GloVe Based Language Modelling for Job to Candidate Matching

Submitted by

Name: Md. Ashikur Rahman

ID:**183-35-2627**

Department of Software Engineering

Daffodil International University

Supervised by

Name: Musabbir Hasan Sammak

Designation: Lecturer

Department of Software Engineering

Daffodil International University

This Thesis report has been submitted in fulfillment of the requirements for the Degree of Bachelor of Science in Software Engineering

APPROVAL

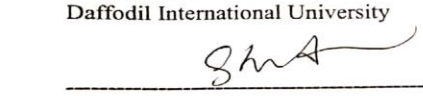
APPROVAL

This thesis titled on **"GloVe Based Language Modelling For Job To Candidate Matching"**, submitted by **Md.Ashikur Rahman (ID: 183-35-2627)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

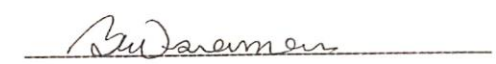
Chairman

Dr. Imran Mahmud
Head and Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University



Internal Examiner 1

Md. Shohel Arman
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University



Internal Examiner 2

Khalid Been Badruzzaman Biplob
Lecturer (Senior)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Md. Tanvir Quader
Senior Software Engineer
Technology Team
a2i Programme

DECLARATION

This document confirms that I completed this thesis while working under the supervision of Musabbir Hasan Sammak, Lecturer in the Department of Software Engineering at Daffodil International University. It is noted that neither this thesis nor any portion of it has been submitted to another university for a degree.

Md.Ashikur Rahman

Student ID:183-35-2627

Batch:27

Department of software Engineering

Faculty of Science & Information Technology

Daffodil International university

Musabbir Hasan Sammak

Lecturer

Department of software Engineering

Faculty of Science & Information Technology

Daffodil International university

ACKNOWLEDGEMENT

First, to Almighty God, I express my heartiest thanks and gratefulness for His divine blessing in making it possible to complete the final year thesis successfully.

I am grateful and wish our profound indebtedness to Musabbir Hasan Sammak, **Lecturer**, Department of Software Engineering, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Deep Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, and constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

I would like to express my heartiest gratitude to Dr. Imran Mahmud, Head In-Charge of the Software Engineering faculty for his kind help to finish our project and also to other faculty members and the staff of the SWE department of Daffodil International University.

Finally, I must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

In the present era, technology has become very advanced. And we can say that now a days everything is online- based. we can also find jobs online. And online job market is widely known all over the world. As example, up work, fiver,freelancer.com are popular online job market. But whenever they offer any job opportunity, a lot of people drop their cv. So the client face problem to find the right worker for the job and end up getting poor quality than their requirement. Which is hampering the reputation of online job markets. I Have decided to make a ranking of the cv of the workers according to their skills.so that the deserving candidates can get fare chances and clients also get freelancers as per their desire. I have used two models here. they are word embedding and glove vector. I have read the literature review and learnt that the accuracy of the models is up to 98%. And they are easier to work with this process, clients and workers both will be satisfied and we will be able to save a lot of time. But we could not complete the work yet. We Only finished word embedding. But in future, we will make a ranking of the cv of workers.

Contents

Table Of Contents

CHAPTER 1

1.INTRODUCTION

1.1 BACKGROUND

We are now Living in a modern era where technology plays a huge role in our life. Our life is now a lot dependent on Technology. In today's world, we Don't have to wait for a long time for updates of all over the world, we can catch the updates of the entire world, from our home. And technology has made this impossible thing possible. After waking up in the morning, we all read newspapers, which is a great invention of Technology. We have television's in our home, from this we get live updates and entertainment. Another great invention of technology is mobile phone. It's a blessing of science. We can talk, face time with our friend and relatives through this device. Even who lives totally different countries, we can also connect them instantly. We Don't have to wait for months to connect them through letters like before. So, technology has made our life so much easier today. Modern world cannot do anything without it. People can now do everything from their home without getting out. They can even work, shopping, consulting a doctor from home. There are a lot of advantage of Technology. Online job market is also a part of technology which is widely known now a day. Fiver, up work, and freelancer.com are some options of online job market. it is a very popular way of making money and also a great way to reduce unemployment. Through online job market, people don't have to go out of their house for work. so, people who cannot go outside for their living such as, old people, disabled people and women's with newborn child will be very benefitted from the system. They will be able to make money and use their skill from their house. Because, online job market has the opportunity of working from home. But, this amazing system is being affected for some unskilled workers. It's happening for the clients who often make mistakes and hired somebody who does not meet The requirements. As a result, clients do not get their desired result. For the

mistake of choosing wrong worker, online job market reduces client reliability in some cases. Inexperienced and new workers are mostly responsible for this. People apply for work without properly reading the job description. It confuses the client because they get way more proposals than they need. It causes the main problem. The client gets confused to choose the right person for the work. Sometimes, they fail to choose the experienced person for the work and then inexperienced people get chance. Experienced people don't get enough work that they deserve.

To avoid this kind of situation, we can get a fare solution using a glove vector to get rid of these problems. Which will help the client very much to hire skilled worker. According to the design of the model, people who don't have enough experience, cannot submit proposals. Only those people can apply who match the freelancer's profile with the skills required by the client. In this method, clients will easily find Their desired worker with the skill and requirement they Have. Hiring the deserving and experienced worker will be much easier than before. The clients and the applicant both will get their desired result. This process can save a lot of time. Client will be able to choose the right worker with the help of the ranking. They Won't get confused who is the perfect one for the job. And it will also be a blessing for workers too. Now a day's people with less skills get hired several times. So the workers who have enough skill and quality is neglected by the client. In this process it will not happen anymore because we are making rankings according to skills and quality. So the deserving ones will get more priority. I believe it will be more advantageous for both client and worker

1.2 MOTIVATION OF THE RESEARCH

I noticed in the data lab that all the cv of the students has been collected and a ranking has made on the basis of their qualification. In this way, things do not messed-up and everything happens nicely. It seems so perfect to me. I like this idea very much. I am feeling much more motivated to see this and I also came up with a plan. We all know that various online platforms such as up work, the platforms are known as online job market. Here, clients advertise job offers with their personal requirement for workers who meet their requirement. I can say it is a great place for people with unemployment. The idea also helps to reduce unemployment. But there are also some problems. There a lot of people drop their cv after clients offer a job opportunity. A huge number of cv has been dropped for one offer. Some people are not even fit for the job and Don't have enough skill and quality. But they often drop their cv. For this, there create so much crowd that the real deserving candidates cannot get chance more often. As usual clients get confused with so much candidates to choose the right one for the job. And it is also very time consuming to check all the cv one by one and find the deserving one. And naturally, clients don't have so much time to check all the cv one by one. As a result, the client ends up choosing a worker who Doesn't fit for the job and do not get their desired result. On the other hand, deserving candidates do not get chances sometimes. And for this enormity, the reputation of online job markets is reducing day by day. People cannot trust on it anymore. So, I thought about ranking the applicants on the basis of their skills. That is to say, those people will be prioritize whose skill matches with the job requirement. If the plan can be implemented, clients will get the proper skilled freelancer according to their requirement, and will be able to get the best results, on the other side, deserving candidates will also get chances. In this way, unnecessary procrastination will be avoided, and the entire process will consume less-time than before and will be completed in a trouble-free way. I am optimistic about the project so much. I believe that it will make things much easier than before and the reputation of online job market will regain.

1.3 PROBLEM STATEMENT

In today's world, Technology is more advanced than Our imagination. In fact, it will not wrong if I say that the entire world is largely based on Technology. We can find the usage of Technology everywhere. The time from we wake up and go back to sleep, we depend and use technology. Technology has been used each and every part of our life including job sector. And there are plenty of job

opportunities now a day and it is a blessing. People can work from anywhere. There are a lot of advantages of online jobs. It has made our life so much easier. Job opportunities have increased a lot now a day which is reducing unemployment problem in a large number.

but the number of inexperienced people also increased at the same time. They don't have enough skills but still they apply for job and cannot deliver their work on time and they cannot maintain the quality. As a result, clients get disappointed. So, clearly it is a huge

challenge to find an ideal applicant for everyone. But it is more difficult than it looks. People drop a lot of cv for one job. As example if there are only 3 seats for one job, a large number of people applying for only 3 seats. And of course, they all are not enough experienced for the job. But actual deserving candidates are not getting enough chance in so much candidates. clients also get confused about the choosing. And also they Don't have so much time to take more than 1 thousand interviews for only 3 seats. So, hr. Department cannot Control the crowd of application and they cannot take all the interviews. They end up selecting inexperienced people who doesn't fit for the job. On the other hand, people with experience cannot get fare chances more often. In the light of above circumstances, we thought about ranking the cv of the workers according to Their skill and experience to solve this kind of problems.

1.4 RESEARCH QUESTIONS

- Q1: Whether or not we can match suitable candidate profiles with job descriptions.
- Q2: Whether or not we can find suitable candidates with human-level accuracy.

1.5 RESEARCH OBJECTIVE

- To collect and annotate job descriptions and corresponding candidate profiles.
- To develop a deep learning-based architecture to match suitable candidates with job descriptions.

1.6 RESEARCH SCOPE

In the first chapter, a particular part on online job market place and its usage, the background behind the work, motivation of the research, problem statement, research questions, research objective are discussed. The other parts related to our research are as below:
In the next chapter I will discuss, the literature review where we can see some researcher's studies which have already been done on the same field. their used methodology, lacking and on the base of their work comparison among my work and their work. In their chapter, we will discuss the methodology of our work. In the methodology of my work, I will discuss data collection, pre-processing of data and will analyze our work. The results of the methodology will be discussed, in chapter four. The last chapter is the ending chapter. Here I will give the conclusion part where there will be a total summary of my work. Here I have discussed what work I will do in the future for the betterment of the work.

CHAPTER 2

2.1 INTRODUCTION

In a literature review, a researcher reviews the previous work, research, conference paper, books, article, etc. with that one can find out what work has already been done on the topic, summarize the whole topic, find out what lacking in the work. after analyzing they can work on limitations and overcome the limitations to get better results.

2.2 PREVIOUS LITERATURE

Sisay Chala and et al [1] they find The purpose of this study is to identify methods that measure the skills, expertise and experience of a job seeker and to investigate importance of using social networking data as input to user modeling that determines the strength of skills to be used for recommending matching job vacancies. The technologies used to perform the analysis and matching are i) machine learning – to extract, analyze, and present the result; ii) web mining – to scrap data online; and iii) user interaction – to allow the user to add or remove features that characterize him/her. They collect data from online recruitment system and provides the role of social networking data.

Chuan Qin and et al [2] They proposed a novel end-to-end Ability-aware Person-Job Fit Neural Network (APJFNN) model, which has a goal of reducing the dependence on manual labor and can provide better interpretation about the fitting results. They used real-world data set. Convolutional Neural Network (CNN) and Recurrent

Neural Network (RNN) are two representatives and widely-used architectures, which can provide effective ways for NLP problems from different perspectives.

Joscha Gruger and et al [3] They used TF-IDF method. this method, 75.33% of the known skills can be assigned correctly. In the future, the semantic similarity recognition process could be improved. The inclusion of external data sources and training of word embedding's on job ads could improve the assignment. German online job exchanges for jobs in the IT sector. The online job exchanges monster.de, stepstone.de and stellenanzeigen.de were used as data sources.

Norhaslinda Kamaruddin and et al [4] In this paper they proposed an automated approach to match the graduates' and employers' needs using a hybrid of text mining and visualization approach to facilitate jobseekers' task of relevant job application. Data collected for five web-related jobs, namely; web developer, software developer, software engineer, PHP developer and .Net developer and used TF-IDF method.

Kanyanut Homsapaya and et al [5] The objective is to help match older job seeker profiles to increase their employability rate and remove job mismatches. This research could be extended by considering more factors for evaluating the matching of applicants to job. The dataset was gathered using both survey tools and interviews with responders. The original dataset contained 118 resumes and information on 16 hiring companies with open older candidate role positions . they used neural network model method.

Md. Sabir Hossain and et al [6] In this paper using logistic regression and support vector machine and they got 83.40% (Logistic Regression) and 84.03% (Linear SVM) accuracy. Personality types of the workers have not yet considered for assigning the job this research. This work can be extended by applying classification algorithms to the posted jobs. real dataset collected from different online marketplaces. they used Logistic Regression, Linear SVM model.

Chao Wang and et al [7] In this paper, they proposed a novel time-aware career trajectory prediction (TACTP) framework for jointly predicting the three key elements in career trajectory, i.e., timing, company, and position. A unique perspective of TACTP is that we can generate time-aware predictions according to the varying duration of the current job owing to our proposed temporal encoding mechanism. Specifically, we first developed a unified time-aware sequential model based on recurrent networks to map. The real-world dataset in this paper was collected from LinkedIn, one of the most famous commercial professional social networks, which has served hundreds of millions of users to share their career experience and professional resumes. They used TACTP method.

Amit Verma and et al [8] These clearly defined skills can be valuable for the hiring process as well as to revamp existing course curricula to cater to the increasing market demand. Future research could involve a localized study of some specific US states. They used Neural Network method. they used data from a popular online job portal, Indeed.com. The collected data, based on US job advertisements, was analyzed using content analysis (Neuendorf, 2016), a qualitative method that enables the tasks of extracting relevant information by counting the number of occurrences of specific keywords.

Evanthia Faliagka and et al [9] they used for semantic search skills. The proposed system extracts a set of objective criteria from the applicants' LinkedIn profile, and compares them semantically to the job's prerequisites. they used the data of LinkedIn profile and subjective data.

Yu Deng and et al [10] they used for SVM, Tree-Based Model, DNN. our model outperforms shallow models, like SVM and Random Forests, in effectiveness and accuracy. In the future, with more information to be snatched from website we will try to extend our work in multiple languages. They used the data of Data castle.

Sisay Adugna Chala and et al [11] they used automatic bidirectional matching system. The research results may not guarantee high quality of recommendation and maturity of matching results. Research is required to test if the proposed system

works for other domains as well as more diverse data sets. They used the data of the integration of occupational standards, web survey data, and social networking data into user profile collection.

Pombo and et al [12] they used word2vec. We believe our results could be even further improved by extending the idea of semantic embedding to other features and by finding candidates with similar job preferences with the target candidate and building upon that a richer presentation of the candidate profile. They used the data of Landing. Jobs. Michael C.

Knaus0F and et al [13] they used IPW. we show the potential of easy-to-implement program participation rules for improving average employment effects of these active Labor market program. This study abstracts from the questions about an optimal program for a particular unemployed person, which is also relevant because of the usually rich program structure of ALMPs. Such a modified goal raises several additional statistical issues that may be addressed in future research. They used the data of unemployment insurance databases (AVAM/ASAL) and social security records (AHV).

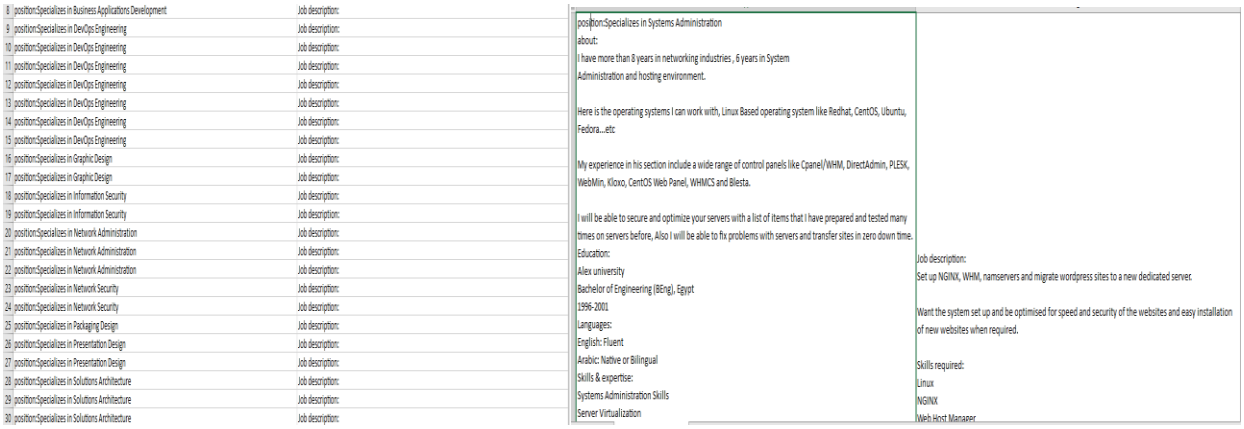
CHAPTER 3

3. RESEARCH METHODOLOGY

3.1 DATA COLLECTION

Data collection is the process of acquiring and analyzing information on relevant variables in a predetermined, methodical way so that one can respond to specified research questions, test hypotheses, and assess results.

Using a web scraper, we gathered information from the Up work online job board. Our dataset is divided into two components. They are 1. Freelancer Details and 2. Job Description. Freelancer Details has some parts such as freelancer overview, Language Expertise, Education, skill and expertise and on the other hand job description has some features such as job Details and skill required. Total number of freelancer is 874 and total number of jobs is 874 and total number of documents is 1748 We saved the collected data in CSV format after collection.



8	position:Specializes in Business Applications Development	Job description:
9	position:Specializes in DevOps Engineering	Job description:
10	position:Specializes in DevOps Engineering	Job description:
11	position:Specializes in DevOps Engineering	Job description:
12	position:Specializes in DevOps Engineering	Job description:
13	position:Specializes in DevOps Engineering	Job description:
14	position:Specializes in DevOps Engineering	Job description:
15	position:Specializes in DevOps Engineering	Job description:
16	position:Specializes in Graphic Design	Job description:
17	position:Specializes in Graphic Design	Job description:
18	position:Specializes in Information Security	Job description:
19	position:Specializes in Information Security	Job description:
20	position:Specializes in Network Administration	Job description:
21	position:Specializes in Network Administration	Job description:
22	position:Specializes in Network Administration	Job description:
23	position:Specializes in Network Security	Job description:
24	position:Specializes in Network Security	Job description:
25	position:Specializes in Packaging Design	Job description:
26	position:Specializes in Presentation Design	Job description:
27	position:Specializes in Presentation Design	Job description:
28	position:Specializes in Solutions Architecture	Job description:
29	position:Specializes in Solutions Architecture	Job description:
30	position:Specializes in Solutions Architecture	Job description:

position:Specializes in Systems Administration
about:
I have more than 8 years in networking industries , 6 years in System Administration and hosting environment.

Here is the operating systems I can work with, Linux Based operating system like Redhat, CentOS, Ubuntu, Fedora...etc

My experience in his section include a wide range of control panels like Cpanel/WHM, DirectAdmin, PLESK, WebMin, Kloxo, CentOS Web Panel, WHMCS and Blesta.

I will be able to secure and optimize your servers with a list of items that I have prepared and tested many times on servers before, Also I will be able to fix problems with servers and transfer sites in zero down time.
Education:
Alex university
Bachelor of Engineering (BEng), Egypt
1996-2001
Languages:
English: Fluent
Arabic: Native or Bilingual
Skills & expertise:
Systems Administration Skills
Server Virtualization

Job description:
Set up NGINX, WHM, nameservers and migrate wordpress sites to a new dedicated server.

Want the system set up and be optimised for speed and security of the websites and easy installation of new websites when required.

Skills required:
Linux
NGINX
Web Host Manager

Figure 3.1.1: All type sample data

3.2 DATA PREPROCESSING

Data preprocessing is the process of transforming raw data into something that can be used by a machine learning model. It is the initial and most important step in developing a machine learning model. Not often do we have access to clean, well-formatted data while developing a machine learning project.

Tokenization is a technique used in natural language processing to break down sentences and paragraphs into simpler pieces that are easier to interpret. The collecting of data (a sentence) and its breakdown into comprehensible components are the first steps of the NLP process (words).

Tokenization and one hot vector are used for data preprocessing. Preprocessing categorical features for machine learning models is frequently done using one hot encoding. This kind of encoding generates a new binary feature for each potential category and gives each sample's feature a value of 1 for the category it originally belonged to.

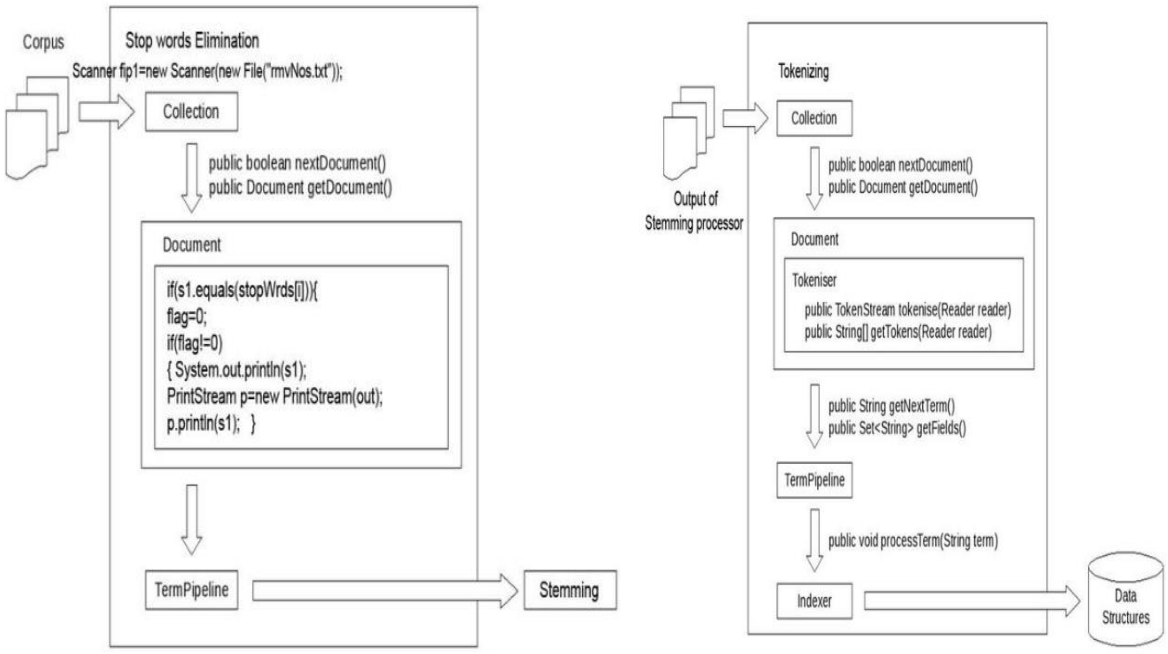


Figure3.2.2:TokenizationDiagram

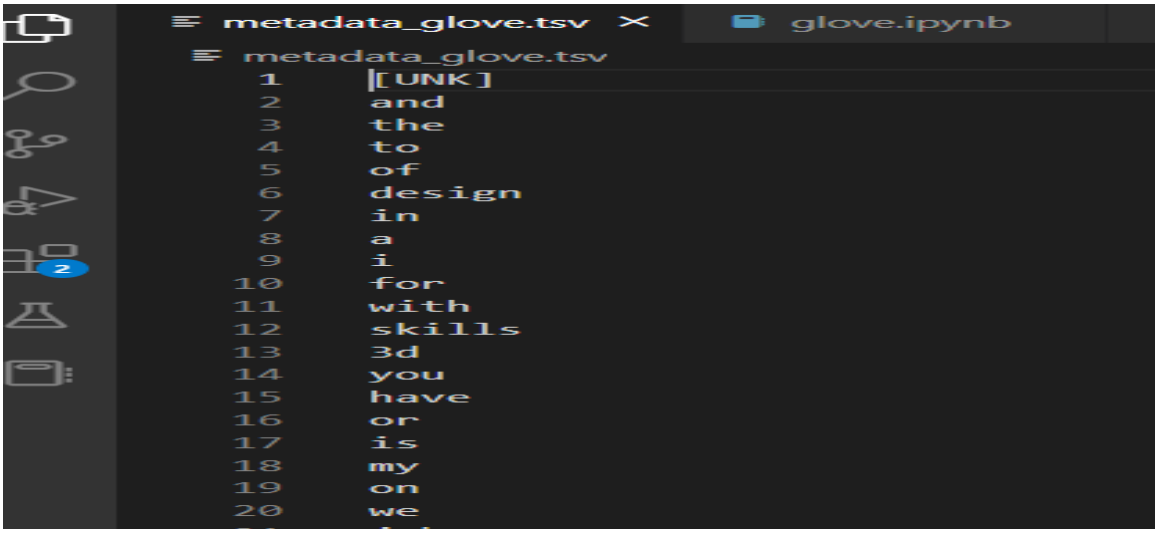


Figure 3.2.3: Tokenization output

3.3 LANGUAGE MODELLING

Glove Vector:

We have applied Glove Vector.

For the purpose of producing word vector representations, Glove is an unsupervised learning algorithm. The representations produced by the training process highlight intriguing linear substructures of the word vector space, which are based on aggregated global word-word co-occurrence statistics from a corpus.

Global Vectors for Word Representation is known as Glove. Researchers at Stanford University created an unsupervised learning method with the goal of creating word embedding by combining global word co-occurrence matrices from a particular corpus.

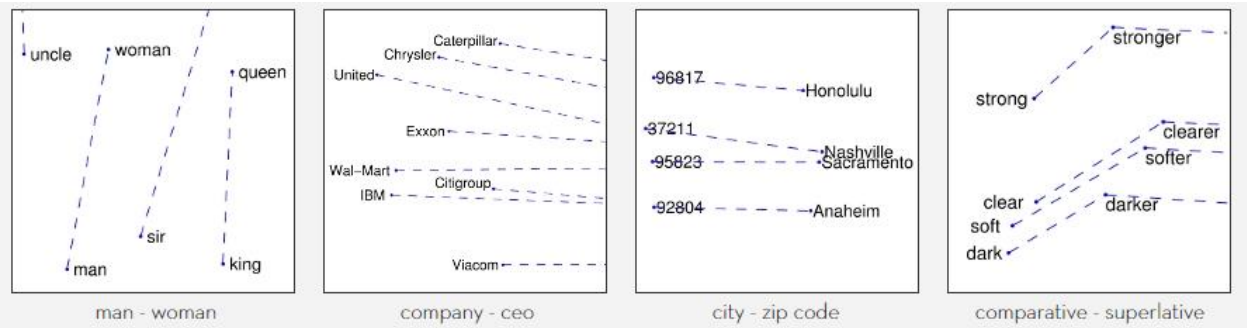
The foundation of the glove model is the use of global word-to-word co-occurrence counts throughout the entire corpus. On the other hand, Word2vec makes use of co-occurrence in local context (neighboring words). However, both of these models do many jobs similarly in practice.

The global word-word co-occurrence matrix, which tracks how frequently words occur alongside one another in a particular corpus, contains non-zero entries that are used to train the Glove model. It takes a single trip through the full corpus to gather the statistics needed to populate this matrix. This step can be computationally expensive for large corpora, but it only has a one-time upfront cost. Because there are often fewer non-zero matrix entries than there are words in the corpus, subsequent training iterations are completed considerably faster.

Linear substructures:

A single scalar that measures how closely two words are connected is produced by the similarity metrics used for nearest neighbor evaluations. The problem with this simplicity is that two given words nearly always reveal more complex relationships than can be expressed by a single number. Man and woman, for instance, could be seen as being similar in that they both describe human beings, yet they are frequently seen as being opposites because they draw attention to a key way in which people differ from one another.

It is necessary for a model to correlate more than one number with the word pair in order to adequately represent in a quantitative fashion the subtlety required to distinguish between men and women. The vector difference between the two word vectors is a straightforward and natural choice for an expanded set of discriminative numbers. Glove was created so that these vector changes would best be able to capture the meaning that was specified by the placement of two words together.



Other word pairs, such as king and queen or brother and sister, can also be used to indicate the fundamental idea that separates man from woman, namely sex or gender. To express this insight numerically, one may anticipate that the vector differences between brothers and sisters, kings and queens, and people would be nearly equal. The aforementioned collection of visualizations shows this feature as well as other intriguing patterns.

CHAPTER 4

4. RESULTS AND DISCUSSION

4.1 INTRODUCTION

the techniques and models for matching jobs with candidates. After gathering and preparing the data, I outlined the models I used for the Word embedding technique. In that post, I'll go into detail about the model's outcomes that I've used.

4.2 RESULT DISCUSSION

Today's technology surpasses our capacity for imagination. Additionally, there are many work options available today. Anyone can work remotely. The number of job opportunities has significantly expanded, but so too has the number of untrained individuals. So it is obvious that finding the perfect applicant for each position is quite difficult. But it's harder than it seems. People put a lot of work into finding the right person. Numerous systems, such as word2vec, the gloves system, and others, have been employed by researchers and authors worldwide. Some techniques also make use of cutting-edge technology. They collect data using a variety of systems. While some users used widely available worldwide data, others used newly developed databases.

After implementing the word embedding models and glove vector on our dataset, we have gotten the model’s accuracy and model’s loss plotting graph. Training accuracy and validation accuracy are compared for gaining the graph of all the model’s accuracy and model’s loss. our model accuracy 0.9879 and epoch loss 0.0868.

We thought about making an advanced feature which meets client’s requirement. We desired to make a ranking in which applicants skills matches with the job

requirement. As example, we will place the most deserving candidate at the upper side of the ranking. In this process deserving candidates will get priority and clients also will be able to choose their desired ones. Things will be a lot easier if we can implement this plan. But it is undoubtedly time consuming process so we could not make it yet. We have only finished word embedding. But we are consciously working on it. Maybe we will be successes to execute our plan in future, where we will make a ranking of deserving candidates whose skills matches with the client's requirements. We believe it Will be a lot easier and hassle-free process.

4.3 Result

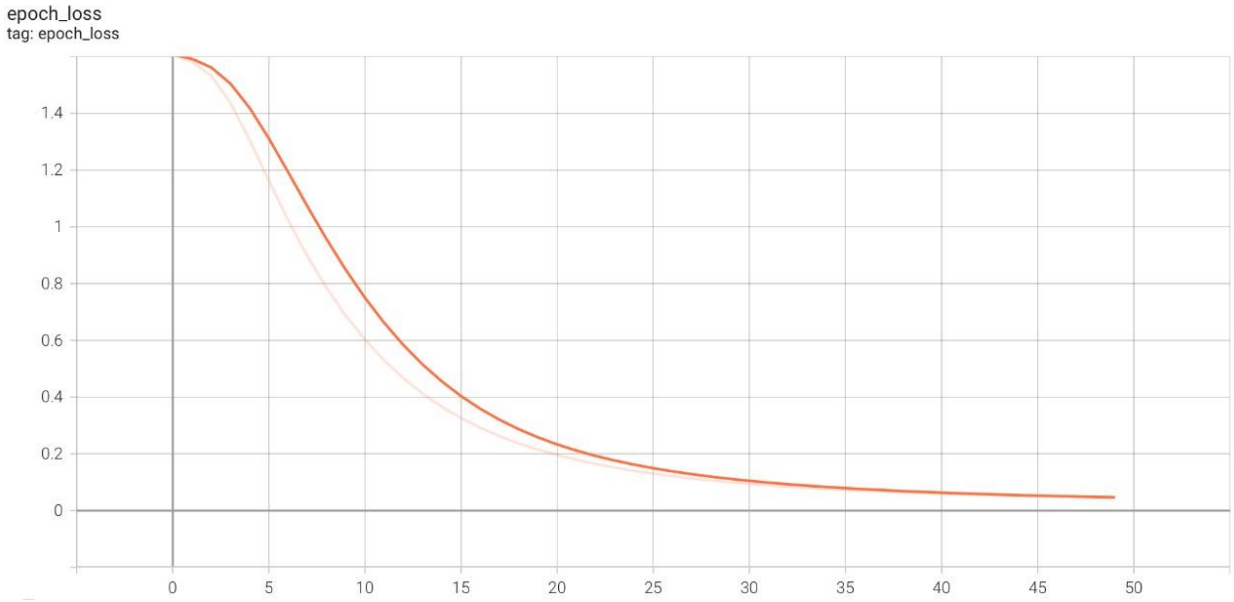


Figure 4.3.1: Model epoch loss graph

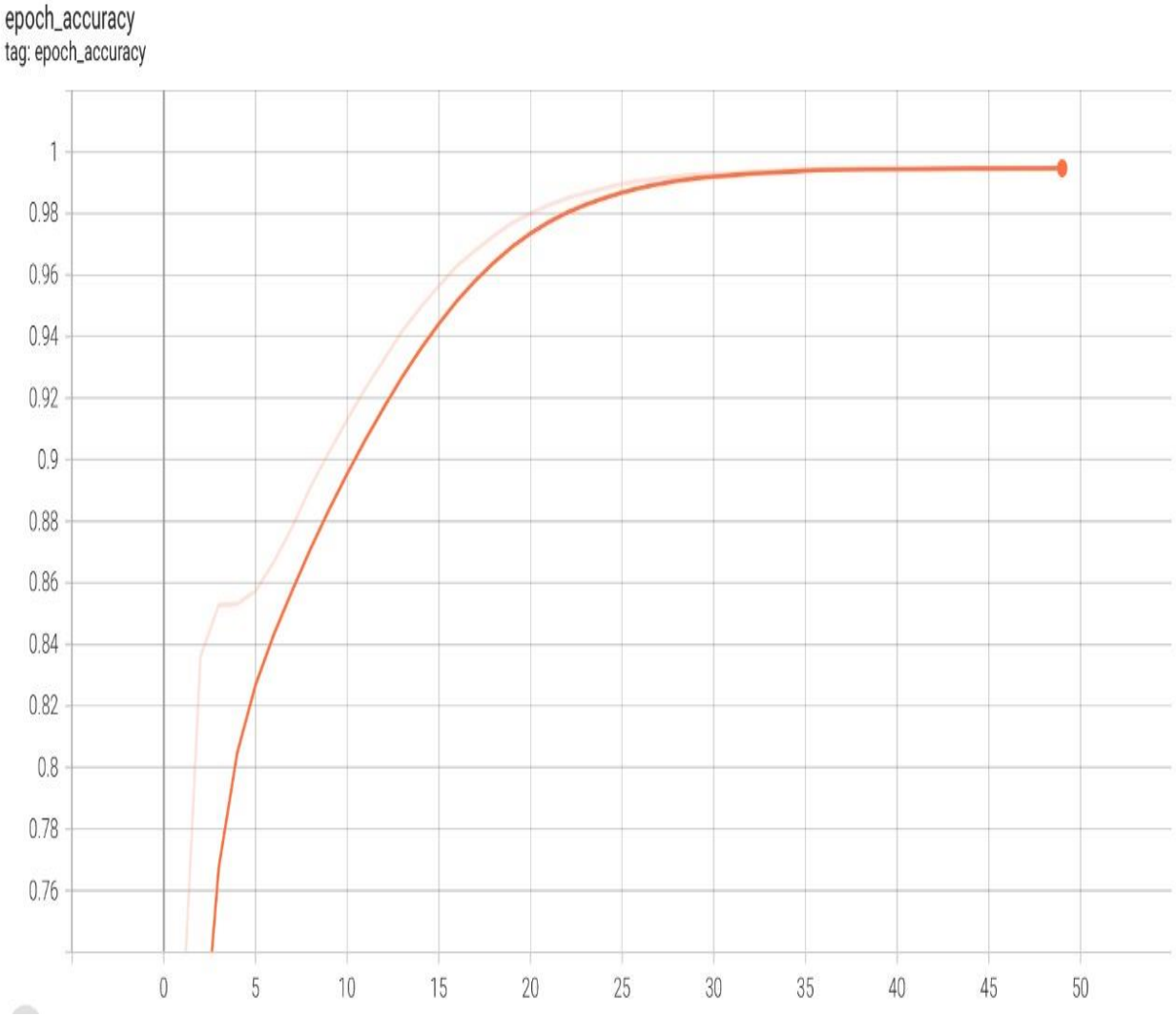


Figure 4.3.2: Model epoch accuracy graph

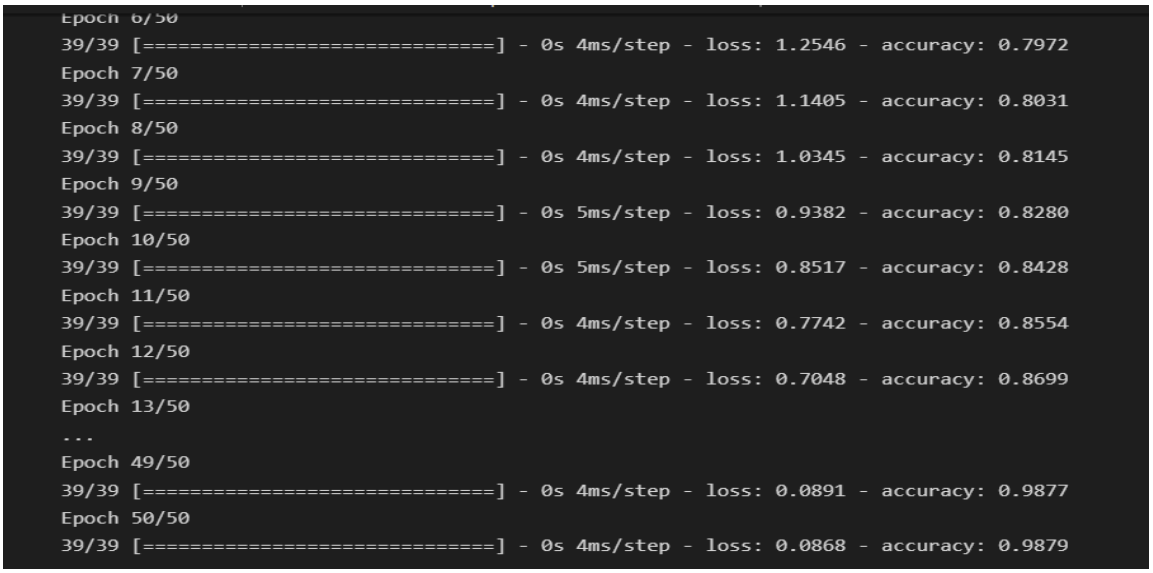


Figure 4.3.3: Model accuracy glove vector

CHAPTER 5

5. CONCLUSION AND LIMITATIONS

5.1 CONCLUSION

The vast amount of information available today presents a dilemma in the age of the internet. This is also true of the job market. On the one hand, the sheer volume of job advertisements that are published online every day makes it difficult and time-consuming for candidates to narrow down their job choices. On the other hand, given the deadlines for businesses, the effort of recruiting is made particularly difficult by the hundreds or thousands of applications that are sent in response to numerous job advertising. Our attempt in this work was to use machine learning to address these issues. The transactional database data from the job portal upwork is the source data that we use in this project. It is made up of numerous tables that contain data about applicants' outcomes as well as information about people's profiles (position, about, education, skill and expertise).

5.2 LIMITATIONS

We have already working on word embedding. But we could not do Ranking of the cv yet. We Will work on it later. If we will collect more data implement it will be very helpful.

CHAPTER 6

6. REFERENCES

1. Chala, S., & Fathi, M. (2017, March). Job seeker to vacancy matching using social network analysis. In *2017 IEEE International Conference on Industrial Technology (ICIT)* (pp. 1250-1255). IEEE.

2. Qin, C., Zhu, H., Xu, T., Zhu, C., Jiang, L., Chen, E., & Xiong, H. (2018, June). Enhancing person-job fit for talent recruitment: An ability-aware neural network approach. In *The 41st international ACM SIGIR conference on research & development in information retrieval* (pp. 25-34).

3. Grüger, J., & Schneider, G. J. (2019, September). Automated Analysis of Job Requirements for Computer Scientists in Online Job Advertisements. In *WEBIST* (pp. 226-233).

4. Kamaruddin, N., Rahman, A. W. A., & Lawi, R. A. M. (2019). Jobseeker-industry matching system using automated keyword selection and visualization approach. *Indonesian Journal of Electrical Engineering and Computer Science*, *13*(3), 1124-1129.

5. Homsapaya, K., & Budsara, K. (2020, June). Machine Learning for Older Jobseeker and Employment Matching. In *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 427-430). IEEE.

6. Hossain, M. S., & Arefin, M. S. (2019). Development of an Intelligent Job Recommender System for Freelancers using Client's Feedback Classification and Association Rule Mining Techniques. *J. Softw.*, *14*(7), 312-339.

7. Wang, C., Zhu, H., Hao, Q., Xiao, K., & Xiong, H. (2021, April). Variable interval time sequence modeling for career trajectory prediction: Deep collaborative perspective. In *Proceedings of the Web Conference 2021* (pp. 612-623).

8. Verma, A., Lamsal, K., & Verma, P. (2021). An investigation of skill requirements in artificial intelligence and machine learning job advertisements. *Industry and Higher Education*, 0950422221990990.

9.Faliagka, E., Iliadis, L., Karydis, I., Rigou, M., Sioutas, S., Tsakalidis, A., & Tzimas, G. (2014). On-line consistent ranking on e-recruitment: seeking the truth behind a well-formed CV. Artificial Intelligence Review, 42(3), 515-528.

10.Deng, Y., Lei, H., Li, X., & Lin, Y. (2018, May). An improved deep neural network model for job matching. In 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD) (pp. 106-112). IEEE.

11.Chala, S. A., Ansari, F., Fathi, M., & Tijdens, K. (2018). Semantic matching of job seeker to vacancy: a bidirectional approach. International Journal of Manpower.

12.Pombo, J. L. F. D. M. (2019). Landing on the right job: a machine learning approach to match candidates with jobs applying semantic embeddings (Doctoral dissertation).

13.Knaus, M. C., Lechner, M., & Strittmatter, A. (2022). Heterogeneous Employment Effects of Job Search Programs A Machine Learning Approach. Journal of Human Resources, 57(2), 597-636.

7. PLAGARISM REPORT

11/6/22, 3:13 PM   Turnitin - Originality Report - 183-35-2627

Turnitin Originality Report

[illegible]

183-35-2627 By Md. Ashikur Rahman

Similarity Index	Similarity by Source	
30%	Internet Sources:	23%
	Publications:	17%
	Student Papers:	14%

5% match (Internet from 26-Oct-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/9480/181-35-2297.pdf?isAllowed=y&sequence=1

2% match (Internet from 10-Jun-2022)
https://bigdata.uestc.edu.cn/paper_pdf/2021/Chao-Wang-WWW.pdf

2% match (student papers from 19-Sep-2022)
Submitted to Kingston University on 2022-09-19

2% match (Internet from 17-Sep-2022)
https://www.grafiati.com/uk/literature-selections/jobseekers/journal/

1% match (Internet from 05-Jan-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5693/171-35-1870%20%2824_%29.pdf?isAllowed=y&sequence=1

1% match (Kanyanut Homsapaya, Kridsada Budsara. "Machine Learning for Older Jobseeker and Employment Matching", 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2020)
Kanyanut Homsapaya, Kridsada Budsara. "Machine Learning for Older Jobseeker and Employment Matching", 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2020

1% match (Internet from 19-Jul-2020)
https://medium.com/@eggon.arif/linkedin-job-opportunities-in-relation-with-social-network-analysis-6bdd8634b187

1% match (Internet from 15-Jan-2020)
https://scitepress.org/Papers/2019/80682/pdf/index.html

1% match (Amit Verma, Kamal Lamsal, Payal Verma. "An investigation of skill requirements in artificial intelligence and machine learning job advertisements", Industry and Higher Education, 2021)
Amit Verma, Kamal Lamsal, Payal Verma. "An investigation of skill requirements in artificial intelligence and machine learning job advertisements", Industry and Higher Education, 2021

1% match (Sisay Chala, Madjid Fathi. "Job seeker to vacancy matching using social network analysis", 2017 IEEE International Conference on Industrial Technology (ICIT), 2017)
Sisay Chala, Madjid Fathi. "Job seeker to vacancy matching using social network analysis", 2017 IEEE International Conference on Industrial Technology (ICIT), 2017

1% match (Michael C. Knaus, Michael Lechner, Anthony Strittmatter. "Heterogeneous Employment Effects of Job Search Programs", Journal of Human Resources, 2022)
Michael C. Knaus, Michael Lechner, Anthony Strittmatter. "Heterogeneous Employment Effects of Job Search Programs", Journal of Human Resources, 2022

1% match (student papers from 11-Aug-2022)
Submitted to University of Limerick on 2022-08-11

1% match (Internet from 23-Jan-2022)
https://core.ac.uk/download/187234802.pdf

1% match (student papers from 12-Sep-2022)
Submitted to University of Surrey on 2022-09-12

1% match (Internet from 09-Oct-2022)
http://www.jsoftware.us/vol14/189-JSW15309.pdf

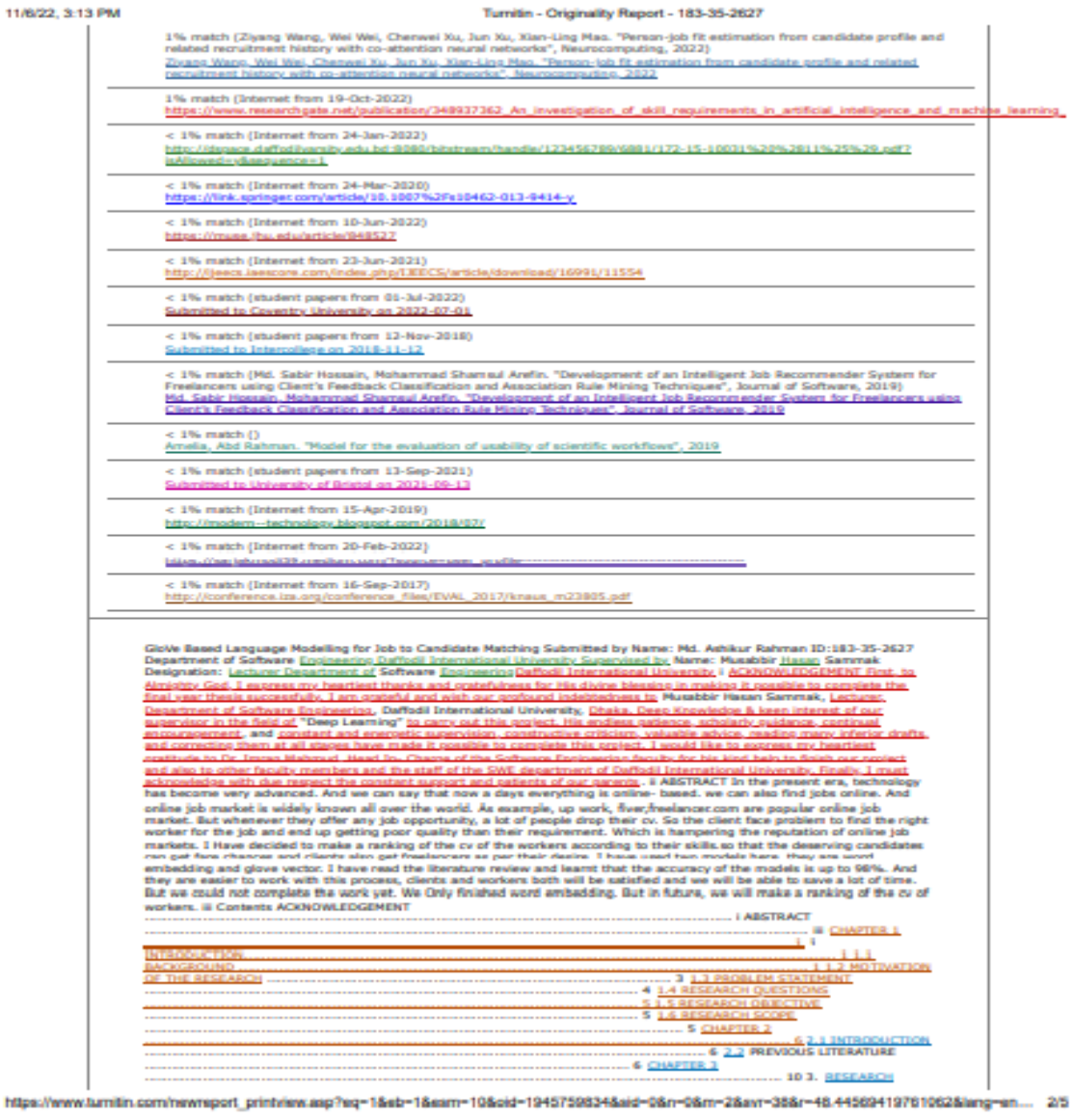
1% match (student papers from 04-Sep-2022)
Submitted to Liverpool John Moores University on 2022-09-04

1% match (student papers from 04-Sep-2022)
Submitted to Gitam University on 2022-09-04

1% match (Yu Deng, Hang Lei, Xiaoyu Li, Yiou Lin. "An improved deep neural network model for job matching", 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), 2018)
Yu Deng, Hang Lei, Xiaoyu Li, Yiou Lin. "An improved deep neural network model for job matching", 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), 2018

1% match (Internet from 01-Feb-2022)
https://deepai.org/publication/enhancing-person-job-fit-for-talent-recruitment-an-ability-aware-neural-network-approach

https://www.turnitin.com/newreport_printview.asp?eq=1&eb=1&esm=10&oid=1945759834&sid=0&n=0&m=2&svr=38&r=48.44569419761062&lang=en... 1/5



11/6/22, 3:13 PM Turnitin - Originality Report - 183-35-2627

1% match (Ziyang Wang, Wei Wei, Chenwei Xu, Jun Xu, Xian-Ling Mao. "Person-job fit estimation from candidate profile and related recruitment history with co-attention neural networks", Neurocomputing, 2022)
Ziyang Wang, Wei Wei, Chenwei Xu, Jun Xu, Xian-Ling Mao. "Person-job fit estimation from candidate profile and related recruitment history with co-attention neural networks", Neurocomputing, 2022

1% match (Internet from 19-Oct-2022)
https://www.researchgate.net/publication/348917362_An_investigation_of_skill_requirements_in_artificial_intelligence_and_machine_learning_

< 1% match (Internet from 24-Jan-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/6881/172-15-10031%20%2811%25%29.pdf?isAllowed=y&sequence=1

< 1% match (Internet from 24-Mar-2020)
https://link.springer.com/article/10.1007%2Fs10462-013-9414-y

< 1% match (Internet from 10-Jun-2022)
https://mpra.ub.uni-muenchen.de/article/948527

< 1% match (Internet from 23-Jun-2021)
http://ijaecs.iaescore.com/index.php/IJEECS/article/download/16991/11554

< 1% match (student papers from 01-Jul-2022)
Submitted to Coventry University on 2022-07-01

< 1% match (student papers from 12-Nov-2018)
Submitted to Interonline.e on 2018-11-12

< 1% match (Md. Sabir Hossain, Mohammad Shamsul Arefin. "Development of an Intelligent Job Recommender System for Freelancers using Client's Feedback Classification and Association Rule Mining Techniques", Journal of Software, 2019)
Md. Sabir Hossain, Mohammad Shamsul Arefin. "Development of an Intelligent Job Recommender System for Freelancers using Client's Feedback Classification and Association Rule Mining Techniques", Journal of Software, 2019

< 1% match ()
Amelia, Abd Rahman. "Model for the evaluation of usability of scientific workflows", 2019

< 1% match (student papers from 13-Sep-2021)
Submitted to University of Bristol on 2021-09-13

< 1% match (Internet from 15-Apr-2019)
http://modern--technology.blogspot.com/2018/07/

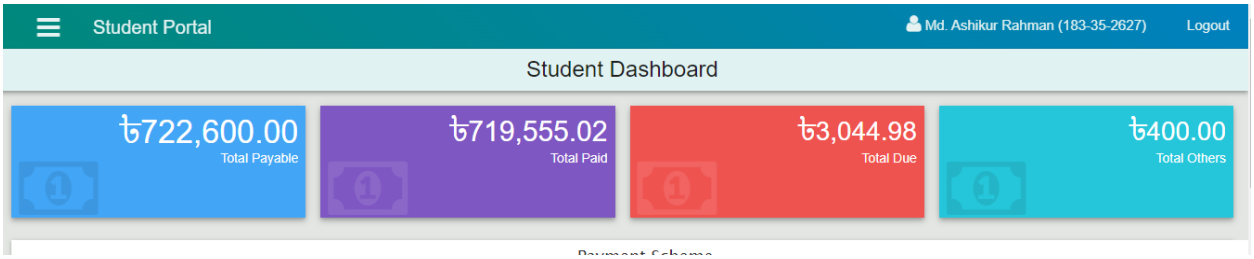
< 1% match (Internet from 20-Feb-2022)
https://www.researchgate.net/profile/

< 1% match (Internet from 16-Sep-2017)
http://conference.iza.org/conference_files/EVAL_2017/knaus_m23805.pdf

GloVe Based Language Modelling for Job to Candidate Matching Submitted by Name: Md. Ashikur Rahman ID:183-35-2627 Department of Software Engineering Daffodil International University Supervised by Name: Musabbir Hasan Sammak Designation: Lecturer Department of Software Engineering Daffodil International University. i ACKNOWLEDGEMENT First, to Almighty God, I express my heartiest thanks and gratefulness for His divine blessing in making it possible to complete the final year thesis successfully. I am grateful and wish our profound indebtedness to Musabbir Hasan Sammak, Lecturer, Department of Software Engineering, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Deep Learning" to carry out this project. His endless patience, scholarly guidance, continual encouragement, and constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project. I would like to express my heartiest gratitude to Dr. Imran Mahmud, Head In- Charge of the Software Engineering faculty for his kind help to finish our project and also to other faculty members and the staff of the SWE department of Daffodil International University. Finally, I must acknowledge with due respect the constant support and patients of our parents. ii ABSTRACT In the present era, technology has become very advanced. And we can say that now a days everything is online- based. we can also find jobs online. And online job market is widely known all over the world. As example, up work, fiverr,freelancer.com are popular online job market. But whenever they offer any job opportunity, a lot of people drop their cv. So the client face problem to find the right worker for the job and end up getting poor quality than their requirement. Which is hampering the reputation of online job markets. I Have decided to make a ranking of the cv of the workers according to their skills.so that the deserving candidates can get fare chances and clients also get freelancers as per their desire. I have used two models here, they are word embedding and glove vector. I have read the literature review and learnt that the accuracy of the models is up to 98%. And they are easier to work with this process, clients and workers both will be satisfied and we will be able to save a lot of time. But we could not complete the work yet. We Only finished word embedding. But in future, we will make a ranking of the cv of workers. iii Contents ACKNOWLEDGEMENT ..... i ABSTRACT ..... iii CHAPTER 1 ..... 1 1 INTRODUCTION ..... 1 1.1 BACKGROUND ..... 1 1.2 MOTIVATION OF THE RESEARCH ..... 3 1.3 PROBLEM STATEMENT ..... 4 1.4 RESEARCH QUESTIONS ..... 5 1.5 RESEARCH OBJECTIVE ..... 5 1.6 RESEARCH SCOPE ..... 5 CHAPTER 2 ..... 6 2.1 INTRODUCTION ..... 6 2.2 PREVIOUS LITERATURE ..... 6 CHAPTER 3 ..... 10 3. RESEARCH

https://www.turnitin.com/newreport_printview.asp?eq=1&esb=1&eam=10&oid=1945759834&sid=0&n=0&m=2&svr=38&r=48.44569419781062&lang=en... 2/5

8.ACCOUNT CLEARENCE



9. LIBRARY CLEARENCE