RESUME SCREENING

A
Minor Project Work
Submitted in Sixth Semester of Bachelor of Technology (BTech)

In
Computer Science & Engineering
(Session: 2022-23)

Submitted By
Muskan Patel
Enroll No: 0205CS201069
Nikita Nanhe
Enroll No: 0205CS201075
Rashmi Uikey
Enroll No: 0205CS201094

Under the Guidance of
Prof. SHWETA KRIPLANI
(Computer Sc. & Engg.)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
Shri Ram Institute of Technology, Jabalpur (M.P.)

Rajeev Gandhi Proudyogiki Vishwavidyalaya, Bhopal (M.P.)

SHRI RAM INSTITUTE OF TECHNOLOGY

Approved by AICTE New Delhi & Govt. of M.P.
(Affiliated to R.G.P.V.V. - University of Technology of Madhya Pradesh)
ISO 9001 : 2000 Certified Institution
Near ITI, MADHOTAL, JABALPUR - 482 002 (M.P.)
Ph. No. : 0761 - 2640291, 94 Fax No. : 2640294 Mobile - 9300104815

CCERTIFICATE

This is to certify that

Muskan Patel (Roll no-0205CS201069)
Nikita Nanhe (Roll no-0205CS201075)
Rashmi Uikey(Roll n0-0205CS201094)

students of 6th semester, Computer Sc. & Engg. S.R.I.T., Jabalpur have duly completed their Minor Project entitled " **Resume Screening"** for the Completion of 6th Semester examination under the requirement for the degree of Bachelor of Engineering as per R.G.P.V., Bhopal.

They have successfully implemented and tested this project, which meets all the requirements specified under my guidance.

Ms. Shweta Kriplani
(Project Guide)
Computer Sc & Engg
Deptt, SRIT

Mr. Brajesh Patel
(H.O.D.)
Computer Sc. &
Engg Deptt., SRIT

SHRI RAM INSTITUTE OF TECHNOLOGY
Approved by AICTE New Delhi & Govt. of M.P.
(Affiliated to R.G.P.V.V. - University of Technology of Madhya Pradesh)
ISO 9001 : 2000 Certified Institution
Near ITI, MADHOTAL, JABALPUR - 482 002 (M.P.)
Ph. No. : 0761 - 2640291, 94 Fax No. : 2640294 Mobile - 9300104815

Preface

We are still working on the project **"Resume Screening"** for more possible outcomes enhancement of the project.

Resume screening is the process of reviewing a resume to determine if the candidate is qualified for the position. It's one part of the larger pre-employment screening process. The decision of whether the candidate is qualified or not is based on education, experience, skills, and any other relevant information that may appear on the resume. It requires Machine learning Algorithms.

SHRI RAM INSTITUTE OF TECHNOLOGY
Approved by AICTE New Delhi & Govt. of M.P.
(Affiliated to R.G.P.V.V. - University of Technology of Madhya Pradesh)
ISO 9001 : 2000 Certified Institution
Near ITI, MADHOTAL, JABALPUR - 482 002 (M.P.)
Ph. No. : 0761 - 2640291, 94 Fax No. : 2640294 Mobile - 9300104815

Acknowledgement

I would like to express my deep sense of gratitude and sincere thanks to my guide D Dr. Shweta Kriplani Ma'am, Department of Computer Science & Engineering at SRIT, Jabalpur for his valuable and ever willing precious guidence, technical support and constant encouragement during the course of this project work. It was pleasure and unique experience for me to work under their guidance.

I am grateful to Dr. Ruchi. K. Patel, Head of Department of Computer Science & Engineering and other staff members of Computer Science and Engineering Department for providing the necessary facilities for the successful completion of this work.

Finally, my greatest thanks to my family for their patience and understanding.

Student Name - Muskan Patel (Enroll No.-0205CS201069)
Student Name - Nikita Nanhe (Enroll No.-0205CS201075)
Student Name - Rashmi Uikey (Enroll No.- 0205CS201094)

SHRI RAM INSTITUTE OF TECHNOLOGY
Approved by AICTE New Delhi & Govt. of M.P.
(Affiliated to R.G.P.V.V. - University of Technology of Madhya Pradesh)
ISO 9001 : 2000 Certified Institution
Near ITI, MADHOTAL, JABALPUR - 482 002 (M.P.)
Ph. No. : 0761 - 2640291, 94 Fax No. : 2640294 Mobile - 9300104815

Declaration

We hereby declare that the project entitled "Resume Screening "which is being submitted in partial fulfillment for the 6th semester of Bachelor of Engineering in Computer Science and Engineering under "Rajeev Gandhi Proudyogiki Vishwavidyalya, Bhopal "is the authentic record of our work carried under the guidence of Ms. Shweta Kriplani Asst. Prof., Department of Computer Science & Engineering, Shri Ram Institute of Technology, Jabalpur. The matter reported in this project has not been submitted earlier for any project in RGPV curriculum.

Muskan Patel(0205CS201069)
Nikita Nanhe (0205CS201075)
Rashmi Uikey(0205CS201094)

Index

Introduction

Resume screening is the process of reviewing a resume to determine if the candidate is qualified for the position. It's one part of the larger pre-employment screening process. The decision of whether the candidate is qualified or not is based on education, experience, skills, and any other relevant information that may appear on the resume.

The hiring process can get messy though, especially when you're using multiple platforms to post your job listings. We suggest using an all-in-one HR platform like Eddy to streamline all of your hiring and onboarding processes so resumes don't get lost in the void.

Step 1: Screening resumes based on minimum qualifications

Minimum qualifications are the mandatory qualifications that a candidate must meet to be able to do the job. A simple example of a minimum qualification is whether the candidate is legally able to work in the country.

These types of qualifications are often considered knockouts because either the candidate has it and can move forward or they don't and gets screened out of the process.

Candidates that meet the minimum qualifications move onto the second step of screening resumes.

Step 2: Screening resumes based on preferred qualifications

Preferred qualifications are non-mandatory characteristics that would make someone a stronger candidate for the job. A common example of a preferred qualification is whether the candidate has prior related work experience.

These types of qualifications are often called *nice-to-haves* and are generally more qualitative than minimum qualifications (e.g. strong communication skills).

Candidates that meet both the minimum and preferred qualifications move onto the shortlisting step of resume screening.

Step 3: Shortlisting candidates based on minimum and preferred qualifications

Deciding which candidates gets shortlisted for the interview phase depends on your recruiting needs.

For high volume recruitment, generally all candidates that meet the minimum qualifications move forward to the interview process. For low volume recruitment, generally only the top few candidates that meet both the minimum and preferred qualifications receive an interview.

You can determine how many candidates you should shortlist using your recruitment conversion rates.

Based on recruiting data, the average recruitment conversion rates are:

- 12% for application to interview
- 17% for interview to offer
- 89% for offer to acceptance

That means for every 100 candidates you screen, you need to shortlist 12 of them to interview, two of them will receive an offer, and one candidate will accept to result in one successful hire.

The important thing to remember is that your screening process is applied consistently and objectively across all resumes.

Analysis

Objective of project

- For each recruitment, companies take out the resume, referrals and go through them manually.

- Companies often received thousands of resumes for every job posting.
- When companies collect resumes then they categorize those resumes according to their requirements and then they send the collected resumes to the Hiring Team's.
- It becomes very difficult for the hiring teams to read the resume and select the resume according to the requirement, there is no problem if there are one or two resumes but it is very difficult to go through 1000’s resumes and select the best one.
- To solve this problem, we will screen the resume using machine learning and NLP using Python so that we can complete days of work in few minutes.

5.1 Requirement Gathering

- Availability of contact information
- Preference to work in the given location
- Education
- Certification
- Experience
- Work history
- Professional
- Achievements
- Relevant skills

5.2 Hardware Requirement

Operating system: Window 11 or later
Processor: Intel Pentium 4 or later
Memory: 128MB Minimum (256 MB recommended)
Internet connection: required

5.3 Software Requirement

Jupyter Notebook

Libraries

- Numpy
- Pandas
- Matplotlib
- Seborn

5.5 Feasibility Study

"Resume Screening" is an important, but time consuming, part of the selection process. A feasibility study is an assessment that determines the likelihood of a proposed project being successful, such as a new product line or technical system. The study analyzes the project's relevant factors, such as technical, economic and legal considerations, to assess whether the project is worth an investment. The study can also identify potential issues and problems that could arise from pursuing the project.

A. Technical Feasibility

In Technical Feasibility current resources both hardware software along with required technology are analyzed/assessed to develop project. This technical feasibility study gives report whether there exists correct required resources and technologies which will be used for project development. Along with this, feasibility study also analyzes technical skills and capabilities of technical team, existing technology can be used or not, maintenance and up-gradation is easy or not for chosen technology etc.

B. Operational Feasibility

In Operational Feasibility degree of providing service to requirements is analyzed along with how much easy product will be to operate and maintenance after deployment. Along with this other operational scop are determining usability of product, Determining suggested solution by software development team is acceptable or not etc.

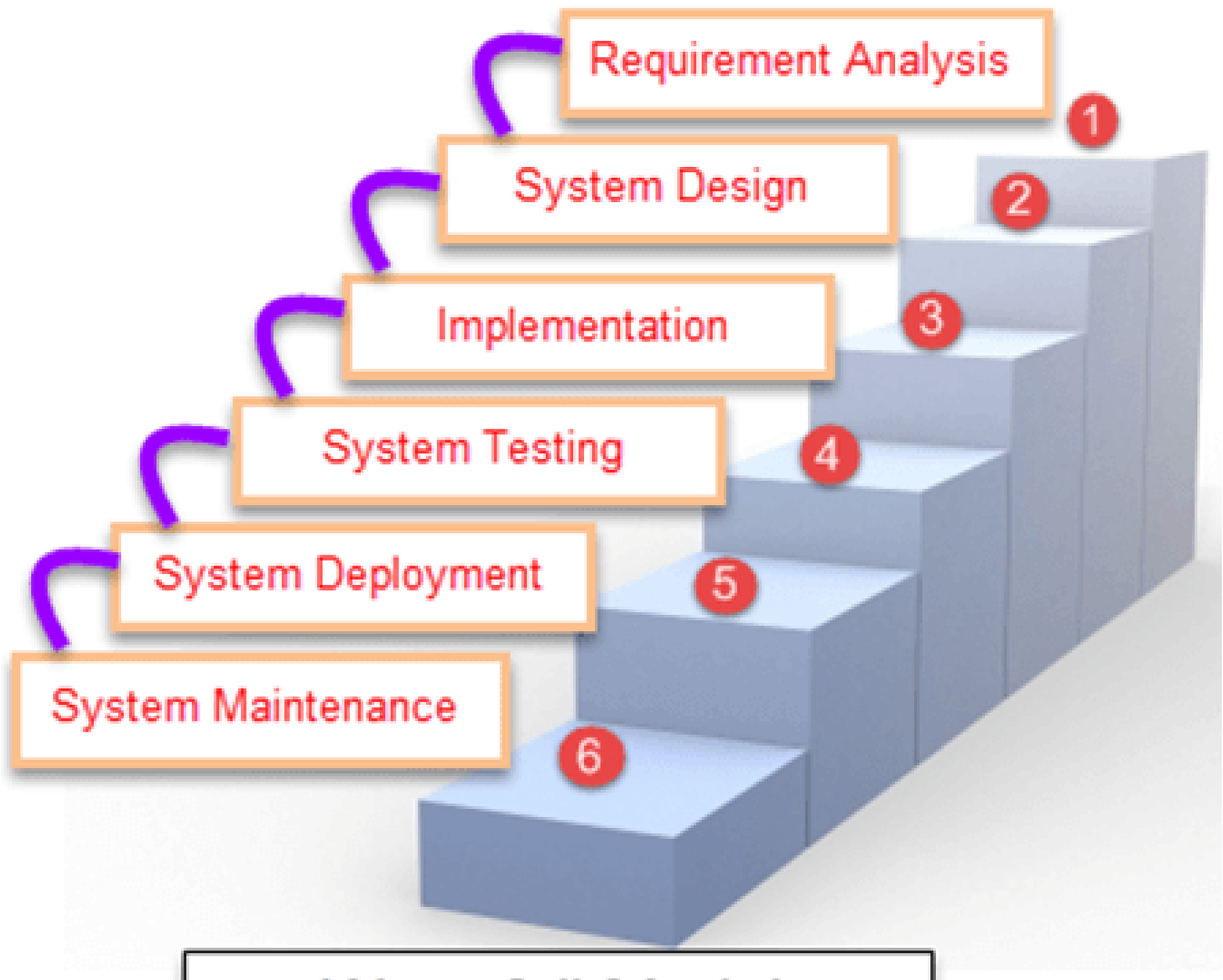
C. Economic Feasibility

In Economic Feasibility study cost and benefit of the project is analyzed. Means under this feasibility study a detail analysis is carried out what will be cost of the project for development which includes all required cost for final development like hardware and software resource required, design and development cost and operational cost and so on. After that it is analyzed whether project will be beneficial in terms of finance for organization or not.

5.6 SOFTWARE MODEL

Waterfall Model

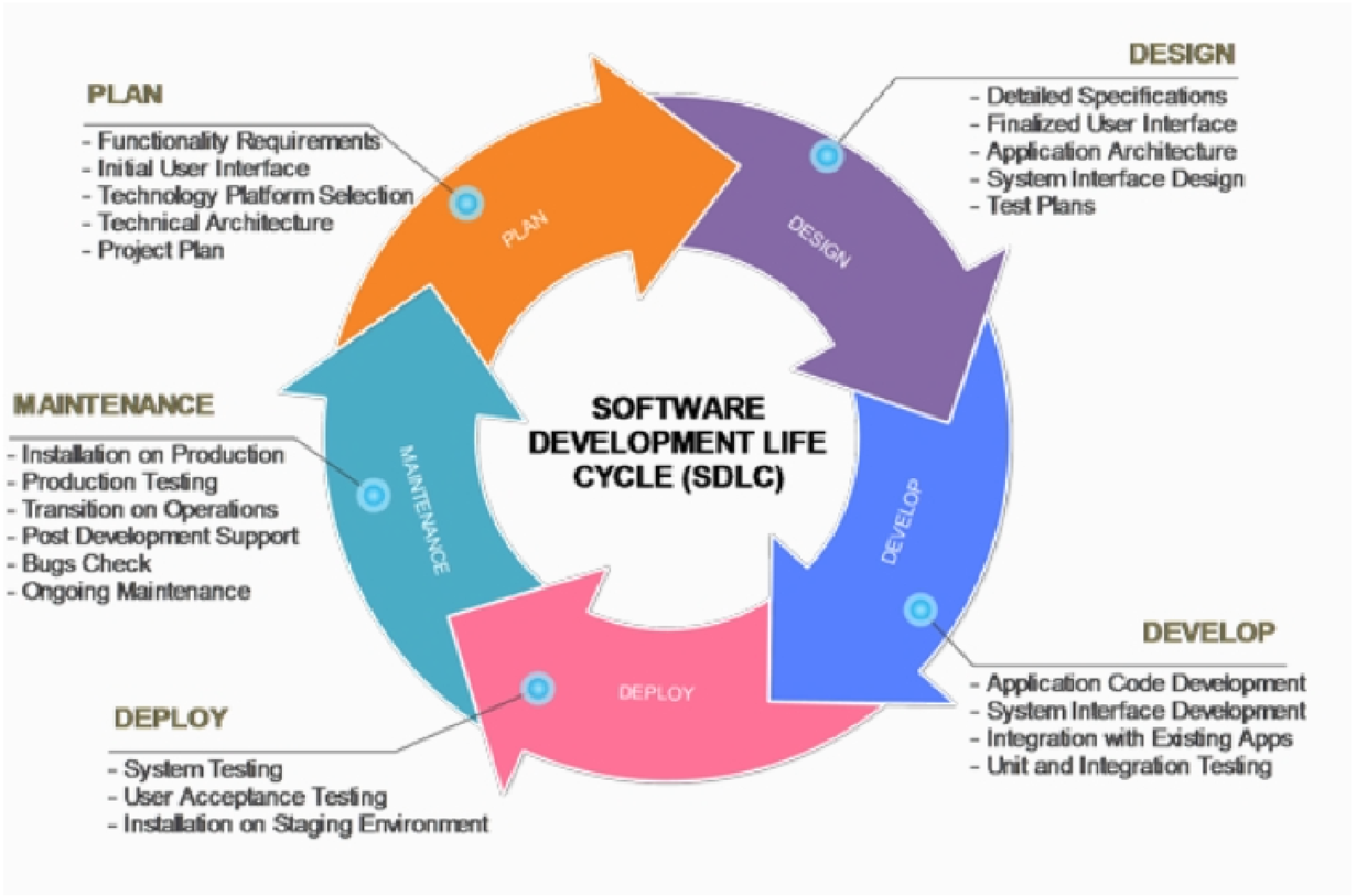
The waterfall model was the first process model to be introduced. It is very simple to understand and use. In a waterfall model, each phase must be completed before be next phase can begin and there is no overlapping in the phases. The waterfall model is the earliest SDLC approach that was use for software development.



Waterfall Model

5.7 SOFTWARE DEVELOPMENT LIFE CYCLE

SDLC is a process followed for a software project, within a software organization. It consists of a detailed plan describing how to develop, maintain, replace and after or enhance specific software. The life cycle defines a methodology for improving the quality of software and the overall development process.



5.8 COST ESTIMATION

cost. There are a number of factors, which are considered, that can affect the ultimate cost of the Software cost comprises a small percentage of overall computer-based system software such as - human, technical, Hardware and Software availability etc. The main point that was considered during the cost estimation of **project** was its sizing. In spite of complete software sizing, function point and approximate lines of code were also used to "size" each element of the Software and their costing. The cost estimation done by me for **Project** also depend upon the baseline metrics collected from past projects and these were used in conjunction with estimation variables to develop cost and effort projections.

We have basically estimated this project mainly on bases –

1) Effort Estimation - This refers to the total man-hours required for the development of the project. It even includes the time required for doing documentation and user manual.

2) Hardware Required Estimation - This includes the cost of the PCs and the

hardware cost required for development of this project.

Cocomo Model:

Cocomo (Constructive Cost Model) is a regression model based on LOC, i.e. number of Lines of Code. It is a procedural cost estimate model for software projects and often usedas a process of reliably predicting the various parameters associated with making a project such as size, effort, cost, time and quality. It was proposed by Barry Boehm in 1970 and is based on the study of 63 projects, which make it one of the best-documented models.

The key parameters which define the quality of any software products, which are also an outcome of the Cocomo are primarily Effort & Schedule:

- Effort: Amount of labour that will be required to complete a task. It is measured inperson-months units.
- Schedule: Simply means the amount of time required for the completion of the job,which is, of course, proportional to the effort put. It is measured in the units of timesuch as weeks, months.

Different models of Cocomo have been proposed to predict the cost estimation at differentlevels, based on the amount of accuracy and correctness required. All of these models can be applied to a variety of projects, whose characteristics determine the value of constant to be used in subsequent calculations. These characteristics pertaining to different systemtypes are mentioned below

Boehm's definition of organic, semidetached, and embedded systems:

- Organic – A software project is said to be an organic type if the team size required is adequately small, the problem is well understood and has been solved in the past and also the team members have a nominal experience regarding the problem.

- Semi-detached – A software project is said to be a Semi-detached type if the vital characteristics such as team-size, experience, knowledge of the various programming environment lie in between that of organic and Embedded. The projects classified as Semi-Detached are comparatively less familiar and difficult to develop compared to the organic ones and require more experience and better guidance and creativity. E.g.: Compilers or different Embedded Systems can be considered of Semi-Detached type.

- Embedded – A software project with requiring the highest level of complexity, creativity, and experience requirement fall under this category. Such software requires a larger team size than the other two models and also the developers need to be sufficiently experienced and creative to

develop such complex models.

Types of Models:

COCOMO consists of a hierarchy of three increasingly detailed and accurate forms. Any of the three forms can be adopted according to our requirements.

These are types of COCOMO model:

- Basic COCOMO Model
- Intermediate COCOMO Model
- Detailed COCOMO Model

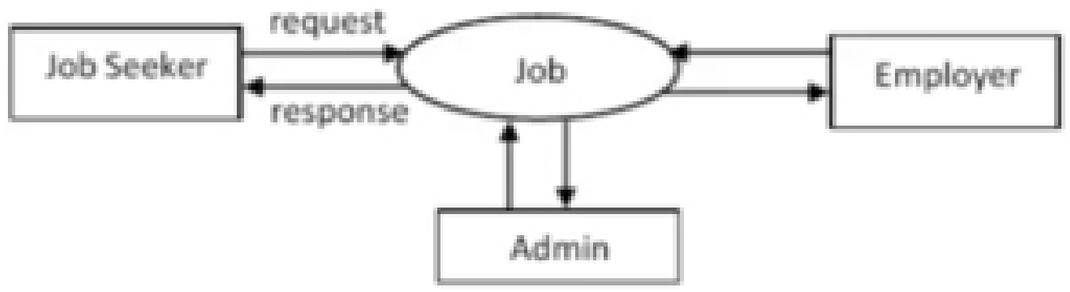
The first level, Basic COCOMO can be used for quick and slightly rough calculations of Software Costs. Its accuracy is somewhat restricted due to the absence of sufficient factorconsiderations.

Intermediate COCOMO takes these Cost Drivers into account and Detailed COCOMO additionally accounts for the influence of individual project phases, i.e. in case of Detailed it accounts for both these cost drivers and also calculations are performed phasewise henceforth producing a more accurate result

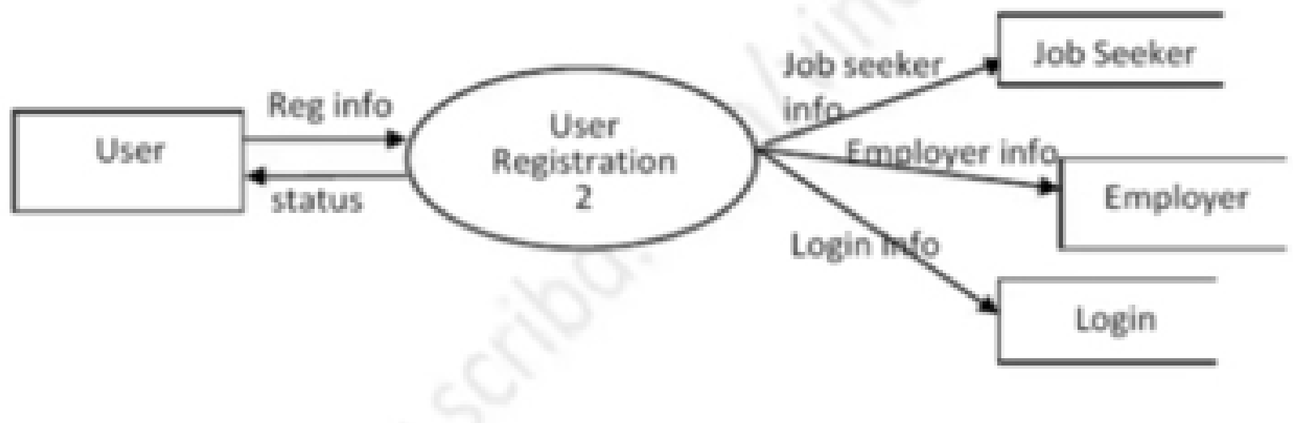
DESIGN

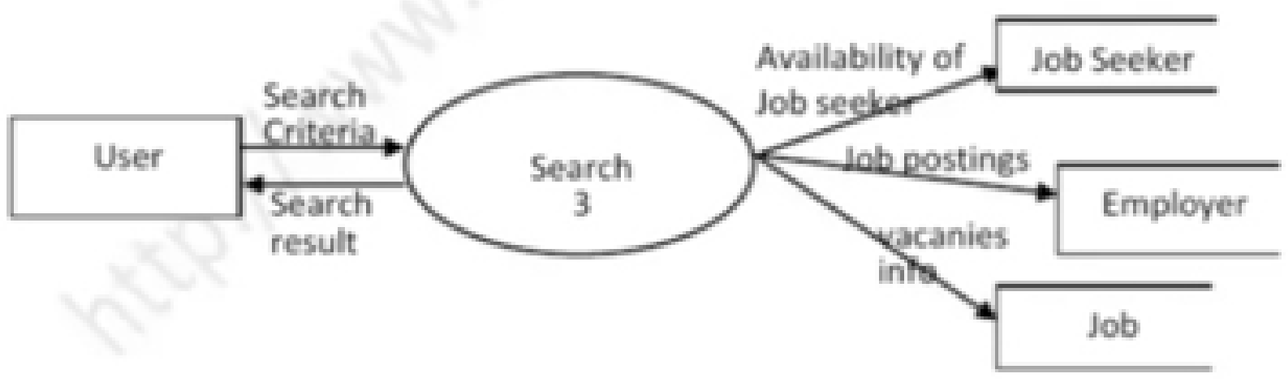
6.1 DATA FLOW DIAGRAM / OBJECT DIAGRAM

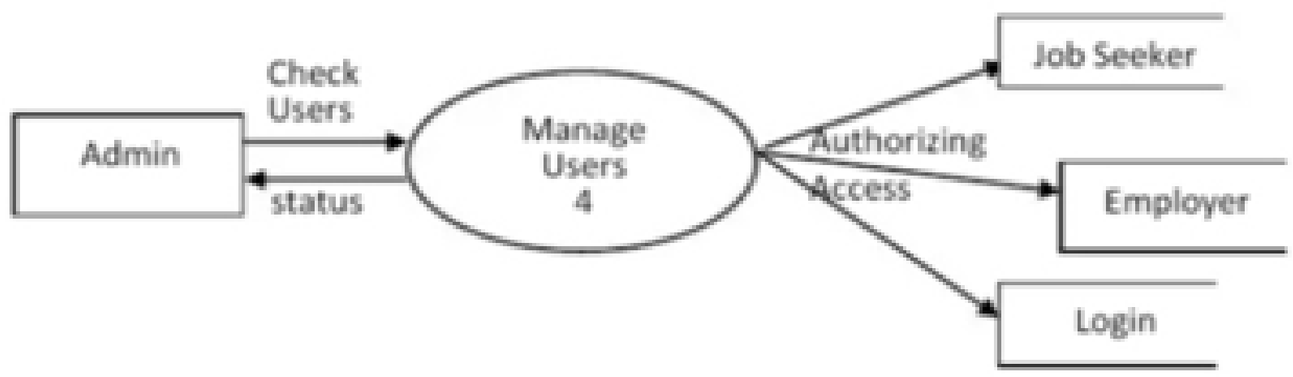
Level 0



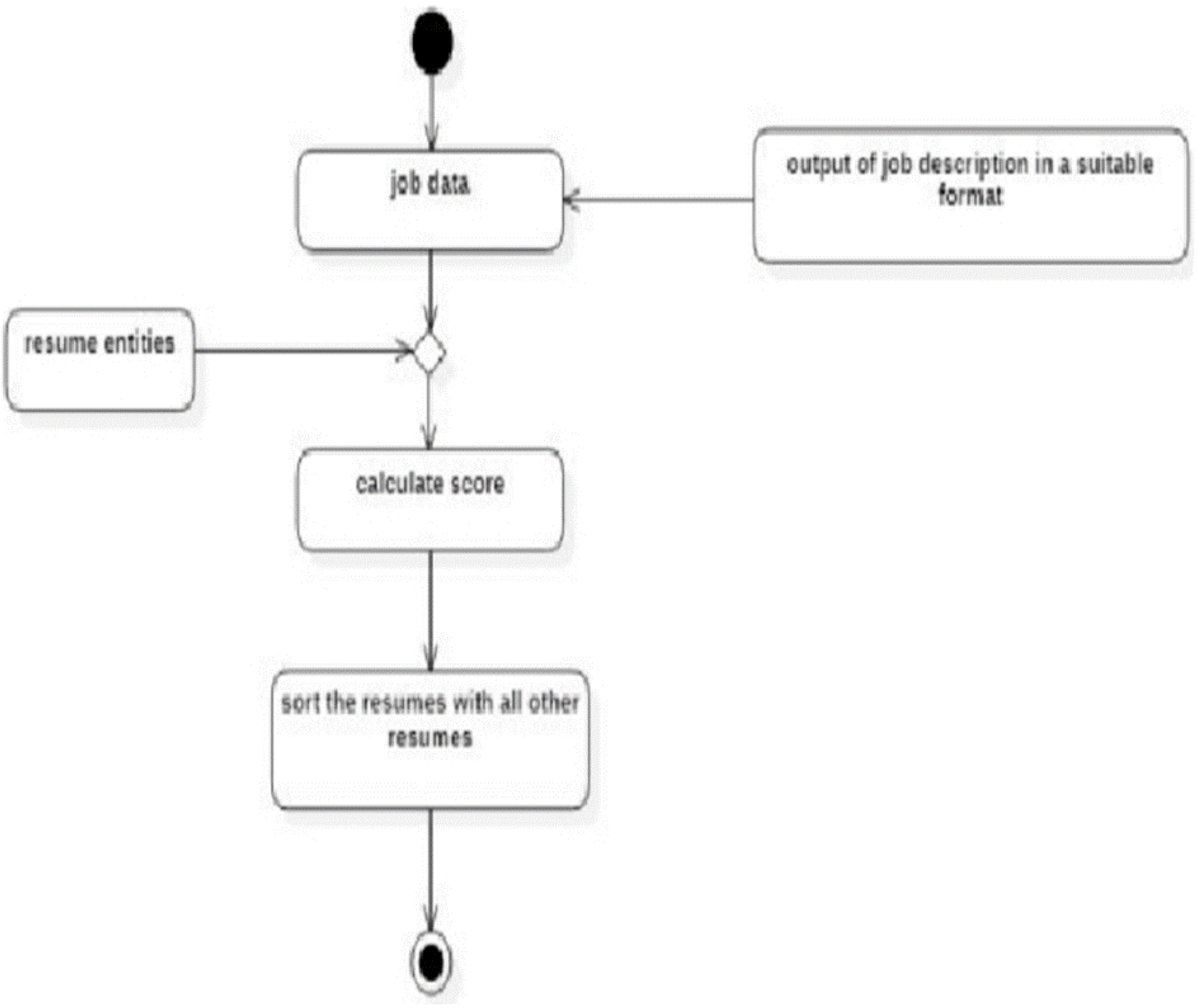
Level 1



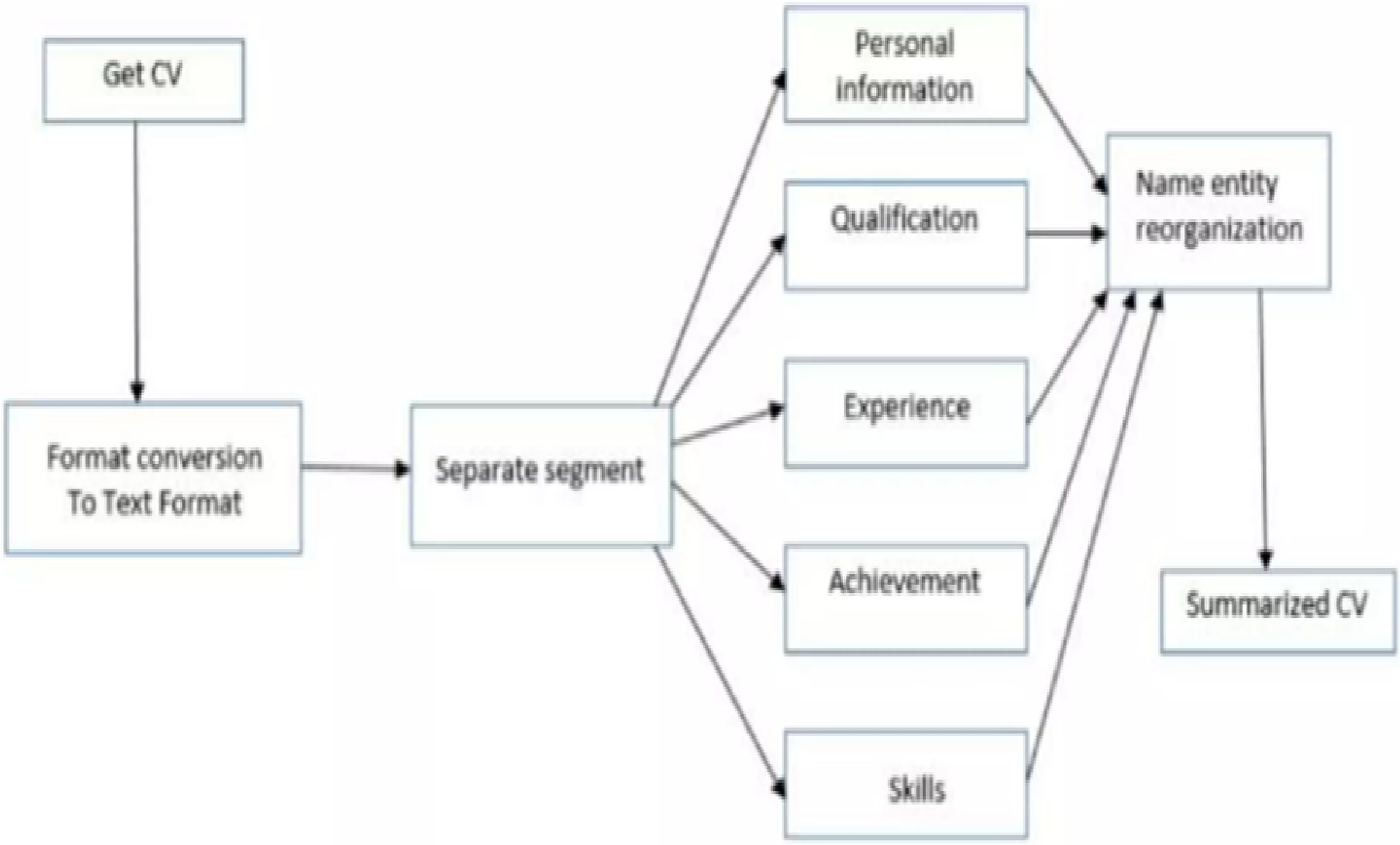




6.2 CONTROL FLOW DIAGRAM



6.3 SYSTEM FLOW DIAGRAM


Get CV
Format conversion
To Text Format
Separate segment
Personal
Information
Qualification
Experience
Achievement
Skills
Name entity
reorganization
Summarized CV

7. TOOLS USED

- Python

Models

- KNN - It’s supervised technique, used for classification. “K” in the KNN represent the number of nearest neighbours used to classify or predict in case of continuous variable.

8. CODING

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
data=pd.read_csv('UpdatedResumeDataSet.csv')
data.head(20)
print(data['Category'].unique())
print("total unique category: {}". format(len(data['Category'].unique())))
print(data['Category'].value_counts())
import seaborn as sns
plt.figure(figsize=(20,15))
sns.countplot(y="Category",data=data)

from matplotlib.gridspec import GridSpec
count=data['Category'].value_counts()
label=data['Category'].value_counts().keys()

plt.figure(1, figsize=(25,25))
grid=GridSpec(2,2)
cmap=plt.get_cmap('coolwarm')
color=[cmap(i) for i in np.linspace(0, 1, 5)]
plt.subplot(grid[0,1], aspect=1, title='Distribution')

pie=plt.pie(count, labels=label, autopct='%1.2f%%')
plt.show()
```

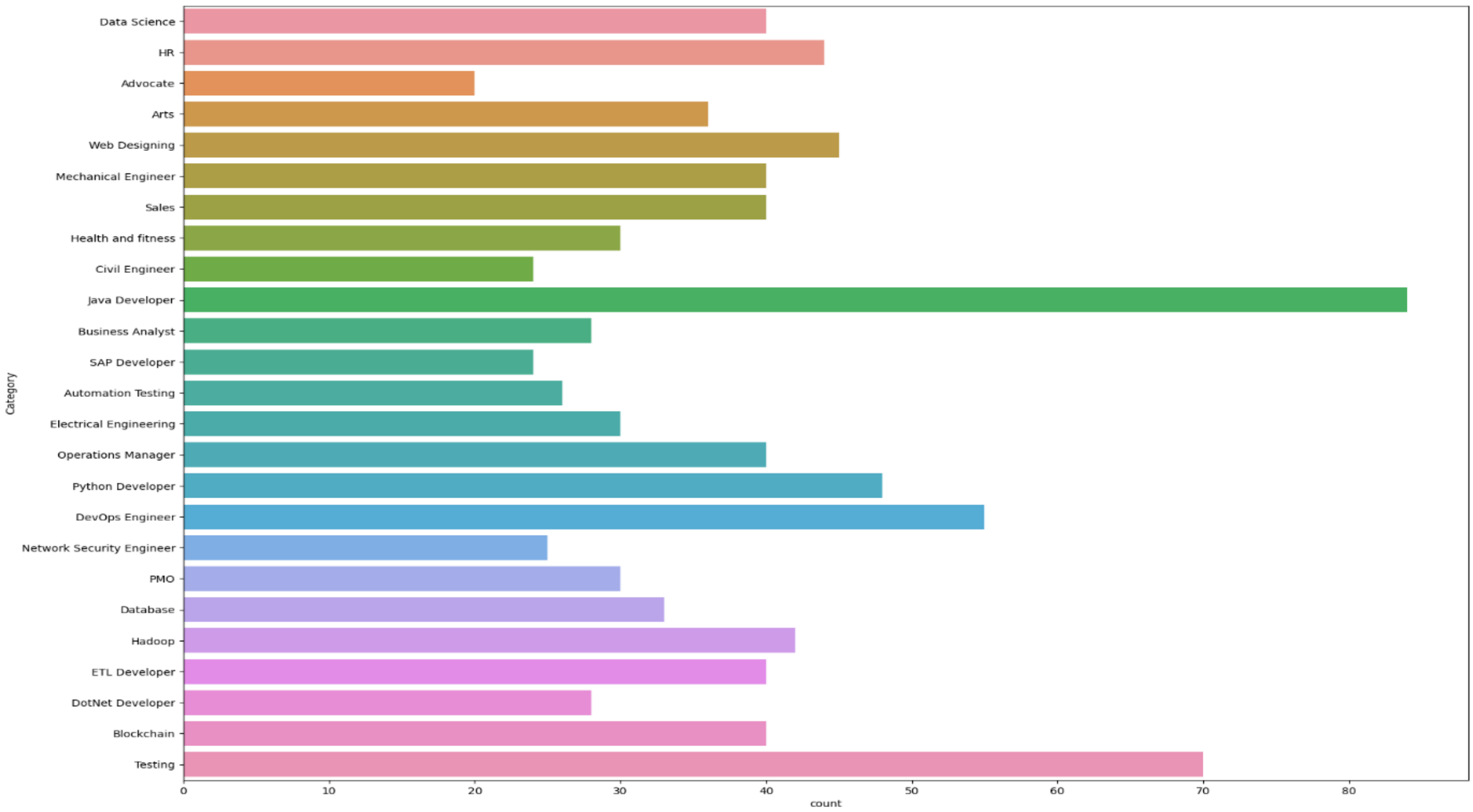
```
import re
def clean(text):
    text=re.sub('http\S+\s*', ' ', text)
    text=re.sub('RT|cc', ' ', text)
    text=re.sub('#\S+', '', text)
    text=re.sub('@\S+', '', text)
    text=re.sub('[%s]' % re.escape("""!"#$%&'()*+,-./:;<=>?@[\]^_`{|}~"""), ' ', text)
    text=re.sub('\s+', ' ', text)
    text=re.sub(r'[^\x00-\x7f]', r' ', text)
    return text

data['clean text']=data.Resume.apply(lambda x: clean(x))
data['clean text']
from sklearn.preprocessing import LabelEncoder
var=['Category']
le=LabelEncoder()

for i in var:
    data[i]=le.fit_transform(data[i])
data
```

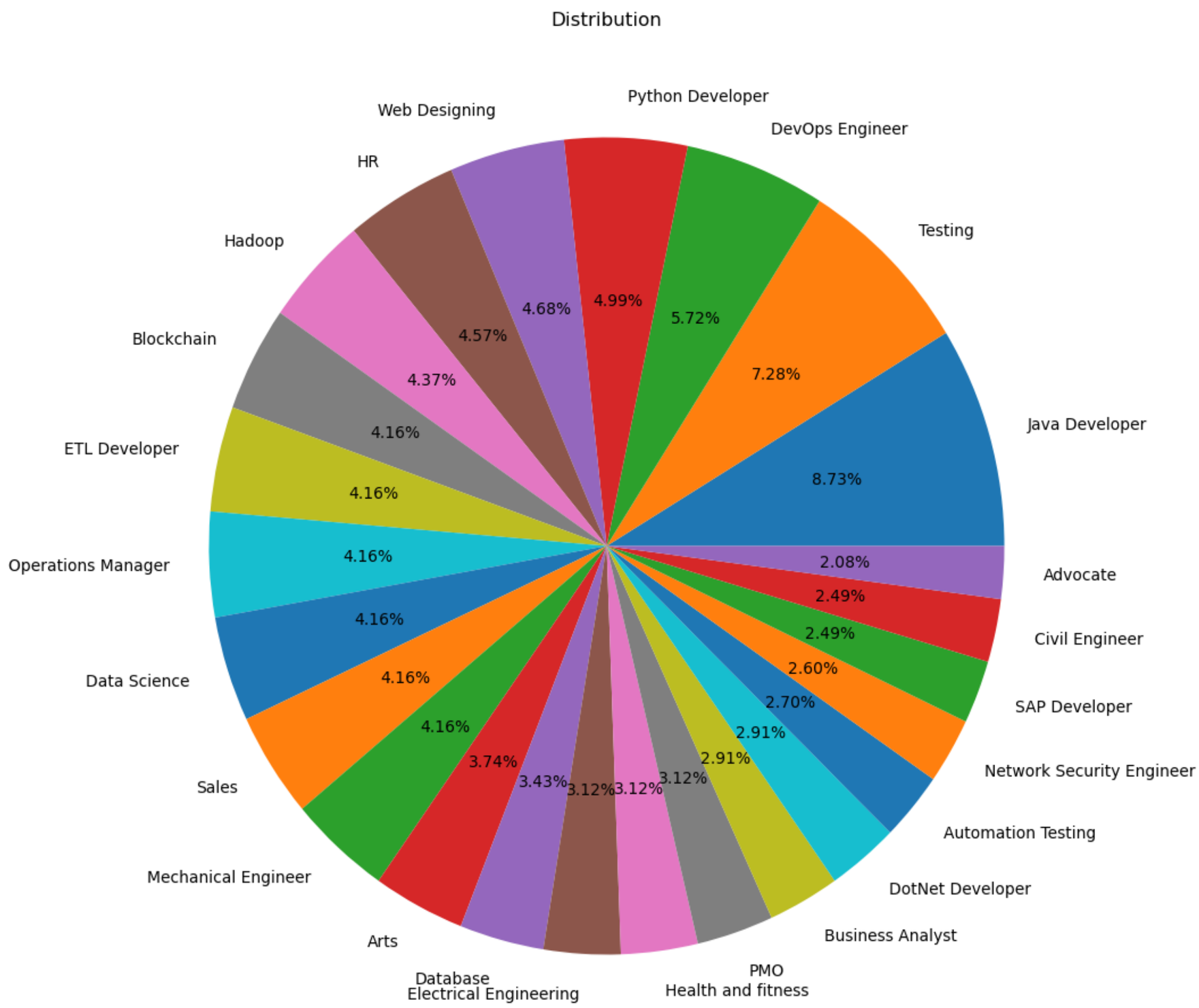
TESTING

Pre-employment tests measure skills, intellect, personality or other charcteristics of candidates. According to SHRM, the equal employment opportunity commission (EEOC) states any employment requirement set by an employer is considered to be a test. These tests include cognitive ability tests that measure intelligence, physical ability tests, aptitude tests and personality tests. Before implementing pre-employment testing, the employment testing, the employer needs to determine which tests are necessary, what skills, knowledge and who will monitor the test. Pre-employment testing can be time-consuming but can also provide valuable information.

OUTPUT

```
['Data Science' 'HR' 'Advocate' 'Arts' 'Web Designing'
 'Mechanical Engineer' 'Sales' 'Health and fitness' 'Civil Engineer'
 'Java Developer' 'Business Analyst' 'SAP Developer' 'Automation Testing'
 'Electrical Engineering' 'Operations Manager' 'Python Developer'
 'DevOps Engineer' 'Network Security Engineer' 'PMO' 'Database' 'Hadoop'
 'ETL Developer' 'DotNet Developer' 'Blockchain' 'Testing']
total unique category: 25
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
<Axes: xlabel='count', ylabel='Category'>
```


Data Science
HR
Advocate
Arts
Web Designing
Mechanical Engineer
Sales
Health and fitness
Civil Engineer
Java Developer
Business Analyst
SAP Developer
Automation Testing
Electrical Engineering
Operations Manager
Python Developer
DevOps Engineer
Network Security Engineer
PMO
Database
Hadoop
ETL Developer
DotNet Developer
Blockchain
Testing
Category
0
10
20
30
40
50
60
70
80
count


Distribution
Java Developer 8.73%
Testing 7.28%
DevOps Engineer 5.72%
HR 4.99%
Hadoop 4.68%
Blockchain 4.57%
ETL Developer 4.37%
Operations Manager 4.16%
Data Science 4.16%
Sales 4.16%
Mechanical Engineer 4.16%
Arts 4.16%
Database 4.16%
Electrical Engineering 4.16%
Advocate 2.08%
Civil Engineer 2.49%
SAP Developer 2.49%
Network Security Engineer 2.60%
Automation Testing 2.70%
DotNet Developer 2.91%
Business Analyst 2.91%
PMO 3.12%
Health and fitness 3.12%
3.12%
3.43%
3.74%

```
0      Skills Programming Languages Python pandas num...
1      Education Details May 2013 to May 2017 B E UIT...
2      Areas of Interest Deep Learning Control System...
3      Skills     R     Python     SAP HANA     Table...
4      Education Details MCA YMCAUST Faridabad Haryan...
                            ...
957    Computer Skills     Proficient in MS office Wo...
958        Willingness to a ept the challenges     Po...
959    PERSONAL SKILLS     Quick learner     Eagernes...
960    COMPUTER SKILLS SOFTWARE KNOWLEDGE MS Power Po...
961     Skill Set OS Windows XP 7 8 8 1 10 Database MY...
Name: clean text, Length: 962, dtype: object
```

	Category	Resume	clean text
0	6	Skills * Programming Languages: Python (pandas...	Skills Programming Languages Python pandas num...
1	6	Education Details \r\nMay 2013 to May 2017 B.E...	Education Details May 2013 to May 2017 B E UIT...
2	6	Areas of Interest Deep Learning, Control Syste...	Areas of Interest Deep Learning Control System...
3	6	Skills â¢ R â¢ Python â¢ SAP HANA â¢ Table...	Skills R Python SAP HANA Table...
4	6	Education Details \r\n MCA YMCAUST, Faridab...	Education Details MCA YMCAUST Faridabad Haryan...
...	...	...	...
957	23	Computer Skills: â¢ Proficient in MS office (...	Computer Skills Proficient in MS office Wo...
958	23	â– Willingness to accept the challenges. â– ...	Willingness to a ept the challenges Po...
959	23	PERSONAL SKILLS â¢ Quick learner, â¢ Eagerne...	PERSONAL SKILLS Quick learner Eagernes...
960	23	COMPUTER SKILLS & SOFTWARE KNOWLEDGE MS-Power ...	COMPUTER SKILLS SOFTWARE KNOWLEDGE MS Power Po...
961	23	Skill Set OS Windows XP/7/8/8.1/10 Database MY...	Skill Set OS Windows XP 7 8 8 1 10 Database MY...

9. CONCLUSION

"Resume Screening" is an important, but time consuming, part of the selection process. It is important that this phase of the staffing process be carefully organized and guided by application of the standards of professional ethics and legal constraints concerning discrimination. How well these tasks are accomplished directly effects the ultimate quality of staff employed in the organization.

10. REFRENCE

- https.wikipedia.com
- https://www.kaggle.com

GitHub link