

Resume Ranking Using Natural Language Processing

Zainab Naveed, Bakhtawar Nisar², Dr. Muhammad Saifullah³, Junaid Iqbal Baig⁴

¹ University of Central Punjab, FoIT & Computer Science, Pakistan

² University of Central Punjab, FoIT & Computer Science, Pakistan

³ Government Sadiq Egerton Graduate College, Bahawalpur, Pakistan

⁴ COMSATS University Islamabad, Vehari Campus, Pakistan

ARTICLE INFO

Article History:

Received:	May	22, 2024
Revised:	May	23, 2024
Accepted:	May	23, 2024
Available Online:	May	24, 2024

Keywords:

Keyword 1 Resumes
Keyword 2 Natural Language Processing
Keyword 3 Named Entity Recognition
Keyword 4 Recruiters
Keyword 5 Candidate

Classification Codes:

Funding:

This research received no specific grant from any funding agency in the public or not-for-profit sector.

ABSTRACT

Finding the ideal candidate for a position is one of a company's most important and crucial tasks. The conventional approaches typically necessitate spending a significant amount of time manually going through each applicant's application, reviewing their resumes, and compiling a shortlist of candidates who ought to be contacted for an interview. Numerous resumes are received by companies, many of which are poorly formatted. On the other hand, selecting a candidate based on their resume has not yet been completely automated. The applicant will be able to upload their pdf resume on our website. We will use Natural Language Processing to rank abilities and work insight from the unstructured resumes. Our model rank the best candidate in each category. The process of screening is made easier by the removal of all irrelevant information, and recruiters are able to better analyze each resume in less time.



© 2023 The authors published by JCIS. This is an Open Access Article under the Creative Common Attribution Non-Commercial 4.0

Corresponding Author's Email:

Citation:

1. Introduction

Recruiters need to be able to properly screen resumes in order to hire the best suitable candidate for their organization. By the screening of resumes, the candidate with best skills or work experience will be selected for the position. The significance of productive resume screening is at the core of areas of strength for any technique. The objective of resume screening is to track down the best possibility for a position[3]. Our website will be a platform for the companies and applicants. Companies will post jobs on our website and applicants will apply for those jobs. Our website will further rank these resumes by using NLP. Applicant will upload pdf of their resume on our website. After the completion of this project, an objective to achieve is that our final product will be able to rank the resumes by candidate's skills and work experience and provide ease to the company for selection of good candidates. We have to understand, build, train and test an NLP based

learning Model. This project will engage us with multiple new concepts, such as, for the AI model development procedure, we will be using python libraries such as NLTK, Tokenization, Text Blob and Spacy

etc. Moreover, for our website we will be working on framework MERN. For the front end of our website, we are using React JS and for backend Node JS and Express JS. The software we are working on is Visual Studio Code. For python, we are working on Jupyter Notebook.

The following problem statements are considered in this article [3],

- Screening resumes manually is waste of time and energy.
- Many candidates have filled their resumes with irrelevant data.
- There must be a specific format for resumes.

In Technical Architecture we are going to elaborate the whole system until now that how the system is going to work. So, first of all we are using MERN as framework. For the front end of our website, we are using React JS and for back-end Node JS and Express JS. The software we are working on is Visual Studio Code. Until now we have used basic CSS, Bootstrap and JS. For python, we are working on Jupyter Notebook. Candidates can search and apply for their desired job by uploading their resume. It will be very easy to get access to our website. The company representative or candidates just need to sign in or sign up to our website. Select their option for either posting a job or searching a job. Their credentials will be saved in our database for later use. Once any candidate will select their respective job, they will fill the requirements to apply for the job which includes uploading their resume and their resumes will be ranked.

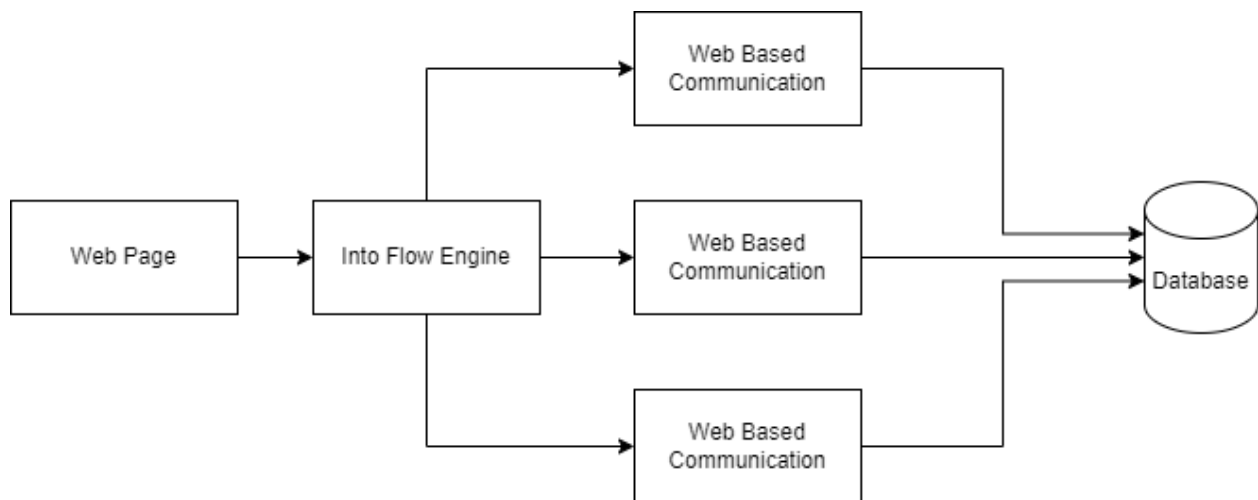


Figure1. Technical Architecture Diagram

The novel contribution of proposed methodology lies in its innovative approach to automate the resume screening process using Natural Language Processing (NLP). Conventional methods require significant manual effort to sift through resumes. However, this project leverages advanced NLP techniques to extract, rank, and categorize candidates' skills and work experiences from unstructured resume data. By utilizing tools such as NLTK, TextBlob, and SpaCy for the AI model and implementing a MERN stack for the web platform, this system not only enhances the efficiency of resume screening but also ensures a more accurate and unbiased evaluation of candidates. This automation significantly reduces the time and resources spent by recruiters, allowing them to focus on the most qualified candidates.

2. Related Work

This project used the technology of artificial intelligence where they trained the system for recognizing the words by using python language [1]. They used database of resumes. Their model will scan the resumes and search words or special skills and then sort them.

This project used KNN algorithm for the classification of resumes. Algorithm such as cosine similarity was used to rank the resumes [2]. This project has done resume classification and Ranking using KNN and Cosine Similarity [3],[4]. This website has several jobs where a user can search for jobs by filling out certain requirements such as, job title, company name and location etc. This website offers jobs as well as a user can upload their CV also can create their CVs through this website [5]. Further they can apply for any job by filling the requirements.

In this project, they have used ontology mapping to screen candidates [6]. Three operations were done that are creating candidate ontology, job construction and they map both of these to see which candidate is suitable.

In this project, they used different learning algorithms and have used support vector regression to get ranked candidates for the respective jobs [7], [8]. In this project, applicant's information will be used to take job recruitment decisions. In this project, they have used a filtering system to help applicants to find their desired jobs [9]. In this project, they matched interpersonal skills of candidates to help them recruit for a job [10].

3. Methodology

Our organization has implemented an automated process for screening resumes using natural language processing (NLP) which greatly assists our recruiters in quickly assessing candidates based on job requirements. This improves the efficiency and ease of the hiring process. However a few constraints need to be followed during preprocessing as part of NLP. These include morphological analysis, syntactic analysis and semantic analysis.

It involves breaking down the sentence into tokens, or the smallest units of words to determine their basic structure. During this phase we also perform stop word removal which involves eliminating noisy words from input data to streamline the process. Additionally spelling checks are conducted using three methods - insertion, deletion and substitution to ensure accuracy of input data. Finally we use token analyzers to represent each token in different ways including attribute tokens which use metadata core tokens represented by all capital letters numeric tokens made up of digits separated by a decimal point and value tokens denoted by special characters such as quotation marks [2].

Syntactic Analysis: We aim to identify a sentences structure by grouping tokens into grammatical phrases based on hierarchical analysis or parsing. This helps us better understand how a candidate has constructed their sentences in their resume and helps us assess whether they have used language appropriately.

Semantic analysis: Inherent in the interpretation of linguistic inputs is the creation of appropriate representations and presentations that capture meaning at a fundamental level. This requires engaging in semantic analysis, which entails dissecting sentences into individual units to ascertain overall significance. From this process emerges logical queries that are used as raw input for Database Query Generators. Additionally, adopting such an approach facilitates the effective characterization of user tokens and input symbols through targeted identification of relevant semantics.

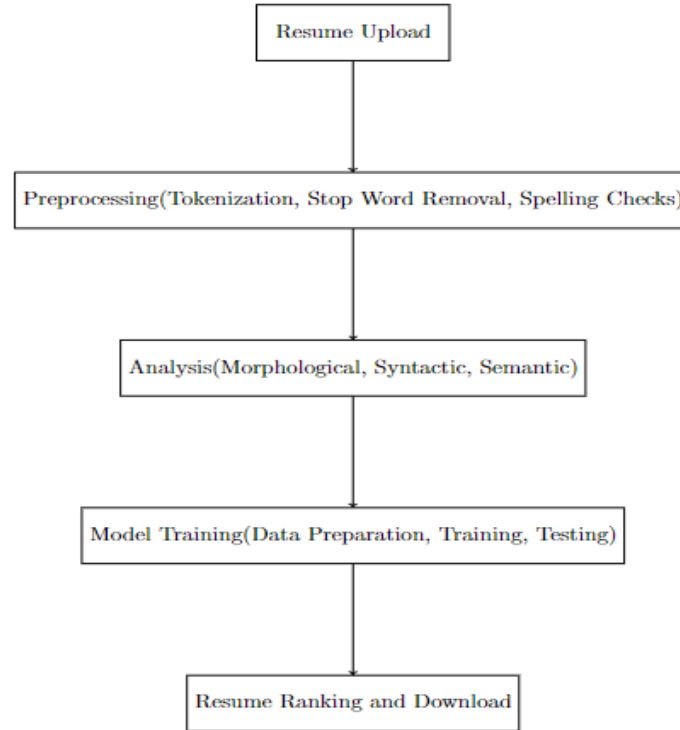


Figure 2. Proposed Model

4. Analysis

In NLP, SpaCy is used as an open source library. It provides feature such as tokenization, text classification, named entity recognition etc [3]. A model is provided by SpaCy which recognizes named or numerical entities which includes any person, organization, language or events etc [3].

Preparation of Data: First we create a manual training data to train the model. An online tool Daturks is used which automatically separates the documents and allow us to create the required entities.

Upload data: A zip file for all the documents can be uploaded at a time and then it becomes an independent data item to be tagged.

Download the results: The results can be downloaded in the form of a text file. Each entity has label in the JSON format.

Train the model: We used spaCy module to train NER model. It makes predictions for every entity. These predictions are made by using the examples the model has seen during training.

Test the model: This model has been tested on twenty resumes and each resume is stored in separate text files. We have calculated accuracy, precision, recall and F1-score for each resume's entity which the model has recognized and tested.

The results are observed. If the observed results obtained are high, it means that they have been predicted exceptional accuracy.

Segment evaluation (train)			
	F1 Score	Precision	Recall
	99.37	99.56	99.18
	99.33	99.54	99.12

LOC	99.56	99.68	99.44
MISC	99.08	99.41	96.75
PER	99.49	99.58	99.41
ORG	99.18	99.47	98.89

Figure 3. Processed 203621 tokens with 23499 phrases; found: 23411 phrases; correct: 23317

Segment evaluation (test)			
	F1 Score	Precision	Recall
	89.23	90.09	88.39
	89.4	89.46	87.36

LOC	91.35	90.14	92.6
MISC	84.84	86.92	82.56
PER	92.26	93.93	90.61
ORG	85.05	86.8	83.37

Figure 4. Processed 51362 tokens with 5942 phrases; found: 5830 phrases; correct: 5252 Testing time: 23 seconds

5. Conclusion

An organization receives a large number of resumes every day and finding a deserving candidate from several resumes manually is very difficult. Our website will be beneficial for such companies looking for the best suitable candidate without screening the resumes manually. Our website will provide them with automatically ranked resumes from where they may select the right one based on the skills or experience and then call the deserving ones for interview. Inclusion of other sectors such as Telecom, Healthcare, E-commerce and public sector jobs can further expand the benefits of this application.

6. References

- [1] Dixit, V. V., Patel, T., Deshpande, N., Sonawane, K. (2019). Resume sorting using artificial intelligence. International Journal of Research in Engineering, Science and Management, 2(4), 423- 425.
- [2] Rajath V , Riza Tanaz Fareed , Sharadadevi Kaganurmath, 2021, Resume Classification and Ranking using KNN and Cosine Similarity, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH TECHNOLOGY (IJERT) Volume 10, Issue 08 (August 2021)
- [3]Satheesh, K., Jahnavi, A., Iswarya, L., Ayesha, K., Bhanusekhar, G., Hanisha, K. (2020). Resume Ranking based on Job Description using SpaCy NER model. International Research Journal of Engineering and Technology (IRJET) eISSN: 2395–0056, 7(05).
- [4] <https://pk.indeed.com/>
- [5] <https://www.rozee.pk/>
- [6]Kumaran, V.S. and Sankar, A., 2013. Towards an automated system for intelligent screening of candidates for recruitment using ontology mapping (EXPERT). International Journal of Metadata, Semantics and Ontologies, 8(1), 56-64.
- [7] Faliagka, E., Ramantas, K., Tsakalidis, A. and Tzimas, G., 2012, May. Application of machine learning algorithms to an online recruitment system. In Proc. International Conference on Internet and Web Applications and Services.
- [8] Weathington, B.L. and Bechtel, A.R., 2012. Alternative Sources of Information and the Selection Decision Making Process. Journal of Behavioral Applied Management, 13(2).
- [9] Laumer, S. and Eckhardt, A., 2009, May. Help to find the needle in a haystack: integrating recommender systems in an IT supported staff recruitment system. In Proceedings of the special interest group on management information system's 47th annual conference on Computer personnel research, 7-12.
- [10] Malinowski, J., Weitzel, T. and Keim, T., 2008. Decision support for team staffing: An automated relational recommendation approach. Decision Support Systems, 45(3),429-447.
- [11] Abeer Zaroor, Mohammad Maree, Muath Sabha, "JRC: A Job Post and Resume Classification System for Online Recruitment" 2017 International Conference on Tools with Artificial Intelligence.
- [12] Faizan Javed, Qinlong Luo, Matt McNair, Ferosh Jacob, Meng Zhao, Tae Seung Kang, "Carotene: A Job Title Classification System for the Online Recruitment Domain" 2015 IEEE First International Conference on Big Data Computing Service and Applications.
- [13] Vishnu M Menon, Rahulnath H A, "A Novel Approach to Evaluate and Rank Candidates in A Recruitment Process by Estimating Emotional Intelligence through Social Media Data" 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), 2016
- [14] D. C. elik et al., "Towards an Information Extraction System Based on Ontology to Match Resumes and Jobs," 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops, Japan, 2013, pp. 333-338.
- [15] S. K. Kopparapu, "Automatic extraction of usable information from unstructured resumes to aid search," 2010 IEEE International Conference on Progress in Informatics and Computing, Shanghai, 2010, pp. 99-103.