

# **An Automated CV Sorting System**

**BY**

**Md.Saieedul Hasan Mahin  
ID:161-15-6878**

**Md.Naimur Raihan Emon  
ID: 161-15-7312**

**And**

**Abu Sazid Md. Shihab  
ID:161-15-6827**

This Report Presented in Partial Fulfillment of the Requirements for the  
Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

**Md.Firoz Hasan**  
Lecturer  
Department of CSE  
DaffodilInternational University



**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

**DECEMBER 2019**

## APPROVAL

This Research titled “**An Automated CV Sorting System**”, submitted by Abu Sazid Md. Shihab, ID No: 161-15-6827, Md. Saieedul Hasan Mahin, ID No:161-15-6878, and Md. Naimur Raihan Emon, ID No: 161-15-7312 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation had been held on December 07, 2019.

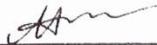
### BOARD OF EXAMINERS



**Dr. Syed Akhter Hossain**  
**Professor and Head**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Chairman**



**Nazmun Nessa Moon**  
**Assistant Professor**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

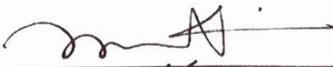
**Internal Examiner**



**Gazi Zahirul Islam**  
**Assistant Professor**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**



**Dr. Mohammad Shorif Uddin**  
**Professor**

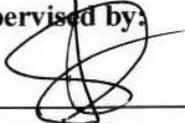
Department of Computer Science and Engineering  
Jahangirnagar University

**External Examiner**

## DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Md. Firoz Hasan, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

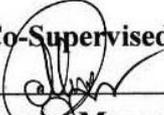
Supervised by:



**Md. Firoz Hasan**

Lecturer  
Department of Computer Science and Engineering  
Daffodil International University

Co-Supervised by:



**Dewan Mamun Raza**

Lecturer  
Department of Computer Science and Engineering  
Daffodil International University

Submitted by:



**Md. Saifedul Hasan Mahin**

ID: 161-15-6878  
Department of Computer Science and Engineering  
Daffodil International University



**Md. Naimur Raihan Emon**

ID: 161-15-7312  
Department of Computer Science and Engineering  
Daffodil International University



**Abu Sazid Md. Shihab**

ID: 161-15-6827  
Department of Computer Science and Engineering  
Daffodil International University

## ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year projects successfully.

We really grateful and wish our profound our indebtedness to **Md.Firoz Hasan, Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Data Mining*” to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Dr.Syed Akhter Hossain, Professor, and Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

## **ABSTRACT**

Our project is developing an important data retrieval framework. Here you can make a list of the most useful information by integrating various information into a single structure and formatting. We have done this in several parts. First of all, we provide a CV-sized template to all users. There all the users fill in all information such as their educational qualifications, work skills, experience etc. Secondly, we bring the information we fill in to the text file. The data is then automatically extracted through data mining. Where we have used it in Natural Programming Language. Now in the third section we publish a list based on merit. And this information is subsequently provided to the desired institution or industry. So that they can hire good and efficient staff. This whole process can be done in a very short time based on accurate results of 95%. All of this was done manually. Where it took a lot of time and many people needed to find the right candidate. That has just become easier with this project.

# TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
Approval	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
List of Figures	viii
List of Abbreviations	ix
<b>CHAPTER 1: INTRODUCTION</b>	<b>1-3</b>
1.1 Introduction	1
1.2 Motivation	1
1.3 Related Work	2
1.4 Outcome	2
1.5 Report Layout	3
<b>CHAPTER 2: BACKGROUND</b>	<b>4-5</b>
2.1 Introduction	4
2.2 CV/Resume Analysis	4
2.3 Natural Language Processing Approach	5
<b>CHAPTER 3: Requirement &amp; Specification</b>	<b>6-9</b>
3.1 Functional Requirement	6
3.2 Non-Functional Requirement	6
3.3 Data Flow Diagram	7
3.4 Sequence Diagram	8
3.5 DB Diagram	9

<b>CHAPTER 4: Analysis and Implementation</b>	<b>10-20</b>
4.1 Graphical Interphase	10-12
4.2 Database	13
4.3 Data Collection	14
4.4 Extract Data and Analyze	15-17
4.5 Extracting Keyword & Accuracy	18
4.6 Weight/Mark Distribution	18-20
<b>CHAPTER 5: Conclusion &amp; Future Work Scope</b>	<b>21-22</b>
5.1 Conclusion	21
5.2 Future Works Scope	21-22
<b>Reference</b>	<b>23-24</b>

## LIST OF FIGURES

<b>FIGURES</b>	<b>PAGE NO</b>
Figure 3.3. Data flow Diagram	7
Figure3.4. Sequence Diagram	8
Figure 3.5.DB diagram	9
Figure 4.1.1. Log in Form	10
Figure 4.1.2. Candidate Home Page	11
Figure 4.1.3. Submitting Information	11
Figure 4.1.4. Submitting Information	12
Figure 4.1.5. Submitting Information	12
Figure 4.1.6. Submitting Information	13
Figure 4.1.7. Submitting Information	13
Figure 4.2.1. Database	14
Figure 4.2.2. User Log in Database	15
Figure 4.4.1. Extract Raw Data	16
Figure 4.2.2. Marking Data Using NLP	17
Figure 4.2.3. After Marking Text File	18
Figure 4.6.1. Weight/Marks distribution	20
Figure 4.6.2. Output	21

## **List of Abbreviations**

**NLP** – Natural Language Processing

**CV** – Curriculum vitae

**HTML** – Hyper Text Markup Language

**PHP** – Hypertext Pre Processor (earlier called personal home page)

**ML** – Machine Language

# CHAPTER 1

## INTRODUCTION

### 1.1 Introduction

A student ends his / her study phase and goes out at his / her desired level. And that desired stage is a job. A CV carries all the educational qualifications, experience, skills of a student. There are some students who have done a lot of work, many skills in education. Which helps him get a job. But the way CV is monitored in every company or industry, many qualified candidates may be excluded even if they are qualified. Manually, the task at hand is to find the most suitable candidate. Because many candidates apply for a position. Because of this, finding qualified candidates is very difficult and time consuming. So, we set up a project, where candidates will have the opportunity to fill a semi-structured CV and provide all their information there. Our system will be able to automatically find qualified candidates through data mining. And it's a much simpler method.

### 1.2 Motivation

Currently there is a lot of work being done to make this task much easier manually. Some work solutions, but some have gone wrong. We have created a CV Semi Structured Format to solve this. Where candidates will provide their information. Natural Language Processing (NLP) and Regular Expression (RegEx) are used to bring that information into a structured format. Currently, a lot of work is done on Natural Language Processing (NLP) and Regular Expression (RegEX). And many researches and many important problems have been solved. Nowadays, NLP and RegEx in the field of e-mail, online shopping, online banking, etc.

### **1.3 Related Works**

There have been lots of work done for job searching process. Whereas, the process of selecting a candidate based on their CV / Resume has not been completely automated.

To solve this problem, an approach combined with Natural Language Processing (NLP) and Regular Expression.

Most importantly, these two topics are used in day to day life almost every day while using mail, online shopping etc.

Although there were some research to automate the process in some other way and there were some research to make the process less boring and easier at the same time, but there is still some room for improvement. Research shows that, 90% of all CVs/ Resumes are checked for less than 2 minutes [2] by the employers. This implies that, in most of the cases the employers only look at the bits of important parts or the points of interest in the CV/ Resumes and ignores the rest.

The specific segmentation scheme of a general CV/ Resume makes it far easier to analyze and understand the necessary information.

Therefore, the first objective was to segment the CV/ Resume into parts and then separate them in order to figure out the topics of each sentence through analyzing the keywords of each segment.

### **1.4 Outcome**

In our project, various companies will supply their specific requirements for their vacant positions in the company. Accordingly, the candidates will give information in our Semi Structured CV format. We can create a list of eligible candidates by mining data through our automatic NLP and RegEx. So that companies can get qualified candidates easily. Here our outcome is to find qualified candidates in less time. And it is possible to find the right qualified candidate up to 95%. Because we are able to extract the original information from the information given by the candidate by processing any common language.

## **1.5 Report Layout**

In this chapter we have discussed about the introduction of pharmacovigilance study, motivation, rationale of the study and the outcome of the thesis. Later followed by the report layout.

In Chapter 2, In this chapter we will describe the literature review of related & background works in this particular field.

In Chapter 3, In this Chapter we will discuss about requirement and specifications

In Chapter 4, In this Chapter we will show the experimental analysis on this topic and it also analyzes the system performance.

In Chapter 5, In this Chapter we will discuss about the conclusion and future work.

## **CHAPTER 02**

### **BACKGROUND**

#### **2.1 Introduction**

Generally, the data we receive from the CV is not structured or in our CV format we accept semi structured data. Structuring this semi structured data is a challenging thing. There are many types and many formats available in CV / Resume. Many people create their own unique and livelihoods through their unique style. Their CVs are different and highlighted by many. Many times, it is difficult to extract the eligible CV from all CVs. Because of this we take everyone in a CV / resume in our own format. And in all those CVs / resumes, a sorted list is created by analyzing data in natural programming languages. Which is a great addition to any company or organization. On the other hand, some CVs contain a lot of exceptional information that cannot be verified. So, an automatic method has to be found where the data will be verified. We have implanted the idea of what a human being looks like on a CV. Which will work automatically on each CV. Many people have encountered such problems before. We have taken this project to solve it. Moreover, many took the project to do such an automatic work. There were also some mistakes. So, we took on this challenging task.

#### **2.2 CV/resume analysis:**

CV / resume is monitored manually from past to present. And at this time there are many candidates for a position. It has been found that thousands of applications after a company or industrial organization leave job notification. But the problem is that it is possible to find qualified candidates from these thousands of CVs but there are many problems behind it. For example, many people need to go for thousands of CV verification, many times is wasted with a lot of trouble. If the time spent in a single office just to get a qualified candidate out of the CV, the other work of the office will be postponed. At the present time, many organizations and companies automate some parts

of a CV. For example, when a recruitment notification is given at Daffodil International University, candidates are tested on Varsity's own ERP system, regardless of the results of their past academic life. If it matches, he can upload his CV in varying ways. The only result here is automatic checking. But the results, skills, experiences, etc. of our CV / Resume will be automated which will be of great help as to why the company or the organization. It takes a lot less time to find qualified candidates. And many candidates in the office do not spend extra time in the elections.

### **2.3 Natural language processing approach**

Natural language processing is used to analyze data from a CV / resume. A specific CV/resume will be applied to natural language processing over a specific manner. Natural language processing is very helpful in all of these areas. Natural language processing will be used to extract important information and marking it, and a list will be given, which will allow any employer to select a suitable employer.

## **CHAPTER 03**

### **REQUIREMENT & SPECIFICATION**

#### **3.1 Functional requirement**

functional requirement describe what system should do. In our case we already know on what basis our system carries on its functional work. Our system should minimize the burden of HR department and make possible efficient E-HR system. Where companies create a job title and provide us the minimum requirements of certain job. Then candidate can visit our site and can create his account. After creating account candidate can fill job form of certain job position which he informed from companies and get job id. While filling the form he should give the correct job id he is applied for, which he gets from company's circular by paying the fee. He will go through many question sections asking about basic info and educational qualification. After providing all information about qualities he has and professional area, he can submit the form to our system. Our system will then mark his information with weight according to company requirements. And we will automatically review all the submitted form data by the help of data mining and Regular expression. Finally, we will rank up all the form according to requirements.

#### **3.2 Non-functional requirement**

Non-functional requirements tell us how system works. Basically, we have two parts, collecting data and providing result data and analyze the data. On collecting data path, we collect information using web page. Which build using PHP and Bootstrap. While collecting data we categories the data on basis how we need them. Such as personal information, educational information, skills, interested area, out of academic course, experience and many more related questions. So, by doing this we automatically turned the unstructured data to semi structured. Then we export certain job id candidates all

information to text file to analyze them and also mark them. We do the analyze part using regular expression and Natural Language Processing (NLP) on python.

### 3.3 Data Flow Diagram

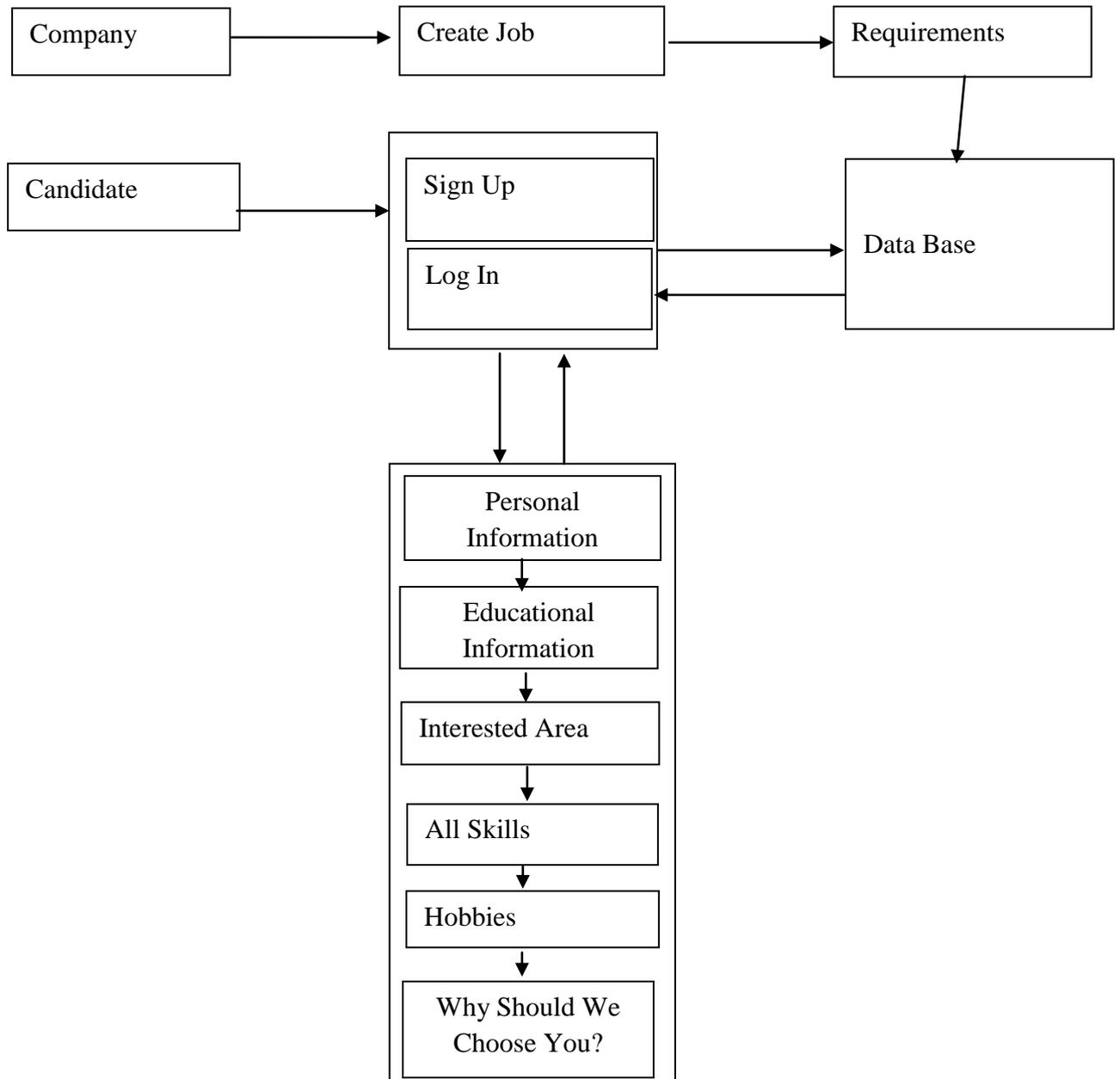


Figure3.3: Data Flow Diagram

### 3.4 Sequence Diagram

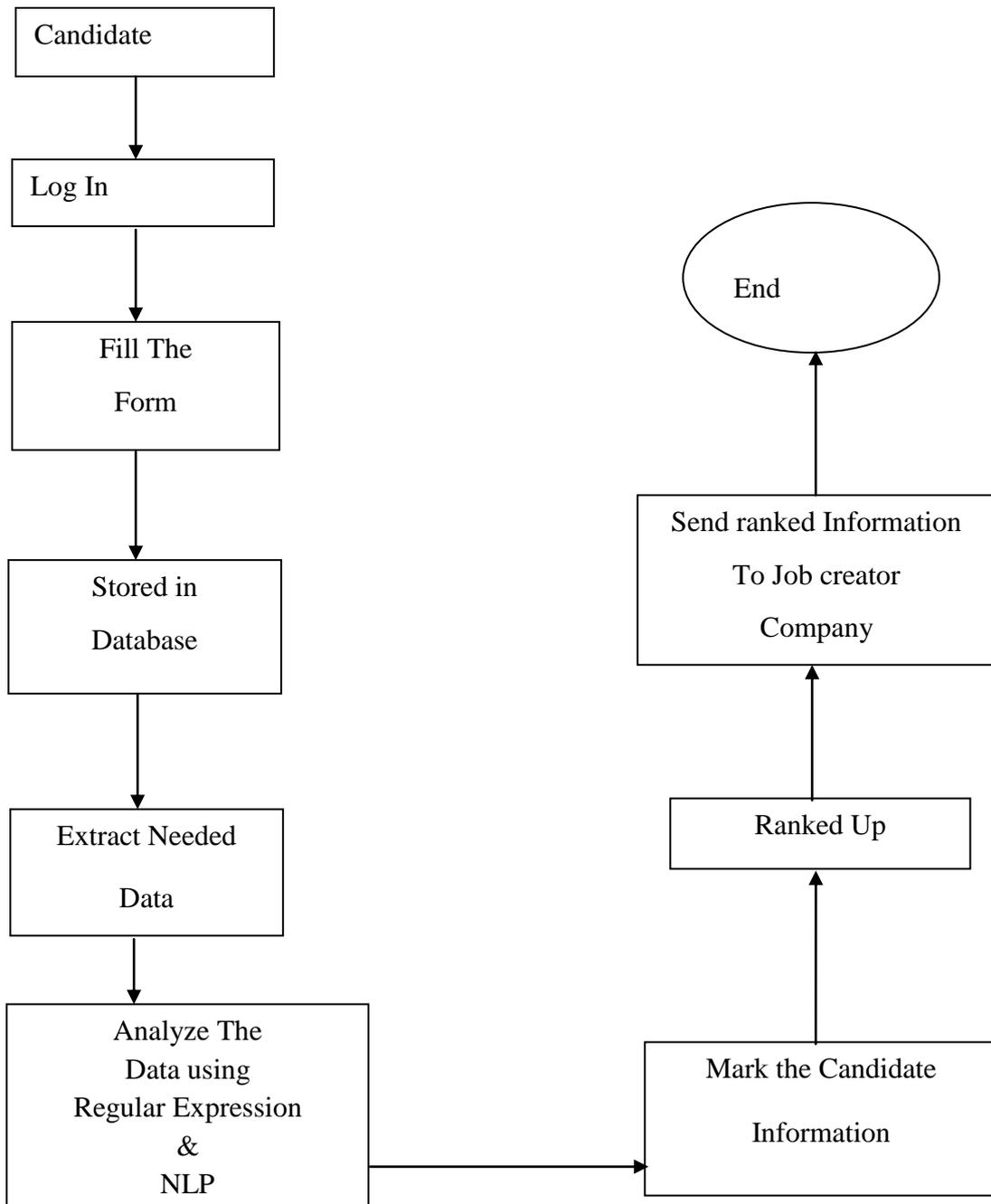


Figure 3.4: Sequence Diagram

### 3.5 DB Diagram

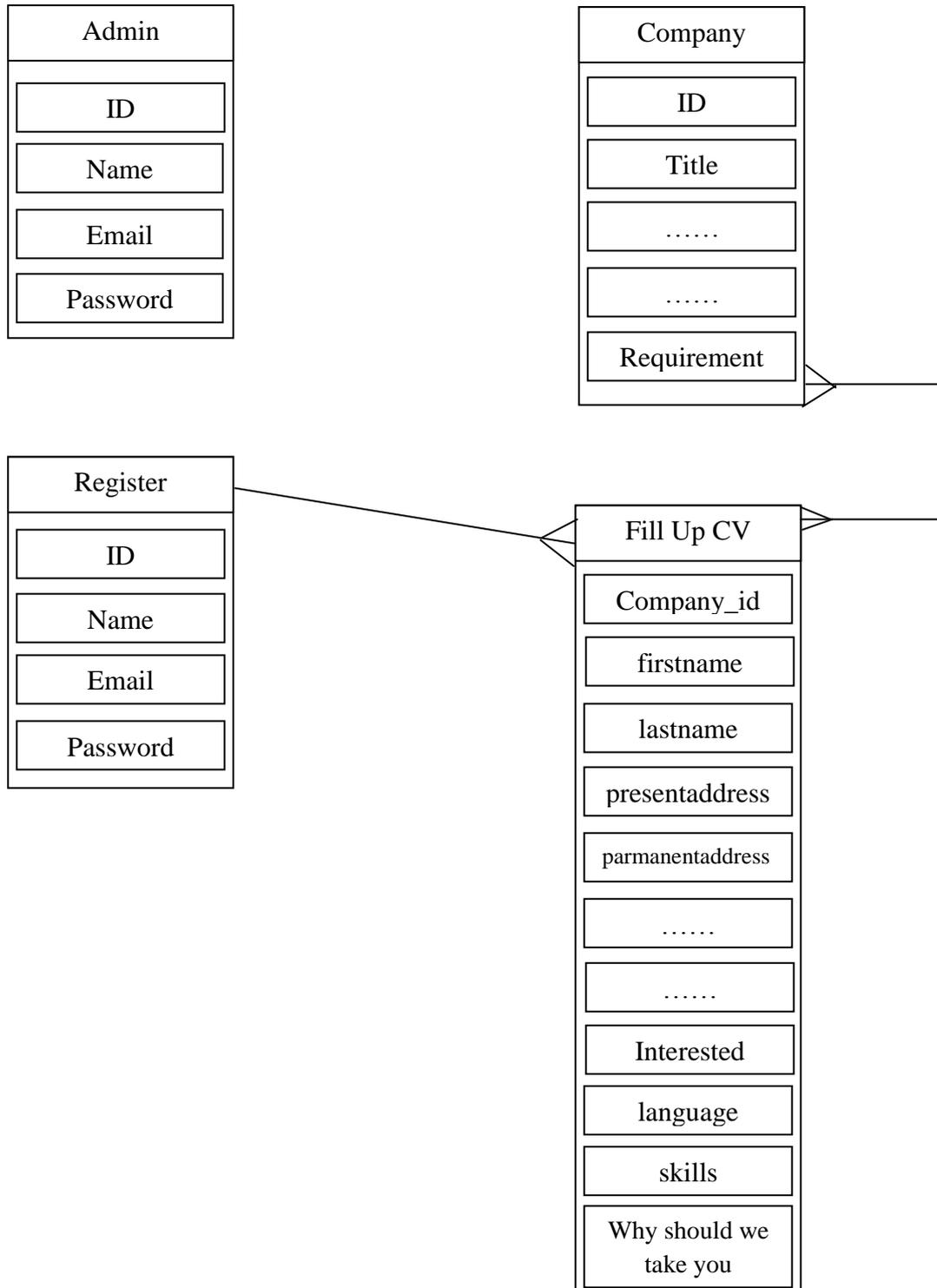


Figure 3.5: DB Diagram

# CHAPTER 04

## ANALYSIS AND IMPLEMENTATION

### 4.1 Graphical Interphase

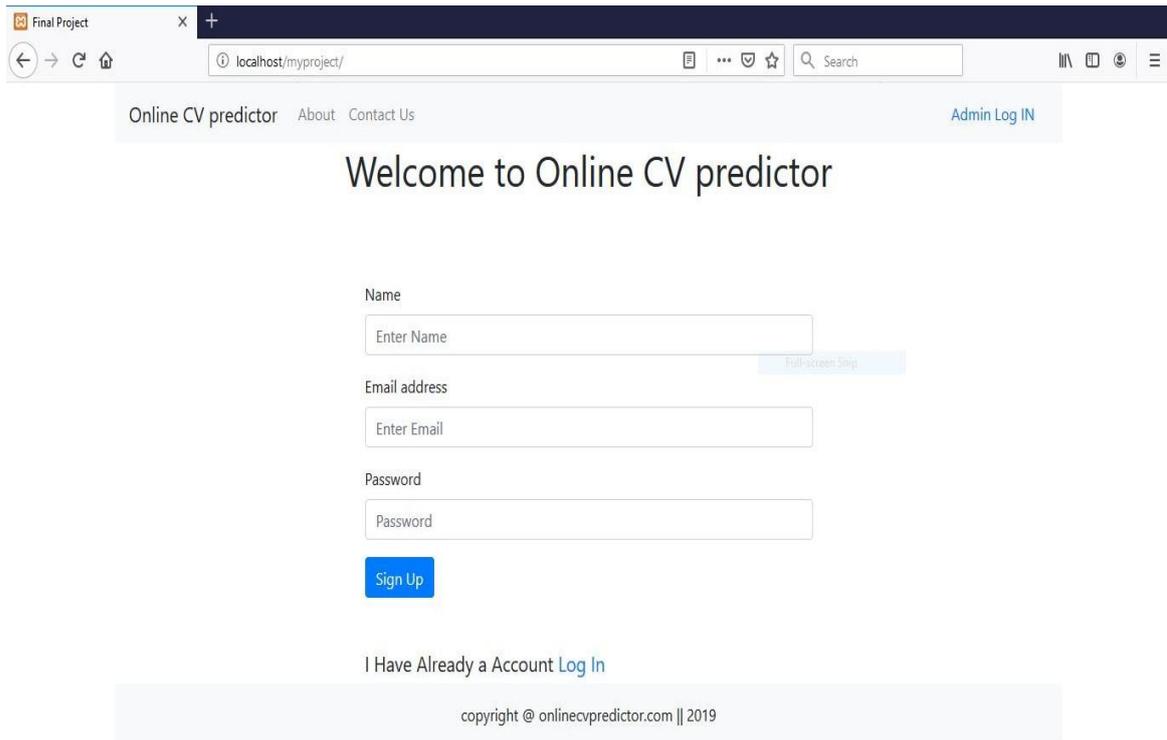
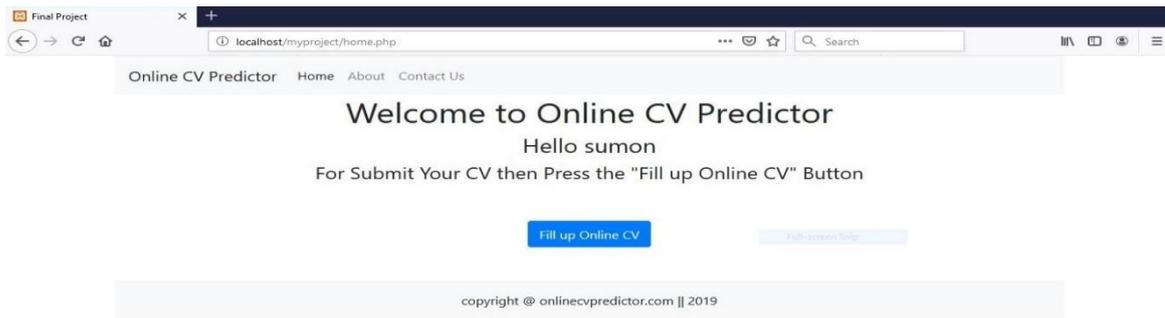


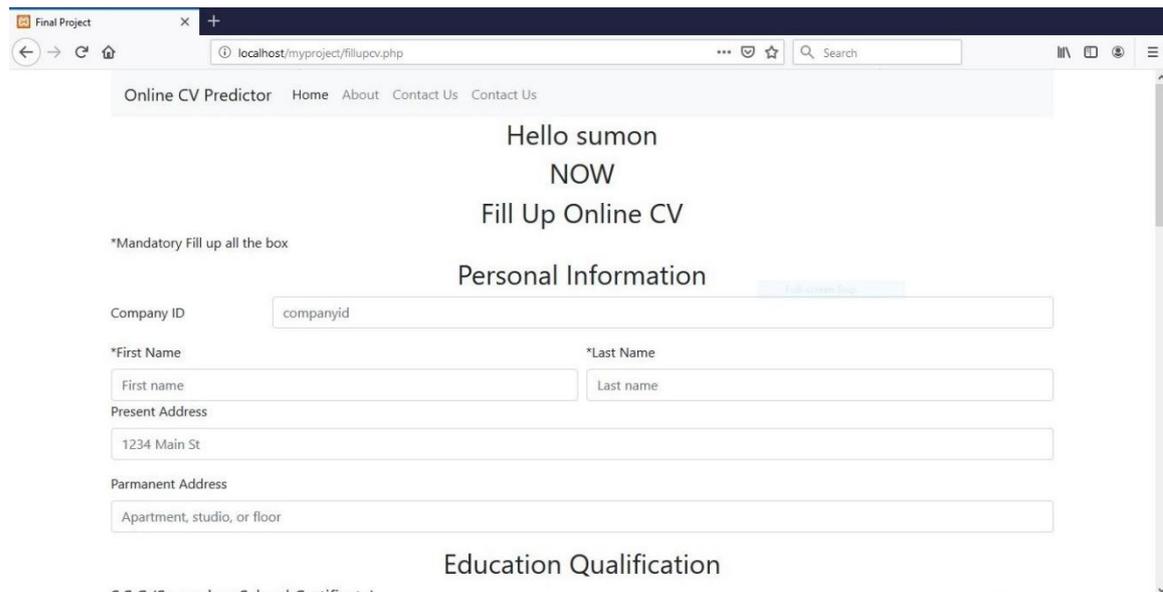
Figure 4.1.1: Log in Form

When any candidate enter the URL provided by the recruiter, this is the page they will face first. Then have to sign up first by using Name, Email & Password. If anyone having account before they just have log in using previous Email & Password.



**Figure 4.1.2: Candidate Home Page**

After successfully log in / sign up, This the interphase candidate will get. Now if candidate want to submit CV (Curriculum Vitae) they just have to “Click Fill up Online.



**Figure 4.1.3: Submitting Information**

CV” Button

Final Project x +

localhost/myproject/fillupcv.php

Subject: Computer Science And Engineering

CGPA: 3.8

Graduation Institute: Daffodil International University

Complete Graduation Year: 2016

Graduation Type: Masters of Science (M.S.)

Subject: Computer Science And Engineering

CGPA: 3.75

Next

copyright @ onlinecvpredictor.com || 2019

Figure 4.1.4: Submitting Information

Figure 4.1.5: Submitting Information

This is the first page of CV where candidate have to provide the basic personal

Final Project x +

localhost/myproject/fillupcv2.php

Online CV Predictor Home About Contact Us

connected

Form 2

Based on your previous form you graduate from :  
**Computer Science & Engineering**

Now Type your Most 3 interested Course Name and Described

Course 1

Course 1

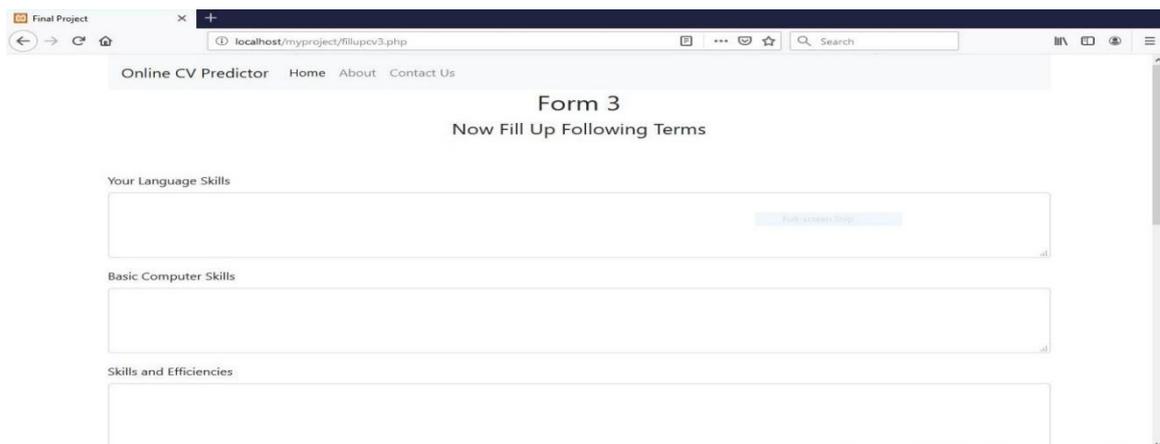
Write about this Course 1

Course 2

Course 2

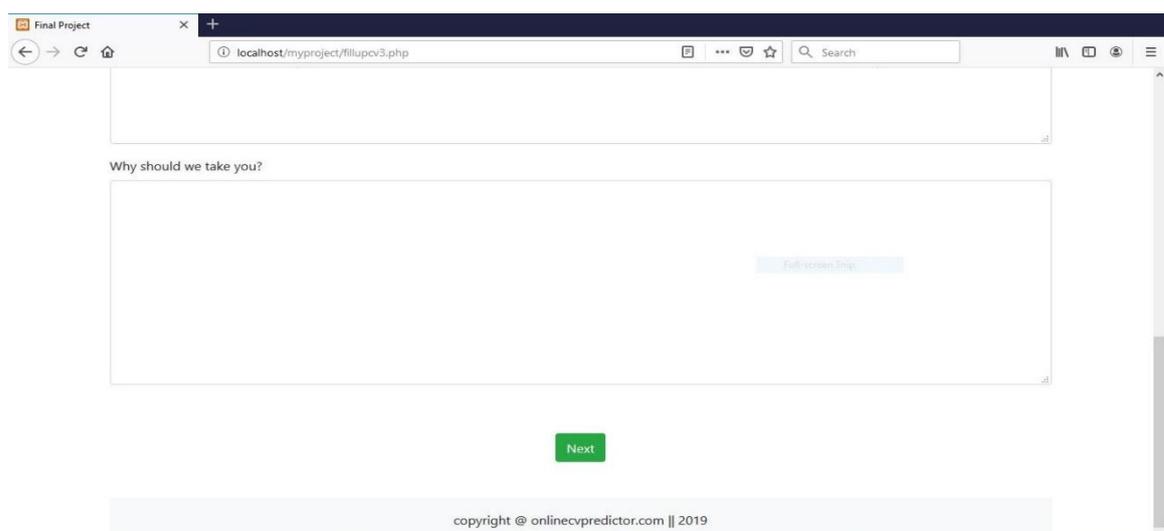
information (First name, Last name, Address. Educational qualification etc). Here system will check

Figure 4.1.6: Submitting Information



If grade does not fulfill the minimum grade asked by job creator/company candidate will be notified by message saying not eligible for this job. If successful, he will go to the next page.

Figure 4.1.7: Submitting Information



On this page, the candidate will explain which course from his background education (Undergraduate/Graduate) he prefers most.

After submitting the 2<sup>nd</sup> form, the system will take him to the third page. Here, the candidate will describe about his skills in computer, language, and other areas. At last, he faces the question, "Why should we choose you!". In this part, the candidate describes about his

professional skills regarding to the job circular and if his skills match the job requirement skills he will be rewarded with points.

## 4.2 Database

companyid	id	firstname	lastname	presentaddress	parmanentaddress	sscInstitute	scopyear	sscgroup	sscgpa	ssc
com124	13	Waheduzzaman	Mizan	sukrabad,dhaka-1000	Kushtia,Jessore	Kushtia govt. School	2007	Science	5.00	Je
com123	14	naimur		Sobhanbagh,Dhaka-1000	chandpur	B.N school and college	2013	Group	5.00	Rz
com124	15	ti	thi	dhanbondi	B.Baniagular Sntp	B.Bania girls school	2011	Business Studies	4.7	Cc
com123	16	akash	btash	chad	mongol	sokro high school	2011	Arts	2.00	Mh
com124	17	amin	khan	Sobhanbagh,Dhaka-1000	jamalpur	jamalpur zilla school	2010	Science	5.00	Rz
com123	18	kajol	kajollille	uttara	pirojpur	pirojpur high school	2008	Business Studies	5.00	Be
com123	19	shihab	jannat	cantorment	gajipur,dhaka	daffodil international school	2010	Science	3.00	Dh
com123	20	nijan	mijan	sukrabad,dhaka-1000	gajipur,dhaka	gajipur school	2001	Science	5.00	Dh
com124	21	online	offline	u		Passing Year		Group		Bc
com124	22	asif	iqbal	cantorment	Kushtia,Jessore	daffodil international school	2010	Arts	5.00	Dh
com124	23	scifon	khan	cantorment	Kushtia,Jessore	scifon school	2007	Science	4.9	Dh

Figure 4.2.1: Database

This is our database where data is stored after submitting the form/CV. Here we store data on 4 table. One for Admin, One for Job (while company creating any job), One for candidate Registration and One for Submitting Information (describe on Chapter 3.5). Registration table is relational to Information table by one to many relations. Job table links to Information table by many to many relations. Information all will be stored here as semi structured as we pre-defined many data. So, it will help us to structure the data more efficiently. For now, we use PhpMyAdmin to stored data on local machine what we get form our site.

			id	name	email	password	
<input type="checkbox"/>				1		d41d8cd98f00b204e9800998ecf8427e	
<input type="checkbox"/>				2	mahin	d41d8cd98f00b204e9800998ecf8427e	
<input type="checkbox"/>				3	Noman	d41d8cd98f00b204e9800998ecf8427e	
<input type="checkbox"/>				4	nahid	d41d8cd98f00b204e9800998ecf8427e	
<input type="checkbox"/>				5	hasan	hasan@gmail.com	81dc9bdb52d04dc20036dbd8313ed055
<input type="checkbox"/>				6	mahin	mahin@gmail.com	81dc9bdb52d04dc20036dbd8313ed055
<input type="checkbox"/>				7	Nahid	nahid@gmail.com	81dc9bdb52d04dc20036dbd8313ed055
<input type="checkbox"/>				8	hridoy	hridoy@gmail.com	81dc9bdb52d04dc20036dbd8313ed055
<input type="checkbox"/>				9	ratul	ratul@gmail.com	81dc9bdb52d04dc20036dbd8313ed055
<input type="checkbox"/>				10	shanto	shanto@gmail.com	81dc9bdb52d04dc20036dbd8313ed055
<input type="checkbox"/>				11	abu	abu@gmail.com	81dc9bdb52d04dc20036dbd8313ed055
<input type="checkbox"/>				12	akil	akil@gmail.com	e10adc3949ba59abbe56e057f20f883e
<input type="checkbox"/>				13	sumon	sumon@gmail.com	e10adc3949ba59abbe56e057f20f883e
<input type="checkbox"/>				14	shishir	shishir@gmail.com	e10adc3949ba59abbe56e057f20f883e

Figure 4.2.2: User Log in Database

To secure we use encryption method to covert entered password to string so that candidate password remains safe.

### 4.3 Data Collection

We collect data from Department of Computer Science & Engineering 43th batch G section to study our system. In near future we will collect more data if our system demand for justification of system process.



```

I am good with server site work. Also completed external course on it. I am also good with robotics. Also have five /5 years of working experiences.
Good with graphical work. Also have basic knowledge on java and c.
Good with python, R. Also learning java.
Know python, R.
I completed my honrs and master degree on economy. Have good knowledge on math and english too.
Know c, c#, java.
Good communication skill, good problem solving skills
Good knowledge and experiences on Python, java script, PHP.
Good experiences with object oriented programming . Such (C#, java, python.
Basic programming knowledge on c, python, java.
completed courses on CDMA, networking site experiences is what i have.
Because i have quality to take the responsibility for that post
Good programming skill, five years of experiences.
Good knowledge and experiences on Python, java script, PHP.
I think what you demand do match perfectly with me.
Miss u
I am good in c#, i have done lots of c# problem.
I am good in java, c#
Lots of work python, and database.
I'm experience with Python . I have done lot's of project on python . Also I know C and java
Python
Php,C
Skill attain in English
I want to build to good carrier & serve our country. Good knowledge on c++, java, python.
I'm experience with Python . I have done lot's of project on python . Also I know C and java
I'm experience with Python . I have done lot's of project on python . Also I know C and java
I know this language java Android Laravel
I know Java . I'm done lots of project with java. also i know good at python web frame work Django
I know Java . I'm done lots of project with java. also i know good at python web frame work Django
I know Java . I'm done lots of project with java. also i know good at python web frame work Django
I'm experience with Python . I have done lot's of project on python . Also I know C and java
I'm experience with Python . I have done lot's of project on python . Also I know C and java
I know Laravel Django Html
I'm experience with Python . I have done lot's of project on python . Also I know html and php
I know html php . I'm done lots of project with php. also i know good at python web frame work Django
I know java. I do a lot of project in java. I also make 3 websites....

```

```

File Edit View Insert Cell Kernel Widgets Help
In [7]: fh=open("D:\\Final year project\\data1.txt","r")
s=""
T= ['python','c','java','c#','five']
store = []
count=0
count2=0

while(s):
    s=fh.readline()
    count2=count2+1
    s=s.lower()
    table=str.maketrans("!.?\\", "6** ")
    s=s.translate(table)
    L=s.split()
    for object in T:
        if object in L:
            print(object)
            count=count+1

print(count)
store.append(count)
count=0
print(store)

fh=open('d:\\Final year project\\marks.txt','w')
for char in store:

```

```

File Edit View Insert Cell Kernel Widgets Help
In [4]: from textblob import TextBlob
fh=open("D:\\Final year project\\data1.txt","r")
s=""
g=""
store=[]

while(s):
    s=fh.readline()
    blob=TextBlob(s)
    g=blob.sentiment.polarity
    print(g)
    store.append(g)

fh=open('d:\\Final year project\\sentiment.txt','w')
for char in store:
    h=char
    fh.write('%f' %h)
    fh.write('\n')
fh.close()

```

```

0.46666666666666666
0.35
0.7

```

```

File Edit View Insert Cell Kernel Widgets Help
while(s):
    s=fh1.readline()
    count=count+1

count=count-3
print(count)

fh1.close()
fh2.close()

fh1=open("D:\\Final year project\\marks.txt","r")
fh2=open('D:\\Final year project\\sentiment.txt','r')

while(count<=count):
    marks=fh1.readline()
    sentiment=fh2.readline()

    sum=(float(marks)*1.75) + (float(sentiment)*3)
    store_sum.append(sum)
    print(sum)
    count1=count+1

fh=open('d:\\Final year project\\sum.txt','w')
for char in store_sum:
    h=char
    fh.write("%f" %h)
    fh.write('\n')

```

**Figure 4.4.2: Marking Data Using NLP**

Here education qualification marks calculated by our site auto pre define numbering system and we sum all together. But analyze part is unstructured textual data. We first detached them from qualification data and store them in another text file sequentially. When any candidate explain why we should choose you/What quality you have according to job requirement field on texture format this is natural language. So, we process it using NLP. We split the texture form into tokenized form. Then we search into this if candidate used a word which explain if he has the quality that recruiter asked for that job. From job recruiter specific word, we match candidate text token word. For every matching

candidate get a point. If any job required quality can explain by 5 keywords and candidate also define those keywords on his text, he will get full marks for keyword matching. But this is not all, we will also process the text further and find out the polarity of it. Where his attitude is positive or negative on job requirement qualification.

1	0.466667	3.150001
2	0.350000	4.550000
2	0.700000	5.600000
1	0.000000	1.750000
0	0.700000	2.100000
3	0.000000	5.250000
0	0.700000	2.100000
2	0.700000	5.600000
3	0.350000	6.300000
3	0.000000	5.250000
0	0.000000	0.000000
0	0.000000	0.000000
1	0.700000	3.850000
2	0.700000	5.600000
0	1.000000	3.000000
0	0.000000	0.000000
1	0.700000	3.850000
2	0.700000	5.600000
1	0.000000	1.750000
2	0.000000	3.500000
1	0.000000	1.750000
1	0.000000	1.750000
0	0.000000	0.000000
2	0.000000	5.600000
3	0.700000	5.250000
3	0.000000	5.250000
1	0.000000	1.750000
2	0.000000	5.600000
2	0.700000	5.600000
2	0.700000	5.600000
3	0.700000	5.600000
3	0.000000	5.250000
0	0.000000	5.250000
1	0.000000	0.000000
1	0.000000	1.750000

Figure 4.4.3: After Marking Text File

This sentiment analysis also finds out whether his quality on certain/required job field is poor, basic or good. Polarity valued -1 to 1. And then we altogether assign them as 10 marks. Where 70% from key tag and 30% sentiment marks. This 10 will be added with qualification marks.

## 4.5 Extracting Keyword & Accuracy

When we step to find out certain keyword that will cover any candidate skill or positive side for certain job, we have to consider the precision and recall of data. Precision and recall both indicate accuracy of data model. While we make decision for any candidate act on derive information from extract data is not 100% accurate. Accuracy may fluctuate from 75-85%. This is the back draw of automated system. So, we have to consider this. Suppose, if we want to extract skills from unstructured data and we have this text

I prefer Python over Java and I have five years experiences on python. I also good at PHP and Java Script. I also now working as junior software engineer on Data Science and I am dealing my work using R. I also developing my skill on Django framework for last 1 years and also Laravel.

If we simply extract skills from above text, we will have python, java, php, java script and R. But in real life skills are python, php, java script, R, Django and Laravel. So here precision is  $4/5$ . And recall is  $4/6$ . So, decision and recall value is important for any algorithm or any method to get exact result.

## 4.6 Weight/Marks Distribution

Suppose for any job company demand,

SSC 4.5 out 5, HSC 4.5 out of 5 and Undergraduate 3.2 out of 4 as minimum grade. Also 2 years working experiences. Need deep knowledge on python, SQL and PHP. General skills on java, java script. Also, deep computer skills.

Now our system will balance the weight marks of required. And suppose any candidate have following data and our system marked it or,

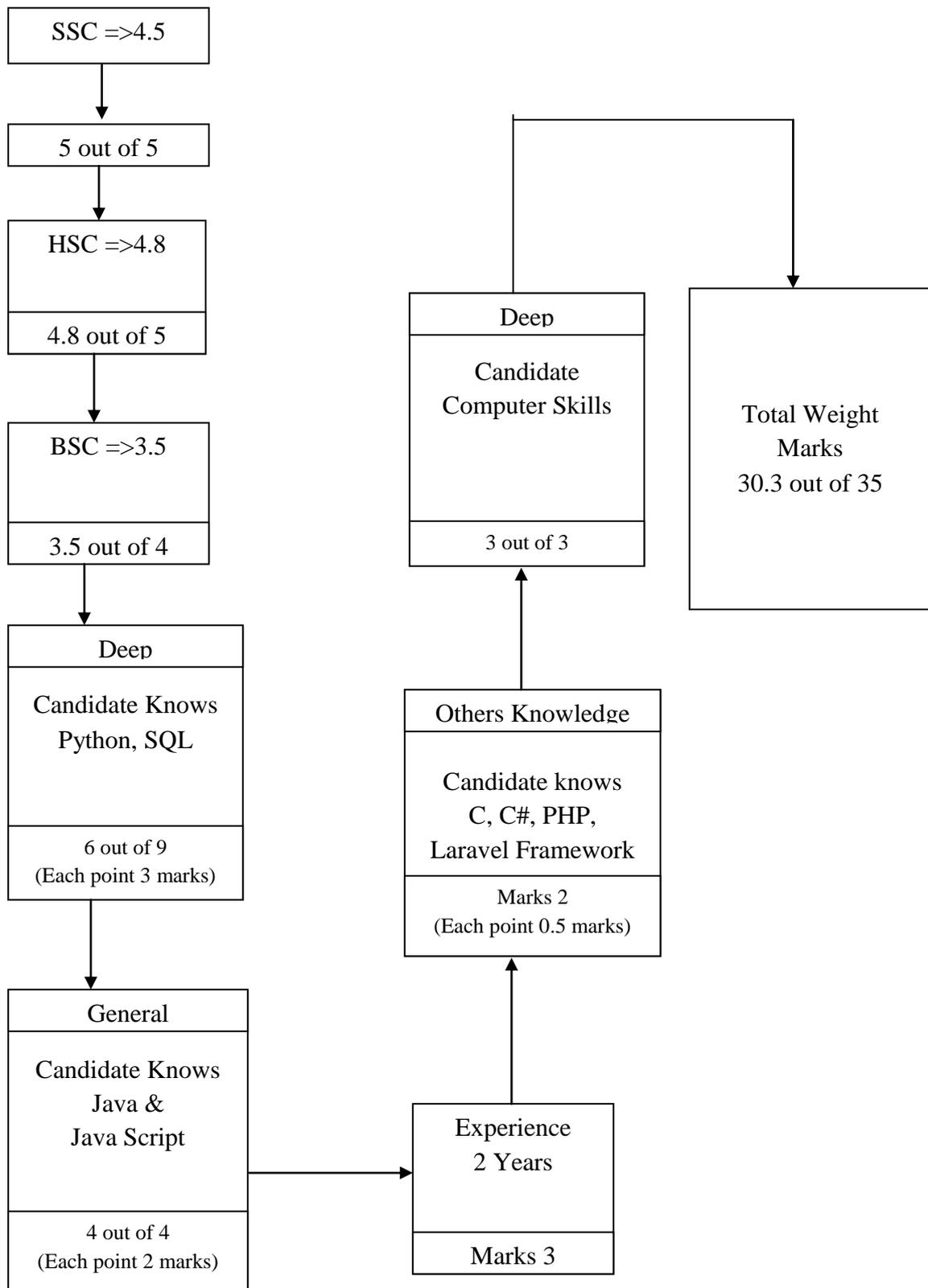


Figure 4.6.1: Weight/Marks Distribution

Job Vacancies List ▾

### Total Overview of Candidates

#### Final Result of Candidates

No.	Job ID	Name	Mobile Number	Education Based	skill Based	Requirements & Sentiment	Total Number
1	com123	Sazid	01856575	13.4/18	1/2	3.15/10	17.55/30
2	com123	sabbir	01856575	16.11/18	1/2	4.55/10	21.66/30
3	com123	Lima	01305698	16/18	2/2	5.60/10	23.6/30
4	com123	Khan	01765698	14.6/18	1/2	1.75/10	17.35/30
5	com123	Monsur	01616356	13.78/18	2/2	2.10/10	17.88/30
6	com123	Nahid	01631175	11.88/18	0/2	2.10/10	13.98/30
7	com123	Mehedi	01620369	11.58/18	2/2	5.25/10	18.83/30

#### Final Result of Candidates

No.	Job ID	Name	Mobile Number	Education Based	skill Based	Requirements & Sentiment	Total Number
1	com123	Sazid	01856575	13.4/18	1/2	3.15/10	17.55/30
2	com123	sabbir	01856575	16.11/18	1/2	4.55/10	21.66/30
3	com123	Lima	01305698	16/18	2/2	5.60/10	23.6/30
4	com123	Khan	01765698	14.6/18	1/2	1.75/10	17.35/30
5	com123	Monsur	01616356	13.78/18	2/2	2.10/10	17.88/30
6	com123	Nahid	01631175	11.88/18	0/2	2.10/10	13.98/30
7	com123	Mehedi	01620369	11.58/18	2/2	5.25/10	18.83/30
8	com123	Kaniz	01632568	16.4/18	1/2	2.10/10	19.5/30
9	com123	Mainul	01898756	13.33/18	2/2	5.60/10	20.93/30
10	com123	Saieedul	01716502	12.9/18	2/2	6.30/10	21.2/30
11	com123	Noor	01670290	16.6/18	2/2	5.25/10	23.85/30
12	com123	Sabbir	01856575	13.75/18	1/2	0.000/10	13.75/30

**Figure 4.6.2: Output**

After calculating educational qualification marks/weight, key skill weight, keytag requirements and sentiment analysis weight marks we add them together and maximum sum will be 30. Here we demonstrate a dummy data where 47 volunteers submit their data. They have various background study. Some from CSE, EEE, Economics and etc. And we unnumbered them where we finding a candidate who should have good skill on Python, c, C#, java and five years experiences. So, marks may vary.

## **CHAPTER 05**

### **CONCLUSION & FUTURE WORK SCOPE**

#### **5.1 Conclusion**

This project is used to extract important information from CV / resume and to do a single ranking on candidate by analyzing the data which is very important and relevant for the company or the organization. Companies or organizations can provide requirement skills which is suitable for certain job. There will be a unique id for the job. Based on that unique id, the candidate will fill CV / Resume. Then the filtered data will be analyzed and on basis of that analyze marks will be distributed. And final ranking will be done among the applicant.

Which will only be transferred to the job creators. The biggest advantage here would be for Job Seeker because it becomes very difficult for an undergraduate when he makes a CV / resume. But this project will have a specific template where Job Seekers can fill up their information and apply to many types of companies through our system. And for the job creator, the benefits will be less on their time and they will be able to recruit employers very quickly and efficiently.

#### **5.2 Future Works Scope**

So far, we have discussed with our project how this system will work. Nevertheless, there are some limitations to our system. Many times, it is not possible to verify all the information. This is why our system cannot provide 100% accurate information. Its volume is up to 70-80%. We will further improve our system verification so that the amount can be up to 90-95%. When we started as a final project, we encountered some problems. Such as CV format, data mining, analyzing processes etc. Gradually we learned about them and implanted this project with a small amount of data. So, it can be said that when we work with more data in the future, our system algorithm will be changed further so that more qualified candidates can be verified. The semi structured CV format that we had was pretty much fixed. There are many houses for data collection.

So, to solve this problem, we have thought to make the template in the front easier. So that candidates can easily provide their information. And they can also select their own information-based information. This will further benefit us in verifying. It can be easily evaluated with specific information. And the chances of finding the right candidate will increase as it will take even less time. It may be that the candidate's general information will be stored in advance. The candidate will then be able to update his stored information whenever he wants. And the data will go directly to whichever position you want to apply. And his application will be completed very easily. Then he will not have to fill the CV with waiting time. When a lot of data is stored and we will add learning to the machine in the future we will add it to our system. So that a lot of data can be worked easily without any problems. Currently, our system is that when a company is looking for an officer there is a job request for it. Then the candidates apply and, in our system, we verify the list of candidates based on the list of candidates and send it to the company. They take the next step on that list. But in the future when we have a lot of candidate CV reserves, including valuation, we can easily provide the companies with a list of why the company wants to directly evaluate the selected candidates. This made the company easily eligible candidates on an urgent basis. And through this, the candidates will be able to get jobs without qualification.

## REFERENCE

1. "Language-Check 0.8: Python Package Index," Pypi.python.org. N.p., 2016. Web. 15<sup>th</sup> October 2019.
2. "Machine Learning: What It Is and Why It Matters," Sas.com. N.p., 2016. Web. 15<sup>th</sup> October 2019..
3. McCallum, A., & Nigam, K. (1998, July). A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization* (Vol. 752, pp.41-48) 16<sup>th</sup> October 2019.
4. "An Overview of Current Research on Automated Essay Grading", S. Valenti, F. Neri and A. Cucchiarelli Blood, Ian. "Automated Essay Scoring: A Literature Review". Working Papers in TESOL and Applied Linguistics 11.(2011) (2016): n. pag. Web. 16<sup>th</sup> October 2019.
5. Automated Essay Grading" P. Reddy and G. Jambagi, University of Illinois, Chicago, USA, 2003 "Machine Learning: What It Is and Why It Matters," Sas.com. N.p., 2016. Web. 16<sup>th</sup> October 2019.
6. Cornelißen, T., (2008). The Interaction of Job Satisfaction, Job Search, and Job Changes .An Empirical Investigation with German Panel Data, Published online: 18 March 2008 Springer Science+Business Media B.V. 2008.
7. Albus, J. E., Anderson, R. H., Brayer, J. M., DeMori, R., Feng, H. Y., Horowitz, S. L., ... & Vamos, T. (2012). Syntactic pattern recognition, applications (Vol. 14). Springer Science & Business Media.
8. Bray, T. (2014). The javascript object notation (json) data interchange format.
9. Cios, K. J., & Sztandera, L. M. (1992, March). Continuous ID3 algorithm with fuzzy entropy measures. In *Fuzzy Systems, 1992., IEEE International Conference on* (pp.469-476). IEEE.
10. Jones, E., Oliphant, T., & Peterson, P. (2014). {SciPy}: open source scientific tools

for {Python}.

11. Ruggieri, S. (2002). Efficient C4. 5 [classification algorithm]. *IEEE transactions on knowledge and data engineering*, 14(2),438-444.
12. Ichihashi, H., Shirai, T., Nagasaka, K., & Miyoshi, T. (1996). Neuro-fuzzy ID3: a method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning. *Fuzzy sets and systems*, 81(1),157-167.
13. Helmut Berger and Dieter Merkl, A Comparison of Text-CategorizationMethodsApplied to N-gram Frequency Statistics, InAustralianJointConference on Artificial Intelligence,2004

Cv

ORIGINALITY REPORT

5%

SIMILARITY INDEX

1%

INTERNET SOURCES

0%

PUBLICATIONS

4%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Daffodil International University

Student Paper

3%

2

Submitted to University of Greenwich

Student Paper

1%

3

documents.mx

Internet Source

1%

Exclude quotes On

Exclude matches Off

Exclude bibliography On