JobSelect: A custom Job Recommendation System

Ansh Mittal
IIIT Delhi
ansh19351@iiitd.ac.in

Harshita Gupta
harshita19467@iiitd.ac.in
IIIT Delhi

Kartikey Gupta
IIIT Delhi
kartikey19427@iiitd.ac.in

Raghav Bhalla
IIIT Delhi
raghav19379@iiitd.ac.in

Rishita Chauhan
IIIT Delhi
rishita19383@iiitd.ac.in

MD Zaid
IIIT Delhi
zaid19433@iiitd.ac.in

KEYWORDS

Job Recommendation System, Information Retrieval, Machine Learning, Natural Language Processing

ACM Reference Format:

Ansh Mittal, Harshita Gupta, Kartikey Gupta, Raghav Bhalla, Rishita Chauhan, and MD Zaid. 2023. JobSelect: A custom Job Recommendation System. In *Proceedings of (Information Retrieval Course Project)*. ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 PROBLEM STATEMENT

This project aims to create a system that recommends suitable job opportunities to job seekers based on their skills and preferences. This will help job seekers find jobs that match their qualifications and career goals more quickly and easily while also helping employers find more qualified and motivated candidates. The system will use natural language processing and machine learning algorithms to analyze job seeker data, such as resumes and job applications, and suggest relevant job postings.

2 MOTIVATION

The process for job search is simplified for job seekers and the recruitment procedure is improved for employers using the Job recommendation system. Job seekers often have to go through a large number of job postings to find relevant opportunities, which can be time-consuming and overwhelming. It is time-consuming for prospective employees to find relevant opportunities, so they have to go through a large number of job openings on various platforms like Linkedin, Indeed, Naukri.com, etc. Candidates often want to build their resume according to a specific job role so they can tweak their profile until their recommended job profile matches their choice. Using machine learning and NLP techniques, a job recommendation system can provide personalized job profile recommendations to job seekers based on their skills, experience, and preferences, making it easier for them to find suitable jobs. At the same time, this system can help employers by matching them with more qualified and motivated candidates. Ultimately, the job recommendation system can benefit job seekers and employers by making the job search and recruitment processes more efficient and effective.

3 LITERATURE REVIEW

The research on job recommendation systems has gained significant attention in recent years.

3.1 Learning-Based Matched Representation System for Job Recommendation

Alsaif, S.A. and Sassi Hidri, M. developed a system that recommends the top-n jobs to the job seekers by analyzing and measuring the similarity between the job seeker's skills and explicit features of job listing using content-based filtering. The system used Natural Language Processing (NLP) to match skills between resumes and job descriptions. Finally, the job offer is recommended, similar to the users' skills on their resumes using Collaborative Filtering. [1]

3.2 A smart Geo-Location Job Recommender System Based on Social Media Posts

There was another study by A. Mughaid and I. Obeidat which aimed to match the best vacancy for the exact job seeker by way of mining social media networks, such as Facebook and twitter. This system involved intensive data mining techniques and used Natural Language Processing (NLP) classifiers such as Support Vector Machines, Naive Bayes, and Random Forest to locate the best job seekers and job locations based on their social media posts history. [5]

3.3 Investigating Natural Language Processing Techniques for a Recommendation System to Support Employers, Job Seekers and Educational Institutions

Koen Bothmer and Tim Schlippe built a Skill Scanner application that outputs the missed and covered skills for all three users,i.e., employers, job seekers, and educational institutions. In their prototype, they retrieved skills from the applicant's Resume, then compared the retrieved skills with the Market (Employers and Educational Institutions). The research focused on only one job type, i.e., data scientist, and scrapped the data from Kaggle and Indeed.com. They used Word2Vec, GloVe, and Sentence-BERT to represent the skills in vectorized form. They used a clustering-based approach for the recommendation to the three users. K-Means clustering was used with K equal to 31. [4]

Information Retrieval Course Project, IIIT Delhi, India,

https://doi.org/XXXXXXX.XXXXXXX

3.4 NLP-Based Bi-Directional Recommendation System: Towards Recommending Jobs to Job Seekers and Resumes to Recruiters

Suleiman Ali Alsaif and Minyar Sassi Hidri built an NLP-based bi-directional Job recommendation system for job seekers and employers. They scrapped five job profile datasets from Indeed.com. The data extracted were preprocessed using Clean Tags, Tokenization, Lemmatization, and Stop words removal. Then the Bag of Words model converted the textual data into a vector representation. Further, Named Entity Recognition was used using spaCy to extract the named entities. Word2Vec model retrieves similar terms, and then cosine similarity is used to find the similarity. [2]

3.5 JobFit: Job Recommendation using Machine Learning and Recommendation Engine

The study intends to develop a job recommendation system, JobFit, which predicts the best candidates for a position using machine learning techniques, a recommender system, and previous data. The system generates a JobFit score indicating how well-suited a particular candidate is for a specific position depending on the applicant's profile and job requirements. The output is a ranked list of candidates best suited and most qualified for the position. This ensures that HR concentrates on interviewing a small group of top prospects, as recommended by the JobFit, without caring that the top candidates will be overlooked. The recommendation system developed in this study combines several different machine learning models with a collaborative filtering recommendation system, whose output is then fed to a final machine learning model through which an applicant's JobFit score is obtained for a specific job position. [3]

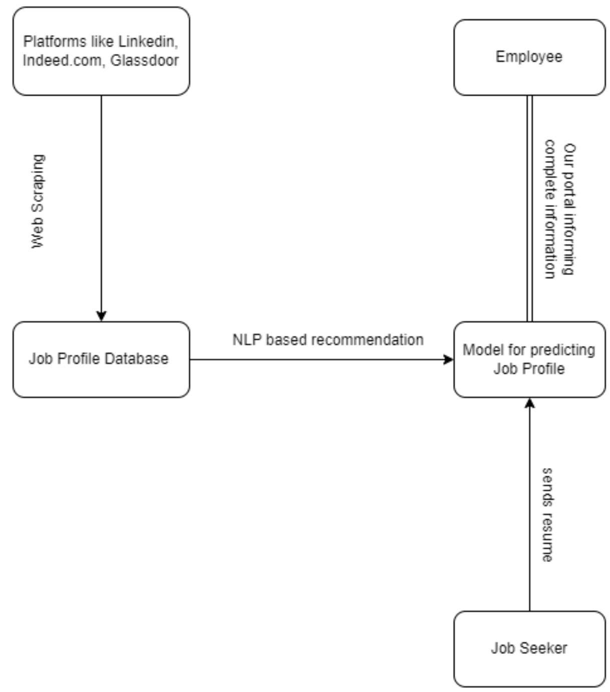

Figure 1: Flow Chart for NLP based Job Recommendation system

4 OUR APPROACH

We have created our database using web scraping from Job Profile website Linkedin. The database is divided into two parts for testing and training. Training dataset includes job posting details like Job ID, Job Title, Company, Job Description, and Job Profile of the candidate. On the other hand testing data will contain users' Linkedin Resume extracted from linkedIn profiles which includes details like name, about, experience, education, courses, skills, etc. thus; the aim is be to feed rich data to make the model highly robust. We have assigned the top 3 job profile categories to every user resume according to their detail in the test dataset. Then we have incorporated various NLP techniques (preprocessing and word embeddings) to convert the textual job profile data into embeddings. Then techniques like multiclass classification using various machine learning algorithms are used to retrieve the top three relevant job profiles to the Linkedin Resume. We will also inform the user to what extent his/her job profile matches our set of job profiles. Hence, the use of rich data and information regarding expected and present skillset deviation will make the recommendation system more meaningful to the user.

5 TECHNIQUES USED

- **Dataset Creation :** Job postings and user resume data has been extracted from LinkedIn. Selenium has been used for scraping the data from the website. Job seekers are expected to provide the resume from which the information (skillset) will be extracted.
- **Data cleanup :** After data extraction all the redundant information is removed from the data to make sure that only relevant data is present.
- **Natural Language Processing :** NLP techniques are used for the processing of the scrapped data like tokenization, stop words removal, lemmatization and stemming to retrieve the key terms/skills from resume and job profiles. Word Net and Porter algorithms are used for the lemmatization and stemming respectively.
- **Job recommendation :** NLP models are used to create word embeddings from the textual training and testing data. We have considered TF-IDF vectoriser and Count vectoriser models. Classification techniques like Random Forest, Stacking Classifier are used for recommendation of most suitable jobs based on matching between candidate resume data (embeddings) and job profile.

6 DATA ANALYSIS

Various methods and techniques were incorporated for analysing the training and testing data prepared above. The following is

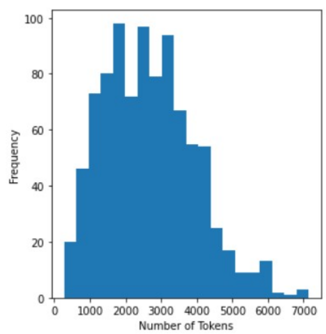

Figure 2: Distribution of the number of tokens in Training Data

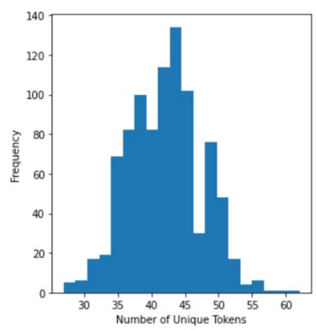

Figure 3: Distribution of the number of unique tokens in Training Data

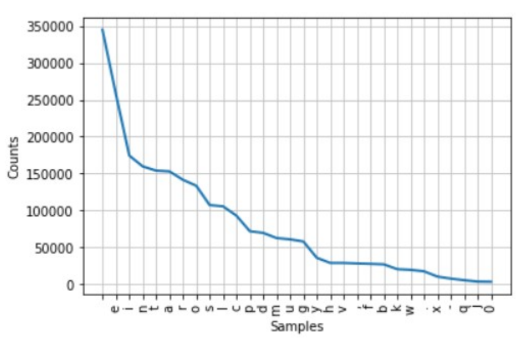

Figure 4: Frequency Distribution of tokens in Training Data

Figure 5: Word Cloud of Training Data

Figure 6: Word Cloud of Testing Data

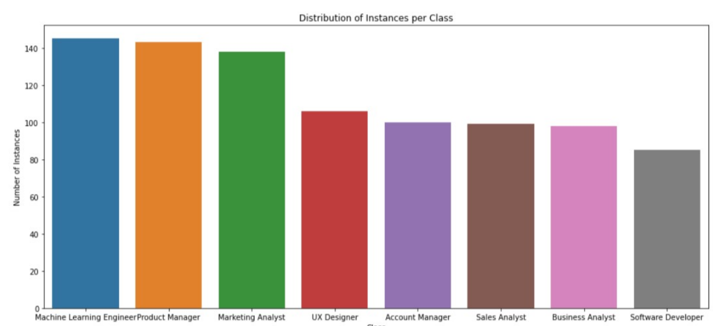

Figure 7: Plot of Job Profile Class Value Count in Training Data

the analysis performed on the datasets. Figures 1 to 7 show the data analysis performed on the training and testing dataset. The following were the main techniques for data analysis:

- **Distribution of number of Tokens per sample (Fig 1,2 and 3)**
- **Frequency Distribution of Tokens (Fig 4)**
- **Word Cloud (Fig 5 and 6)**
- **Class Value Count of Job Profile Class (Fig 7 and 8)**

7 EVALUATION

To evaluate the performance of our recommender system, we have used classification metrics like decision support metrics (precision and recall) and precision@k. These classification metrics evaluate the decision making capacity of our recommender system without taking ranks/relevance of recommendations into account. We have also used evaluation metrics that are used for rank based recommendation systems like Mean Average Precision (MAP).

Machine Learning Engineer	145
Product Manager	143
Marketing Analyst	138
UX Designer	106
Account Manager	100
Sales Analyst	99
Business Analyst	98
Software Developer	85

Figure 8: Class Value Counts per Job Profile in Training Data

8 BASELINE RESULTS

For the Baseline Results, we have used Count Vectorizer to create features or word embeddings of the training database i.e Job Profiles data scraped from Linkedin. We have used a Random Forest classifier with n_estimators as 100 for predicting or recommending the job profiles of the resume of a candidate. To get the top 3 job recommendations, probabilities are used for each of the job classes and the highest 3 probabilities are then selected. Label encoding is done for both the training and testing set for probability prediction.

Following are the metrics which we considered for baseline evaluation:

- **recall@2:** Measures the fraction of relevant jobs (matches the ground truth values) that are retrieved by the system for the first 2 recommendations. For example, if the system returns 2 relevant recommendations out of 4 total relevant jobs then the recall@2 for that system would be 0.5 (50%), because it successfully retrieved half of the relevant recommendations within 2.
- **precision@3:** Measures how many of the 3 jobs recommended by the system are a good match for the user's interests or skills (matches the ground truth values). For instance, if the system suggests 3 jobs and 2 of them are relevant to the user's interests, then the precision@3 for that system would be 0.67 (67%), because it accurately identified 2 out of the 3 most relevant jobs for the user.
- **Mean Average Precision (MAP):** Measure that combines both recall@k and precision@k (where k is a particular number of recommended items) to provide an overall evaluation of a job recommender system's performance. For our evaluation we computed MAP for k=3.

Recall@3 and Precision@3 will be similar because of the fact that the denominator in both the formulas will result in the same value.

This can be shown as:
Number of relevant documents = (No. of queries) x 3 (3 is the number of relevant documents in each query)
Number of retrieved documents = (No. of queries) x 3 (3 is the number of retrieved documents in each query)

Since, number of documents is same in both the denominators of the formula, and numerator is same due to relevancy, the value of Precision@3 and Recall@3 results the same. However, this is not true for the case of Precision@2 and Recall@2 since the denominator of these formulas will be different.

9 RESULTS & ANALYSIS

The NLP models used for creating word embeddings are Count Vectorizer and TF-IDF. The classification models used for recommending job profiles to user are Random Forest and Stacking Classifier (consisting of Random Forest and XGBoost classifier). The Results with these combinations are provided as follows:

Models	Precision@3	Recall@2	Mean Average Precision (MAP)
Count Vectorizer + Random Forest	0.645	0.763	0.681
Count Vectorizer + Stacking Classifier	0.694	0.852	0.736
TF-IDF + Stacking Classifier	0.659	0.802	0.684

Table 1: Performance Analysis

Based on the results, the Stacking Classifier method with Count Vectorizer appears to perform the best in terms of precision, recall, and mean average precision. This suggests that this model is the most accurate in predicting job recommendations based on the job description and profile. The models are trained to analyze job descriptions and recommend the most suitable job profiles. The Count Vectorizer method breaks down the job description into smaller pieces and analyzes them, while the Stacking Classifier method combines the results of multiple models to make a more accurate prediction. The results suggest that the Stacking Classifier with Count Vectorizer is the most accurate method for predicting job recommendations.

10 PROPOSED METHOD

Till now, we have implemented models (Count Vectorizer + Random Forest Classifier, Count Vectorizer + Stacking Classifier and TF-IDF Vectorizer + Stacking Classifier) that are trained to analyze job descriptions and recommend the most suitable job profiles. Further we are planning to adopt more sophisticated techniques as follows:

- Embeddings of the job profiles data will be done using more advanced NLP models like Word2Vec, GloVe (Global Vector for word representation), BERT (Bidirectional Encoder Representations from Transformers).
- More sophisticated and advance classification and recommendation models will be used like Naive Bayes, LSTM and RNN.
- User Interface will be created for our use-case. The interface will be used as a portal to recommend the most related job to the user when he/she uploads his/her Resume.
- The information will be extracted from the uploaded resume of the user and will be given to trained model on the job profiles dataset extracted from Linkedin.

```
Mean Average Precision: 0.6017025089605734
Recall@2: 0.7217741935483871
Precision@3: 0.6227598566308243
```

Figure 9: Result of the Baseline Model

11 CONTRIBUTIONS

- **Data Webscraping :** Rishita Chauhan
- **Resume Information Extraction :** Rishita Chauhan and Harshita Gupta
- **Data cleanup :** Harshita Gupta
- **Natural Language Processing :** Raghav Bhalla and Md. Zaid
- **Classification Algorithms and Evaluation :** Kartikey Gupta and Ansh Mittal

12 ACKNOWLEDGMENT

We are grateful to our instructor Dr. Rajiv Ratn Shah for providing us with the opportunity to work on this project under his guidance. We would like to thank him for his patience, motivation and immense knowledge throughout the project. We would also like to express our gratitude to all the TA's who have directly or indirectly guided us in doing this project.

REFERENCES

[1] Suleiman Ali Alsaif, Minyar Sassi Hidri, Hassan Ahmed Eleraky, Imen Ferjani, and Rimah Amami. 2022. Learning-Based Matched Representation System for Job Recommendation. *Computers* 11, 11 (2022). https://doi.org/10.3390/computers11110161

[2] Suleiman Ali Alsaif, Minyar Sassi Hidri, Imen Ferjani, Hassan Ahmed Eleraky, and Adel Hidri. 2022. NLP-Based Bi-Directional Recommendation System: Towards Recommending Jobs to Job Seekers and Resumes to Recruiters. *Big Data and Cognitive Computing* 6, 4 (2022). https://doi.org/10.3390/bdcc6040147

[3] Kevin Appadoo, Muhammad Bilaal Soonnoo, and Zahra Mungloo-Dilmohamud. 2020. Job Recommendation System, Machine Learning, Regression, Classification, Natural Language Processing. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*. 1–6. https://doi.org/10.1109/CSDE50874.2020.9411584

[4] Koen Bothmer and Tim Schlippe. 2022. Investigating natural language processing techniques for a recommendation system to support employers, job seekers and educational institutions. In *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium: 23rd International Conference, AIED 2022, Durham, UK, July 27–31, 2022, Proceedings, Part II*. Springer, 449–452.

[5] Ala Mughaid, Ibrahim Obeidat, Bilal Hawashin, Shadi AlZu'bi, and Darah Aqel. 2019. A smart Geo-Location Job Recommender System Based on Social Media Posts. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. 505–510. https://doi.org/10.1109/SNAMS.2019.8931854