A review paper on resume screening using machine learning techniques

1st Rishabh Singh Department of Engineering and

Technology Sharda University Greater Noida, India

[2019564439.rishabh@ug.sharda.ac.in](mailto:2019564439.rishabh@ug.sharda.ac.in)

2nd Mallika Pal Department of Engineering and

Technology Sharda University Greater Noida, India

[2019565330.mallika@ug.sharda.ac.in](mailto:2019565330.mallika@ug.sharda.ac.in)

3rd Vishal Tomar Department of Engineering and

Technology Sharda University Greater Noida, India

[2019007585.vishal@ug.sharda.ac.in](mailto:2019007585.vishal@ug.sharda.ac.in)

***Abstract*— It is difficult to find qualified candidates for available positions, especially when there are a lot of candidates. Finding the right person at the right time can positively affect the success of the team. The laborious process of fair selection and pre-selection can be greatly simplified by an automated system for "sorting and comparing resumes", which will also speed up the selection and decision-making process. To create a summary form for each application, our system uses natural language processing to extract relevant data from unstructured resumes, such as skills, education, experience, etc. The selection work is simplified when all unnecessary materials are eliminated and recruiters can examine each resume in more detail in less time. Once this text extraction procedure is completed, the proposed solution uses a routing model with cosine similarity to match each resume to the job description.**

***Keywords—resume screening, job screening, classification, One vs Rest, KNeighborsClassifier***

1. INTRODUCTION

How to find the right people with the minimum number of online resources and in the shortest possible time is the main challenge facing the entire industry. Making a short list is a difficulty for the HR department, because currently there are many different vacancies available, as well as a large number of applications that come in. The situation is aggravated by the lack of a wide range of skills and thematic knowledge to improve efficiency. To make the whole process more efficient, there are three important obstacles that must be eliminated. Choose the best candidates from the crowd. Understand the candidates ' resumes and check that the candidates are suitable for the position before hiring them. Our project is aimed at providing a solution to the above problems by automating the process. The solution will help to find the right resume from large resume dumps; it will not depend on the format in which the resume was created and will provide a list of resumes that best match the job description given by the recruiter, presenting an automated form based on machine learning. The form uses the characteristics extracted from the candidate's resume as an input and finds his categories, then, based on the required job description, compares the classified resume and recommends to the HR department the most suitable profile for the candidate. Our system uses natural language processing to extract relevant information, such as skills, education, experience, etc., From unstructured resumes and thus create a summary form for each application. By removing all non- essential information, the selection task is simplified and simplified and recruiters can better analyze each resume in less time. After completing this text extraction process, the proposed solution uses the orientation model and uses cosine similarity to match each resume with the job description.

There are several resume analyzers which are present in the internet but they hidden fact behind them is their irregularity with classification of the resume which hinders on finding the correct resume of the candidates.

1. RELATED WORK

Prof suggested that every workplace receives a lot of applications, many of which are suitable for such a position, as the number of job seekers increases over time.[1] this presents a major challenge for hiring managers, who must limit resumes to the most qualified candidates. [2] Zhang provided a detailed review that included several protocols that the researcher has used in recent years for the recommendation system. They discussed how the recommendation system is widely used in real-time applications. The ontological mapping proposed by Kumaran as a method of selecting candidates was proposed in the presented expert works. It consists of three operational stages: the development of the ontology for the candidate, [3] the development of the ontology document for the job criteria, and finally the comparison of the two to identify suitable candidates for a particular position. In 2012, an automated system for selecting vacancies was proposed. .[4] analyzes various machine learning algorithms and uses Support Vector regression to create a list of qualified candidates for a particular position.[5] Witherington proposed to submit another document describing how information about candidates on social media (such as LinkedIn, Facebook, LinkedIn etc.) will be presented.) Can be used to make hiring decisions. [6] Eckhardt, in another approach proposed by the document, described a system based on collaborative filtering to recommend the most suitable candidates for the position. We also looked at the job part where hiring decisions were based on how compatible the team members were with the potential employer. .[7] Malinovsky in order to comply with the job instructions for candidates, we extract the appropriate skills and criteria from the resume material, which differ from other assignments. [8] Huang proposed a recommendation service mainly consisting of four types of collaborative filtering, content-based filtering, knowledge- based and hybrid approaches, and discussed in detail all kinds of different recommendation methods with their working principle.[9] the expectation maximization algorithm (M) was used by Malinovsky and al.to recommending a job takes into account both the candidate's resume and the job description of the enterprise. Giovanni explained that the process of returning modified or derived words to their word position is known as the proposed rule. To achieve this goal.[10] an elementary heuristic procedure is used, consisting of the endings of words, which often involves the removal of derived surnames.

**Table I: Related Work** III. MOTIVATION

|  |  |  |
| --- | --- | --- |
| **Research Paper Title** | **Model used** | **Remarks** |
| [1] The use of biodata for employee selection: Past research and future directions | Gaussian Naïve Bayes, Multinomial Naive Bayes, SVC with linear kernel and SVR with gaussian  kernel | SVR with gaussian kernel performed the best with RMSE of 0.5271 |
| [2] Recommender system application developments: a survey. | Lasso Regression, Random Forest Regression, Ensemble and Regression tree  tuned by depth | Random Forest performed the best with an accuracy of 85.4% on the training data |
| [3] Towards an automated system for intelligent screening of candidates for recruitment using ontology mapping | Simple Linear Regression, Multivariate Regression models, Ridge Regression and Lasso  Regression | A need to create a mix or an ensemble of these models is stated to increase the accuracy. |
| [4] Application of machine learning algorithms to an online recruitment  system | Support Vector Machines, Random Forest, Artificial Neural  Network | Support Vector Machine classifier can be said to be reliable with an  accuracy of 82% |
| [5] Alternative Sources of Information and the Selection Decision Making  Process | Linear Regression | Linear Regression gives out the minimum prediction error of 0.3713 |
| [6] Help to find the needle in a haystack: integrating recommender systems in an IT supported staff recruitment  system | Linear Regression, Boosted Regression, Forest Regression and Neural Networks | The weighted mean of the given techniques was taken into consideration to give most accurate results |
| [7] Matching people and jobs: A bilateral recommendation approach | Linear Regression, Polynomial Regression, Tree, Neural Network and  Support Vector Machine | Tuned Support Vector Machine performed the best with the evaluation ratio of 0.56 |
| [8] A survey of e- commerce recommender systems | Recommendatio n technique and boosting algorithm XG  Boosting | Bagging performed the best with R^2 value 99.4% |
| [9] Matching people and jobs: A bilateral recommendation  approach | Expectation Maximization (EM) algorithm and stemming | XGBoost can be used to deliver superior results |
| [10] A comparative study of stemming algorithms | Least squares support vector regression (LSSVR),  classification  and regression tree (CART) | Call tree should have been performed the best with RMSE value 0.99 |

The motivation of our review paper is to create a real-time problem solution from large companies and recruitment agencies who often receive thousands of resources every day. This situation is further aggravated by the high mobility of workers and in situations of economic distress, as many people are looking to get jobs. Thus, with our review paper screening task will be simplified and recruiters can better analyze each resume in less time. In order to obtain better results for the resume shortlist, it is necessary to investigate more competent approaches to matching the candidate and job description. Our proposed solution will select the best suitable candidates for a specific opportunity by correlating the main features of the applicant profile with the requirements specified in the job description.

* 1. CONCLUSION

In our review thesis we have noticed several issues which the peer researchers were not able to resolve them. Like in [2] “Recommender system application developments: a survey” used Random Forest algorithm in which the authors weren’t able to rectify the overfitting issues which extremely lowers the accuracy of the models and result for the same. Therefore, we get to know from them that using random forest regression technique will not solve the problem. Another research [4] “Application of machine learning algorithms to an online recruitment system”

They also faced the same issue; they were not able to resolve the overfitting issue up to the extent which entirely impacted their result. They got the result as 82% which can be improved further.

Keeping the issues which, the fellow researchers came across in our checklist, we will come up with the system which will be able to overcome several issues present in the literature survey. We will try to maximize our results by reducing the parameters for the error which will be entirely boost our accuracy of the model.

* 1. FUTURE SCOPE

In our review thesis we will develop a system which will automatically choose the best fitted candidates from the crowd of many applicants. The system will choose the candidates based on profile matching with the job description. Thus, the most suitable candidate is chosen over others based on the job description provided by the organization. The project will be based on the Natural Language Processing. We will be using several classifier algorithms for more optimized results which categorize under Natural Language Processing branch. The system will provide automation in screening of the resumes which automizes the process and hence results in less labor in finding good and suitable candidates in the industry. Our outcome will be able to choose the best candidate from the crowd of applicants who apply for the job description in big number and handling such a huge crowd is a hectic task therefore our system will come to rescue which can work in such circumstance and will be beneficial in saving managers efforts and time and they can rely on our systems result as it comes up with very good accuracy and performance.

There will be scope for more optimization and accuracy which could be increased further by using different machine learning algorithms for more better results which can lead to more efficient system with less time and space complexity. There can be more aspects of the project where one can also check the authentication of the candidates’ certificates and courses provided by him in the resume for analyzing candidate more efficiently and thus it will reduce the labor behind selecting the best candidate for the provided job profile.

* 1. SUMMARY

In our review article, we reviewed relevant research articles published in various journals on the subject of resume verification and identification. In the last published research paper [1], the technology used was able to work better than others, but still did not provide good relative value. Thus, these papers offer an approximate solution to the product problem with the results of average accuracy: in [2] “Recommender system application developments: a survey” used Random Forest algorithm in which the authors weren’t able to rectify the overfitting issues which extremely lowers the accuracy of the models and result for the same. Therefore, we get to know from them that using random forest regression technique will not solve the problem.

Other articles include how to detect applicants based on job descriptions [3]. Another research [4] “Application of machine learning algorithms to an online recruitment system”

They also faced the same issue; they were not able to resolve the overfitting issue up to the extent which entirely impacted their result. They got the result as 82% which can be improved further. So, it will be our priority to enhance the results in the near future and to resolve the issues as much as possible.

* 1. ACKNOWLEDGMENT

We would like to thank and express gratitude towards our mentors Mrs. Rani Astya and Dr. Anuj Kumar for helping us in all the research related work and guide us throughout the journey of it. Lastly, I would like to thank my family for their continuous support and efforts.

VII. REFERENCES

1. Breaugh, J.A., 2022. The use of biodata for employee selection: Past research and future directions. Human Resource Management Review 19, 219–231.
2. Lu, J., Wu, D., Mao, M., Wang, W., Zhang, G., 2021. Recommender system application developments: a survey. Decision Support Systems 74, 12–32.
3. Kumaran, V.S. and Sankar, A., 2021. 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.
4. Faliagka, E., Ramantas, K., Tsakalidis, A. and Tzimas, G., 2020, May. Application of machine learning algorithms to an online recruitment system. In Proc. International Conference on Internet and Web Applications and Services.
5. Weathington B.L. and Bechtel, AR, 2020. Alternative Source of Information and the Selection Decision Making Process Journal of Behavioral and Applied Management. 13(2).
6. Laumer, S. and Eckhardt, A., 2020, 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.
7. Malinowski, J., Keim, T., Wendt, O. and Weitzel, T., 2019, January. Matching people and jobs: A bilateral recommendation approach. In Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), 6, 137c-137c. IEEE
8. Wei, K., Huang, J., Fu, S., 2019. A survey of e-

commerce recommender systems, in: 2007 international conference on service systems and service management, IEEE. pp. 1–5.

1. Malinowski, J., Keim, T., Wendt, O., Weitzel, T., 2019. Matching people and jobs: A bilateral recommendation approach, in: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS’06), IEEE. pp. 137c–137c
2. Jivani, A.G., 2019. A comparative study of stemming algorithms. Int. J. Comp. Tech. Appl, 2(6), 1930-1938.
3. Paparrizos, I., Cambazoglu, B.B., Gionis, A., 2018. Machine learned job recommendation, in: Proceedings of the fifth ACM Conference on Recommender Systems, ACM. pp. 325–328.
4. Resnick, P., Varian, H.R., 2018. Recommender systems. Communications of the ACM 40, 56–59
5. Faliagka, E., Ramantas, K., Tsakalidis, A. and Tzimas, G., 2018, May. Application of machine learning algorithms to an online recruitment system. In Proc. International Conference on Internet and Web Applications and Services.
6. Thimma Reddy Kalva, Utah State University, 2013. Skill-Finder: Automated Job-Resume.
7. Based Framework for automatic resume quality Suhjit Amin, Fr.Conceicao Rodrigues Institute of Technology, 2019. Web Application for Screening resume,
8. Ashwini K, Umadevi V, Shashank M Kadiwal,Revanna, Design and Development of e Learning based Resume Ranking.
9. Riza tana Fareed, rajah V, and Sharadadevi kaganumat “Resume Classification and Ranking using KNN and Cosine Similarity” In 2021 International Journal of Engineering.
10. Sujit Amin, Nikita Jayakar, Sonia Sunny, Pheba Babu,

M. Kiruthika, Ambarish Gurjar, Web Application for Screening Resume, 2019 International Conference on Nascent Technologies in Engineering (ICNTE), DOI: 10.1109/ICNTE44896.2019.8945869.

1. Suhas H E, Manjunath AE, “Differential Hiring using Combination of NER and Word Embedding”, In 2020 International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277- 3878, Vol.9
2. Centre for Monitoring Indian Economy Pvt Ltd. (CMIE),2022. The unemployment rate in India.
3. Howard, J.L., Ferris, G.R., 1996. The employment interview context: Social and situational influences on interviewer decisions Xavier Schmitt, Sylvain Kubler, Jer my Robert, Mike Papadakis, Yves LeTraon University of Luxembourg, Luxembourg Replicable Comparison Study NER Software: StanfordNLP, NLTK, OpenNLP, SpaCy, Gate.
4. Y. Luo, Y. Wen, T. Liu, and D. Tao,“Transferring knowledge fragments for learning distance metric from a heterogeneous domain,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.