APPENDIX 1

TITLE OF PROJECT

Job Recommendation System

END TERM REPORT

By

Aman, Gurleen, Vandana, kanika

NAME OF THE CANDIDATE(S)

Section	Rollno.
K18JF	22
K18JF	28
K18JF	61
K18JF	64

Department of Intelligent Systems

School of Computer Science Engineering

Lovely Professional University, Jalandhar

April-2020

APPENDIX 2

Student Declaration

This is to declare that this report has been written by us. No part of the report is copied from other sources. All information included from other sources have been duly acknowledged. We aver that if any part of the report is found to be copied, we shall take full responsibility for it.

Name of Students:-

Aman,Gurleen,

Vandana, Kanika

Roll No:-
22,28,61,64

APPENDIX 3

TABLE OF CONTENTS

TITLE **PAGE NO.**

Introduction

1. Background

The Internet-based recruiting platforms become a primary recruitment channel in most companies. While such platforms decrease the recruitment time and advertisement cost, they suffer from an inappropriateness of traditional information retrieval techniques like the Boolean search methods. Consequently, a vast amount of candidates missed the opportunity of recruiting. The recommender system technology aims to help users in finding items that match their personnel interests; it has a successful usage in e-commerce applications to deal with problems related to information overload efficiently. In order to improve the e-recruiting functionality, many recommender system approaches have been proposed. This article will present a survey of e-recruiting process and existing recommendation approaches for building personalized recommender systems for candidates/job matching.

Problem statement:

To build a highly personalized job recommendation system which suggests specific jobs postings to users based on their skill set.

To try different models and come up with the best possible model for recommendations and validation.

Datasets used:

Stack Overflow dataset:

https://www.kaggle.com/stackoverflow/stack-overflow-2018-developer-survey

US Technology Jobs dataset:

https://www.kaggle.com/PromptCloudHQ/us-technology-jobs-on-dice.com

Stack overflow Dataset:

- This dataset was obtained from a survey conducted by stack overflow in 2018.
- It contains information about users, their coding period and their experiences in various domains of information technology like languages, Frameworks, platforms etc.

US Technology Jobs dataset:

- This dataset contains various IT job postings in the United States.
- Contains job title, company, work location, job description and the type of employment.

Our Approach:

1. Content based filtering

a. Assumption: People are likely to choose jobs that are in their domain of expertise.
2. Collaborative filtering
    a. Recommending jobs based on user user system.

Validation:

- Clustering on both the profiles independently.
- Similarity between the profiles to identify a suitable job cluster for each user cluster.
- Similarity between clustered output and the recommended output.

1.1 Outcome of the project.

After going through this project we come to know about many concepts of data science and machine learning.

We learnt how to fetch data from kaggle and use this data for machine to learn and then give the outcome on basic of pervious data analysis.

1.2 Goals and objectives.

The aim of recommender systems is to assist users in finding their way through huge databases and catalogues, by filtering and suggesting relevant items taking into account or inferring the users" preferences (i.e., tastes, interests, or priorities).

Based on this objectives for Job Recommendation systems are

1. Study of matching technique to match the suitable candidate for job position.

2. Use of Collaborative Filtering to find best suitable match.

3. Use content-based approach that takes into consideration an organization's needs and the skills of candidate.

4. Comparison of matching technique with Collaborative Filtering and Content Based Recommendation.

2. Description of Project.

There are major requirements presented in literatures that should be derived when recommending candidates for a specific job
1. The matching of individuals to job depends on skills and abilities that individuals should have.
2. Recommending people is a bidirectional process that needs to take into account the preferences not only of the recruiter but also of the candidate.

3. Recommendations should be based on the candidate attributes, as well as the relational aspects that determine the fit between the person and the team members with whom the person will be collaborated.

4. Individual is considered to be unique; we cannot choose a single person several times such as a movie or book. Job recommendation problem is bidirectional recommendation between job-seeker and job.
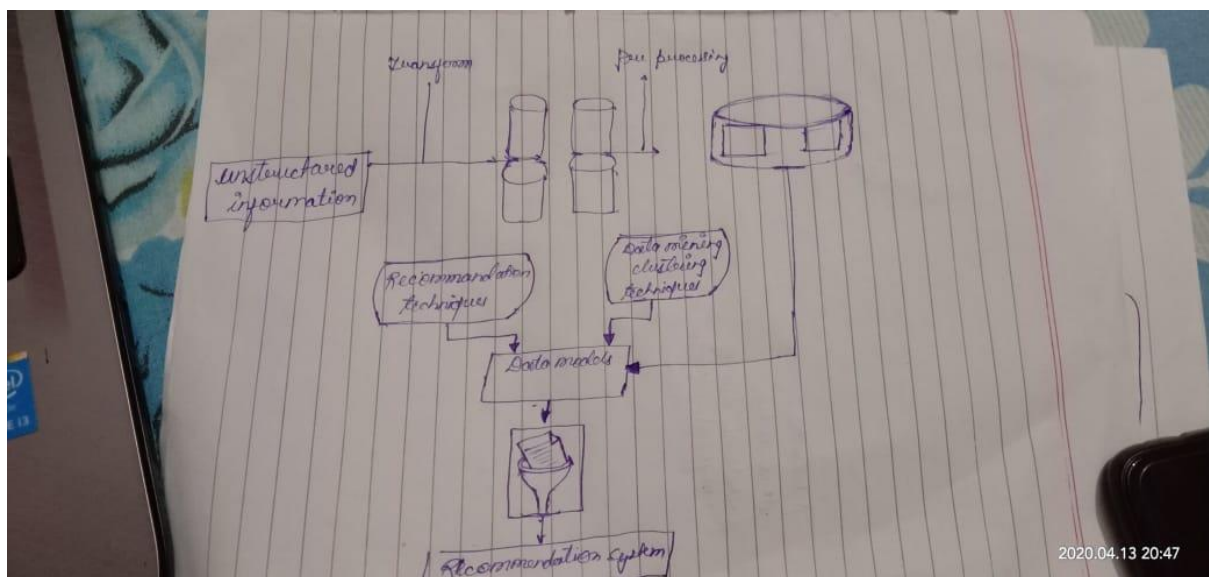
Job description
(eg. responsibilities, constraints on skills, experience level & degree)

candidate attributes
: analytical skills, certificates, professional experience

Job Recommendation requirements

Relational attributes
(eg. interpersonal compatibility skills such as trust)

3. Roles of the students.

- Aman Suryavanshi : I have fetch the files from kaggle and helped in coding.
- Gurleen Kaur : I have created all the programs.
- Vandana Kumari : I have created a report and built functions in the program.
- Kanika Thakur: I have created a report and linked all the files in the program.

4. Implementation of scheduled work of Project.

JOB RECOMMENDATION SYSTEM

i) Code for importing libraries.
ii) For fetching the details form the downloaded csv files.

	advertiserurl	company	employmenttype_jobstatus	jobdescription
5874	https://www.dice.com/jobs/detail/Python-Develo...	ICONMA	Full Time, full time	Python Developer Location: Houston, TXDuratio...
3299	https://www.dice.com/jobs/detail/Full%2526%252...	Cypress Group	Full Time, Permanent	My client is a world-class media and entertain...
7358	https://www.dice.com/jobs/detail/Full-Stack-Ja...	SolTech, Inc	Full Time	SOLTECH Staffing SolutionsAt SOLTECH, we work ...
11702	https://www.dice.com/jobs/detail/Full-Stack-De...	Saicon Consultants Inc.	Contract W2, 12 Months	Proficient in RESTful web servicesProficient i...
4009	https://www.dice.com/jobs/detail/System-Admini...	Samsung SDS America Inc	Full Time	Responsibilities:Responsible for deployment an...
8515	https://www.dice.com/jobs/detail/System-Admini...	Strivector	Full Time	"The best way to predict the future is to crea...
9981	https://www.dice.com/jobs/detail/Application-A...	Aria Systems	Full Time, PerYear	Title: Application AdministratorDepartment: Pr...

	https://www.dice.com/jobs/detail/Full-Stack-Ja...	Inc	Full Time	SolutionsAt SOLTECH, we work ...
11702	https://www.dice.com/jobs/detail/Full-Stack-De...	Saicon Consultants Inc.	Contract W2, 12 Months	Proficient in RESTful web servicesProficient I...
4009	https://www.dice.com/jobs/detail/System-Admini...	Samsung SDS America Inc	Full Time	Responsibilities:Responsible for deployment an...
8515	https://www.dice.com/jobs/detail/System-Admini...	Strivector	Full Time	"The best way to predict the future is to crea...
9981	https://www.dice.com/jobs/detail/Application-A...	Aria Systems	Full Time, PerYear	Title: Application AdministratorDepartment: Pr...
17522	https://www.dice.com/jobs/detail/Application-A...	Aria Systems	Full Time, PerYear	Title: Application AdministratorDepartment: Pr...
18190	https://www.dice.com/jobs/detail/System-Admini...	Strivector	Full Time	"The best way to predict the future is to crea...
21051	https://www.dice.com/jobs/detail/Full-Stack-En...	Universal Software Corporation	Full Time	We are seeking a Full Stack Engineeer to join ...

OpenSource	No
Country	Italy
Student	Yes, full-time
Employment	Not employed, and not looking for work
FormalEducation	Some college/university study without earning ...
UndergradMajor	Computer science, computer engineering, or sof...
CompanySize	NaN
DevType	Student
YearsCoding	3-5 years
YearsCodingProf	NaN
JobSatisfaction	NaN
CareerSatisfaction	NaN
HopeFiveYears	NaN
JobSearchStatus	I am not interested in new job opportunities
LastNewJob	I've never had a job
AssessJob1	1
AssessJob2	9
AssessJob3	5
AssessJob4	2
AssessJob5	7
AssessJob6	4
AssessJob7	8
AssessJob8	3
AssessJob9	10
AssessJob10	6
AssessBenefits1	2
AssessBenefits2	3
AssessBenefits3	11
AssessBenefits4	9
...	
EthicsResponsible	NaN
EthicalImplications	NaN
StackOverflowRecommend	NaN

```
AssessBenefits3                                    3
AssessBenefits4                                   11
                                                   9
EthicsResponsible                      ...
EthicalImplications                              NaN
StackOverflowRecommend                           NaN
StackOverflowVisit                               NaN
StackOverflowHasAccount                          NaN
StackOverflowParticipate                         NaN
StackOverflowJobs                                NaN
StackOverflowDevStory                            NaN
StackOverflowJobsRecommend                       NaN
StackOverflowConsiderMember                      NaN
HypotheticalTools1                               NaN
HypotheticalTools2                               NaN
HypotheticalTools3                               NaN
HypotheticalTools4                               NaN
HypotheticalTools5                               NaN
WakeTime                                         NaN
HoursComputer                                    NaN
HoursOutside                                     NaN
SkipMeals                                        NaN
ErgonomicDevices                                 NaN
Exercise                                         NaN
Gender                                           NaN
SexualOrientation                                NaN
EducationParents                                 NaN
RaceEthnicity                                    NaN
Age                                              NaN
Dependents                                       NaN
MilitaryUS                                       NaN
SurveyTooLong                                    NaN
SurveyEasy                                       NaN
Name: 2, Length: 128, dtype: object
```

```
8193   https://www.dice.com/jobs/detail/Data-Scientis...
10053  https://www.dice.com/jobs/detail/DATA-SCIENTIS...
11722  https://www.dice.com/jobs/detail/Data-Scientis...
143    https://www.dice.com/jobs/detail/Lead%2526%252...
911    https://www.dice.com/jobs/detail/Lead-Data-Eng...
5271   https://www.dice.com/jobs/detail/Data-Architec...
8192   https://www.dice.com/jobs/detail/Data-Scientis...
8361   https://www.dice.com/jobs/detail/Dillard%2527s...
8643   https://www.dice.com/jobs/detail/Scientist-Cyb...
8911   https://www.dice.com/jobs/detail/Scientist-Cyb...

                             company  \
8193                     INSYS Group
10053           Quantig iBizlline Inc
11722     Intelliswift Software Inc
143                 BayOne Solutions
911               InfoVision, Inc.
5271                   TRIAD Group
8192                   INSYS Group
8361                      Dillards
8643            CyberData Technologies
8911            CyberData Technologies

                         employmenttype_jobstatus  \
8193    Full Time, Contract Corp-To-Corp, Contract Ind...
10053   Contract Corp-To-Corp, Contract Independent, C...
11722   Full Time, Contract Corp-To-Corp, Contract Ind...
143                                         Full Time
911                                   Full Time, perm
5271    Full Time, Contract Corp-To-Corp, Contract Ind...
8192                                Full Time, permanent
8361                                        Full Time
8643                                        Full Time
8911

                                    jobdescription          jobid  \
8193    We are working with a prestigious Telecommunic...   Dice Id : insysus
                                                             Dice Id : 10371829
```

```
                                    jobdescription          jobid  \
8193    We are working with a prestigious Telecommunic...   Dice Id : insysus
10053   DATA SCIENTIST ROLE . Duration - 6 months with...   Dice Id : 10371829
11722   While applying please mention "DY_Data_Science...   Dice Id : 10108150
143     Title: Lead Data Scientist (Applied Researcher...   Dice Id : 10494547
911     Hello ,                                      ...   Dice Id : infotx
5271    Data Architectlocation: Seattle, WACompensatio...   Dice Id : TRIADW
8192    Insys is working directly with the Hiring Mana...   Dice Id : insysus
8361    DATA SCIENTISTThe Advanced Analytics and Strat...   Dice Id : 10138961
8643    CyberData Technologies, Inc., an established t...   Dice Id : RTX146efa
8911    CyberData Technologies, Inc., an established t...   Dice Id : RTX146efa

            joblocation_address                                jobtitle  \
8193                  Plano, TX                            Data Scientist
10053        Philadelphia, PA                          DATA SCIENTIST .
11722           Sunnyvale, CA                            Data Scientist
143         San Francisco, CA                   Lead/Sr Data Scientist
911                Waltham, MA                      Lead Data Engineer
5271               Seattle, WA                           Data Architect
8192               Warren, NJ                            Data Scientist
8361          Little Rock, AR   Dillard's Business Intelligence Data Scientist
8643              Sterling, VA                                Scientist
8911              Sterling, VA                                Scientist

            postdate                                       shift  \
8193       1 day ago    Telecommuting not available|Travel not required
10053      1 week ago   Telecommuting not available|Travel not required
11722      10 hours ago Telecommuting not available|Travel not required
143        5 hours ago  Telecommuting not available|Travel not required
911        2 weeks ago  Telecommuting not available|Travel not required
5271       1 week ago   Telecommuting not available|Travel not required
8192       1 day ago    Telecommuting not available|Travel not required
8361       1 day ago    Telecommuting not available|Travel not required
8643       3 hours ago  Telecommuting not available|Travel not required
8911       3 hours ago  Telecommuting not available|Travel not required

            site_name                                       skills  \
8193            NaN                                      Data Scientist
10053           NaN   R (SparkR), Spark mllib, H20, Python , Data Sc...
```

```
           postdate                                         shift  \
8193      1 day ago   Telecommuting not available|Travel not required
10053     1 week ago  Telecommuting not available|Travel not required
11722    10 hours ago Telecommuting not available|Travel not required
143       5 hours ago Telecommuting not available|Travel not required
911       2 weeks ago Telecommuting not available|Travel not required
5271      1 week ago  Telecommuting not available|Travel not required
8192      1 day ago   Telecommuting not available|Travel not required
8361      1 day ago   Telecommuting not available|Travel not required
8643      3 hours ago Telecommuting not available|Travel not required
8911      3 hours ago Telecommuting not available|Travel not required

          site_name                                         skills  \
8193            NaN                                    Data Scientist
10053           NaN   R (SparkR), Spark mllib, H2O, Python , Data Sc...
11722           NaN   Data scientist, Statistics, R programming, Python
143       www.dice.com                    Machine Learning, Big Data, Java
911       www.dice.com  1+ years of experience engineering ingestion a...
5271            NaN   Data, SQL, ETL, AWS, MPP, Reshift, Terdata, Ve...
8192            NaN                 digital data science statistical
8361            NaN   data analytics, business intelligence, data mi...
8643            NaN   Scientist, Meteorology, thermodynamic principl...
8911            NaN   Scientist, Meteorology, thermodynamic principl...

                              uniq_id
8193      4534e8231eeaee03bdf626804bc51698
10053     ad372b0e20ed9fcc3511275df26fc6ef
11722     07835301c44a2d7acda29297321a35d1
143       a1e1bd1243cdd15a3219e0edc8a93131
911       41dc8a866b6351ff07abdd3bfa778d04
5271      33b9d97b3cfad10b6e038b4980aa8bd1
8192      056dbfe00756ac85618a02dd05b6916b
8361      9eb58eb1421f01b105803fd6ed7ab793
8643      b08b2ba58739c65dd8ac4aec76259c7c
8911      8be9971d84f98ba538f73e61d909d84e
```

```
#print(stopwords)
#stopwords_list=stopwords
#words=[x for x in tokens if x not in stopwords_list]
#print(tokens)
extracted_skills[job_id]=[]
extracted_skills[job_id].extend(tokens)
#print(extracted_skills[job_id])
print(extracted_skills)
```

Out[33]:

.net c	aurelia	bottle	cakephp	cassandra	catalyst	...	vaadin	vanilla	vert.x	vue.js	wicket	xamarin	yarn	yii	zend	zope
√	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Part A: Building the content analyzer

In [ ]:
```
#Identifying various categories in job postings

#Method1: Applying TF-IDF on the dataset
count=0
docs=[]
for i in range(len(job_description[:100])):
    #print(job_description[i])
    if(job_description[i]==np.nan):
        continue
    doc=[x for x in job_description[i].split(" ") if x not in stopwords_list]
    docs.append(" ".join(doc))
print(len(docs))
```

iii) Identifying various categories in job posting code.

IV) Applying tf-idf on datasets.

V) To separate the domain fields in dataset.

VI) For extracting years of experience.

Vii) linkedin collab

```
files=glob.glob("../scraped profiles/*.json")
#print(files)
for file in files[:1]:
    f=open("../scraped profiles/"+file)
    data=json.load(f)
    print(data[0:2])
```

[{'positions': [{'title': 'Owner', 'is-current': True, 'company-name': 'AMBUS VENTURES'}], 'public-profile
-url': '/pub/amark-a/39/565/222', 'location': 'Israel', 'first-name': 'Amark', 'num-connections': '13', 'l
ast-name': 'A', 'industry': 'Wholesale'}, {'skills': ['Channel Management', 'Team Management', 'Key Accoun
t Management', 'People Management', 'Customer Relations'], 'positions': [{'title': 'Relationship Manager-
WM', 'start-date': '2009-04-01', 'is-current': True, 'company-name': 'INGVysya Bank Ltd'}, {'title': 'Mana
ger-Sales', 'end-date': '2009-03-01', 'start-date': '2008-05-01', 'company-name': 'Ikya Human Capital Solu
tions Pvt Ltd'}, {'title': 'Sales Manager', 'end-date': '2008-04-01', 'start-date': '2007-08-01', 'company
-name': 'HDFC Bank ltd'}, {'title': 'Manager- Life insurance', 'end-date': '2007-08-01', 'start-date': '20
03-08-01', 'company-name': 'ICICI Bank'}], 'public-profile-url': '/pub/amarnath-reddy-a/8/52a/280', 'locat
ion': 'Hyderabad Area, India', 'first-name': 'Amarnath Reddy', 'num-connections': '84', 'educations': [{'s
chool-name': 'Jawaharlal Nehru Technological University', 'end-date': '2002-12-31', 'start-date': '1998-01
-01', 'degree': 'B tech'}, {'school-name': 'Jayaprakash Narayan College Of Engineering', 'end-date': '2002
-12-31', 'start-date': '1998-01-01'}], 'last-name': 'A', 'industry': 'Banking'}]

Viii) For splitting the data for test.

```
         65272

In [26]: test.shape[0]

Out[26]: 16318

In [27]: df1.shape[0]

Out[27]: 81590

In [28]: test.to_csv("test_user.csv")

In [29]: train.to_csv("train_user.csv")

  In [ ]:
```

ix) For analysing user profile.

```
       Unnamed: 0  Respondent  Engineering manager  QA or test developer  \
0               0           1                    1                   NaN
1               1           3                  NaN                   NaN
2               2           4                  NaN                   NaN
3               3           5                  1.0                   NaN
4               4           7                  NaN                   NaN
                                                                    1.0

   Designer  Mobile developer  System administrator  \
0       NaN               NaN                   NaN
1       NaN               NaN                   1.0
2       NaN               NaN                   NaN
3       NaN               NaN                   NaN
4       NaN               NaN                   NaN

   Game or graphics developer  Product manager  Data or business analyst  \
0                          NaN              NaN                       NaN
1                          NaN              NaN                       NaN
2                          NaN              NaN                       NaN
3                          NaN              NaN                       NaN
4                          1.0              NaN                       1.0

      Mac OS  Raspberry Pi  Heroku  Google Home  Serverless  \
0  ...    NaN          NaN     NaN          NaN         NaN
1  ...    NaN          NaN     NaN          NaN         NaN
2  ...    NaN          NaN     NaN          NaN         NaN
3  ...    NaN          NaN     NaN          NaN         NaN
4  ...    NaN          NaN     NaN          NaN         NaN

   Windows Phone  Arduino  Linux  Drupal  ESP8266
0            NaN      NaN    1.0     NaN      NaN
1            NaN      NaN    1.0     NaN      NaN
2            NaN      NaN    NaN     NaN      NaN
```

x) For job posting preprocessing.

```
                                          advertiserurl  \
0  https://www.dice.com/jobs/detail/AUTOMATION-TE...
1  https://www.dice.com/jobs/detail/Information-S...
2  https://www.dice.com/jobs/detail/Business-Solu...
3  https://www.dice.com/jobs/detail/Java-Develope...
4  https://www.dice.com/jobs/detail/DevOps-Engine...

                          company  \
0      Digital Intelligence Systems, LLC
1      University of Chicago/IT Services
2                   Galaxy Systems, Inc.
3                          TransTech LLC
4                       Matrix Resources

                      employmenttype_jobstatus  \
0  C2H Corp-To-Corp, C2H Independent, C2H W2, 3 M...
1                                      Full Time
2                                      Full Time
3                                      Full Time
4                                      Full Time

                                     jobdescription           jobid  \
0  Looking for Selenium engineers...must have sol...  Dice Id : 10110693
1  The University of Chicago has a rapidly growin...  Dice Id : 10114469
2  GalaxE.SolutionsEvery day, our solutions affec...  Dice Id : CXGALXYS
3  Java DeveloperFull-time/direct-hireBolingbrook...  Dice Id : 10113627
4  Midtown based high tech firm has an immediate ...  Dice Id : matrixga

   joblocation_address                                       jobtitle  \
0         Atlanta, GA                        AUTOMATION TEST ENGINEER
1         Chicago, IL                   Information Security Engineer
2      Schaumburg, IL                    Business Solutions Architect
3     Bolingbrook, IL  Java Developer (mid level)- FT- GREAT culture,...
4         Atlanta, GA                                 DevOps Engineer

        postdate                                          shift site_name  \
0    1 hour ago  Telecommuting not available|Travel not required        NaN
```

xi) For error validation

     0.00000000e+00,   2.71050543e-20,   0.00000000e+00,   0.00000000e+00,
     5.42101086e-20,   0.00000000e+00,   0.00000000e+00,   1.62630326e-19,
     5.42101086e-20,   2.71050543e-20,   1.62630326e-19,   1.62630326e-19,
     0.00000000e+00,   5.42101086e-20,   8.19672131e-03,   0.00000000e+00,
    -6.93889390e-18,   0.00000000e+00,   5.42101086e-20,   4.09836066e-02,
     0.00000000e+00, -2.16840434e-18,   9.01639344e-02,   2.71050543e-20,
     1.63934426e-02,   0.00000000e+00, -2.16840434e-18, -6.93889390e-18,
    -3.90312782e-18, -1.56125113e-17,   2.71050543e-20,   2.71050543e-20,
     2.71050543e-20,   6.50521303e-19,   5.74374707e-02,   0.00000000e+00,
     3.52935110e-02,   6.32491681e-02,   1.77073116e-02,   0.00000000e+00,
    -1.73472348e-18,   0.00000000e+00,   0.00000000e+00,   0.00000000e+00,
     4.40421653e-02,   3.97217594e-02,   2.74489931e-02,   1.93859536e-02,
     7.09272081e-02,   2.60747162e-02,   5.16097226e-02,   7.63854935e-02,
     3.24712829e-02,   5.45044177e-02,   3.07272174e-01,   0.00000000e+00,
     4.02854804e-02,   3.61831723e-02,   0.00000000e+00,   6.55737705e-02,
     4.09836066e-02,   4.09836066e-02,   1.63934426e-02,   2.71050543e-20,
     2.71050543e-20,   5.42101086e-20,   8.19672131e-03,   3.25260652e-19,
    -6.50521303e-18,   1.62630326e-19, -1.08420217e-18,   5.42101086e-20,
     8.19672131e-03,   2.45901639e-02,   1.08420217e-19,   5.42101086e-20,
     8.13151629e-20,   6.50521303e-19, -2.16840434e-19, -3.25260652e-18,
     8.13151629e-20,   8.19672131e-03,   3.25260652e-19,   8.13151629e-20,
     8.19672131e-03,   2.45901639e-02,   5.42101086e-20,   0.00000000e+00,
     2.71050543e-20,   6.93889390e-18,   5.42101086e-20, -3.25260652e-18,
     2.71050543e-20,   8.19672131e-03,   5.42101086e-20,   2.71050543e-20,
     5.42101086e-20,   2.71050543e-20,   1.62630326e-19,   2.45901639e-02,
     1.64798730e-17, -3.03576608e-18,   8.19672131e-03,   5.42101086e-20,
     5.42101086e-20,   2.71050543e-20,   5.42101086e-20,   4.77048956e-18,
     5.42101086e-20,   5.42101086e-20,   1.62630326e-19,   2.71050543e-20,
     2.71050543e-20,   5.42101086e-20,   2.71050543e-20, -3.90312782e-18,
     1.63934426e-02, -3.03576608e-18,   8.19672131e-03, -5.42101086e-19,
     5.42101086e-20,   8.13151629e-20,   0.00000000e+00,   5.42101086e-20,
     5.42101086e-20,   2.71050543e-20,   1.08420217e-19,   5.42101086e-20,
     1.62630326e-19, -3.90312782e-18, -3.90312782e-18,   5.42101086e-20,
    -1.62630326e-18,   2.71050543e-20,   1.63934426e-02,   1.04083409e-17,

array([0.00000000e+00,   0.00000000e+00,   0.00000000e+00,   0.00000000e+00,
     0.00000000e+00,   2.71050543e-20,   0.00000000e+00,   0.00000000e+00,
     5.42101086e-20,   0.00000000e+00,   0.00000000e+00,   1.62630326e-19,
     5.42101086e-20,   2.71050543e-20,   1.62630326e-19,   1.62630326e-19,
     0.00000000e+00,   5.42101086e-20,   8.19672131e-03,   0.00000000e+00,
    -6.93889390e-18,   0.00000000e+00,   5.42101086e-20,   4.09836066e-02,
     0.00000000e+00, -2.16840434e-18,   9.01639344e-02,   2.71050543e-20,
     1.63934426e-02,   0.00000000e+00, -2.16840434e-18, -6.93889390e-18,
    -3.90312782e-18, -1.56125113e-17,   2.71050543e-20,   2.71050543e-20,
     2.71050543e-20,   6.50521303e-19,   5.74374707e-02,   0.00000000e+00,
     3.52935110e-02,   6.32491681e-02,   1.77073116e-02,   0.00000000e+00,
    -1.73472348e-18,   0.00000000e+00,   0.00000000e+00,   0.00000000e+00,
     4.40421653e-02,   3.97217594e-02,   2.74489931e-02,   1.93859536e-02,
     7.09272081e-02,   2.60747162e-02,   5.16097226e-02,   7.63854935e-02,
     3.24712829e-02,   5.45044177e-02,   3.07272174e-01,   0.00000000e+00,
     4.02854804e-02,   3.61831723e-02,   0.00000000e+00,   6.55737705e-02,
     4.09836066e-02,   4.09836066e-02,   1.63934426e-02,   2.71050543e-20,
     2.71050543e-20,   5.42101086e-20,   8.19672131e-03,   3.25260652e-19,
    -6.50521303e-18,   1.62630326e-19, -1.08420217e-18,   5.42101086e-20,
     8.19672131e-03,   2.45901639e-02,   1.08420217e-19,   5.42101086e-20,
     8.13151629e-20,   6.50521303e-19, -2.16840434e-19, -3.25260652e-18,
     8.13151629e-20,   8.19672131e-03,   3.25260652e-19,   8.13151629e-20,
     8.19672131e-03,   2.45901639e-02,   5.42101086e-20,   0.00000000e+00,
     2.71050543e-20,   6.93889390e-18,   5.42101086e-20, -3.25260652e-18,
     2.71050543e-20,   8.19672131e-03,   5.42101086e-20,   2.71050543e-20,
     5.42101086e-20,   2.71050543e-20,   1.62630326e-19,   2.45901639e-02,
     1.64798730e-17, -3.03576608e-18,   8.19672131e-03,   5.42101086e-20,
     5.42101086e-20,   2.71050543e-20,   5.42101086e-20,   4.77048956e-18,
     5.42101086e-20,   5.42101086e-20,   1.62630326e-19,   2.71050543e-20,
     2.71050543e-20,   5.42101086e-20,   2.71050543e-20, -3.90312782e-18,
     1.63934426e-02, -3.03576608e-18,   8.19672131e-03, -5.42101086e-19,
     5.42101086e-20,   8.13151629e-20,   0.00000000e+00,   5.42101086e-20,
     5.42101086e-20,   2.71050543e-20,   1.08420217e-19,   5.42101086e-20,
     1.62630326e-19, -3.90312782e-18, -3.90312782e-18,   5.42101086e-20,
    -1.62630326e-18,   2.71050543e-20,   1.63934426e-02,   1.04083409e-17,
     8.13151629e-20,   1.62630326e-19, -6.07153217e-18,   6.55737705e-02,

```
     Respondent  Back End  C-suite executive  Data Scientist  \
0             1       0.0                0.0             0.0
1             3       0.0                0.0             0.0
2             4       0.0                0.0             0.0
3             5       0.0                0.0             0.0
4             7       0.0                0.0             0.0

   Data or business analyst  Database Administrator(DBA)  Designer  DevOps  \
0                       0.0                          0.0       0.0     0.0
1                       0.0                          1.0       0.0     1.0
2                       0.0                          0.0       0.0     0.0
3                       0.0                          0.0       0.0     0.0
4                       1.0                          0.0       0.0     0.0

   Educator or academic researcher  \
0                              0.0
1                              0.0
2                              0.0
3                              0.0
4                              0.0

   Embedded applications or devices developer  ...  mac os  mainframe  \
0                                          0.0  ...     0.0        0.0
1                                          0.0  ...     0.0        0.0
2                                          0.0  ...     0.0        0.0
3                                          0.0  ...     0.0        0.0
4                                          0.0  ...     0.0        0.0

   predix  raspberry pi  salesforce  serverless  sharepoint  \
0     0.0           0.0         0.0         0.0         0.0
1     0.0           0.0         0.0         0.0         0.0
2     0.0           0.0         0.0         0.0         0.0
3     0.0           0.0         0.0         0.0         0.0
4     0.0           0.0         0.0         0.0         0.0

   windows desktop or server  windows phone  wordpress
0                        0.0            0.0        0.0
1                        0.0            0.0        0.0
```

```
                              uniq_id  amazon dynamodb  amazon rds/aurora  \
0  418ff92580b270ef4e7c14f0ddfc36b4                0                  0
1  8aec88cba08d53da65ab99cf20f6f9d9                0                  0
2  46baa1f69ac07779274bcd90b85d9a72                0                  0
3  3941b2f206ae0f900c4fba4ac0b18719                0                  0
4  45efa1f6bc65acc32bbbb953a1ed13b7                0                  0

   amazon redshift  apache hbase  apache hive  aster data  cassandra_x  \
0                0             0            0         0.0            0
1                0             0            0         0.0            0
2                0             0            0         0.0            0
3                0             0            0         0.0            0
4                0             0            0         0.0            0

   elasticsearch  firebird  ...  mac os  mainframe  predix  \
0              0       0.0  ...     0.0        0.0       0
1              0       0.0  ...     0.0        0.0       0
2              0       0.0  ...     0.0        0.0       0
3              0       0.0  ...     0.0        0.0       0
4              0       0.0  ...     0.0        0.0       0

   raspberry pi  salesforce  serverless  sharepoint  \
0           0.0         0.0           0         0.0
1           0.0         0.0           0         0.0
2           0.0         0.0           0         0.0
3           0.0         0.0           0         0.0
4           0.0         0.0           0         0.0

   windows desktop or server  windows phone  wordpress
0                          0            0.0        0.0
1                          0            0.0        0.0
2                          0            0.0        0.0
3                          0            0.0        0.0
4                          0            0.0        0.0

[5 rows x 300 columns]
```

```
   Back End  C-suite executive  Data Scientist  Data or business analyst  \
0       0.0                0.0             0.0                       0.0
1       0.0                0.0             0.0                       0.0
2       0.0                0.0             0.0                       0.0
3       0.0                0.0             0.0                       0.0
4       0.0                0.0             0.0                       1.0

   Database Administrator(DBA)  Designer  DevOps  \
0                          0.0       0.0     0.0
1                          1.0       0.0     1.0
2                          0.0       0.0     0.0
3                          0.0       0.0     0.0
4                          0.0       0.0     0.0

   Educator or academic researcher  \
0                              0.0
1                              0.0
2                              0.0
3                              0.0
4                              0.0

   Embedded applications or devices developer  Engineering manager  ...  \
0                                         0.0                  0.0  ...
1                                         0.0                  0.0  ...
2                                         0.0                  1.0  ...
3                                         0.0                  0.0  ...
4                                         0.0                  0.0  ...

   mac os  mainframe  predix  raspberry pi  salesforce  serverless  \
0     0.0        0.0     0.0           0.0         0.0         0.0
1     0.0        0.0     0.0           0.0         0.0         0.0
2     0.0        0.0     0.0           0.0         0.0         0.0
3     0.0        0.0     0.0           0.0         0.0         0.0
4     0.0        0.0     0.0           0.0         0.0         0.0

   sharepoint  windows desktop or server  windows phone  wordpress
0         0.0                        0.0            0.0        0.0
```

```
   amazon rds/aurora  amazon redshift  apache hbase  \
0                  0                0             0
1                  0                0             0
2                  0                0             0
3                  0                0             0
4                  0                0             0

   apache hive  aster data  cassandra_x  elasticsearch  firebird  \
0            0         0.0            0              0       0.0
1            0         0.0            0              0       0.0
2            0         0.0            0              0       0.0
3            0         0.0            0              0       0.0
4            0         0.0            0              0       0.0

   google bigquery  ...  mac os  mainframe  predix  raspberry pi  \
0                0  ...     0.0        0.0       0           0.0
1                0  ...     0.0        0.0       0           0.0
2                0  ...     0.0        0.0       0           0.0
3                0  ...     0.0        0.0       0           0.0
4                0  ...     0.0        0.0       0           0.0

   salesforce  serverless  sharepoint  windows desktop or server  \
0         0.0           0         0.0                          0
1         0.0           0         0.0                          0
2         0.0           0         0.0                          0
3         0.0           0         0.0                          0
4         0.0           0         0.0                          0

   windows phone  wordpress
0            0.0        0.0
1            0.0        0.0
2            0.0        0.0
3            0.0        0.0
4            0.0        0.0

[5 rows x 299 columns]
```

```
[0, array([ 0.00000000e+00,  0.00000000e+00,  0.00000000e+00,  0.00000000e+00,
        0.00000000e+00,  2.71050543e-20,  0.00000000e+00,  0.00000000e+00,
        5.42101086e-20,  0.00000000e+00,  0.00000000e+00,  1.62630326e-19,
        5.42101086e-20,  2.71050543e-20,  1.62630326e-19,  1.62630326e-19,
        0.00000000e+00,  5.42101086e-20,  8.19672131e-03,  0.00000000e+00,
       -6.93889390e-18,  0.00000000e+00,  5.42101086e-20,  4.09836066e-02,
        0.00000000e+00, -2.16840434e-18,  9.01639344e-02,  2.71050543e-20,
        1.63934426e-02,  0.00000000e+00, -2.16840434e-18, -6.93889390e-18,
       -3.90312782e-18, -1.56125113e-17,  2.71050543e-20,  2.71050543e-20,
        2.71050543e-20,  6.50521303e-19,  5.74374707e-02,  0.00000000e+00,
        3.52935110e-02,  6.32491681e-02,  1.77073116e-02,  0.00000000e+00,
       -1.73472348e-18,  0.00000000e+00,  0.00000000e+00,  0.00000000e+00,
        4.40421653e-02,  3.97217594e-02,  2.74489931e-02,  1.93859536e-02,
        7.09272081e-02,  2.60747162e-02,  5.16097226e-02,  7.63854935e-02,
        3.24712829e-02,  5.45044177e-02,  3.07272174e-01,  0.00000000e+00,
        4.02854804e-02,  3.61831723e-02,  0.00000000e+00,  6.55737705e-02,
        4.09836066e-02,  4.09836066e-02,  1.63934426e-02,  2.71050543e-20,
        2.71050543e-20,  5.42101086e-20,  8.19672131e-03,  3.25260652e-19,
       -6.50521303e-18,  1.62630326e-19, -1.08420217e-18,  5.42101086e-20,
        8.19672131e-03,  2.45901639e-02,  1.08420217e-19,  5.42101086e-20,
        8.13151629e-20,  6.50521303e-19, -2.16840434e-19, -3.25260652e-18,
        8.13151629e-20,  8.19672131e-03,  3.25260652e-19,  8.13151629e-20,
        8.19672131e-03,  2.45901639e-02,  5.42101086e-20,  0.00000000e+00,
        2.71050543e-20,  6.93889390e-18,  5.42101086e-20, -3.25260652e-18,
        2.71050543e-20,  8.19672131e-03,  5.42101086e-20,  2.71050543e-20,
        5.42101086e-20,  2.71050543e-20,  1.62630326e-19,  2.45901639e-02,
        1.64798730e-17, -3.03576608e-18,  8.19672131e-03,  5.42101086e-20,
        5.42101086e-20,  2.71050543e-20,  5.42101086e-20,  4.77048956e-18,
        5.42101086e-20,  5.42101086e-20,  1.62630326e-19,  2.71050543e-20,
        2.71050543e-20,  5.42101086e-20,  2.71050543e-20, -3.90312782e-18,
        1.63934426e-02, -3.03576608e-18,  8.19672131e-03, -5.42101086e-19,
        5.42101086e-20,  8.13151629e-20,  0.00000000e+00,  5.42101086e-20,
        5.42101086e-20,  2.71050543e-20,  1.08420217e-19,  5.42101086e-20,
        1.62630326e-19, -3.90312782e-18, -3.90312782e-18,  5.42101086e-20,
```

```
In [ ]: #Identifying
```

Part B: Profile Learner

```
In [ ]: #Explicit/Implicit ways of identifying user preferences
        #1) Use a like/dislike method to indicate user preference:
        #Optimal method: Model based user-preference method to infer a score fo
        worked for in the past
```

```
In [17]: #LinkedIn Collab
         files=glob.glob("../scraped profiles/*.json")
         #print(files)
         for file in files[:1]:
             f=open("../scraped profiles/"+file)
             data=json.load(f)
             print(data[0:2])
```

5. Technologies and Framework to be used.

• We uses Artificial Intelligence and Machine Learning.

  This dataset contains various IT job posting in the United States.\

• Stack Overflow dataset:

• US Technology Jobs dataset:

Platform :- W uses Kaggle, Notepad,

Dada Files:-

databases_profile_job.csv

databases_profile_user.csv

databases_job_profile.csv

domain_user_profile.csv

frameworks_profile_job.csv

languages_proifile_user.csv

platforms_profile_job.csv

platforms_job_profile.csv

preprocessed_df.csv

recommendations.csv

databases_profile_job.csv

databases_profile_user.csv

dice_com-job_us_sample.csv

domain_job_profile.csv

domain_user_profile.csv

frameworks_profile_job.csv

frameworks_proifile_user.csv

languages_profile_job.csv

languages_proifile_user.csv

platforms_profile_job.csv

platforms_proifile_user.csv

6. SWOT

This project is use to get a job according to the skills and qualification of a person by using is project the will find the opportunity to get the desired job.