iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52183

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Job Recommendation System Using NLP

Sneha Kumari[1], Gittika Agarwal[2], Dr. Sheshappa S.N[3], Mr. Vijaykumara.Y.M.[4], Prof. Mr. Byre Gowda B.K[5]
BE Students, Associate Professor, Assistant Professor, Department Information Science and Engineering, SIR M. Visvesvaraya Institute of Technology Hunasamaranahalli, Bengaluru-562157

Abstract: ***Recommender systems is one of the most successful machine learning applications. We are exposed to the products/information recommended by these systems everywhere in our daily life. This project will introduce several recommender systems in NLP, specifically in the domain of online recruitment. I will explain the important ideas behind the recommender systems in our web application, it might be also similar with those in recruiting platforms like LinkedIn, Xing. One of my main jobs is to design and develop recommender systems on the skill-based matching app. Before getting into this specific topic, I would like to go over the general concepts and methods of current recommender system in different domains. We have provided a brief overview of the approaches, techniques, and applications of recommendation systems in this paper. One significant application of recommendation systems is in the field of job recruitment, where candidates are chosen using online job recruitment portals based on their profiles, employment histories, and behavioral tendencies. According to a recent survey, this field has not been thoroughly investigated up to this point, and the current job recommender system has many flaws. They analyze resumes, profiles, and job descriptions, but because of the "cold start" problem, new job postings and candidate profiles are not properly matched. In some cases, potential candidates lose their jobs as a result of inadequate job descriptions and education information in the ontology.***

Keywords: NLP, Gensim, Bag of Words, Naïve Bayes, Word2V3ec , Tableau, Flask

I. INTRODUCTION

Today, social media is a very popular medium to share information and discuss current events. Internet users may access a wealth of information regarding online learning, social user behavior, and commerce thanks to the extensive usage of various internet sources, including mobile phones and smart gadgets. An individual user may experience numerous information overload issues that make it difficult for them to make the best judgments when data volume and diversity rise dramatically. Information overload is the term used for this framework. A novel approach to a recommender system uses visuals to address the issue of information overload for consumers. A recommender system can efficiently identify users' likely needs and select interesting items from a large pool of application data to solve a variety of difficulties. A job recommendation system using NLP (Natural Language Processing) is an automated tool that utilizes various techniques of NLP to analyze job descriptions, user profiles, and other relevant data to provide personalized job recommendations. This system can be used by job seekers to find suitable job opportunities and by employers to identify qualified candidates for open positions.

The NLP-based job recommendation system analyzes the job descriptions and user profiles to identify relevant skills, experience, education, and other relevant information. It then uses this information to match job seekers with suitable job openings and employers with qualified candidates. The system can also provide insights into job trends and requirements, allowing job seekers to better understand the job market and make informed decisions about their career. Overall, an NLP-based job recommendation system has the potential to greatly enhance the job search process, making it more efficient, personalized, and effective for both job seekers and employers.

II. MOTIVATION

Job recommendation using natural language processing (NLP) can be motivated by several factors, including:

1) *Personalization:* NLP can be used to analyse a candidate's resume and job preferences to recommend jobs that match their skills and interests. This personalization can make the job search process more efficient and effective for both the candidate and the employer.
2) *Efficiency:* Job recommendation using NLP can save time and effort by automating the process of filtering through job postings. This can be especially helpful for recruiters who receive hundreds of resumes for a single job opening.
3) *Accuracy:* NLP algorithms can use advanced language processing techniques to accurately identify the skills and experience required for a specific job. This can lead to more accurate job recommendations that better match a candidate's qualifications.

©IJRASET: All Rights are Reserved

4) *Data-driven:* Job recommendation systems using NLP can use data-driven approaches to identify trends and patterns in job postings and candidate resumes. This can help employers and recruiters better understand the job market and make more informed hiring decisions. Overall, job recommendation using NLP can provide a range of benefits for both candidates and employers, including personalization, efficiency, accuracy, and data-driven insights

III. RELATED WORKS

A lot of research work has been carried out in the field of recommender systems. F.O. Isinkaye et al. [1] discuss various aspects of the recommender system. It provides more emphasis on the filtering techniques of the recommender system. It uses statistical accuracy metrics for the filtering techniques, distinguishing actual ratings from predicted ratings. Gediminas Adomavicious et al. [2] present an overview of various domains of recommender systems. It provides brief information on current approaches and techniques of the recommender system including the limitations and extensions. Shuo Yang et al. [3] discuss the various aspects related to content and collaborative filtering techniques.

It combines both techniques to get the hybrid technique. The model is built by using statistical relational learning (SRL) for the representation of probabilistic dependencies. Sidahmed Benabderrahmane et al. [4] proposed a decision-making tool to provide a path to recruiters that helps them get efficient job seekers. It represented a Doc2Vec embedded system which stores the information on the clickstream history of job seekers in the database.

It also uses deep neural networks to predict future clicks using various job board databases. It overcomes the problem of overfitting using dropout layers. Nedra Mellouli et al. [5] in a similar context represented a smart4job recommender system that connects job aspirants with job offers that match their profiles the most. The model works on temporal prediction and domain knowledge analysis. Miao Jiang et al.[6] further improved the model by introducing email mode. In this system, the candidates provide their resumes to get email alerts about new job offers.

But these above models partially explain the context. To resolve this issue, Priscila Valdiviezo Diaz et al. [7] used the Bayesian model for the recommendation process. This model is based on a collaborative filtering method that uses user-item correlation. The collaborative filtering technique builds a database of users' interests. It then matches those relevant interests with other similar users to make an accurate recommendation. This kind of user makes a group consisting of more users. Shabbir Ahmed et al. [8] introduce an online recruitment system that will use by recruiters to get efficient candidates. It uses a hybrid filtering technique between training and testing datasets.

It selects top job offers as the recommended list and makes the selection based on that. Shiqiang Guo et al. proposed a system named as Resume Matcher. It matches the resumes with the corresponding job offers. It uses machine learning techniques to get effective results.

This paper also deals with natural language processing (NLP) for the recommendation process. Although the above models have achieved significant improvements as compared to traditional models using various filtering techniques. But there is a need to recommend socially which includes friends of similar interests. Mamadou Diaby et al. [10] use Facebook and LinkedIn datasets for the implementation of their work.

IV. METHODOLOGY

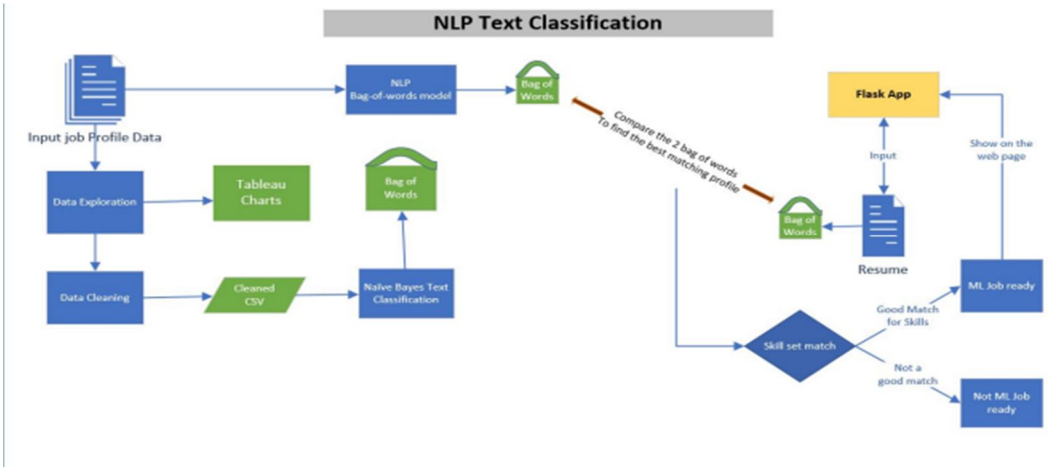
A. *Bags of Words*

Bags of words is a common natural language processing (NLP) technique used for text analysis. The basic idea behind bags of words is to break down a text into individual words and count their frequency. This creates a "bag" of words where each word is treated as an independent feature.

The bags of words approach are used in a wide range of NLP tasks, such as sentiment analysis, text classification, and topic modeling. Here is how the bags of words approach work:

1) Text Preprocessing: The raw text data is preprocessed to remove stop words, punctuation, and other irrelevant information. The remaining text is then converted into lowercase and tokenized into individual words.
2) Feature Extraction: The bags of words approach are used to extract features from the preprocessed text. The system creates a vocabulary of all the unique words present in the dataset and assigns a numerical value to each word based on its frequency in the dataset. This vocabulary serves as a dictionary of features that will be used in the machine learning algorithm.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



1) *Machine Learning Algorithm:* A machine learning algorithm such as Naive Bayes, Support Vector Machines, or Random Forest is trained on the dataset using the extracted features. The algorithm learns to predict the likelihood of a certain outcome based on the features present in the input text.
2) *Text Analysis:* Once the machine learning model is trained, it can be used to analyze new text data. The input text is first preprocessed and then transformed into a bag of words representation using the same vocabulary that was used during training. The machine learning algorithm then makes predictions based on the features present in the input text.

Overall, the bags of words approach is a powerful and widely used technique in NLP for text analysis. Its simplicity and effectiveness make it a popular choice for a variety of applications.

B. Natural Language Processing

In this study, natural language processing plays a vital role. The data collected by web scraping and the user data collected from stack overflow contain descriptive fields. As both sets of data have no history of previous interaction, the study continued with the approach of analyzing the explicit properties of the content. The data set is a file full of text data, which is unstructured, while on the other hand, user data that was collected from stack overflow is structured. The current study needed a method to analyze text data to categorize all the job listings into different categories, and also to find the similarity between the vector of words from the user data and the vector of words from the job listing data. In the current study, the Natural Language Processing package is used instead of NLTK. Both NLTK and SpaCy are popular NLP tools available in python. While both tools can accomplish required NLP tasks, each one excels in particular scenarios. SpaCy is developed from the perspective of developers in a real- time production environment. While NLTK provides all the access to algorithms to complete the task, SpaCy provides accurate syntactical analysis of all NLP libraries available in NLTK with better performance, and it also provides access to the largest word vector corpus called en_vectors_web_lg. While NLTK is not it's not known for robustness and fast results because of its approach to handling inputs and output. On the other hand, SpaCy is completely built on Python from the ground up, which contributes to its fast execution time and better performance (Kakarla, 2019). In this current study, SpaCy is used in extracting Named entities from the job description and used to check the similarity between the user profile. The similarity function in the spaCy package allows us to compare two vectors of words and provide how similar two words are to one another. This similarity function utilizes a word vector, which is represented in a multidimensional representation of the word. These word vectors are generated using a word embedding algorithm.

C. Word2Vec

Word embedding is the method to translate the words or phrases from the corpus into vectors of a real number, as shown in figure 3.2. It's used for language modelling and feature learning methods as well. The corpus here is taken from SpaCy's one of the largest word models, i.e., en_vectors_web_lg. The Word2vec word embedding algorithm takes a large corpus en_vectors_web_lg of text as input and produces a word vector space, which has several hundred dimensions. These models learn based on the two-layered or shallow neural network method to perform the required task. Word vectors that are in the same vector space are bound to be similar and share the same context.

©IJRASET: All Rights are Reserved

This allows us to use this word embedding method in our study to find similarities between two-word vectors. There are two different learning models available in the word2vec algorithm: the continuous bag-of-words (CBOW) model and the continuous skip-gram model. The CBOW model tries to learn word embedding by predicting the current based on surrounding words or context. Whereas, Continuous skip-gram models learn by predicting the context of the surrounding words based on the current word. The advantage of using the word2vec word embedding method is that it can learn efficiently with high-quality word embeddings, which in turn allows us to learn from bigcorpora like en_vectors_web_lg of SpaCy (Brownlee, 2017).

D. Tableau

Tableau is a data visualization and business intelligence software that allows users to connect, visualize, and share data in a way that helps organizations make informed decisions. The software includes a range of tools and features that enable users to create interactive dashboards, charts, and reports from various data sources.

Some of the key features of Tableau include:

1) *Data Connection:* Tableau can connect to a wide range of data sources, including spreadsheets, databases, cloud services, and big data platforms.
2) *Data Visualization:* The software provides a range of tools and options for creating interactive visualizations, including charts, graphs, maps, and dashboards.
3) *Data Analysis:* Tableau allows users to perform ad-hoc analysis on their data, with features such as data blending, calculations, and forecasting.
4) *Collaboration and Sharing:* Tableau allows users to publish and share their visualizations with others, including the ability to embed them in web pages or presentations.
5) *Mobile Access:* Tableau provides mobile access to its dashboards and reports, allowing users to access their data from anywhere.

Overall, Tableau is a powerful tool for businesses and organizations that need to make data-driven decisions. Its intuitive interface and powerful features make it a popular choice for data analysts, business users, and decision-makers**.**

E. Gensim

Gensim is an open-source Python library for natural language processing (NLP) and machine learning. It is designed to handle large, complex datasets and provides a wide range of tools and algorithms for topic modeling, document similarity, text summarization, and other NLP tasks.

Some of the key features of Gensim include:

1) *Topic Modeling:* Gensim provides tools for unsupervised topic modeling, including Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA). These algorithms can help identify latent topics within a collection of documents.
2) *Document Similarity:* Gensim includes algorithms for computing document similarity, such as cosine similarity, Jaccard similarity, and Euclidean distance. These algorithms can be used to find similar documents within a dataset.
3) *Text Summarization:* Gensim provides tools for text summarization, including the TextRank algorithm and the Gensim Summarization module. These algorithms can be used to generate summaries of long documents or articles.
4) *Word Embeddings:* Gensim provides tools for word embeddings, including Word2Vec and Doc2Vec. These algorithms can be used to learn dense, low-dimensional vector representations of words and documents, respectively.
5) *Scalability:* Gensim is designed to handle large, complex datasets and can process data in parallel on multicore machines.

Overall, Gensim is a powerful and flexible tool for NLP and machine learning tasks. Its wide range of tools and algorithms make it a popular choice for researchers, data scientists, and developers working with text data.

V. PROPOSED WORK

In our proposed system we have improvised and modified the recommendation systems. This Employment Recommendation System has considered many parameters like ratings, experience, location etc. For implementing algorithm, python is used along with Naïve Baiyes algorithm and some machine learning libraries. Machine learning has been improvising the recommendation systems, also it brings more possibilities to improve performance of recommendation system. In future work we will develop android application and provide multiple language facility for workers for searching jobs in regional language. We will also add ratings feature at workers end so contractor can easily finalize workers based on the ratings and feedback. Our system is useful for government sites as they need worker.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

A. Proposed Architecture and System Flow Diagram

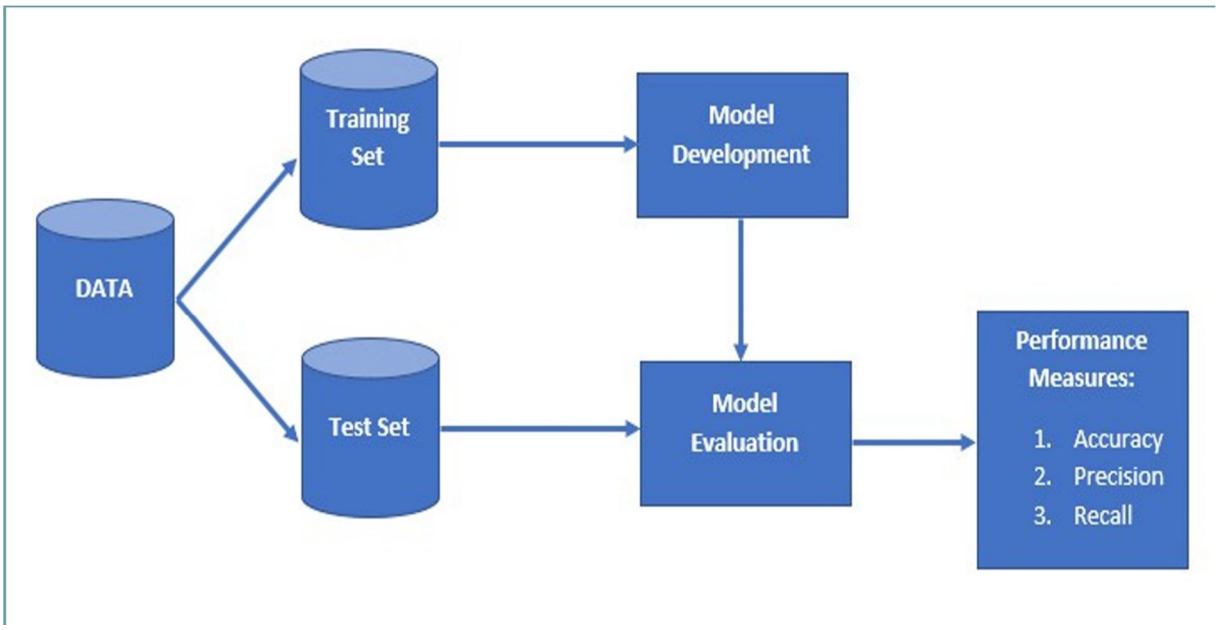
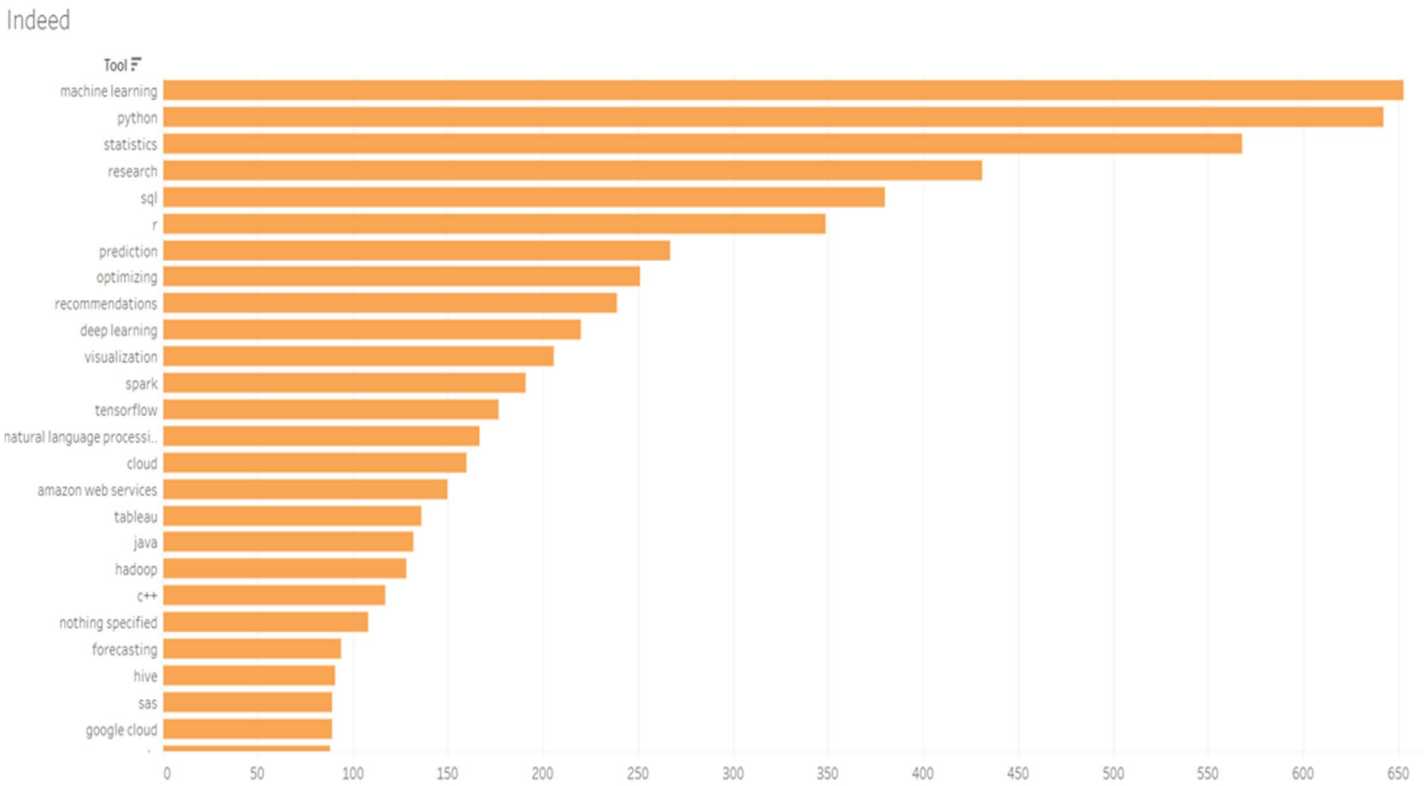


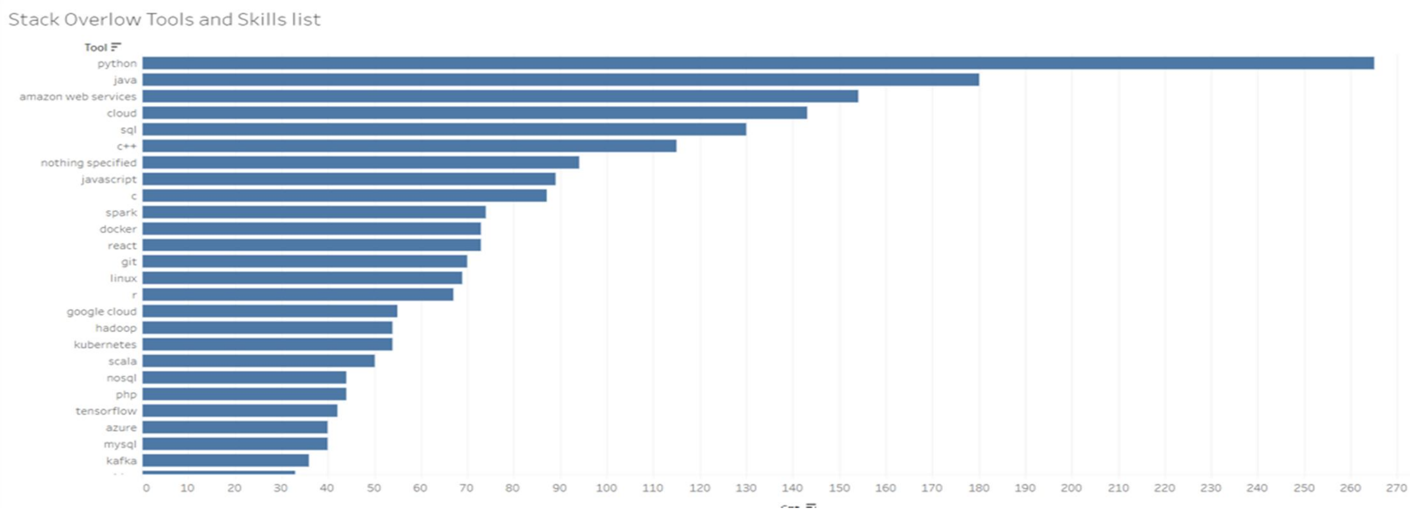
Fig 7. Overall Architecture

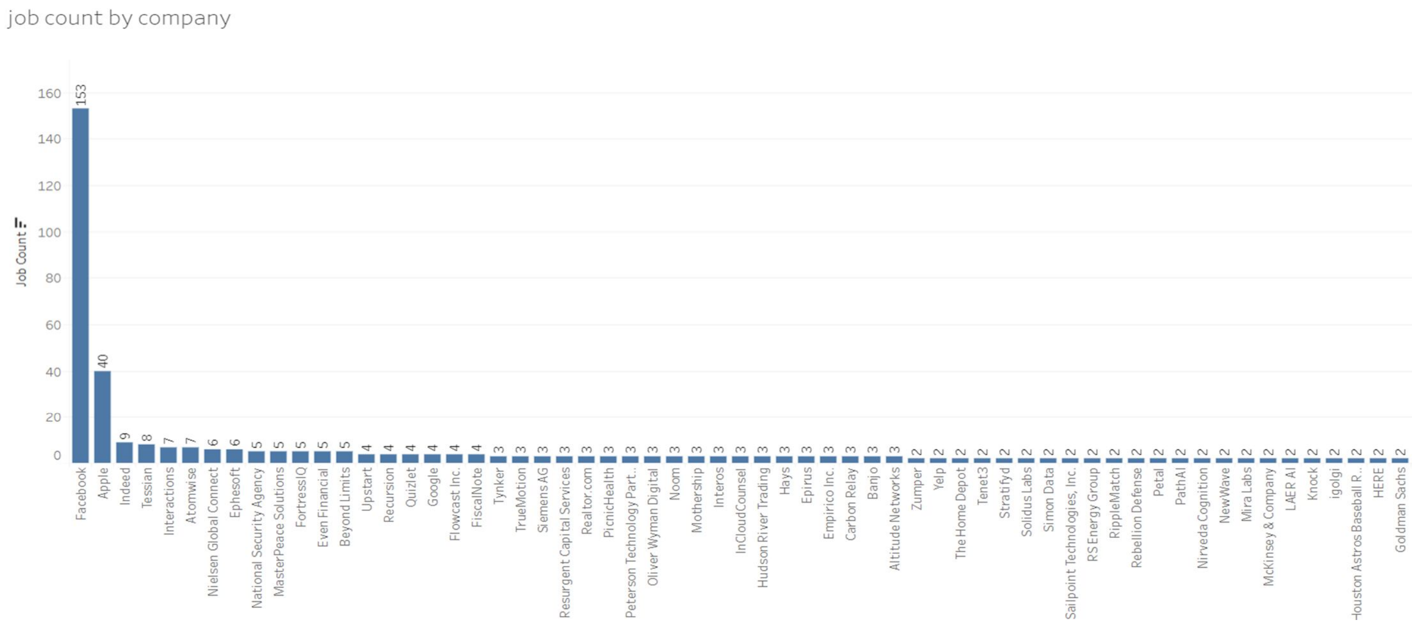
VI. EXPERIMENTAL RESULTS

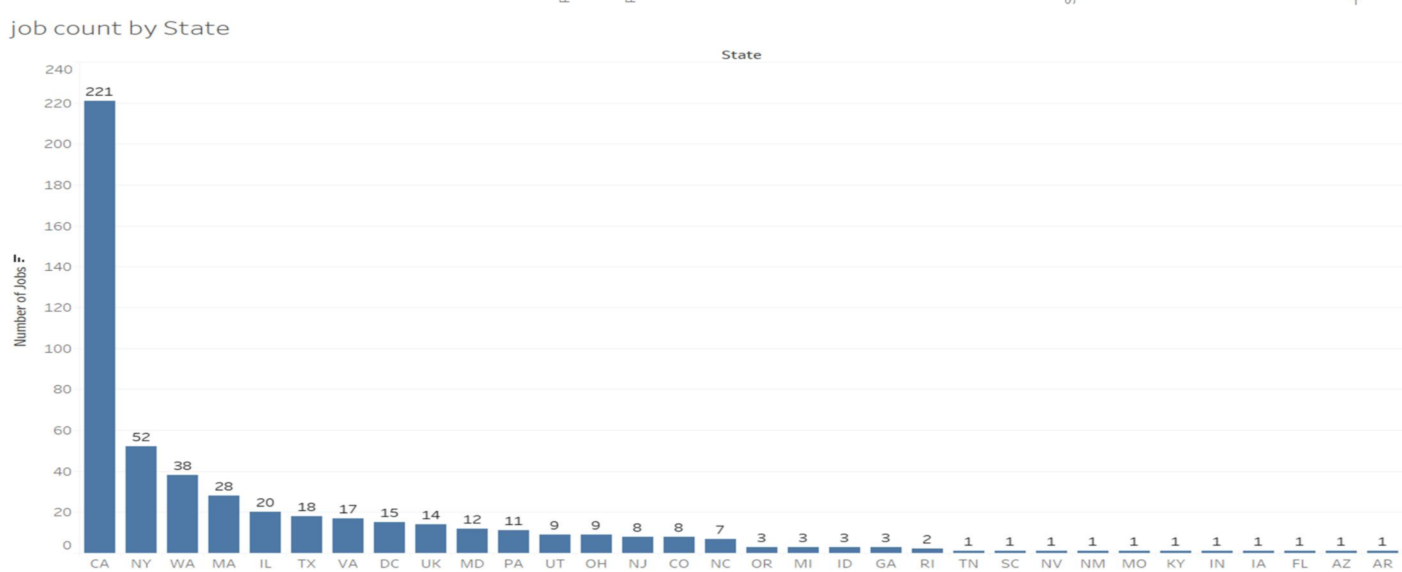
`

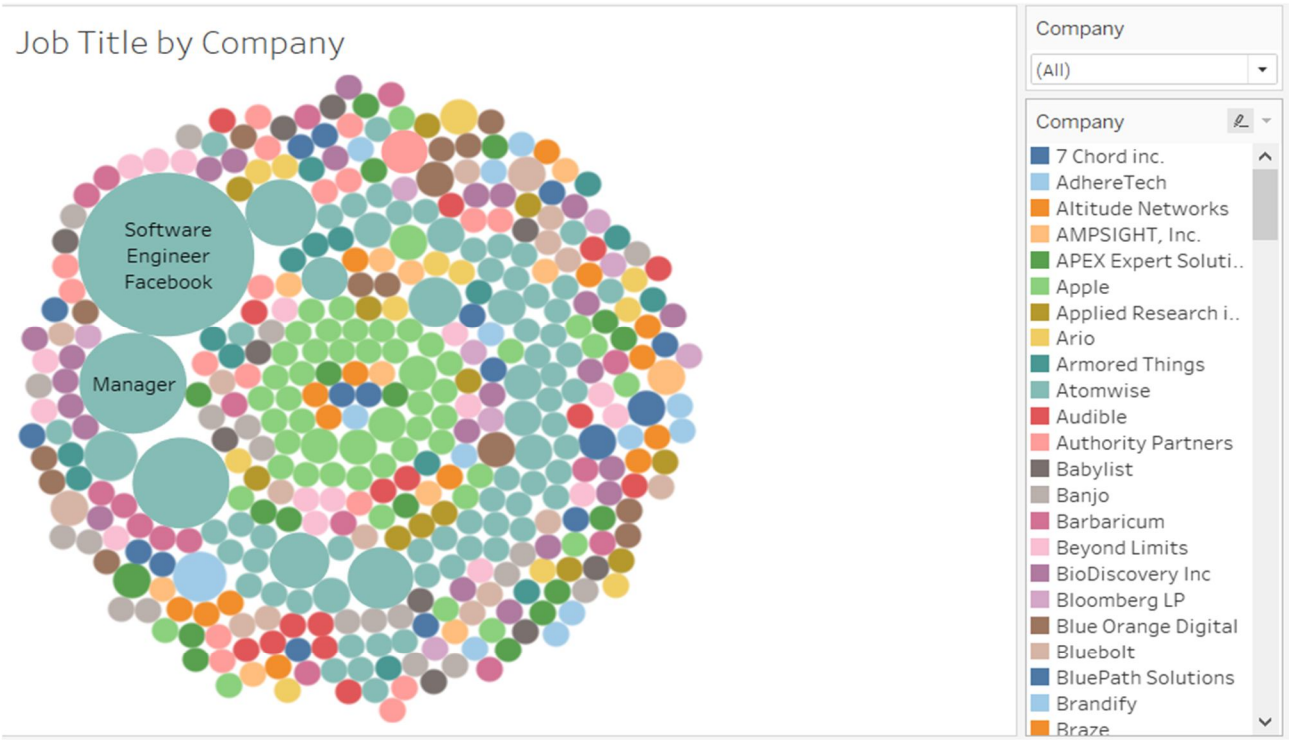


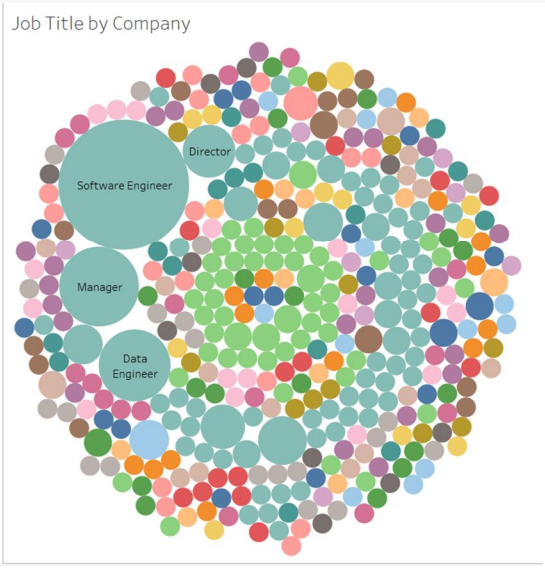
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |


Stack Overlow Tools and Skills list
Tool
python
java
amazon web services
cloud
sql
c++
nothing specified
javascript
c
spark
docker
react
git
linux
r
google cloud
hadoop
kubernetes
scala
nosql
php
tensorflow
azure
mysql
kafka
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270


job count by company
Job Count
0 20 40 60 80 100 120 140 160
Facebook 153
Apple 40
Indeed 9
Tessian 8
Interactions 7
Atomwise 7
Nielsen Global Connect 6
Ephesoft 6
National Security Agency 5
MasterPeace Solutions 5
FortressIQ 5
Even Financial 5
Beyond Limits 5
Upstart 4
Recursion 4
Quizlet 4
Google 4
Flowcast Inc. 4
FiscalNote 4
Tynker 3
TrueMotion 3
Siemens AG 3
Resurgent Capital Services 3
Realtor.com 3
PicnicHealth 3
Peterson Technology Part.. 3
Oliver Wyman Digital 3
Noom 3
Mothership 3
Interos 3
InCloudCounsel 3
Hudson River Trading 3
Hays 3
Epirus 3
Empirico Inc. 3
Carbon Relay 3
Banjo 3
Altitude Networks 3
Zumper 2
Yelp 2
The Home Depot 2
Tenet3 2
Stratifyd 2
Solidus Labs 2
Simon Data 2
Sailpoint Technologies, Inc. 2
RS Energy Group 2
RippleMatch 2
Rebellion Defense 2
Petal 2
PathAI 2
Nirveda Cognition 2
NewWave 2
Mira Labs 2
Mckinsey & Company 2
LAER AI 2
Knock 2
igolgi 2
Houston Astros Baseball R.. 2
HERE 2
Goldman Sachs 2


job count by State
State
Number of Jobs
0 20 40 60 80 100 120 140 160 180 200 220 240
CA 221
NY 52
WA 38
MA 28
IL 20
TX 18
VA 17
DC 15
UK 14
MD 12
PA 11
UT 9
OH 9
NJ 8
CO 8
NC 7
OR 3
MI 3
ID 3
GA 3
RI 2
TN 1
SC 1
NV 1
NM 1
MO 1
KY 1
IN 1
IA 1
FL 1
AZ 1
AR 1





VII. CONCLUSION

The job recommendation system using NLP bag of words can be an effective way to match job seekers with relevant job openings. The bag of words approach involves representing job descriptions and job seekers' resumes as vectors of word counts, which can then be used to calculate the similarity between them. One of the benefits of using the bag of words approach is its simplicity and efficiency, making it easy to implement and scale. However, it has limitations such as not being able to capture the semantic meaning of words and the context in which they are used, which can result in inaccurate recommendations. To overcome these limitations, more advanced NLP techniques such as word embeddings, topic modelling, and deep learning can be used. These techniques can capture more complex relationships between words and improve the accuracy of job recommendations.

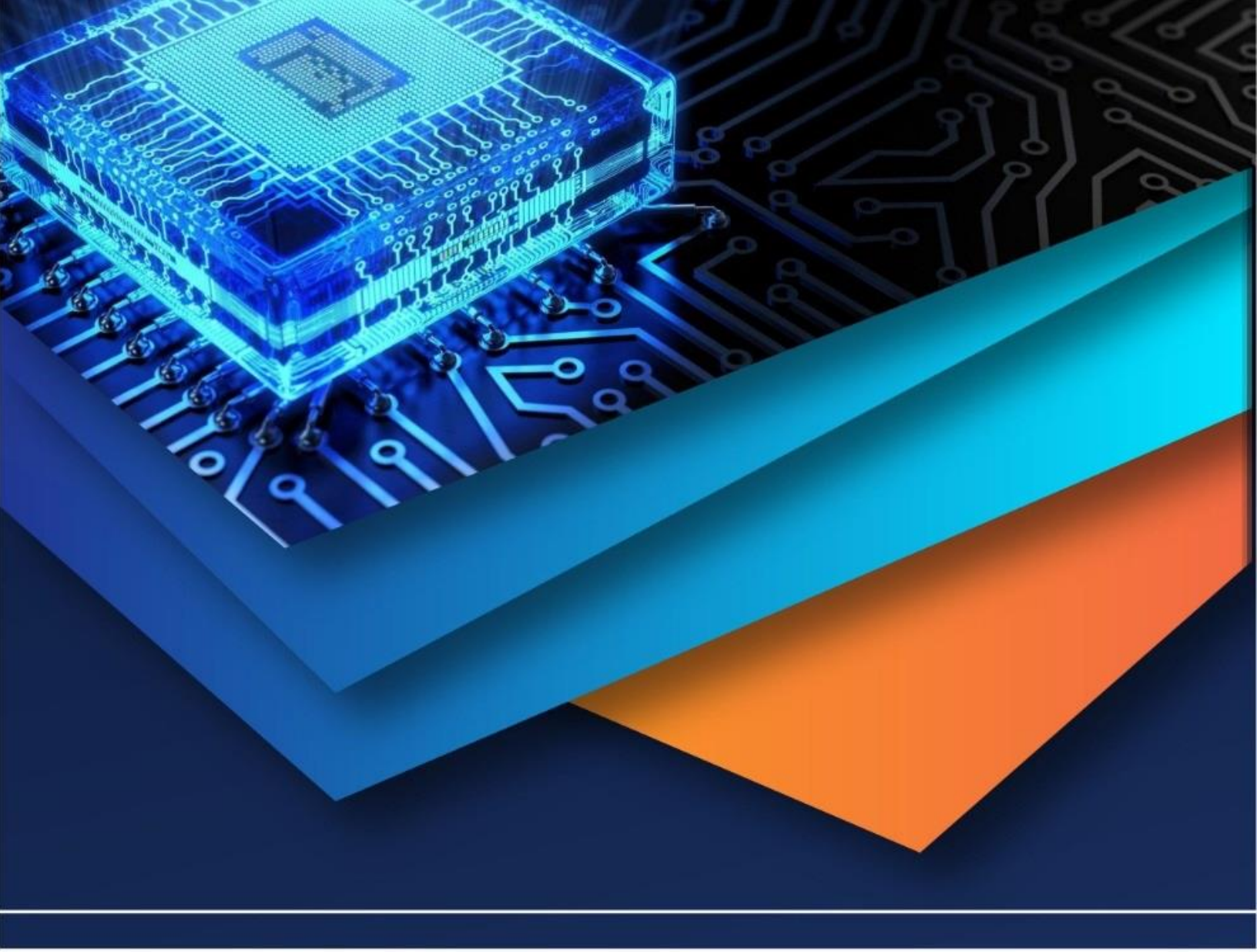
Overall, the job recommendation system using NLP bag of words can be a good starting point, but it is important to consider more advanced techniques to improve the accuracy and relevance of recommendations

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Al-Otaibi, S.T. and Ykhlef, M. (2012) Job recommendation systems for enhancing erecruitment process in: Proceedings of the International Conference on Information and Knowledge Engineering (IKE) p. 1 The Steering Committee of The World Congress in Computer Science, Computer
[2] Barrón-Cedeno, A., Eiselt, A. and Rosso, P. (2009) Monolingual text similarity measures: A comparison of models over wikipedia articles revisions ICON 2009, pp. 29–38
[3] Barzilay, R. and Elhadad, N. (2003) Sentence alignment for monolingual comparable corpora in: Proceedings of the 2003 conference on Empirical methods in natural language processing pp. 25–32 Association for Computational Linguistics
[4] Brants, T. (2003) Natural language processing in information retrieval. in:CLIN Citeseer
[5] Brownlee, J. (2017) What are word embeddings for text?
[6] Burke, R. (2002) Hybrid recommender systems: Survey and experimentsUser modeling and user-adapted interaction 12(4), pp. 331–370
[7] Burke, R. (2007) Hybrid web recommender systems in: The adaptive webpp. 377–408 Springer
[8] Dhameliya, J. and Desai, N. (2019) Job recommender systems: A survey in: 2019 Innovations in Power and Advanced Computing Technologies (i-PACT) vol. 1 pp. 1–5 IEEE
[9] Herlocker, J.L., Konstan, J.A., Borchers, A. and Riedl, J. (1999) An algorithmic framework for performing collaborative filtering in: 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1999 pp. 230–237 Association for Computing Machinery, Inc

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)