***Twelfth International Conference on Advances in Computer Science and Application – CSA 2023***

An Efficient and Automated System of Resume Screening Using Machine Learning Algorithms

Mallika Pal1, Rishabh Singh2, Rani Astya3, Vishal Tomar4

1-4Sharda University, School of Engineering and Technology, Greater Noida, Uttar Pradesh, India

Email: [2019565330.mallika@ug.sharda.ac.in](mailto:2019565330.mallika@ug.sharda.ac.in),{2019564439.rishabh }@sharda.ac.in, {rani.astya}@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. In this, we employed the one vs rest classifier and KNeighborsClassifier and got an accuracy of approximately 99%. Hence our proposed model performs substantially better than others and therefore it could be used for resume screening in industries as well as by individuals.**

***Index Terms*— *resume screening, job screening, classification, One vs Rest, KNeighborsClassifier***

I. 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 many 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 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.

*Grenze ID: 01.JMMB.1.1.23*

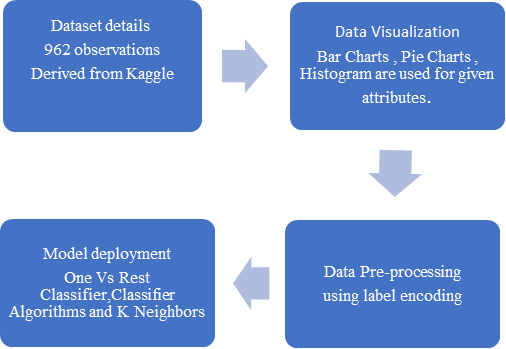
*© Grenze Scientific Society, 2023*

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 considers 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 in cutting off the endings of words, which often involves the removal of derived surnames.

III. PROPOSED METHODOLOGY

The aim of this work is to find the right candidates resume from the pool of resumes. To achieve this objective, we have developed a machine learning based solution. The complete work process can be divided into 4 segments. These are: Data Pre-processing, Data Visualization, Model deployment and Evaluating model's performance.



# Figure I. Methodology Overview

1. **Dataset Details**

The dataset has been derived from Kaggle. There are 962 observations we have in the data. Each observation represents the details of each candidate so we have 962 resumes for screening. We will split the dataset into training and testing set. We will use 80% data for training and 20% data for testing.

1. **Data Visualization**

Data Visualization is very important when it comes to machine learning and data science implementation. Data when being read in the form of graphs and charts, it becomes easier to comprehend and derive conclusions from it. Thus, it holds the key for getting accurate results. Python provides us with various visualization tools like matplotlib and seaborn. We’ve made use of numerous graphs like joint plots and box plots to establish key relationships between the attributes.

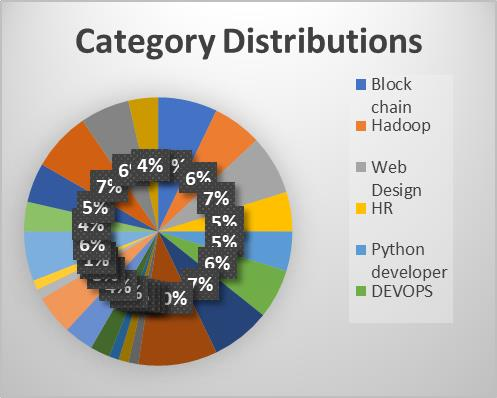
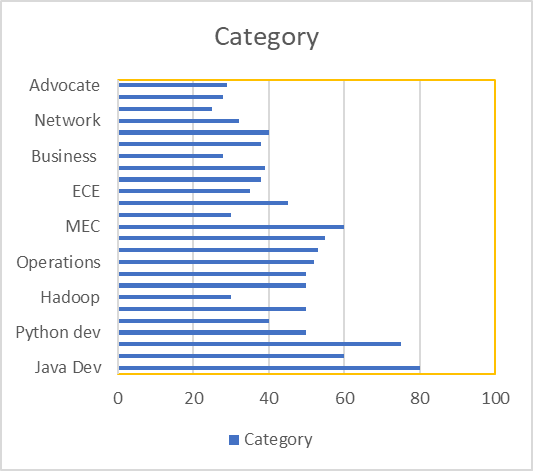


Figure II. Category DistributionsFigure III. Category Distribution Bar Chart

1. **Data Pre-Processing**

Data pre-processing is a technique which is used to convert unprocessed data into meaningful data. There are numerous steps associated with data pre-processing out of which we are going to be mainly focusing on three of them being:

Step 1: First, we will clean the ‘Resume’ column. In this step, we make sure to remove all unnecessary information from resumes such as hashtags, special encoded characters, and URLs.

Step 2: Through Label Encoding, we will encode the ‘Category’ column.

Step 3: For preprocessing the ‘cleaned resume’ we will convert them into vector. For doing such operations we can use ‘word2vik’, 'bag of words' or a combination of all methods.

1. **Model Deployment**

The proposed model has an implementation of the given machine learning techniques namely One vs rest classifier and KNeighbors Algorithms.

1. **One vs Rest Classifier**: It is also known as One vs All algorithm. It is an analytical method for the usage of dual categorization for multiple categorizations. It basically splits the dataset into dual categorization problems. A dual classifier is then instructed on each dual categorization problem and the model with the higher accuracy is used for predictions. This procedure creates n models for n classes and then that class which has the highest score is used for predictions. The scikit-learn library also gives us a discrete One vs Rest Classifier which permits to use the algorithm.
2. **KNeighbors**: It is one of the machine learning algorithms which is based on supervised learning. It is a sluggish algorithm because it does not grasp the information from the dataset rather it does some manipulations in the dataset during classification. It firstly arranges the data into present categories then it accumulates all new data and divides the data into well suitable group.

IV. RESULTS AND OUTPUTS

Evaluating a model’s performance is significant as it helps in making the decision as to which model performs better and then selecting that model. It expresses how a model performs in real time using a particular numerical value. We have used recall score, f1 score, precision score.

1. Precision score: It is the score which predicts the values which come out of positive classes only. The formula of precision score is given below:

**Equation I:** Precision = True positives / (True +False Positives)

1. Recall Score: It is the score which predicts the values or the numbers of positive classes which come out of examples of all the positives values in the dataset. The formula of recall score is given below

**Equation II:** Recall = True positives / (True positives + False Negatives)

1. F1 Score: It is the score which gives or balances the possibility of both the precision and recall in a single score. The formula of F1 score is given below

**Equation III**: F1 Score = 2 \* (((Precision \* Recall))/ (Precision + Recall))

Accuracy is a measurement used to determine the best performing model when it comes to identify the relationship between the features of a dataset. The training accuracy of our proposed model comes out to be 98.96% with the validation accuracy as 96.89%. This clearly shows a very high level of superiority when compared to the pre-existing literature. Given below are the results of 23 observations and based on results an accuracy has been calculated for the same.

|  |  |  |  |
| --- | --- | --- | --- |
| Precision | Recall | F1  Score | Total |
| 0 | 1.00 | 1.00 | 4 |
| 1 | 1.00 | 1.00 | 9 |
| 2 | 1.00 | 1.00 | 6 |
| 3 | 1.00 | 1.00 | 9 |
| 4 | 1.00 | 1.00 | 3 |
| 5 | 1.00 | 1.00 | 3 |
| 6 | 1.00 | 1.00 | 7 |
| 7 | 1.00 | 1.00 | 6 |
| 8 | 1.00 | 1.00 | 0 |
| 9 | 1.00 | 1.00 | 6 |
| 10 | 1.00 | 1.00 | 10 |
| 11 | 1.00 | 1.00 | 8 |
| 12 | 1.00 | 1.00 | 6 |

|  |  |  |  |
| --- | --- | --- | --- |
| Precision | Recall | F1  Score | Total |
| 13 | 1.00 | 1.00 | 6 |
| 14 | 1.00 | 1.00 | 8 |
| 15 | 1.00 | 1.00 | 16 |
| 16 | 1.00 | 1.00 | 7 |
| 17 | 1.00 | 1.00 | 6 |
| 18 | 1.00 | 1.00 | 8 |
| 19 | 1.00 | 1.00 | 8 |
| 20 | 1.00 | 1.00 | 11 |
| 21 | 1.00 | 1.00 | 4 |
| 22 | 1.00 | 1.00 | 9 |
| 23 | 1.00 | 1.00 | 13 |
| Accuracy | 1.00 | 1.00 | 193 |

# 

# Table I. Results

# V. CONCLUSION

In our project we have developed a system which will automatically choose the best fitted candidates from the crowd of many applicants. The system chooses 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 accuracy of the project is about 98 percent. This project is based on the Natural Language Processing. We have used classifier algorithms which are One Vs Rest and K Neighbors for more optimized results which categorize under Natural Language Processing branch. The system provides 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 project chooses 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 there our system comes 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.

VI. FUTURE SCOPE

Our project of resume selection helps in selecting the best and most qualified candidates applying for the job. We incorporated classifier algorithms like One Vs Rest and K Neighbors which gives us a better accuracy rate of about 98 percent. There is a scope of more optimization and accuracy can 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.

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.

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, A.R., 2020. Alternative Sources of Information and the Selection Decision Making Process. Journal of Behavioral & 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.
9. 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
10. Jivani, A.G., 2019. A comparative study of stemming algorithms. Int. J. Comp. Tech. Appl, 2(6), 1930-1938.