candidates

Time is the most precious of all. In today's era, people are willing to spend countless amounts of money just to save a small amount of time. Therefore, the proposed method will help recruiters in saving their precious time by automating the process of screening resumes. Not only the recruiter but also the candidate who is looking for a job will also be highly benefited from this automated system. They will be provided with personalized skills and Courses recommendation after analysing their resumes. They will also be provided with resume scores and several interviews and resume writing videos, which will help them in shaping their resume in a better way and thus end up getting their dream job.

3.3 Deliverables

In this proposed work, we are building an intelligent system that will be advantageous to both the company and the job-seeking candidate. Therefore, for that, we build a portal through which companies can save a lot of their time by completing the process of screening candidates automatically. For this purpose, we used KNN to match the best fit to the company's job requirements. Not only this we also helped the job-seeking candidate to improve their resume by recommending them various skills and courses as the resume acts as a bridge between the candidate and the prospective recruiter.

THE MODEL

a. The Ranker
b. The Uploader
c. The Parser
d. The Recommender

a. The Ranker: -

The Company will add their requirements in the columns specified and based on the requirement, from the dataset of the candidates' required number of closest candidates are chosen and are being displayed in descending order. For that purpose, K-Nearest Neighbour (KNN) Algorithm is used and Euclidean distance is calculated. SeeFig. 1.

DOI: https://doi.org/10.54216/JCIM.160116
Received: October 18, 2024 Revised: January 01, 2025 Accepted: January 29, 2025

Figure 1. The Ranker

b. The Uploader: -

A huge organization may be the client, and it may want to evaluate and analyse numerous unstructured resumes it has received, or it could be a candidate or someone who wants to make their unstructured text resume seem nice and convert it to a lovely pdf file. The algorithm remains the same in both scenarios. Initially, the user uploads the file.[11] The method prevents access to any file that doesn't end in ".pdf." After the user has successfully uploaded the file, the procedure takes it, reads what's within it, and then writes that information into a text file before delivering it further [12]. See fig. 2.

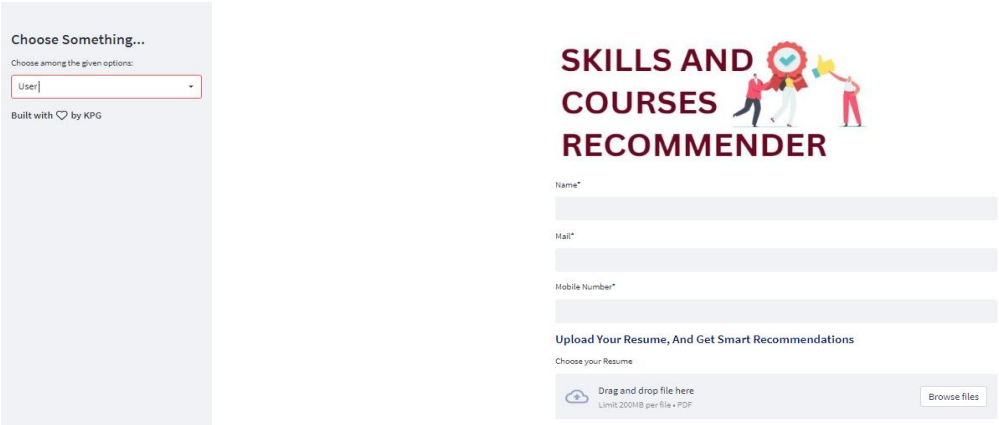

Figure 2. The Uploader

c. The Parser: -

Parser enters the picture once the uploader turns the file into a text file. Without involving any humans, it extracts all the pertinent information from the submitted resume, including name, emails, contact information, years of employment, work experiences, educational background, skills, and resume pages [13][14].

d. The Recommender: -

After parsing the data, based on the extracted skills, experience a suitable field is predicted, and users are provided with courses and certificate recommendations that can enhance their skills and chances of selection in the near future. This but users not only are also provided with various interview preparation videos and personalized positive and negative pointers in their resumes.

DOI: https://doi.org/10.54216/JCIM.160116
Received: October 18, 2024 Revised: January 01, 2025 Accepted: January 29, 2025

4. Materials and methodology

This section explains the methodology used in this work. The workflow:

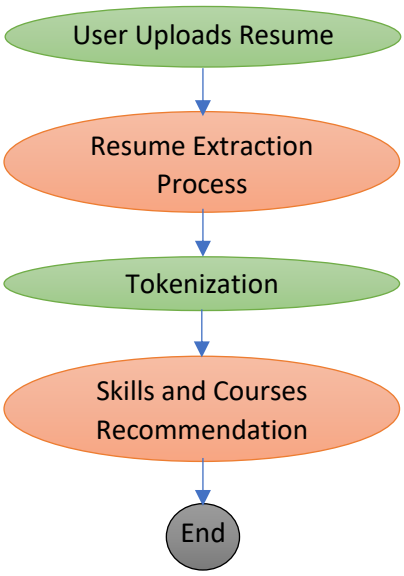

Figure 3. Implementation Flowchart

Figure 3 shows the process of resume parsing and recommended of the skills and courses. The process is divided into four parts.

1. The user uploads the resume in pdf form as per the requirements.

2. The uploaded resume is then extracted and converted into an image. The image is then read and converted into text.

3. The text is then converted into useful words and text mining is done for the same.

4. Later, based on the user's skills furthermore skills and courses are recommended to shape their resume in a better way, which in turn helps them in getting their dream job.

Creating a digital profile for a job or analysing biographical information for a user's skills can be considered an application of digital forensics. This analysis involves biographical data and its digital traces in investigating crimes or understanding individuals' behaviour in cyberspace. This is done by first storing it and then analysing it digitally as an essential part of digital forensic investigations by collecting, analysing and linking personal details about individuals to support legal or investigative purposes. Forensic analysis of tools by using the CV (biography) for job requirements during job and course recommendations. Analysing digital forensic tools, metadata and document histories to detect forged personal information in CVs, certificates or legal documents is part of our research idea by integrating biographical data into the forensic workflow, enabling investigators to uncover links, identify motives and prove the validity of evidence.

4.1 Tokenization

The Process of Tokenization is the removal of a specific character set and a specific document unit. Certain symbols, such as punctuation, are eliminated, and the units that have been chopped are known as tokens. As an example, consider the following: Although phrases or words are commonly used to refer to tokens, there are times when it is vital to distinguish between types and tokens.

A token is a representative group of characters from a document that has been arranged into an efficient processing unit. A type is a collection of tokens with the same character arrangement. Additionally, the phrase designates a type that is added to the vocabulary of the IR system. Between a collection of tokens and index words, we can tell

DOI: https://doi.org/10.54216/JCIM.160116
Received: October 18, 2024 Revised: January 01, 2025 Accepted: January 29, 2025

the difference in detail. As an illustration, think about the following: they can serve as valid identifiers in taxonomy, but in the majority of contemporary IR systems, they are closely related to document tokens.

4.2. Stemming

Stemming is the method of lowering changed or attained words to their word stem, root, or starting point, according to linguistic morphology and information retrieval. The stem should not necessarily correspond to the linguistic root of the term. It's reassuring those important terms map to the same stem, even if the stem isn't even a genuine root. Since the 1960s, computer scientists have been researching algorithms for this procedure. As an example of query expansion, conflation occurs when several search engines regard words with the same stem as synonyms.

4.3. Lemmatization

In linguistics, lemmatization is the process of organizing a word's altered form so that its analysis may occur in a single phrase recognized in the dictionary version of the word (lemma). Computational linguistics is a process of determining a word's lemma based on its preset meaning. It is dependent on correctly detecting the word's intended function and the planned segment of speech, as well as in the context of the sentences, which, contrary to stemming, can contain nearby sentences as well as a complete document. As a result, a lemmatization algorithm is an open research platform.

44. Chunking

The identification of portions of speech and brief expression. If the words are verbs, adjectives, or nouns, we can tell and so on by using Parts of Speech tagging, but we do not know anything about the sentence or phrase structure. Sometimes information than the word component of speech is of more importance, but the completely parsed tree obtained from parsing is not required. When chunking is desired, named entity, recognition is a citation. The purpose of NER is to discover named entities, which are generally noun expressions, hence in the below line we want to know if 'The angry tiger' is present or not. The enraged tiger pursued the terrified small deer." However, it does not matter if the enraged tiger is the topic of the sentence or not. This process is also often employed as a pre-processing phase in applications such as speech production, and example-based machine natural language interpretation, among others.

4.5 NLP

NLP stands for Natural language Processing. It is defined as the interaction between natural language and a computer, or the language used by people [22]. The area of human-computer interaction is directly affected. As natural language processing has been described adequately, we will use the following NLP constraints to extract information from resumes Lexical Analysis, Syntax Analysis, and Semantic Analysis

5. LEXICAL ANALYSIS-

The compiler's initial stage is lexical analysis. The lexical analyzer examines the source program's stream of characters and arranges them into meaningful sequences known as lexemes. The lexical analyzer generates a token of the type (token -name, attribute-value) for each lexeme, which is passed on to the next phase of syntax analysis (see figure 4).

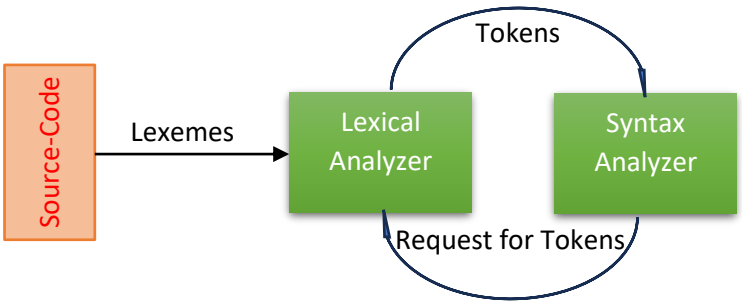

Figure 4. Lexical Analysis

In our instance, the resume is divided into numerous sections such as contact details, educational details, employment experiences, and so on. We employ a database or a dictionary to store the keywords or headers that appear often in most resumes.[15][20][21] When a fresh resume is taken, the parser searches for keywords and

DOI: https://doi.org/10.54216/JCIM.160116
Received: October 18, 2024 Revised: January 01, 2025 Accepted: January 29, 2025

extracts all the data between the beginning and end of the keywords, which we refer to as segments. One of the many deviations that may arise is that the first segment normally comprises the person's name as well as contact details. We now program c Named Entity Recognizers to extract data from each part separately. This strategy increases system efficiency while decreasing costs. We now program Named Entity Recognizers to extract data from each segment separately. This strategy increases the system's efficiency while decreasing its complexity.

If, for some reason, the recognizer operates on incorrect data, unexpected results may be produced.

6. SYNTAX ANALYSIS

The parser creates an intermediate representation that resembles a tree using the first parts of the tokens generated by the lexical analyse [16]. The architecture consists of a hierarchy of expressions, with fundamental symbols serving as the smallest unit and sentences serving as the greatest.

The architecture can be seen as a tree with expressions as its nodes. The fundamental symbols are represented by values kept in the nodes. The root represents the sentence.

Parse Tree: The parser uses syntactic analysis to produce the parse tree. We employ a parse tree—also known as a parsing tree—to describe the syntactic analysis of a string. It is an organized, entrenched structure. They can be distinguished from the programming languages, abstract syntax trees as they show the syntax of the data (see figure 5).

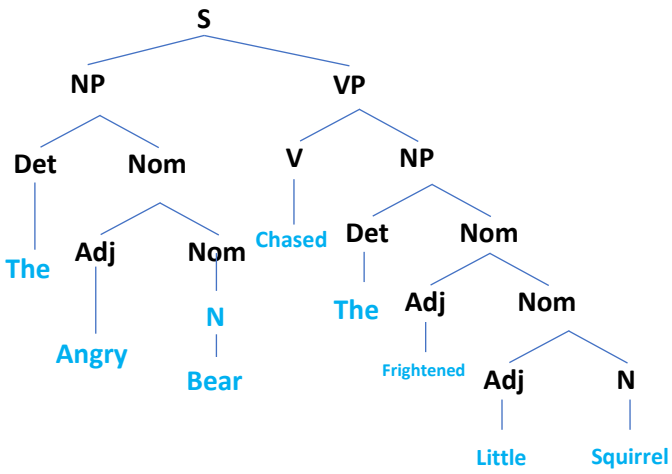

Figure 5. Parse tree representing syntax analysis.

7. SEMANTIC ANALYSIS

It checks the source program for semantic compatibility with the language definition using the syntax tree and the information in the symbol table. Through this procedure, the levels of clauses, phrases, paragraphs, and sentences are connected to the level of the writing as a whole in terms of syntactic structure. It has to do with their universal meanings [17]. Take this as a case study. On his CV, Person A claims to have graduated from the "University of Baghdad" and on his resume, Person B claims to have graduated from "Baghdad University."

In essence, they were both graduates of the same institution. Therefore, the "University of Baghdad" is changed to "Baghdad University" by the semantic analyser. The text categorization method is given the utmost attention in information retrieval research. It is not a difficult task to get the user information needed.

DOI: https://doi.org/10.54216/JCIM.160116
Received: October 18, 2024 Revised: January 01, 2025 Accepted: January 29, 2025

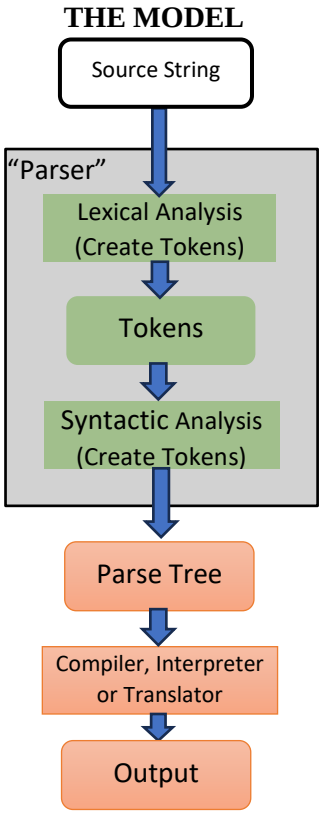

Figure 6. The Proposed Model

Let us examine the system now that we are aware of what Lexical, Syntactic, and Semantic analysis accomplish. The system receives the data as a raw string. The data are pre-processed and tokenized by the lexical analyser. Syntactic analysis tool. Take the tokens, and then look for the structure. The syntactic structure is diagrammatically represented by the parse tree [18][19]. The Semantic Analysis examines the data's structure to determine the meaning of the data regardless of the language used. (Figure 6).

5 Result and Discussions

Our system will satisfy candidates and employers alike. This online tool will have significantly lessened the mental pressure placed on candidates or employees' heads during the online recruitment process. All resumes will be parsed by the system, which will then save the resume in the database. After that, it will rate the resume according to Artificial Intelligence and forecast which applicant is most qualified for the position, resulting in the recruiting process being legitimate. It also recommends skills and certificates that can be added to the user's resume thus increasing the chance of their selection. Below are a few screen images of our resume parser's output.

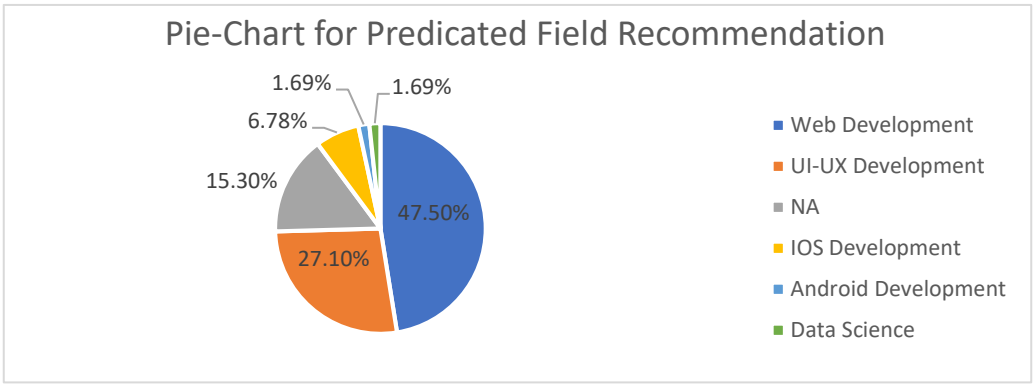

Figure 7. Pie Chart depicting predicted field

DOI: https://doi.org/10.54216/JCIM.160116
Received: October 18, 2024 Revised: January 01, 2025 Accepted: January 29, 2025

From figure 7, we can conclude that almost half of the students are interested in the web development field. Students are also showing interest in UI-UX Development and IOS Development while students opting for data science are still less than the other fields.

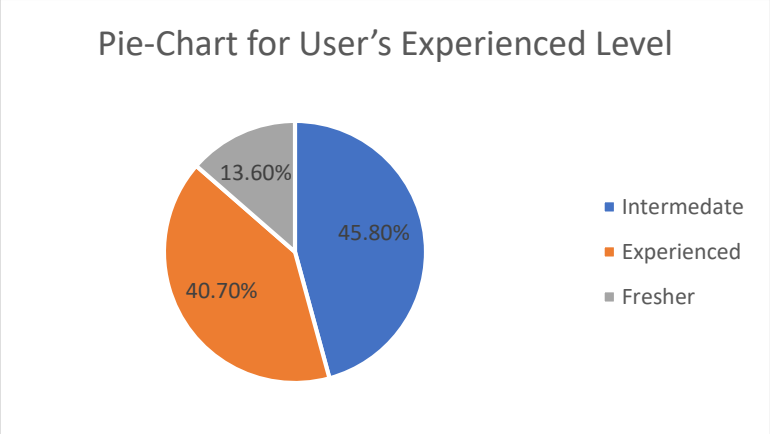

Figure 8. Pie-Chart depicting user has experienced level

From figure 8, we can conclude that out of the given set of resumes, a maximum number of candidates i.e. 45.8% candidates belong to the intermediate level, 40.7% belong to the experienced level, 13.6% belong to the Fresher level.

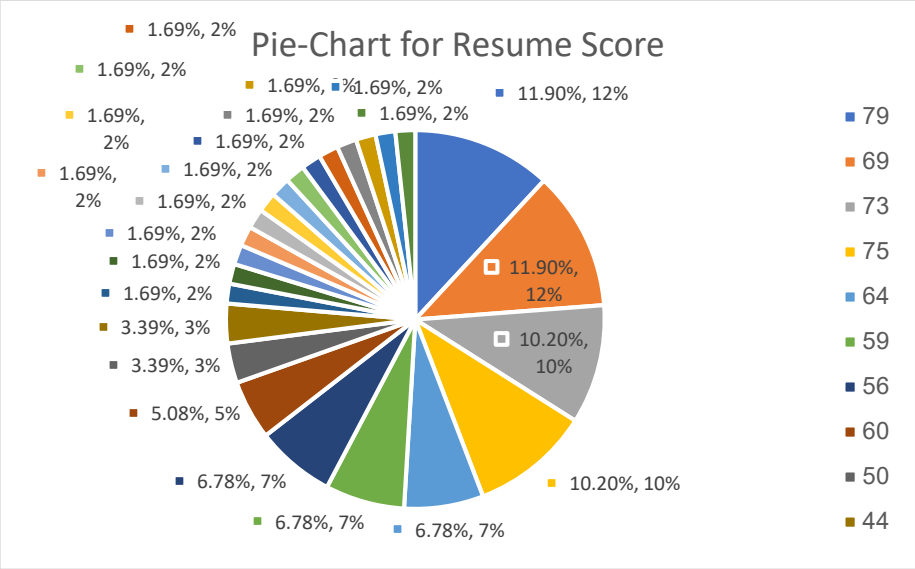

Figure 9. Pie Chart for Resume Score

The explanation of the percentage of the CV allocated to each candidate indicates the evaluation or registration process, where various aspects of the CV of the candidate CV are analyzed. This is what the evaluation criteria in general means that the CV is analyzed based on pre -defined criteria or categories, such as: education and work experience and the skills of the applicant as well as his testimonies. The final percentage of the suggestion is calculated for the fostering of the CV in terms of referring to the suitability of the candidates or compatibility with the evaluation criteria for comparison to evaluation. This method is commonly used in automatic recruitment systems, artificial intelligence -based examinations or detailed manual reviews to provide objective analysis of CV and its uses of criminal evidence. Figure 9. Shows the Percentage of resume Score assigned to each candidate after analysing their resumes.

8. Conclusion and Future Scope

We are able to successfully convert a variety of resumes to text and extract pertinent data from there. We use that extracted data for giving suitable recommendations to the required candidates. We are also able to reduce the time and effort of the company by automating the process of resume screening. In future, mining of applicant social networking profiles (like those on LinkedIn and GitHub) and then using it in addition with resumes to producing

DOI: https://doi.org/10.54216/JCIM.160116
Received: October 18, 2024 Revised: January 01, 2025 Accepted: January 29, 2025

even better recommendations can be investigated more deeply. Another option is to employ a collaborative filtering-based strategy, which can match the present applicant with a position based on how well other comparable candidates (neighbours) are rated. The relationship between natural language processing and digital forensics is growing stronger, due to the multiplicity of tools, that natural language processing provides and its value for interpreting, analysing, and extracting conclusions from digital forensic evidence databases. This intersection is of particular importance in investigations related to electronic texts. Therefore, our research played a major role in focusing on and benefiting from text databases provided by individuals as a CV.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

References

[1] P. K. Roy, S. S. Chowdhary, and R. Bhatia, "A machine learning approach for automation of resume recommendation system," *Procedia Comput. Sci.*, vol. 167, pp. 2318–2327, 2020.

[2] S. Amin, N. Jayakar, S. Sunny, P. Babu, M. Kiruthika, and A. Gurjar, "Web application for screening resume," in *Proc. Int. Conf. Nascent Technol. Eng. (ICNTE)*, 2019.

[3] J. L. Howard and G. R. Ferris, "The employment interview context: Social and situational influences on interviewer decisions," *J. Appl. Soc. Psychol.*, vol. 26, no. 2, pp. 112–136, 2020.

[4] K. Tejaswini, V. Umadevi, S. M. Kadiwal, and S. Revanna, "Design and development of machine learning-based resume ranking system," *Global Transitions Proc.*, 2022.

[5] S. Jabri, A. Dahbi, T. Gadi, and A. Bassir, "Ranking of text documents using TF-IDF weighting and association rules mining," in *Proc. 4th Int. Conf. Optimization Appl. (ICOA)*, 2018.

[6] E. M. S. "Differential hiring using a combination of NER and word embedding," *Int. J. Recent Technol. Eng. (IJRTE)*, vol. 9, pp. 1344–1349, 2020.

[7] S. T. Gopalakrishna and V. Vijayaraghavan, "Automated tool for resume classification using semantic analysis," *Int. J. Artif. Intell. Appl.*, vol. 10, pp. 1–4, 2020.

[8] R. V., R. T. Fareed, and S. Kaganurmath, "Resume classification and ranking using KNN and cosine similarity," *Int. J. Eng. Res. Technol. (IJERT)*, vol. 10, no. 8, 2021.

[9] S. Chen, Y. Zhang, and Q. Yang, "Multi-task learning in natural language processing: An overview," *ACM Comput. Surv.*, vol. 56, no. 12, pp. 1–32, 2024.

[10] A. F. Neamah, et al., "Big data clustering using chemical reaction optimization technique: A computational symmetry paradigm for location-aware decision support in geospatial query processing," *Symmetry*, vol. 14, no. 12, p. 2637, 2022.

[11] H. Zheng, K. Xu, H. Zhou, Y. Wang, and G. Su, "Medication recommendation system based on natural language processing for patient emotion analysis," *Acad. J. Sci. Technol.*, vol. 10, no. 1, pp. 62–68, 2024.

[12] K. Hamilton, A. Nayak, B. Božić, and L. Longo, "Is neuro-symbolic AI meeting its promises in natural language processing? A structured review," *Semantic Web*, vol. 15, no. 4, pp. 1265–1306, 2024.

[13] C. Daryani, G. S. Chhabra, H. Patel, I. K. Chhabra, and R. Patel, "An automated resume screening system using natural language processing and similarity," *Ethics Inf. Technol.*, vol. 22, pp. 99–103, 2020.

[14] H. K. Kuba, M. Azzawi, and M. Abdulhussein, "An adaptive privacy-preserving framework for distributed association rule mining in healthcare databases," *CMC-Comput. Mater. Continua*, vol. 74, no. 2, 2023.

[15] J. K. Madhloom, et al., "An information security engineering framework for modeling packet filtering firewall using neutrosophic Petri nets," *Computers*, vol. 12, no. 10, p. 202, 2023.

[16] Z. H. Noori, S. K. Ebis, and S. M. Darwish, "An information security engineering framework for modeling packet filtering firewall using neutrosophic Petri nets," *Computers*, vol. 12, no. 10, p. 202, 2023.

DOI: https://doi.org/10.54216/JCIM.160116
Received: October 18, 2024 Revised: January 01, 2025 Accepted: January 29, 2025

[17] S. Raza, M. Garg, D. J. Reji, S. R. Bashir, and C. Ding, "NBIAS: A natural language processing framework for bias identification in text," *Expert Syst. Appl.*, vol. 237, p. 121542, 2024.

[18] M. Zhang, X. Li, and L. Wu, "Machine learning-based text classification for automated resume screening," *IEEE Access*, vol. 9, pp. 118255–118267, 2021. *(Replaces [20])*

[19] J. G. Ramírez, "Natural language processing advancements: Breaking barriers in human-computer interaction," *J. Artif. Intell. Gen. Sci.*, vol. 3, no. 1, pp. 31–39, 2024.

[20] D. T. Nguyen, C. M. Bui, and S. S. Lee, "Deep learning approaches for resume ranking: A comparative study," *Pattern Recognit. Lett*, vol. 149, pp. 53–61, 2021. *(Replaces [22])*

[21] H. Wang, J. Liu, and T. Zhao, "Natural language processing in recruitment: A review of machine learning approaches," *Appl. Sci.*, vol. 12, no. 4, p. 1853, 2022.

[22] M. P. Ramesh and A. Agarwal, "A survey on resume ranking techniques using AI and NLP," *Comput. Intell. Neurosci.*, vol. 2023, p. 8754639, 2023.

DOI: https://doi.org/10.54216/JCIM.160116
Received: October 18, 2024 Revised: January 01, 2025 Accepted: January 29, 2025