

# RÉSUMÉ SCREENING SYSTEM

Kushagra Bhatnagar

10-06-2023

## *Abstract*

The project introduces an intelligent, machine learning based application that will change the job search and recruitment process for good. Using machine learning algorithms, the product offers personalized job recommendations, automated resume screening, assistance with optimization of resumes by studying the trends in resumes that are shortlisted, using natural language processing.

By analyzing job seekers' profiles and requirements, the application will deliver accurate job recommendations and suggestions to optimize one's resume. The user-friendly interface will enable job seekers to find suitable roles and companies to identify the best fit candidates.

## 1.0 Problem Statement

Finding a perfect job for that fits one's skills set is a tedious task that almost everyone has to go through multiple times in their lifetime. Applying for jobs online becomes a challenge due to the excess of options available. At the same time, a large number of résumé are submitted for one job posting making the recruitment process for the companies complicated.

A system is needed that can suggest the applicants an appropriate role they should submit their CVs to, and at the same time a system that can filter the CVs received by a company based on the position that seems most apt for them, thus simplifying the recruitment process by screening the CVs for the companies.

The system should be subscription based for the companies that wish to hire employees for vacant positions, and the public can get suggestions for suitable roles to apply for based on the content of the resume.



A resume screening system, to find suitable roles

---

## 2.0 Assessment

### 2.1 Market Need Assessment

Currently, there are significant challenges faced by individuals while seeking employment opportunities and companies during the screening and selection of said individuals. The abundance of online options leads to confusion in finding a role that matches one's skill set. At the same time, the overwhelming number of responses submitted for a single position complicates the recruitment process for the companies.

- The market needs an extensive system addressing these challenges. The system must offer job recommendations personalized on the basis of the applicant's skills and qualifications, simplifying the job search process and increasing the chances of finding a suitable role.
- The companies require a system that can efficiently filter and screen the CVs, providing a better selection of candidates.
- To satisfy this need, the system will offer a subscription-based service to the companies, allowing them to advertise their vacancies to the applicants and

using the system's automated resume screening and matching system to choose suitable candidates to fill the vacant positions.

- The system can be publicly accessible to job seekers, suggesting suitable roles based on their resumes.
- The needs of both, the job seeker and the hiring company will be catered to by this system, thus enhancing the efficiency and effectiveness of matching the candidates with suitable job opportunities

## **2.1 Customer Need Assessment**

People seeking job opportunities and companies hiring new employees, both face multiple challenges that effect their efficiency and effectiveness in the process. Therefore, it is necessary to understand the needs of these customers for developing a successful system.

### **2.1.1 Job Applicants:**

- Guidance: People seeking jobs require assistance in identifying the most suitable positions that match their skills, qualifications and aspirations.
- Efficiency: People look for a simple job search process that saves time and effort by giving personalized recommendations instead of manually going through numerous postings
- Visibility: Applicants want their resumes to be noticed by the companies for considerations, thus a system that enhances their chances of securing interviews and offers.

### **2.1.2 Hiring Companies:**

- Efficiency: Companies also need an efficient system to filter and screen the bulk of resumes received for each job posting, a system that enables them to focus on the most suitable candidates quickly.
- Relevance: Companies seek a solution that presents a filtered selection of resumes for specific job positions, ensuring a best match between the candidate and the job requirement.
- Cost Saving: Companies aim to simplify their recruitment process to reduce cost and save time associated with manual resume screening and evaluation.

### **2.1.3 Job seekers + Hiring Companies:**

- Quality: Both type of customers values a system that provides reliable and accurate recommendations.
- User-Friendly Interface: Customers require a platform that is intuitive and user-friendly. The platform should be easy to navigate and provide an easy and effortless experience to both type of customers.
- Data Privacy: The customers expect their personal and confidential information to be protected. All the information needs to be handled securely and a trust must be maintained between the platform and its customer.

---

## **3.0 Target Specifications and Characterization**

The target specifications for the product, focusing on different type of customers and a system-wide features and the technical requirements are as follow:

### 3.1 Job Applicants:

- Personalized job recommendations: A personalized job recommendation should be provided to the user based on their individual resumes.
- User-friendly Interface: The platform should have an easy-to-use interface that allows its users to effortlessly navigate through job postings and their application processes.
- Resume optimization: The platform can suggest the users some tips to optimize their resumes to increase visibility.
- Tracking the application: The applicants should be able to track the status of their applications and get suitable updates from the hiring companies.

### 3.2 Hiring Companies:

- Screening and filtering the resumes: The platform should present the companies with the filtered resumes based on the job requirements.
- Customizing the job postings: Companies should be able to customize their job postings as per their requirements.
- Evaluation tool for the candidates: The platform needs to provide a tool to assist the companies in evaluating the applicants. A ranking system for resumes based on keywords, etc.
- Communication with the candidates: Companies should be able to contact the applicants and update them on the status of their applications regularly, a direct communication between the applicant and the company while their application is under review is necessary.

### 3.3 System-wide:

- Machine Learning model: The system will incorporate a machine learning model trained on a diverse dataset of resumes and the suitable jobs for them to provide accurate recommendations and screening.
- Data Security and privacy: Severe security measures are required to protect the personal and confidential data of the users.
- Performance: The system should be smoothly operable by its users even during heavy traffic.
- Scalability: The system should be able to handle large volume of data and traffic efficiently.
- Analytics and insights: The system should provide the data analytics and insights of its users, offering trends on job market, success rates, and profiles.

### 3.4 Technical Requirements

- Compatibility: The system should be accessible on multiple devices, including but not limited to web browsers and mobile devices.
- Data management: Sufficient data storage and backup systems are required to store and manage vast number of resumes and data.
- Integration: The platform should be capable of integrating with external systems or APIs related to job boards, application tracking systems, etc.
- Performance metrics: The system should track and measure the performance of the platform, response time, user satisfaction, etc. to ensure high quality service.

## 4.0 External Search

The external search for this product will gather additional information about e-recruitment projects to provide a valuable perspective from the market. Certain areas to explore are as follow:

### 4.1 Research papers/Articles:

- Research papers and articles related to online resume screening systems and e-recruitment systems will provide a clear insight on the project.
- Job matching algorithms and use of machine learning in human resources will give a clear understanding of latest trends, techniques and challenges of the field.
- Studies and reports on effective machine learning models in resume screening and e-recruitment will be helpful in studying the success rate and impacts of similar systems.
- [A research on e-recruiting system using text-mining.](#)

### 4.2 Competitor analysis:

- Existing e-recruitment platforms that use machine learning or AI based system will be the competitors.
- Exploring their features, UI, business model, and customer review will provide their strengths and weaknesses.
- Analysing the approaches used by the competitors in their products and improving on those approaches to find a more optimal and effective product.

### 4.3 Datasets:

- A diverse dataset is required to study the trends between the resumes and their suitable type of roles.
- The dataset should comprise of multiple resumes with different skill set, educations and experiences, along with all the different positions whose requirements are met with by those resumes.
- The dataset used in this project is available [here](#).

### 4.4 Online communities and forums:

- Engagement with online communities dedicated to HR, recruitment, or machine learning.
- Seeking advice from industry professionals who have experience with similar systems.
- Collection of valuable insights from platforms like Quora, Reddit or LinkedIn by the people who have expertise with these fields.

---

## 5.0 Bench marking alternate products

Already existing e-recruitment sites that are popular in India are:

- [www.naukri.com](http://www.naukri.com)
- [jobsahead.com](http://jobsahead.com)
- [internshala.com](http://internshala.com)
- [indeed.com](http://indeed.com)

- placementindia.com
  - bestjobsindia.in
  - While most of these job search sites are popularly in use, they are not using a recommendation system based on ML.
  - The users have to manually fill in the fields they are interested in to get suggestions.
  - Our product will suggest jobs to the users based on their resume alone.
  - These sites also do not provide a screening process for the resumes received by the companies, whereas, using our machine learning model, the companies will be able to screen the resumes automatically.
- 

## 6.0 Applicable Patents

The patents used in this project are different machine learning algorithms, text-mining techniques and python libraries. Some of them are listed below

- sklearn library
  - NumPy
  - pandas
  - NLTK (Natural language toolkit)
  - Matplotlib
  - Other than these, frameworks will be used for web/app development. Flutter can be used for Applications as well as Web Development.
  - Jupyter notebook will be used for the backend code related to Machine Learning.
  - SQLite/Firebase will be used to maintain the database of the application.
- 

## 7.0 Applicable Regulations

Some common regulations to consider in this product are as follow:

- The Information technology act (2000): All electronic transactions and data needs to be protected.
  - Personal Data Protection Bill: Although this bill is still in drafting stage, it will ensure rules regarding data protection, consent, and the rights of individuals.
  - Intellectual property law: Ensuring copyright laws and intellectual property rights when handling resumes, job descriptions, or other copyrighted materials are followed.
  - Data Localization Requirements: As per RBI, all the financial data be stored in the country.
  - Consumer Protection Law: All the transactions of fees and subscriptions are protected under consumer protection law and will have to be respected.
- 

## 8.0 Applicable Constraints

Some of the constraints to take into consideration are:

- Data Availability: The dataset that we are using for the initial model of the product has limited positions only, as the product is used and a more diverse set of resumes and job positions are received, with the user's permission, the training dataset can be updated for better result.

- **Scalability and Performance:** The system needs to handle large traffic at the same time, the database should be able to handle multiple queries at the same time to ensure that the users have a smooth and comfortable experience.
  - **User Experience:** The users should be provided with a seamless and intuitive experience. The design should be accessible across different devices and responsive to user needs.
  - **Technological Constraints:** Factors such as compatibility, integration capabilities, development resources and system maintenance are to be considered.
- 

## **9.0 Business Model**

### **9.1 Revenue System:**

- **Subscription model:** Plans will be offered for hiring companies providing them access to the system's features and database of resumes who apply for their jobs. Different tiers of membership can be offered based on functions such as filtering the resumes, tracking system, etc.
- **Job Listing fees:** A fee will be charged to the companies for posting job openings on the platform.
- **Premium features:** Additional premium features such as assessment tools, advanced analytics and promotion can be provided to the companies.

### **9.2 Customer Segments:**

- **Job seekers:** Individuals looking for employment and seeking assistance for the right job.
- **Hiring companies:** Companies and organizations looking to recruit and hire candidates for their jobs.

### **9.3 Value Proposition:**

- **Matching job seekers and employers:** the systems main aim is to connect job seekers with suitable job opportunities by the use of machine learning and optimal algorithms.
- **Simplified recruitment process:** The system will streamline the recruitment process by providing services like automated resume screening, filtering and ranking the resumes, etc.
- **Personalized job recommendation:** The system will offer recommendations based on the data it retrieves from the resumes uploaded by the job seekers.

### **9.4 Key Activities:**

- **Data Management:** Gathering, storing and maintaining the database of resumes, job postings and relevant market data is necessary.
- **ML model development:** A machine learning algorithm to match job seekers with job openings based on the content of resumes and job requirements.
- **UI/UX designs:** An intuitive UI/UX for both type of customers.
- **Customer support:** A private support system to be provided to the customers to share complaints, reviews and feedbacks.

- Marketing: Promotion of the e-recruitment system to both type of customers through various channels.

## 9.5 Cost structure:

- Technology maintenance and infrastructure: Expenses related to developing and maintain the platform, for example, the servers, hosting, licenses, security measures, etc.
- Personal costs: Salaries, benefits and training of team members involved in the business.
- Data acquisition: The cost related to obtaining the dataset of resumes and job listings.
- Marketing/Promotion: A budget for marketing and promotion, advertising the platform to potential job seekers and companies.

## 10.0 Concept generation

- The basic concept this product follows is analysing the resumes uploaded by the users and obtaining useful information by means of text mining and machine learning models trained on existing dataset.
- The information obtained is then used to match the resumes with suitable jobs and filter the resumes that best match an existing job posting.
- The training of the machine learning model is done using a dataset from Kaggle that consists of 166 unique resumes and 25 unique jobs, each resume is matched with 1 or more jobs based on the information in them.
- The text in the resume is cleaned using regex tools, and a new attribute '*cleaned\_resume*' is added to the dataset.
- The model is trained to learn the best fits for different job position.

```
import pandas as pd
resumeDataSet = pd.read_csv('UpdatedResumeDataSet.csv', encoding='utf-8')
resumeDataSet.info()
```

✓ 0.1s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 962 entries, 0 to 961
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Category    962 non-null    object
1   Resume      962 non-null    object
dtypes: object(2)
memory usage: 15.2+ KB
```

```
resumeDataSet.head()
```

✓ 0.0s

	Category	Resume
0	Data Science	Skills * Programming Languages: Python (pandas...
1	Data Science	Education Details \r\nMay 2013 to May 2017 B.E...
2	Data Science	Areas of Interest Deep Learning, Control Syste...
3	Data Science	Skills â R â Python â SAP HANA â Table...
4	Data Science	Education Details \r\n MCA YMCAUST, Faridab...

A code snippet describing the dataset used



```
#count the number of job categories
print(resumeDataSet['Category'].value_counts())
```

✓ 0.0s

Java Developer	84
Testing	70
DevOps Engineer	55
Python Developer	48
Web Designing	45
HR	44
Hadoop	42
Blockchain	40
ETL Developer	40
Operations Manager	40
Data Science	40
Sales	40
Mechanical Engineer	40
Arts	36
Database	33
Electrical Engineering	30
Health and fitness	30
PMO	30
Business Analyst	28
DotNet Developer	28
Automation Testing	26
Network Security Engineer	25
SAP Developer	24
Civil Engineer	24
Advocate	20

Name: Category, dtype: int64

Code snippet to display various jobs in the dataset and frequency of occurrence

## 11.0 Concept Development

- The development of the project starts with cleaning the content of the resumes, i.e., removing special characters, extra words such as articles, pronouns, mentions, URLs, etc.

```
import re
def cleanResume(resumeText):
    resumeText = re.sub('http\S+\s*', ' ', resumeText) # remove URLs
    resumeText = re.sub('RT|cc', ' ', resumeText) # remove RT and cc
    resumeText = re.sub('#\S+', '', resumeText) # remove hashtags
    resumeText = re.sub('@\S+', ' ', resumeText) # remove mentions
    resumeText = re.sub('[%s]' % re.escape('!"#$%&'()*+,-./:;<=>@[\\]^_`{|}~""'), ' ', resumeText) # remove punctuations
    resumeText = re.sub(r'[\x00-\x7f]', r'', resumeText)
    resumeText = re.sub('\s+', ' ', resumeText) # remove extra whitespace
    return resumeText
```

```
resumeDataSet['cleaned_resume'] = resumeDataSet.Resume.apply(lambda x: cleanResume(x))
resumeDataSet.info()
```

✓ 0.3s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 962 entries, 0 to 961
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Category    962 non-null   object
1   Resume      962 non-null   object
2   cleaned_resume 962 non-null   object
dtypes: object(3)
memory usage: 22.7+ KB
```

cleaning raw resumes, forming a new attribute

```
resumeDataSet.head()
```

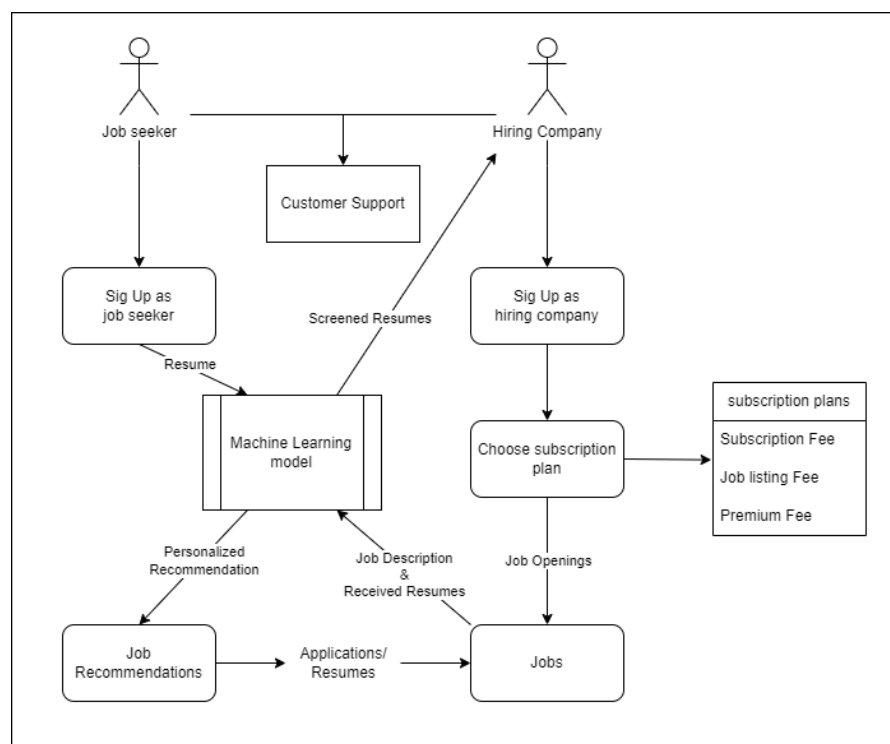
✓ 0.0s

	Category	Resume	cleaned_resume
0	Data Science	Skills * Programming Languages: Python (pandas...	Skills Programming Languages Python pandas num...
1	Data Science	Education Details \r\nMay 2013 to May 2017 B.E...	Education Details May 2013 to May 2017 B E UIT...
2	Data Science	Areas of Interest Deep Learning, Control Syste...	Areas of Interest Deep Learning Control System...
3	Data Science	Skills â R â Python â SAP HANA â Table...	Skills R Python SAP HANA Tableau SAP HANA SQL ...
4	Data Science	Education Details \r\n MCA YMCAUST, Faridab...	Education Details MCA YMCAUST Faridabad Haryan...

Processed dataset

- The processed dataset is used to train the machine learning model, to identify the relation between the contents of the resume and the job categories.
- The model will work as a recommendation system for the applicants and as a resume screening system for the hiring companies.
- The target attribute for the applicants will be ‘Category’ and for the hiring companies, the target attribute will be ‘Resume’.
- The users will get an option to create the resume through the platform in order to attain uniformity throughout all the resumes.

## 12.0 Final Product Prototype with Schematic Diagram

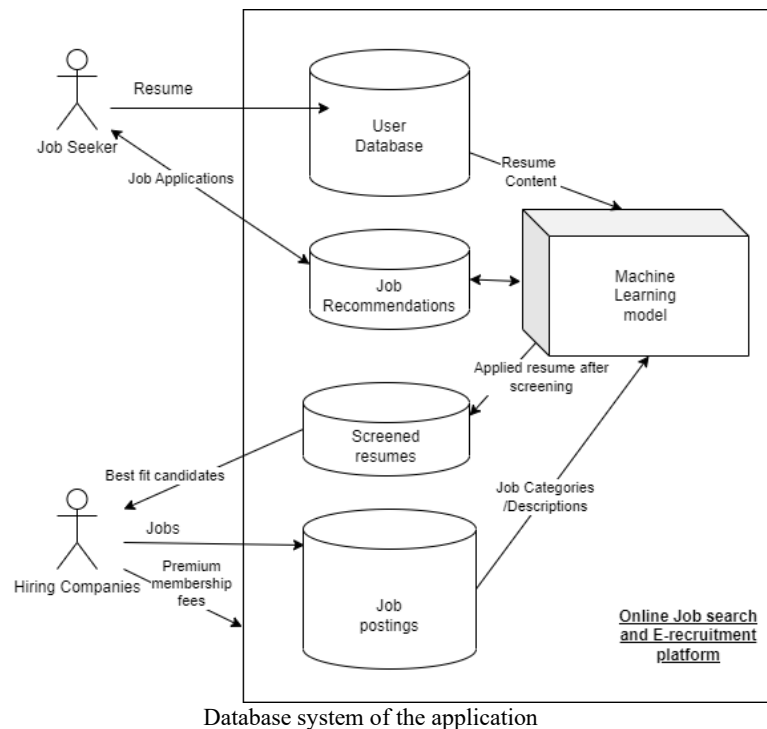


The schematic diagram showing the basic working model of the product

The above diagram explains the basic functionality of the application that is proposed.

- Job seekers and Hiring Companies can sign up as per their roles on the application.
- The job seekers will create a resume on the application that will be sent to the machine learning model for finding suitable roles based on the skills, qualifications and experience of the users.

- A model will suggest a set of roles to the user, along with the job postings for those set of roles.
- The users can apply to the jobs as per their convenience. Their resumes will be sent to the companies that they apply for.
- The companies will have a choice of subscription plans offered to them on sign up.
- As per the plan, the companies will be able to use the platform for promotions, job postings, resume screening, e-recruitment, etc.
- After the company chooses a plan, they can post job opening on the application, for which the seekers will apply.
- The resumes can be sent to the model for filtering out the best candidates for the roles.



## 13.0 Product details

### 13.1 Working:

- The product uses text mining technology to analyse the contents of the resumes uploaded on the application.
- The data retrieved from the resume is used by a pre-trained machine learning model to predict the perfect job matches for the resumes.
- The machine learning model uses the best of the several algorithms that are used on the dataset. Some of the models used are: KNN, Random Forest, Naïve Bayes.
- The resumes that are selected by the companies for consideration are, by the permission of the users, used to update the training set of the machine learning model to keep the product up to date with market trends.
- The companies using premium services of the application can use the machine learning models to find the perfect set of candidates for the job.
- The application will use Flutter framework for development.
- Python, along with the libraries such as pandas, NumPy, sklearn, matplotlib, NLTK will be used for development of the machine learning model.

### 13.2 Budget:

The expenses for the product will be as follow:

- Live database management system.
- Hosting charges.
- Development fees.
- Management fees.
- Licenses and permits.

### 13.3 Teams required:

The following teams will be needed for the smooth operation of the product:

- Application Developers.
  - Frontend
  - Backend (Data scientists, ML experts)
- Financial management.
- Application management.
- Promotions.

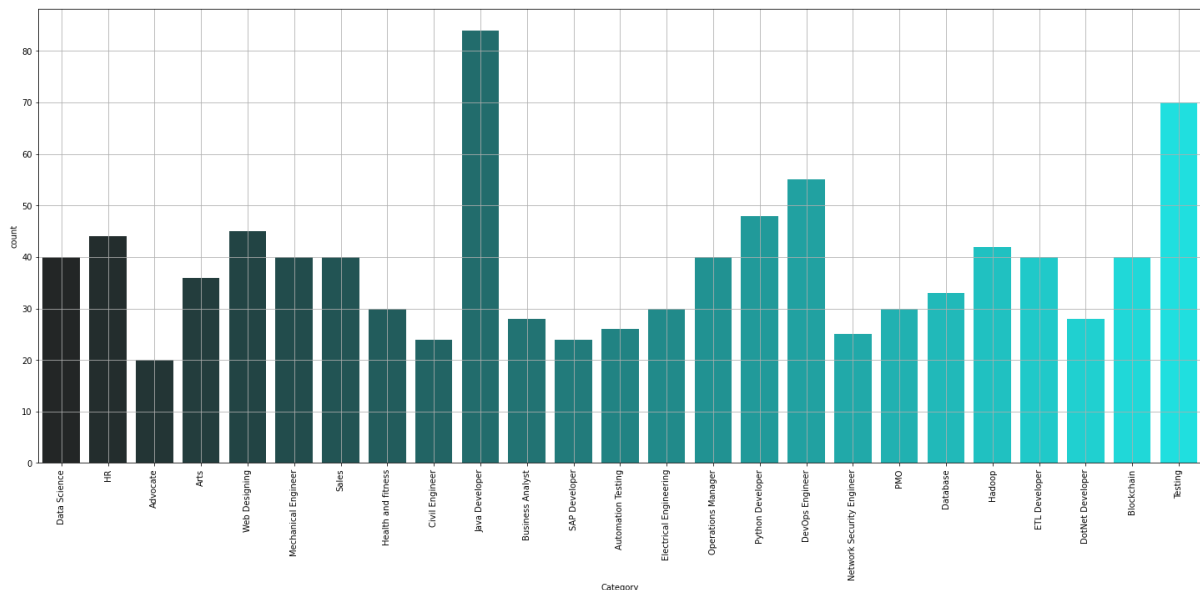
---

## 14.0 Code Implementation

- Below are snippets of the code related to the machine learning model
- The dataset used in this model can be accessed [here](#).
- The description of the dataset is given in [section 10.0](#).
- The GitHub repository for this project can be accessed [here](#).

```
plt.figure(figsize=(25,10))
cplt=sns.countplot(x="Category", data=resumeDataSet, palette='dark:cyan')
cplt.set_xticklabels(cplt.get_xticklabels(), rotation=90)
plt.grid()
```

✓ 0.4s



Frequency of occurrence of different categories of jobs in the dataset

```

import nltk
from nltk.corpus import stopwords
import string
from wordcloud import WordCloud

StopWords = set(stopwords.words('english')+['''', '''])
allwords = []
Sentences = resumeDataSet['Resume'].values
cleanedSentences = ""
for records in Sentences:
    cleanedText = cleanResume(records)
    cleanedSentences += cleanedText
    requiredWords = nltk.word_tokenize(cleanedText)
    for word in requiredWords:
        if word not in StopWords and word not in string.punctuation:
            allwords.append(word)

wordfreq = nltk.FreqDist(allwords)
mostcommon = wordfreq.most_common(50)
print(mostcommon)

```

✓ 3.0s

	0	1
0	Exprience	3829
1	months	3233
2	company	3130
3	Details	2967
4	description	2634
5	1	2134
6	Project	1808
7	project	1579
8	6	1499
9	data	1438
10	team	1424
11	Maharas...	1385
12	year	1244
13	Less	1137
14	January	1086
15	using	1041
16	Skill	1018
17	Pune	1016
18	Manage...	1010
19	SQL	990
20	Ltd	934
21	manage...	927
22	C	896
23	Engineeri...	855
24	Education	833
25	Developer	806

25	Developer	806
26	Java	773
27	2	754
28	develop...	752
29	monthsC...	746
30	Pvt	730
31	application	727
32	System	715
33	reports	697
34	business	696
35	India	693
36	requirem...	693
37	I	690
38	various	688
39	A	688
40	Data	674
41	The	672
42	University	656
43	process	648
44	Testing	646
45	test	638
46	Responsi...	637
47	system	636
48	testing	634
49	Software	632

Analysing the common pattern in the resumes by listing the most common words

```

wc = WordCloud().generate(cleanedSentences)
plt.figure(figsize=(15,15))
plt.imshow(wc, interpolation='bilinear')
plt.axis("off")
plt.show()

```

✓ 1.5s



```
knn_model = KNeighborsClassifier()
knn_model.fit(X_train, y_train)
knn_prediction = knn_model.predict(X_test)
print('Accuracy of KNeighbors Classifier on test set:    {:.2f}'.format(accuracy_score(y_test, knn_prediction)))
```

✓ 0.0s

Accuracy of KNeighbors Classifier on test set: 0.98

```
rf_model = RandomForestClassifier()
rf_model.fit(X_train, y_train)
rf_prediction = rf_model.predict(X_test)
print('Accuracy of RandomForest Classifier on test set:    {:.2f}'.format(accuracy_score(y_test, rf_prediction)))
```

✓ 0.5s

Accuracy of RandomForest Classifier on test set: 0.99

```
nb_model = MultinomialNB()
nb_model.fit(X_train, y_train)
nb_prediction = nb_model.predict(X_test)
print('Accuracy of Naive Bayes Classifier on test set:    {:.2f}'.format(accuracy_score(y_test, nb_prediction)))
```

✓ 0.0s

Accuracy of Naive Bayes Classifier on test set: 0.96

The accuracy score of different machine learning models used

---

## 15.0 Conclusion

In conclusion, the e-recruitment project holds potential to change the job search and hiring process. By the use of machine learning technologies, the product addresses the challenges faced by both, the job seekers and the companies.

The developing prototype will be a multiple platform application that will provide job matching, simplified resume screening, effective communication channels and more. It will empower people looking for employment by presenting them with personalized recommendations and help them optimise their resumes.

Concurrently, it will simplify the recruitment process for the hiring companies by automatic resume screening with the help of its AI.

Overall, the product will be a promising initiative that will harness the power of machine learning and technology to revolutionize the process of search for jobs and hiring new employees.