

**TITLE**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING,**

**SHARDA SCHOOL OF ENGINEERING AND TECHNOLOGY,**

**SHARDA UNIVERSITY, GREATER NOIDA**

**AN EFFICIENT AND AUTOMATED SYSTEM OF RESUME SCREENING USING MACHINE LEARNING ALGORITHMS**

***A project submitted***

***in partial fulfillment of the requirements for the degree of Bachelor of Technology in Computer Science and Engineering***

**by**

**RISHABH SINGH (2019564439)**

**MALLIKA PAL (2019565330)**

**VISHAL TOMAR (2019007585)**

**Supervised by:**

**DR. RANI ASTYA, ASST. PROF (CSE)**

May 2023

# CERTIFICATE

This is to certify that the report entitled **“An Efficient and Automated System of Resume Screening using Machine learning algorithms”** submitted by **“**Rishabh Singh (2019564439) , Mallika Pal (2019565330) and Vishal Tomar (2019007585)**”** to Sharda University, towards the fulfillment of requirements of the degree of **“Bachelor of Technology”** is record of bonafide final year Project work carried out by them in the **“**Department of Computer Science & Engineering, School of Engineering and Technology, Sharda University**”**.

The results/findings contained in this Project have not been submitted in part or full to any other University/Institute for award of any other Degree/Diploma.

### Signature of the Guide

**Name:** Dr. Rani Astya

**Designation:** Asst. Professor (CSE)

**Signature of Head of Department Name:** Prof. (Dr.) Nitin Rakesh

**Place:** Sharda University

### Date:

### Signature of External Examiner Date:

# ACKNOWLEDGEMENT

I would like to express my sincere appreciation to everyone who helped me in completing the final year report. First and foremost, I am grateful to our guide Dr. Rani Astya, for providing us the guidance, expertise, and feedback throughout the project. Her assistance was crucial in helping us overcoming the challenges of the research study.

I would also like to thank Dr. Nitin Rakesh for giving us the opportunity for working in the project reviewing my work and offering constructive criticism that helped improve the quality of this report.

Furthermore, I am thankful to the institution where I completed my studies for providing me with the resources and facilities necessary to pursue my academic goals.

Lastly, I would like to acknowledge the support and encouragement of my family, friends, and classmates, who have been a constant source of motivation throughout my academic journey.

I am grateful to everyone who played a role in making this project a success and for making my academic journey a rewarding experience."

Name and signature of Students

Rishabh Singh (2019564439)

Mallika Pal (2019565330)

Vishal Tomar (2019007585)

# 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.

**CONTENTS**

[TITLE i](#_bookmark0)

[CERTIFICATE ii](#_bookmark1)

[ACKNOWLEDGEMENT iii](#_bookmark2)

[ABSTRACT iv](#_bookmark3)

[LIST OF FIGURES v](#_bookmark4)

[LIST OF TABLES vi](#_bookmark5)

[CHAPTER 1: INTRODUCTION 1](#_bookmark6)

* 1. [Problem Definition 1](#_bookmark7)
  2. [Project Overview 2](#_bookmark8)
  3. [Expected Outcome 3](#_bookmark9)
  4. [Hardware & Software Specifications 4](#_bookmark10)
  5. Other Non-Functional Requirements………………………………………………………...4
  6. Report Outline………………………………………………………………………………..5

[CHAPTER 2: LITERATUR](#_bookmark13)E SURVEY………………………………………………………...6

* 1. Existing Applications………………………………………………………………………...6
  2. Existing Methodologies……………………………………………………………………...9
  3. Proposed System……………………………………………………………………………12
  4. Feasibility System…………………………………………………………………………..12

CHAPTER 3: SYSTEM DESIGN & ANALYSIS……………………………………………...14

* 1. Project Perspective………………………………………………………………………….14
  2. Functionality of the Project…………………………………………………………………15
  3. Performance requirements………………………………………………………………….15
  4. System features……………………………………………………………………………..18
  5. Methodology………………………………………………………………………………..22
  6. Testing Process……………………………………………………………………………..27

CHAPTER 4: RESULTS AND OUTPUTS…………………………………………………….31

* 1. Dataset and Performance Matrix……………………………………………………………31
  2. Web Snapshots of User’s Interface…………………………………………………………32
  3. [Admin Console……………………………………………………………………………..3](#_bookmark27)4
  4. [Multiplayer Game 40](#_bookmark28)

CHAPTER 5: CONCLUSION…………………………………………………………………..37

* 1. System Usability……………………………………………………………………………38
  2. Future Scope………………………………………………………………………………..39

CHAPTER 6: REFERENCES…………………………………………………………………..40

ANNEXURE I……………………………………………………………………………………42

ANEXXURE II…………………………………………………………………………………...43

**List of Figures**

|  |  |  |
| --- | --- | --- |
| **Figure no** | **Figure name** | **Page no** |
| 1 | Tasks performed by the application | 3 |
| 2 | Objective of the resume screener | 14 |
| 3 | Functionality of the resume screener | 16 |
| 4 | End user functions | 17 |
| 5 | Level of admin console | 19 |
| 6 | Resume checking on an organizational level | 20 |
| 7 | Resume checking on an individual level | 21 |
| 8 | Methodology Overview | 23 |
| 9 | Category Distributions | 24 |
| 10 | Bar Chart for Category Distributions | 24 |
| 11 | Snapshots of the project | 32 |
| 12 | Annexure I | 42 |
| 13 | Annexure II | 43 |

**List of tables**

|  |  |  |
| --- | --- | --- |
| **Table no** | **Table name** | **Page no** |
| 1 | Description of the existing applications | 7 |
| 2 | Pre-existing machine learning models | 10 |
| 3 | Results | 26 |

**CHAPTER 1**

**INTRODUCTION**

* 1. **Problem definition**

We all are aware with the fact that the seeking the right job is how much important for anyone in the firm. The selection of good resume is very much important for the business as the good and hardworking employees behind the company are the key to success for it to excel. During lockdown there were no physical interviews, so it was the resume only through which the company officials were able to configure out whether the applicant is right person for the designation. As the world is progressing to digital world, there are several chances of faulty resumes which cannot be captured easily, thus it becomes very difficult for the organization to find the right and deserving candidate for the job role.

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.

As we all know Artificial Intelligence and Machine Learning are ruling in every other industry, so why not in IT Firm. Jobs in Computer Science field is trending nowadays. There are numerous applications for resume screening. The development of new techniques for resume screening are increasing day by day. Natural Language Processing is taking resume screening to an enhanced level with the use of segmentation and categorization with high level of accuracy.

Recently, there has been extensive use of Natural Language Processing algorithms in the IT firm in the form of mobile applications, website and PWA apps. But the common fault in most of the applications/ website is they are not able to provide the correct guidance to the company officials regarding the resume of the applicant. So, our project is aimed at providing a solution to the above problem by automating the process. Hence, an enhanced implementation is discussed to get better results of resume screening.

* 1. **Project Overview**

We are making a website application that will be able to clearly guide the officials to determine whether the candidate is right for the job or not. It will be also be helpful for the candidate too as they can also address the issues present in their resume and can easily rectify them before sending it to company or any organization.

Resume Screening application is the natural language processing tool which will be able to perform following tasks:

* + - *Identify the skills*,

with the help of extraction, skills can be extracted from the mentioned projects (if skills are not mentioned by the candidate).

* + - *Identify the achievements*

with the help of extraction, achievements are extracted and presented in a very concise manner.

* + - *Determines the overall score of the resume*

with the help of NLP, it generates the overall score of the resumes based on the information provided in the resume. It generally ranges between 0 to 80.

* + - *Provides the data about the previously uploaded resumes*

with the help of python libraries, it creates and keep a track of all previously uploaded resumes. It also provides the visual data for the same in the form of pie chart.

Apart from above points it also helps the candidate for better analyzing of the resume by suggesting the improvements in the resumes. Therefore overall, we bring out a unique and useful application for all the needy people which don’t just give the verdict but also presents the data in a very intuitive way so that it can be clearly understood by the person using it.

Tasks performed by the application in the given below in the figure I.

Fig I: Tasks performed by the application.

* 1. **Expected Outcome**

The project revolves around how the screening of resumes can be simplified. It is a real-time application for all the working professionals as well as for students who are freshers. There are two important aspects of the project viz. overview of the previously uploaded resumes as it is a very handy tool for the HR department as from there they can easily compare the candidates on the basis of the score of the resume and skills they possess. It also presents the data in the form of pie chart so that it is easily readable and the user gets the option to download the csv file of it. Second most important aspect is it generates the score of the resume based on the skill set of the candidate which helps the organization in easy selection and elimination. Therefore overall, we bring out a unique and useful application for all the needy people which don’t just give the verdict but also presents the data in a very intuitive way so that it can be clearly understood by the person using it.

* 1. **Hardware and Software Specifications**

For the implementation of the project, following specification will be used.

**Hardware requirements**

* + - Laptop/Desktop that supports Windows / MacOS / Android OS / iOS or any other, with minimum 2GB RAM
    - Phone with android above Android 5.0 (Lollipop)
    - A browser which supports HTML, CSS & JavaScript
    - Stable internet connectivity.

**Software requirements**

* + - Language: Python, IDE: PyCharm or VS Code
    - Database: SQL
    - Python libraries like spacy, matlplotlib etc.
    - NLTK Model.
  1. **Other Non-functional Requirements**

Another non-functional requirement of the project include:

### Safety Requirements

### In the event of significant damage to a large section of the database caused by a catastrophic event like a disk malfunction. The recovery process involves retrieving a previously saved version of the database from storage and recreating a more recent state by executing the operations of authorized transactions from the saved log, up to the point of the failure

**Security Requirements**

Like many other applications, security systems require database storage. However, due to the unique demands of the security industry, vendors must select their database partner cautiously to avoid becoming locked in with a particular vendor.

**Software Quality Attributes**

* Availability: The screening of the resume sessions should be available on time given by the user or whenever the user wants to inspect the resumes.
* Correctness: The screening sessions should be proper.
* Maintainability : The admin should maintain the previously records of the previous resumes so that the user can upload more resumes and also the maintenance of the database is must
* Usability: The resume screening application should be user friendly and the UI should be simple and easy for the user to interact with and should satisfy a maximum number of customer needs of the resume screening.
  1. **Report Outline**

Chapter 2 focuses on the previous work done to highlight the existing applications in the domain of resume screening which make use of Artificial Intelligence and Machine Learning.

Chapter 3 focuses on the proposed model, how it proceeds, what are the requirements, who are the users and what all will be the methodology to build our solution.

Chapter 4 provides the results and experimental analysis of our system.

Chapter 5 finally concludes the whole paper and talks about the future scope of the proposed system.

**CHAPTER 2**

**LITERATURE SURVEY**

* 1. **Existing Applications**

With the increasing trend of Artificial Intelligence and Machine Learning, there have been various applications in the field of Job Sector which are entirely based on AI.

Artificial Intelligence in job sector is undergoing a modernization and revolutionization of its old technology through the incorporation of artificial intelligence. Numerous applications are now being designed to be AI- compatible for advanced usage. Likewise, several well-known applications are transitioning to AI, which is proving to be a game-changer for both app users and owners.

Some of the use case of AI in the Job Sector industry are as follows:

* + 1. **TALEO –** It is a software platform for managing talents which is powered by cloud management.
    2. **Greenhouse –** It is a cloud-based application which uses human Capital Management (HCM) software for hiring the deserving candidate.
    3. **iCIMS –** It is a cloud-based application which uses applicant tracking systems that helps the organization in the hiring of the right candidate.
    4. **Zoho Recruit –** It is a cloud-based application which uses applicant tracking system (ATS) which helps the organizations for their recruitment process and helps in the automation of the hiring workflows.
    5. **Jobvite –** It is a cloud-based recruiting platform which offers a large variety of features out of which ATS, CRM are key aspects of it. It also helps in the hiring of the deserving candidates for the designation.
    6. **Workday –** It is the HR management system which help companies and organization in the workflow of HR processes and helps in improving the HR score. It focusses on talent acquisitions. It is one of the popular choices for the business.

AI powered applications have simplified the work of organization and also helps them to cut down their resources so that it can be utilized in other ways. We will discuss some of the top AI based web and mobile applications which uses AI technology.

Table I: Description of the existing applications

|  |  |  |  |
| --- | --- | --- | --- |
| **S.**  **No.** | **App Name** | **Description** | **Key Features** |
| 1. | Taleo | Cloud based recruitment software | Resume screening, collaborative tools, automated workflows, resume parsing. |
| 2. | Greenhouse | An applicant tracking system | Customizable scorecards, scheduling and interview management. |
| 3. | iCIMS | A recruiting tracking system | Flexible and customizable, integration with HR systems. |
| 4. | Zoho Recruit | Cloud based Applicant tracking system | Seamless flow of data between other HR, integration with other Zoho apps |
| 5. | Jobvite | A recruiting tracking system | Customizable scoring, automatic candidate ranking, report analysis. |
| 6. | Workday | HR Management system | Customizable screening criteria, reporting and analytics. |

**Limitations:** Given below are the limitations of the above-mentioned pre-existing applications for resume screening.

1. **Taleo**
   * + 1. It has the limitations for customization as per the organization’s specific needs. This extent is limited to all other options present within the software.
       2. Taleo is very costly, it is very costly for the medium sized business as well as for some larger organization. They also charge for the additional features which can be essential for the business. The price ranges between $20,000 - $1,00,000 per year.
       3. It is not easy to understand, the UI component appears to be complex. This makes challenging for someone who is not as efficient to use it.
       4. Limited support for the users as there are several users who have reported about the lack of support from Taleo team.
2. **Greenhouse**
3. Limited Integration with other important HR software tools
4. Complex for new users.
5. Limitations in reporting options which make users unable to export data to external tools.
6. Limited customization as the users cannot customize the system according to their needs.
7. Costlier for medium sized business. It starts from $6000 and is increased on the basis of usability.
8. **iCIMS**
9. Limited functionality for the creation of those report which is generated in the level of detail provided.
10. Limited Integrations as users fails to find quality integrations with other HR management platforms.
11. Although the UI of iCIMS is intuitive and easy to understand but still there are various features in which users fails to actually use them.
12. It is very costly and it increases based on the number of users or job applicants. It basically starts from $4000 - $5000 per year which is too much for small and medium sized business.
13. **Zoho Recruit**
14. Limitations in reporting options which make users unable to export data to external tools.
15. Limited support for the users as there are several users who have reported about the lack of support from Taleo team.
16. Learning curve is difficult for the users as they are not familiar with the Zoho protocols.
17. Limited features in Applicant Tracking System makes the software lack behind other software present in the market.
18. As Zoho provide integrations with other HR management tools but still there is more work to do in it by the Zoho team.
19. **Jobvite**
20. It is very costly for the small and medium sized business. Several users have reported that the prices for the software is comparatively higher. It starts from $4800.
21. Learning curve is difficult for the users as they are not familiar with the Zoho protocols.
22. Limited Integrations as users fails to find quality integrations with other HR management platforms.
23. Limitations in reporting options which make users unable to export data to external tools.
24. **Workday**
25. Costing is a major issue as there are several users on the internet who have reported tha the pricing is not transparent, making the business difficult to adjust their budget for the software. Because of the non-transparency we are also not able to collect the correct data from the internet.
26. Limited Support issues as the users finds difficulty in finding the solution to their issues.
27. Limited Customizations.
28. Although the UI of Workday is intuitive and easy to understand but still there are various features in which users fails to actually use them.
    1. **Existing methodologies**

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

Given below is the table which gives the brief overview of all the recent machine learning techniques.

Table II: Pre-existing machine learning models

|  |  |  |
| --- | --- | --- |
| **Research Paper Title** | **Approaches/Models used** | **Remarks** |
| [1] Biodata usage for selection of candidates. | Gaussian Naïve Bayes, Multinomial Naive Bayes, SVC with linear kernel and SVR with gaussian kernel | Gaussian kernel performance works better with 0.5721 RMSE along with SVR. |
| [2] Survey of recommending application system | Lasso Regression, Random Forest Regression, Ensemble and Regression tree tuned by depth | An algorithm which is Random Forest is performing here by giving accuracy of 85.4 percent on the trained data retrieved, |
| [3]Recruiting candidates using mapping of ontology for screening. | Simple Linear Regression, Multivariate Regression models, Ridge Regression and Lasso Regression | For increasing of accuracy there is a need of mixing of data or putting the models together. |
| [4] Machine Learning application for selecting deserving applicants | Support Vector Machines, Random Forest, Artificial Neural Network | A classifier named support vector machine will be more useful if the accuracy would have been up to 82 percent. |
| [5] Gaining information from different sources and making correct decisions | Linear Regression | A model or approach named linear regression gives less prediction of approximately 0.3713 |
| [6]Recommending system in IT industry for choosing candidates | Linear Regression, Boosted Regression, Forest Regression and Neural Networks | The mathematical function mean is taken into account for giving more meaningful predictions and results. |
| [7] Jobs matching with people for choosing right candidate | Linear Regression, Polynomial Regression, Principal Component Analysis, Regression Tree, Neural Network and Support Vector Machine | Vector machine which was supported by tuned best evaluated and performed by 0.56 ratio. |
| [8] E-Commerce recommender system: A Survey | Recommendation technique and boosting algorithm XG Boosting | An approach named bagging performs best with giving accuracy of about 99.4 percent |
| [9] Approach of bilateral recommendation | Expectation Maximization (EM) algorithm and stemming | Algorithms like XG Boost and such type of more algorithms are used to give superior results. |
| [10] Study of different algorithms of stemming | Least squares support vector regression (LSSVR), classification and regression tree (CART) | Algorithm named call tree would have performed with more accurate RMSE values of about 0.99 |

* 1. **Proposed System**

The web application is basically referred to as “Resume Screener” which can be used for the screening of resume for better analysis. It provides the following features to the end users:

### Admin Console

* + It can be managed by any organization representator as an end user by simply creating a login password.

### Uploading the resume

* + The pdf miner helps the user to upload the pdf of the resume in a very easy way. The resume screener will process the resume and extract all the relevant information mentioned in the resume for e.g. Skillset, years of experience, certifications and courses completed, education etc.
  + According to the information obtained from the resume the software will rate the resume based on the excellence of the person mentioned on the resume.

### Report Analysis

* + Based on the different resumes it compares the candidates and show the data in the form of pie chart and bar chart so that there should be a better comparison among different candidates in terms of experience levels, education and skill set.
  1. **Feasibility Study**

As previously discussed in the existing applications, there are several websites/applications that provide us the features of screening of resume and other applicant related system and these are the global apps which have control on the market, but as mentioned they are several issues in which they lack behind. These applications fulfill about 70% requirements which opens the gate for opportunity of improvements in the field. Our product tries to cover the issues present in the internet by proper screening and make the experience of hiring better. Beneficial features which help the users are:

* **Simple and User Friendly:** Our website is simple and easy to use. Even beginners can use the application easily and can enjoy all the functionalities within the system.
* **Resume Screening:** The pdf miner helps the user to upload the pdf of the resume in a very easy way. The resume screener will process the resume and extract all the relevant information mentioned in the resume for e.g. Skillset, years of experience, certifications and courses completed, education etc. According to the information obtained from the resume the software will rate the resume based on the excellence of the person mentioned on the resume.
* **Outcome:** After uploading is done, it will categorize all the data present in the resume and represent the data in a very intuitive way. It will determine the score of the resume based on the data present in the resume uploaded by the user.
* **Report Analysis:** In the admin console, there is complete record of all the resume previously uploaded which is maintained with the help of sql database which helps in the comparison of all the resumes to select which resume can be considered for the selection. The finally it will showcase the data in the form of pie chart which is easily readable as we all know that data can be better understood if the data is visualized. The report then can be easily downloaded for future usage.
* **Enhanced Classification Models:** The proposed model is created in such way that can be used for every person with different body proportions. It proved to provide accurate results to the users.

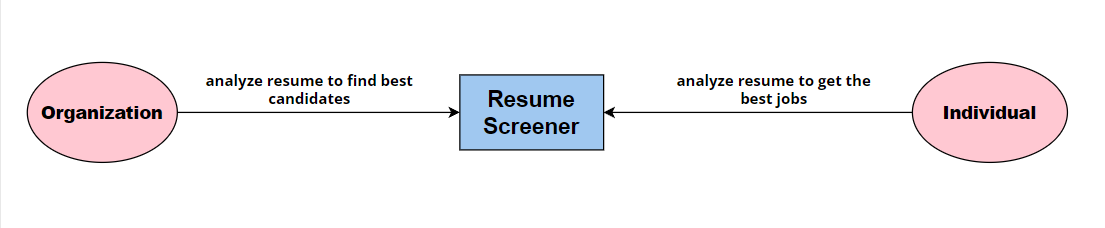
Thus, in this way, our product overcomes the problems faced by other apps in the existing resume screening applications present on the internet. This project also discusses a new approach of natural language processing techniques to make some advancements in the old ones. The new approach captures the user’s entered data present in the resume and comes up with the better accuracy for the same.

# CHAPTER 3

**SYSTEM DESIGN & ANALYSIS**

## **3.1 Project Perspective**

It is very difficult to find the qualified candidate for the organization. This problem is faced by todays IT firms in finding good quality of applicants or candidates who can perform brilliantly in the company and can lead to the betterment of the organization of the company. Our project mainly focuses on those organization which are in the hunt of good candidates. Thus, our project solves real world problem of appointing good candidates in the company. Our project chooses best candidates for the job post by matching their profile with the job description and thus there will be less hard work and labour in selecting the best candidates. 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.

 Fig II: Objective of resume screener

## 

## The overall perspective of resume screener is to find the best candidate among a huge number of candidates in the organizational point of view . In organization it is the most important task to find the best suited candidate for a job profile in order to get the best results for the company which itself will increase the reputation and clients for any organization.

In the individual point of view, it is important for anyone to get the best suited job according to the resume and job profile which will increase the skill set of any individual and also helps in the workplace which is best suited according to the personal interest of any individual. Therefore, our research project solves all the problems and caters to the need of both the organization and individual.

## **3.2 Functionality of the Project**

The admin console can control the login window. A proper system of password and User Id is provided to the controller for the authentication purposes. It means only he or she can control the software window keeping in mind the internal operations hidden of the organization. Thus, reducing the risk of information leak and data breaching. Hence the end user is the main entity to run the software resume screener the following points justify the features obtained by the end user through the software .

Through the software the end user has full control over the resumes submitted by the candidates for the job descriptions. The end user will get the full screening reports after uploading the resume in a simple manner which is very easy for any individual to perform thus it makes the task much easier to perform in less time.

## **3.2** **Performance Requirements**

Understanding the number of people involved while handling the software and the functionality of the project is the first step. Before proceeding further, we must know the actors and the functionality of the system there are mainly an admin console for the end user to obtain full length analysis of the resume.

* **End user:** The user should be able to do the following functions:
  + - * + Log into the application
        + Upload various resumes
        + Obtain complete analysis of resume

**3.2.1 Admin Console**: It can be managed by any organization representator as an end user by simply creating a login password. The resume can then be simply uploaded for complete analysis of the candidates resume based on the description.

The below diagram Fig: depicts in the data flow diagram form the way by which the resume screener functions internally making in use of the trained data which is set to analyze the resume in a better way in order to get the best rated resume and hence it will help the organization.

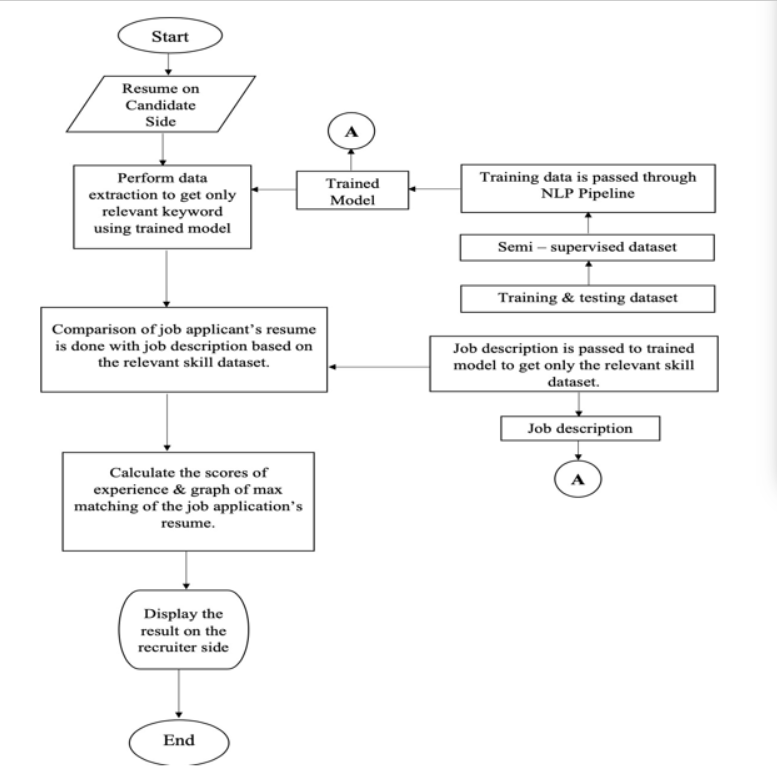
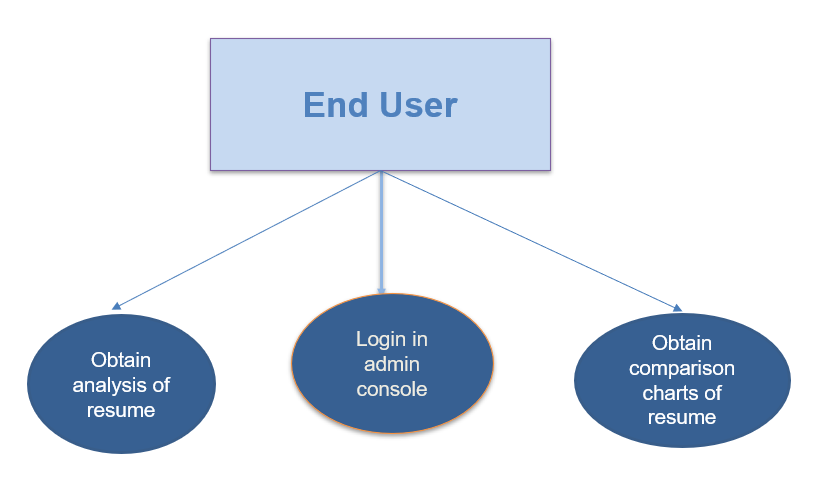


Fig III: Functionality of resume screener

* **End user:** The user should be able to do the following functions:
  + - * + Log into the application
        + Upload various resumes
        + Obtain complete analysis of resume
        + Rating based on experience and level of skills
        + Comparisons charts on various parameters

Following is the illustration for the same given below in the figure IV:

  
 Fig IV: End user functions

## **3.3.2** **Features of Admin console window**

* **Signing in as a user**: It provides the facility to create a profile of any end user by creating a password and a user id so that the software processes remain in safe hands and should be limited to the organization representative.
* **Uploading a resume:**

1. Every time an end user must upload a resume. He/She undergoes few steps in every resume upload session.
2. The end user can directly access the computer’s memory in which the submitted resume of the candidates according to the job profile are saved.
3. The resume which must be analyzed should be directly pick from the source by just clicking on it.
4. Soon the resume will upload on the software,
5. The scanning starts after a few second of the upload which makes it a fast and reliable software for resume screening.
6. The resume screener will process the resume and extract all the relevant information mentioned in the resume for e.g. Skillset, years of experience, certifications and courses completed, education etc.
7. The software will also segregate the information into meaningful information of the utmost importance for the job description.
8. According to the information obtained from the resume the software will rate the resume based on the excellence of the person mentioned on the resume.
9. The software also takes into account different resumes uploaded by the end user.
10. Based on the different resumes it compares the candidates and show the data in the form of pie chart and bar chart so that there should be a better comparison among different candidates in terms of experience levels, education and skill set

## **3.4 System Features**

The system features of our software categorize for two different type of end user so as to provide them an easier and hassle-free way of screening and analyzing resume which will surely help them in their career path.

The first and foremost feature of the system is that from a bunch of resumes submitted by the candidates when passed through the software it handles them properly and without much laborious work it analyzes the resumes in a very less and efficient way by making it easier for the organization to choose the best candidates among thousands of the ones who submits the resumes based on the job profile. And also, it should help to analyze the resume on an individual basis so as to help them in getting the best jobs offered in the market. Thus, our system provides an easier and reliable approach to work on different kinds of resumes.

The pdf miner library helps the user to upload the pdf of the resume in a very easy way. The resume screener will process the resume and extract all the relevant information mentioned in the resume for e.g. Skillset, years of experience, certifications and courses completed, education etc. According to the information obtained from the resume the software will rate the resume based on the excellence of the person mentioned on the resume.

Fig V: Levels of Admin console

In this system the user just must upload the resume from the admin console after logging into his/her profile and then uploads various resumes of the candidates and easily checks out the details of the candidates without looking at the details. The software itself segregates the information and handles the data carefully by observing and depicting meaningful results from the data in the form of pie charts and bar graphs.

It also rates a candidate’s profile by some benchmark’s which the system has set according to the model training done over it. It makes the selection process of the organization for a designation in their company very easy. Thus because of this software the hiring process of the companies will run on a smooth and hassle-free way in which there is a tendency of choosing the best candidates because of the detailed overview and analysis on a person’s profile. Also, the comparison which the software draws by comparing the resumes and profiles of the candidates and depicts it in the form of charts makes it easier to choose the top most efficient candidates for the respective position. Here is also a sequence in which a resume information is obtained after uploading in sequential steps. Illustration for the same is given below in the figure VI.

Fig VI: Resume checking on an organizational level

## **3.4.1 Checking Resume on an individual level**

Many a times a candidate who is preparing his/her resume for any job profile then it becomes difficult to judge the relevance of the entities present in the resumes fulfills the job criteria or not. It can also happen that the resume needs more correction on a personal level to make it more appealing for the organization and to get selected for the job profile positively. This task seems daunting when a person does not have an experienced senior and it is difficult to take the help in such cases. The person loses hope and many at times it has been seen that a person gets disqualified or did not get the preferred job role just because of the resume. Thus, the importance of a resume is so high in the job procedure because of the following reasons:

* A resume is an initial step for any job procedure.
* First selection round in most of the organization is mostly done based on resume.
* Resume builds an identity of a person’s career till the date.
* All the achievements and skills are presented in the resume in a lucrative manner
* A resume is continuously been in the process even after getting a job to get hikes in the salary or to switch a job it is very important to keep updating the resumes.

Thus, the importance of resume is on a high scale in any individuals career path and it always helps a person to gain the best opportunities in a career and to always gets a chance in the race of the job seeking where there is a huge chunk of resumes submitted by the candidates on a daily basis to the organization. There is a lot of competition because of less jobs and more candidates applying for job roles in the market thus there is very less chances of a candidate to get selected because of an average resume.

Problems faced by an individual while making resumes:

* Template of the resumes are easily available on the internet but how to fill the information is the main concern
* An individual always finds it hard to analyze the resume that it fulfills the targeted job descriptions or not.
* The certifications and skills are at par with the job profile or not.
* Rating the resume on the basis of content present in the resume.
* Suggested courses and skills for improvement of the resume.

The above-mentioned problems are targeted and solved with the help of our project software because of the functionality of individual analysis of resume for getting better opportunities is a functionality added in our software to not only limit the usage of software on the organizational level but also to an individual level to get easy analyzed resume and rating with betterment suggestion to grasp the most wanted job easily and also to get a standalone opportunity and to get better with work and to make life easier and hassle free because the resume will get more efficient and better while using the software. Illustration for the same is given below in the figure VII.

Fig VII: Resume checking on an individual level

The above figure shows the steps of the process in which the end user is an individual who is applying for the job but he/she is confused that whether the resume is at par with the job description or not and does it need some changes and improvements but due to less availability of such type of software’s it is difficult to seek help online and thus a person is dependent on only offline sources for example discussing with seniors and analyzing the resume based on external factors.

Our project software works efficiently in searching for tech related job profiles as in today’s digital era when everything is moved to a digital world there are more employee graph of working in technology domain. Therefore, to cater such type of need we knew that the job profiles which a candidate is searching or an organization appoints will be of more for technical roles as that is what going on in the relevance of today’s era. Thus, if a person has prepared the resume on the basis of technical job post, then a greater advantage to them will be provided as the system will analyze better according to the training rounds of the data. The software will be more efficiently caters to the skill sets and experience level of the individual.

The roles for which the system shows much more efficiency are:

* Web Developer
* Application Developer
* Data Analyst
* Network Engineer
* Machine Learning Developer

Thus, a person belonging or applying for such job position can very easily analyze their resume as the software will efficiently rate the resume on the basis of the skill set given in the resume also it can suggest different courses and certifications to pursue so that there should be an increase in the rating of the resume by adding a greater number of courses and skill set.

In the organizational front the system analyzes the resume of different candidates more efficiently and compares them for these job profiles. The system software caters to the different needs of the people according to the job roles and the tendency of the software is to provide the best results efficiently without much efforts. Thus, it makes this daunting task very easy to operate and provides very good accuracy on the data set.

## **3.5 Methodology**

The research methodology lies at the core of our research paper implementation. It can be summarized with the help of flow chart given below. The complete work process can be divided into 5 segments.

The methodology follows the methods of machine learning algorithms in which data is trained and processed so as to obtain meaningful results.

These 5 segments are:

* Introduction to the Dataset
* Data Visualization
* Data Pre-processing
* Model deployment
* Evaluating model's performance.

Overview of the Methodology is given below in the figure VIII.

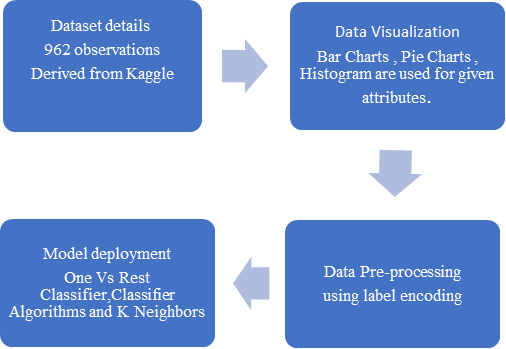


Figure VIII: Methodology Overview

## **3.5.1 Dataset details:**

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

## **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 bar charts and pie charts and histogram which also help in data pre-processing. Illustration for the same is given below in the figure IX and X.

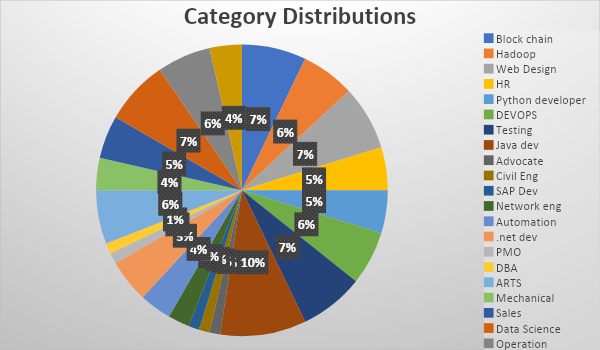


Figure IX: Category Distributions

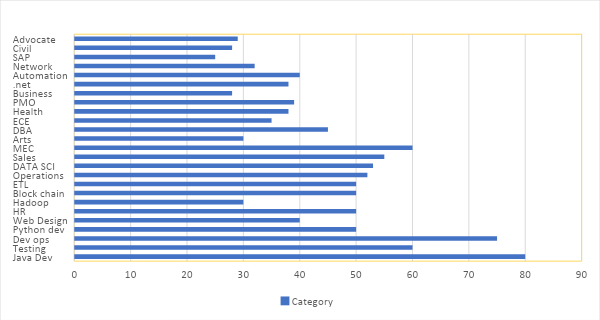


Figure X: Bar Chart for Category Distributions

## **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’re 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 LabelEncoding, 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.

## **Model deployment:**

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

## **Results and Outputs**

**Proposed Model Outputs**

Evaluating a model’s performance is of significant importance 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:

  Precision = True positives / (True positives + False Positive)

**Equation I**

**2. 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:

Recall= True positives / (True positives+False Negatives)

**Equation2**

**3. F1 Score:** It is the score which gives or balances the possibility of both the precision and recall in a single score.

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

**Equation 3**

Accuracy is a measurement used to determine the best performing model when it comes to identify the relationship between the features of a dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |
| 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 |
| 24 | 1.00 | 1.00 | 10 |
| Accuracy | 1.00 | 1.00 | 193 |

## Table III: Results

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.

## **Testing Process**

### 3.6.1 Software Testing

The role of software testing is to ensure that programmers are efficient and accurate. Software testing is an observational science investigation conducted to provide consumers with information regarding a product's quality in the environment in which it is intended to function. This can include but is not limited to running a program or application to detect errors.

### 3.6.2 Unit Testing

In this case, each module is evaluated independently. The standards for defining unit test modules were selected to identify modules that have key functionality. A module may be either an individual or a method.

The unit testing functions that will be tested are as follows:

* + - * Working of admin\_console()
      * Upload resume feature
      * Sign in with password
      * Resume analysis working

### 3.6.3 Integration Testing

Relevant components are integrated and analyzed as a group during integration planning. Integration testing takes unit-tested elements like data, groups them into larger aggregates, applies integration test plan tests to those aggregates, and produces the integrated testing framework.

### 3.6.4 Validation Testing

At the start or end of the production process, this approach is used to determine if the software satisfies the specified specifications.

### 3.6.5 GUI Testing

GUI testing is the process of examining a product's graphical user interface to ensure that it complies with standards, such as retaining navigation between icons/buttons with source code.

**3.7 Test Cases**

**Descriptions:**

The four test cases are described completely in the below tables:

|  |  |
| --- | --- |
| Test Case ID | 1 |
| Test Case Name | Working of admin\_console() |
| Test Case Description | Detection of admin console window in the extreme left corner |
| Steps | 1. Executing the console code 2. Retrieving the data 3. Detecting a slide window |
| Expected Results | Admin console window is showing at the designated place |
| Actual Results | As expected |

|  |  |
| --- | --- |
| Test Case ID | 2 |
| Test Case Name | Upload resume feature |
| Test Case Description | Detection of upload resume option in admin console window |
| Steps | 1. Executing the console code 2. Retrieving the data 3. Detecting a console window 4. Detecting two options in the window |
| Expected Results | Upload resume feature is present in admin console |
| Actual Results | As expected |

|  |  |
| --- | --- |
| Test Case ID | 3 |
| Test Case Name | Sign in with password |
| Test Case Description | Checking whether the user is able to login in admin console |
| Steps | 1. Executing the save password and ID 2. Retrieving the data saved in the system 3. Signup is completed |
| Expected Results | Successfully completed the sign in procedure |
| Actual Results | As expected |

|  |  |
| --- | --- |
| Test Case ID | 4 |
| Test Case Name | Resume analysis working |
| Test Case Description | Checking the analyzation of resume uploaded by the user |
| Steps | 1. Click on the admin console 2. Press the upload resume feature 3. The resume analysis will begin after uploading 4. The analyzation and rating is given at the end of the window 5. Rating is provided along with suggestions |
| Expected Results | Analyzed accurately along with rating and suggestions. |
| Actual Results | As expected |

# CHAPTER 4

# RESULTS AND OUTPUTS

There are four points in which the result & outcome of our report has be divided:

1. Firstly, the data set and performance matrix that had been used in project for evaluating the effectiveness of our automated system of resume screening has been discussed.
2. Secondly, the website for screening resumes has a user interface that consists of three pages, namely the page for uploading resumes, the page for analysing resumes, and the page for providing a score for resumes.
3. Third, the website has a different admin console which consists of three sections namely a page for users data, pie chart for predicted field recommendations & pie chart for users experienced level.

## **DATASET & PERFORMANCE MATRIX**

In this research, we obtained a dataset of 962 resumes from Kaggle, where each resume contained comprehensive details of a candidate. We processed the resumes through our novel system, which implemented machine learning algorithms to screen and promptly discard any resumes that failed to meet the predetermined criteria. Our findings demonstrated the high efficiency and accuracy of our system, as it successfully identified the most competent candidates with the training accuracy rate of 98.96% with the validation accuracy rate of 96.89%. These results serve as evidence of the effectiveness of our approach for screening resumes.

To assess and compare the performance of various models, evaluating their effectiveness is crucial. This assessment provides valuable insights to select the best-performing model for a given task. The evaluation process is typically expressed in numerical terms, providing a quantitative measure of the model's real-time performance. In this research, we employed recall score, f1 score, and precision score to assess the performance of our models.

1. The precision score is a measure of the model's ability to accurately predict positive class values. In other words, the precision score is the ratio of true positive predictions to the total number of positive predictions made by the model. This metric is useful in scenarios where the focus is on minimizing false positive predictions.
2. The recall score is a metric that measures the model's ability to accurately predict positive class values by identifying all relevant cases. The recall score is calculated using the following formula
3. This score represents the ratio of true positive predictions to the total number of actual positive values in the dataset. In other words, the recall score helps to assess how well the model identifies all relevant cases and avoids false negatives.
4. The F1 score is a metric that combines both precision and recall into a single score where Precision and Recall are two separate metrics that measure different aspects of performances.
   1. **WEB SNAPSHOTS OF USER’S INTERFACE**

We are providing visual representations of the basic layout design of our system below:

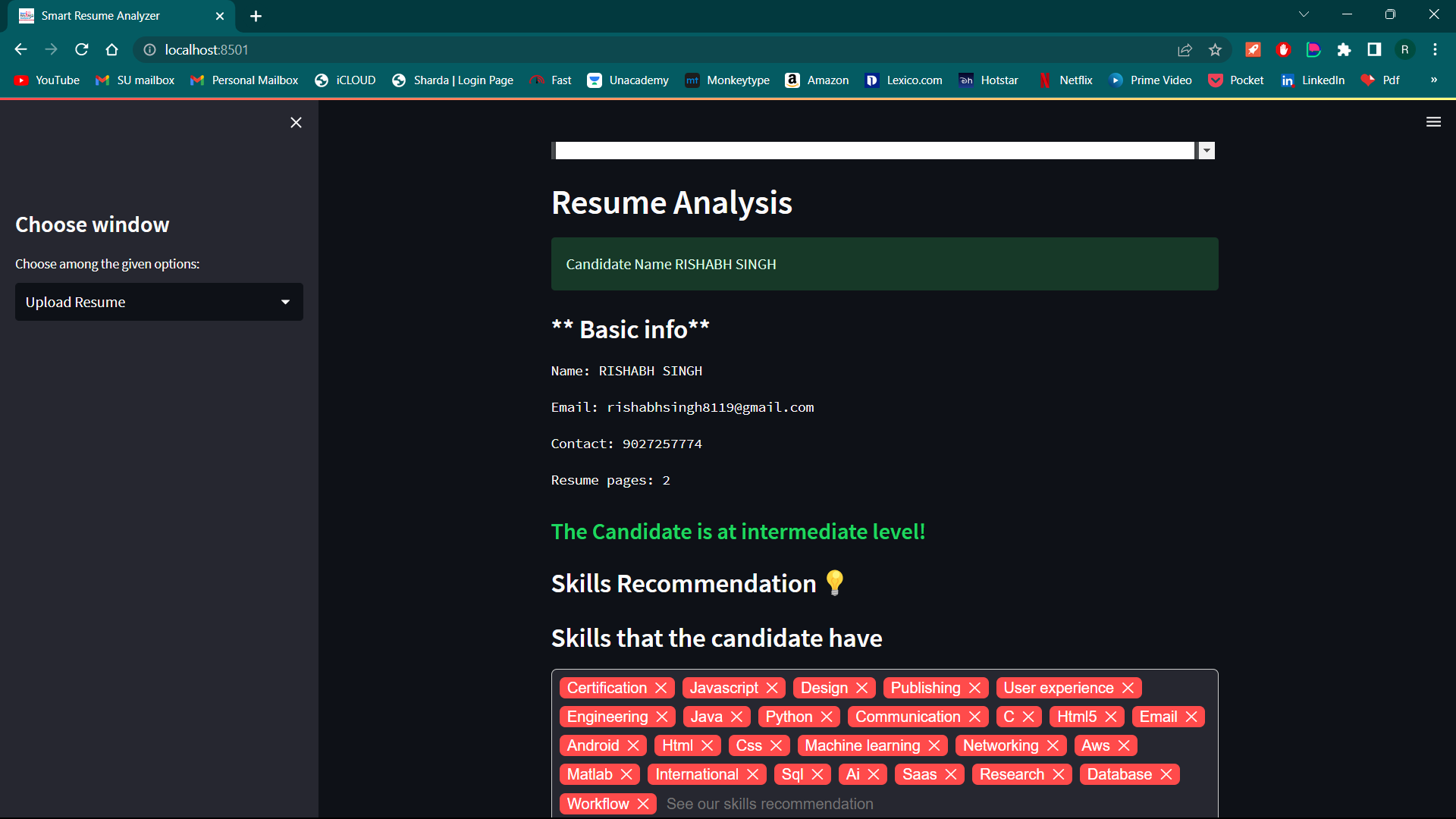
This is the snapshot of the basic homepage which will be visible to the user when he enters the homepage of the resume screening page. Here the user will have the option to upload his resume by selecting it from the files or he will have the option of drag and drop as well.

A screenshot of a computer

Description automatically generated

**Figure XI**

* Now when the user has uploaded the resume and when he will submit his resume. The processing will start and after analyzing his resume an analyzed page of resume will open with his basic information, the level of the candidate also with some skill recommendations for the user which he can learn to enhance his skills.



**Figure XII**

Graphical user interface, text

Description automatically generated

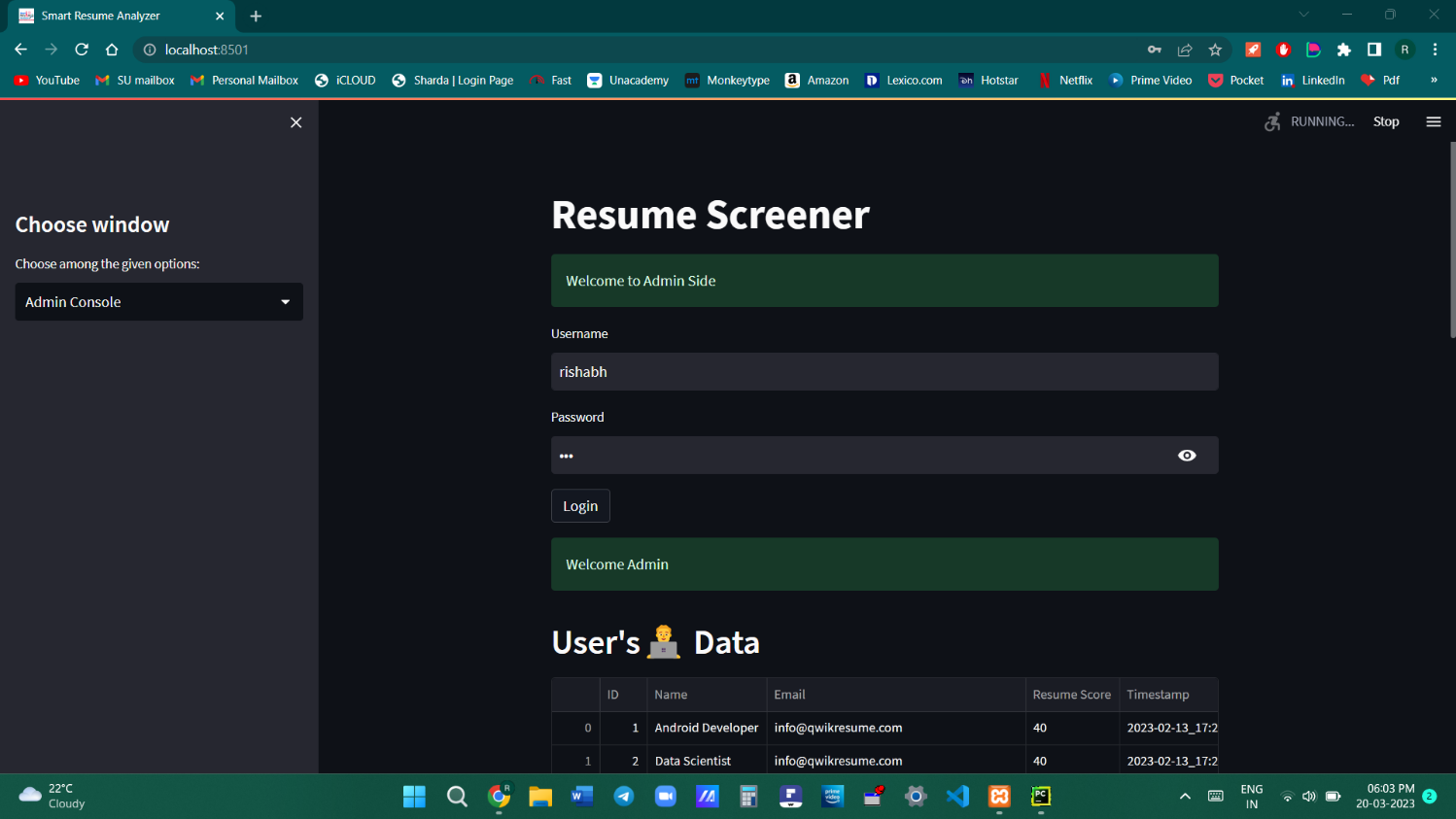
**Figure XIII**

* In the above snapshot the user will be able to see some resume tips and ideas for him, also to make it more attractive we have introduced some interactive points where the user can see his positives and negatives in his resume, and he will also get a resume score of his resume.

## **.1 SNAPSHOTS OF ADMIN CONSOLE**

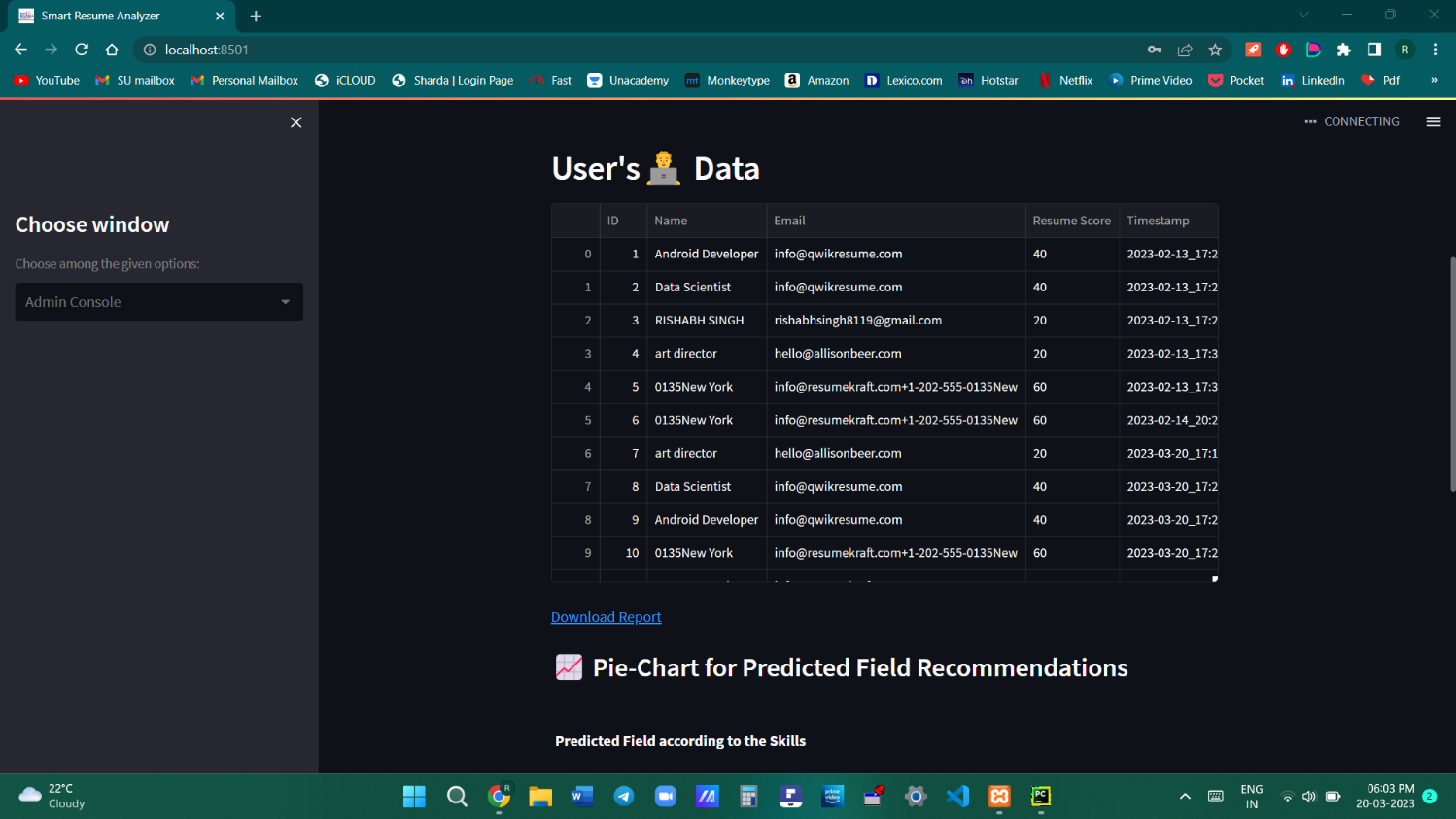
In addition, the admin console contains several features. We have included snapshots of these elements below for your reference.

* 1. In the first tab admin will be asked to login and after login admin will get access to the data available.



**Figure XIV**

1. In the very first section, the admin can see the list of users which have uploaded resume for the specific job profile. Here the admin can also download the list of candidates in the figure XV given below.



**Figure XV**

1. **Now in the second section of the admin console, the recruiter side can see a pie chart predicted field recommendations also in the given figure XVI below.**

A screenshot of a computer

Description automatically generated

**Figure XVI**

1. **There will be a pie chart of the experience level of the user also in the given XVII below.**



**Figure XVII**

# CHAPTER 5

# CONCLUSION

Our system uses advanced algorithms and machine learning techniques to accurately evaluate candidates' skills and qualifications for a specific job. The system employs sophisticated profile matching algorithms to identify the most suitable candidates for each job opening. By analyzing candidate profiles and comparing them to the job description, our system can pinpoint those who possess the necessary skills and experience to excel in the role. The selection process is highly automated, with our system handling the majority of candidate screening and evaluation tasks, thus saving valuable time and resources for HR departments. By analyzing key criteria such as skills, experience, and qualifications, our system identifies the candidates who are the best fit for the job and recommends them to the hiring team.

Our system utilizes advanced machine learning techniques and algorithms to accurately assess the skills and qualifications of job candidates. We employ sophisticated profile matching algorithms to identify the most suitable candidates for each job opening. By analyzing candidate profiles and comparing them to job descriptions, our system can identify individuals who possess the required skills and experience to succeed in the role. This highly automated selection process saves time and resources for HR departments. The system analyzes key criteria such as skills, experience, and qualifications to determine the best fit for each job and recommends the most suitable candidates to the hiring team.

Our project has achieved an impressive success rate of around 98%. This high level of accuracy demonstrates a deep understanding of the underlying concepts and data involved in the project's design and execution. The complexity and scope of the project make this achievement even more noteworthy.

The foundation of our project is Natural Language Processing (NLP), a subfield of Artificial Intelligence that focuses on enabling machines to understand and generate human language. Leveraging NLP techniques, our project aims to extract meaningful insights and information from large volumes of text data. These insights can inform decision-making processes, improve customer experiences, or support research and development efforts.

We utilized One Vs Rest and K Neighbors classifier algorithms to optimize the results categorized under the Natural Language Processing branch. Our automated system screens resumes, making it easier to find suitable candidates and reducing the need for manual labor. The system selects the best candidate from a large number of applicants and handles the task of screening with great accuracy and performance. The system's reliability is beneficial in saving manager's efforts and time.

In summary, this research paper presents an innovative solution for resume screening using machine learning algorithms and natural language processing. This system provides several advantages over traditional manual screening methods, including increased accuracy, speed, and scalability.

The research demonstrates the effectiveness of the proposed system in accurately and efficiently screening resumes, identifying key skills and experiences, and providing actionable insights to recruiters and hiring managers. The system can be customized to meet the specific needs of different organizations and job roles, making it highly versatile and adaptable.

The use of machine learning algorithms and natural language processing represents a significant advancement in the field of resume screening. This research contributes to the growing body of knowledge in this area, suggesting that automated resume screening systems have the potential to transform the hiring process, leading to more efficient and effective recruitment practices.

Overall, this research highlights the importance of leveraging new technologies and innovative approaches to address long-standing challenges in the recruitment process. The proposed system represents a promising step towards a more streamlined and efficient approach to resume screening, and similar systems will likely continue to be developed and refined in the years to come.

# 5.1 System Usability

The automated resume screening system proposed in this research paper has been designed with a high level of usability in mind. It features an intuitive and user-friendly interface, which is clear and well-organized, allowing recruiters and hiring managers to navigate through resumes and access relevant information quickly and easily. The system also provides real-time feedback and suggestions, enabling users to make more informed decisions and improve the overall effectiveness of the screening process. Furthermore, the system's scalability and flexibility make it adaptable to the specific needs of different organizations and job roles. Overall, the system's usability is a key factor in its potential to transform the recruitment process by providing a highly efficient and effective means of screening resumes while reducing the time and resources required for manual screening.

# 5.2 Future Scope

With the rapid advancements in technology, the proposed system of automated resume screening using machine learning algorithms and natural language processing can be further optimized and refined to enhance its accuracy, speed, and efficiency. The system can also be integrated with other HR technologies such as applicant tracking systems and candidate relationship management tools, to provide a seamless and comprehensive recruitment process. The use of the system can promote diversity and inclusion by mitigating bias in the screening process. Additionally, the system can be extended to other HR functions, such as performance evaluation and workforce planning, providing a more holistic view of talent management. With its versatility and adaptability, the system can be deployed in various industries and sectors, including healthcare, finance, and technology, to improve recruitment practices and support organizational goals. In summary, the proposed system holds vast potential and can revolutionize how organizations manage and recruit their talent.

CHAPTER 6

REFERENCE

* Breaugh, J.A., 2022. The use of biodata for employee selection: Past research and future directions. Human Resource Management Review 19, 219–231.
* Lu, J., Wu, D., Mao, M., Wang, W., Zhang, G., 2021. Recommender system application developments: a survey. Decision Support Systems 74, 12–32.
* 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.
* 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
* 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).
* 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.
* 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
* 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.
* 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
* Jivani, A.G., 2019. A comparative study of stemming algorithms. Int. J. Comp. Tech. Appl, 2(6), 1930-1938.
* 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.
* Resnick, P., Varian, H.R., 2018. Recommender systems. Communications of the ACM 40, 56–59
* 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
* Pradeep Kumar Roy, Vellore Institute of Technology, 2019. A Machine learning approach for automation of resume recommendation system, ICCIDS 2019.
* Thimma Reddy Kalva, Utah State University, 2013. Skill-Finder: Automated Job-Resume.
* Based Framework for automatic resume quality Suhjit Amin, Fr.Conceicao Rodrigues Institute of Technology, 2019. Web Application for Screening resume,
* Ashwini K, Umadevi V, Shashank M Kadiwal,Revanna, Design and Development of e Learning based Resume
* Riza tana Fareed, rajah V, and Sharadadevi kaganumat “Resume Classification and Ranking using KNN and Cosine Similarity” In 2021 International Journal of
* 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: 1109/ICNTE44896.2019.8945869.
* 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
* Centre for Monitoring Indian Economy Pvt Ltd. (CMIE),2022. The unemployment rate in
* 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.
* 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.
* Mikheev, Andrei; Moens, Marc; Glover, 1999. “Named Entity Recognition without Gazetteers.” Proceedings of EACL ’99. HCRCLanguage Technology Group, University

# ANNEXURE I

Research Paper for the said project has been **accepted and presented** in Twelfth International Conference on Advances in Computer Science and Application – CSA 2023

**Paper Title:**

An efficient and automated system of resume screening using machine learning algorithms.

**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.

**Authors:**

Mallika Pal, Rishabh Singh, Rani Astya, Vishal Tomar.





# ANNEXURE II

Review Paper for the said project has been **accepted** in GIS Science Journal

**Paper Title:**

A review paper on resume screening using machine learning techniques

### 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.

### Authors:

Mallika Pal, Rishabh Singh, Vishal Tomar, Rani Astya.

