Intelligence Business Model for Skill.jobs with Machine Learning Approaches

By

Zarrin Tasnim
(152-35-1234)

A thesis submitted in partial fulfillment of the requirement for the degree of
Bachelor of Science in Software Engineering

Department of Software Engineering
DAFFODIL INTERNATIONAL UNIVERSITY

Fall – 2019

Approval

This **Thesis** titled "**Intelligence Business Model for Skill.jobs with Machine Learning Approaches**", submitted by **Zarrin Tasnim**, **152-35-1234** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc in Software Engineering and approved as to its style and contents.

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan
Professor and Head **Chairman**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Dr. Md. Asraf Ali
Associate Professor **Internal Examiner 1**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Asif Khan Shakir
Lecturer **Internal Examiner 2**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Prof Dr. Mohammad Abul Kashem
Professor **External Examiner**
Department of Computer Science and Engineering
Faculty of Electrical and Electronic Engineering
Dhaka University of Engineering & Technology, Gazipur

Declaration

It hereby announces that, this **bachelor thesis** under the supervision of **Dr. Shaikh Muhammad Allayear, Associate Professor, Department of Software Engineering, Associate Professor and Head, Department of Multimedia & Creative Technology, Daffodil International University.** It is also declared that neither this thesis nor any part of the thesis has been submitted elsewhere from award of any degree.

Submitted by,

Zarrin Tasnim

Zarrin Tasnim
ID: 152-35-1234
Bachelor of Science in Software Engineering
Department of Software Engineering
Faculty of Science & Information Technology
Daffodil International University

Certified by,

Dr. Shaikh Muhammad Allayear
Associate Professor
Department of Software Engineering
Associate Professor and Head
Department of Multimedia & Creative Technology
Daffodil International University

Abstract

Business intelligence and analytics are data management solutions implemented in companies and enterprises to collect historical and present data, while using statistics and software to analyze raw information, and deliver insights for making better future decisions. In the circumstances of today's world, to survive and established own business need an analytical and find an easiest way or intelligence business model. This study is on "Intelligent business model for skill. Jobs with machine learning approach". The main objective is to examine the performance of various Machine Learning algorithms in order to perform with the system of skill.jobs. This proposed module integrated with three phase such as, the Clusters similar kind of job search phase (CSK) is a way of knowing the demand is to create a visual graph showing clusters of similar kinds of job searched by the job seekers in the website of skill. jobs, the email notifications send phase (ENS) is responsible to send email notifications to the job seekers when a job circular is posted in the website of skill.jobs, extract the job circular phase (EJC) is the way to extract the job circular post from the career section of each of the company's website. The result shows the successful clustering of similar job search, email notification send to specific people and extracts the information from the web.

Acknowledgments

Firstly, I might want to thank my supervisor, Dr. Shaikh Muhammad Allayear, Associate Professor, I owe such a great amount to his motivating direction over the span of this venture, for his recommendations on papers to peruse, and for his endless hours of accommodating exchanges and assessment. He gives me an opportunity to work in Smart Data Science Center (SDSC) for complete my research. SDSC is a computer research laboratory of Daffodil International University. I might likewise want to demonstrate appreciation to my committee, including Department of Software Engineering and my noteworthy our Department Head Professor Dr.Touhid Bhuiyan for their profitable instructions.

All the more by and large, I can't exaggerate the amount Daffodil International University's software engineering offices have helped me develop as an understudy. Uncommon much gratitude goes F.M. Javed Mehedi Shamrat, Research Associate (RA) for putting me on the way to seeking after hypothetical software engineering research and for being a uniquely rousing coach and to lecturer Md. Mushfiq for filling in as my scholarly consultant. I might likewise want to thank lecturer Md Fahad Bin Zamal for teaching one incredible course that truly got me amped up for a few complex factors. I would also like to thank lecturer Ms. Tapushe Rabaya Toma for her incredibly extensive and important input on the evidence of my principle result. I'm particularly appreciative for Research Associate F.M. Javed Mehedi Shamrat for being regular teammates on issue sets. I'm particularly thankful for my paper observers Senior Lecturer Nazia Nishat and lecturer Ms. Lamisha Rawshan for being continuous collaborators on issue sets and their suggestions on research report to peruse or read.

Lastly, I might want to thank my parents for bringing me into this world and making everything conceivable. They were the reason I initially began to look all starry eyed at learning, and I am appreciative consistently for what they have done to raise me up to be simply the best form.

Table of Contents

Table of figures:

Chapter 1: Introduction

While companies are still too busy to become fully digitization in terms of the processes they use, the products and services they offer, and the customer experiences they create, they are already faced with the next wave of disruptive change: the intelligent enterprise. The intelligent enterprise utilizes connectivity (between things, people, and enterprises), data, cloud applications, algorithms, and advanced analytics (instead of hard-coded rules and rigorous procedures). This enables them to come to the right decisions with minimal human involvement even in turbulent, fast-changing environments. To free up knowledge workers from repetitive tasks that a machine can do as well or even better, you need a combination of technologies to collect and process the data (e.g., IoT and cloud applications), artificial intelligence and Machine Learning. We are working on topics that actually a business intellectual model that will optimize the process of job seeking, skill development for job seekers and for human resource department (HRD) to hire candidates. Within three phase we are accomplished the whole process.

1.1 Background

Too often, innovative technological capabilities are used only to improve existing processes consisting of many repetitive, manual tasks and replace them with more automated and adaptive processes. This obviously drives down costs and leads to temporary competitive advantages; however, it neither changes the rules of the game nor brings about innovation. The winners in the era of the intelligent enterprise will be those companies that utilize intelligent technologies to come up with new, intelligent business models.

In order to build a business intellectual model, that can automatically learn and take decision what should do. I would first take a look of the existing system and plan a module of include three phases that are web-crawling, emailing a group of person based of classifying the user's and put aside the number of most searched keywords.

This model will be very useful or vast findings for the skill.jobs company. The business intellectual model can be reduce the effort of the skill.jobs employee's and generate proper graphitic result for the specific category. So, I decided to collect the sample dataset of the skill.jobs and work on it by the intelligent business model that I proposed. If the model successfully implemented, it can be turn over for skill.jobs and most importantly it can reduce the human effort and increase productivity.

1.2 Motivation of the Research

Business Intelligence, BI is a concept that usually involves the delivery and integration of relevant and useful business information in an organization. Companies use BI to detect significant events and identify/monitor business trends in order to adapt quickly to their changing environment and a scenario. If you use effective business intelligence training in your organization, you can improve the decision making processes at all levels of management and improve your tactical strategic management processes.

We developed business intellectual model that will optimize the process of job seeking, skill development for job seekers and for human resource department (HRD) to hire candidates. Sometimes it’s hard to find out the similar job with expertise, and sometimes it’s not feasible for skill.jobs to search job from different web, so there’s need a system model that can automatically crawling the job information from the web and notify the similar jobseeker expertise.

1.3 Problem Statement

Skill.jobs is one of the biggest job searching portal in the country. Job seekers use the website to search for jobs all the time. At the same time, HRD of different company post the job circular in the website to recruit candidates. But, the system is not automated or much efficient, as the job circular data has to be manually provided to the system by the HRD. That the same time, when a job circular is posted, the notification of the job is provided to every person registered in the system even though the job is totally irreverent to them. As a result, frequent irrelevant notifications become a bother for the register users of the system. At the same time, the website does nothing to give the job seekers any idea of what kind of job has more competition in the market and who needs to improve their skills and qualifications in order to get a better job.

1.4 Research Questions

The thesis with titles showing that it’s a business intellectual model that will optimize the process of job seeking, skill development for job seekers and for human resource department (HRD) to hire candidates. There have exactly some research questions and this will enable to understand some features of this thesis.

- Why need proposed module?
- What is the advantage of intelligent business model for skill.jobs with machine learning approach?
- How the module works?
- Is it a new or modify approach?
- How much data can handle by the intelligent business model?
- Is it possible to build the proposed model of an intelligent business model with machine learning approach?
- Is it possible to clustering the similar types of job search?

1.5 Research Objectives

The main objective of this thesis is build machine learning and business intellectual model that can perform on a system to process information, graph generate, data visualization and hold search information.

This research report exhibits about a model. The purpose of model of proposed algorithm is improvement of the scenario of the existing system of the specific company. It can be time consuming by using machine learning approach.

1.6 Research Scope

From the research, an efficient system is expected to be derived. The system will contain phases that will automate the website in gathering data without depending to manual input of data into the database. The system will also have to analyze the data using appropriate machine learning algorithm in order to make decision. The system must also show a visual representation of the data from the database into order to help the users analyze the information themselves.

1.7 Thesis Organization

In Chapter 1, we presented and overview the system and talked about the research questions, objectives and scope. At the same time, the keywords of the thesis are briefly described in this chapter.

In Chapter 2, the literature review is presented that talks about the past research done on the proposed system. It describes the motivation behind the thesis and talks about the algorithms to be implemented for the intelligent business module.

In Chapter 3, we introduced the three phases of the proposed system. Here the entire system in described with a flowchart diagram. How to the implement the three phases of the system with the help of flowchart diagrams are shown as well.

In Chapter 4, we have shown that after implementation of the phases, the result that is received is as expected for the proposed system.

In Chapter 5, we summarized the entire system and talked about all the phases and the purpose each phase serves in the system. It also talks about the future scopes of the system and how it can be improved in the future.

In Reference, we tried to show all the relevant and proper references that we studied to complete the research.

In Appendix, we included the datasets that were used for the research with all the attributes of the dataset.

1.8 Definitions

Intelligence Business Model: According to Forrester Research, business intelligence is "a set of methodologies, processes, architectures, and technologies that transform raw data into meaningful and useful information used to enable more effective strategic, tactical, and operational insights and decision-making.

Machine learning: *Machine learning* (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial *intelligence*. Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

Web Crawling: A web crawler (also known as a web spider or web robot) is a program or automated script which browses the World Wide Web in a methodical, automated manner. This process is called Web crawling or spidering. Many legitimate sites, in particular search engines, use spidering as a means of providing up-to-date data.

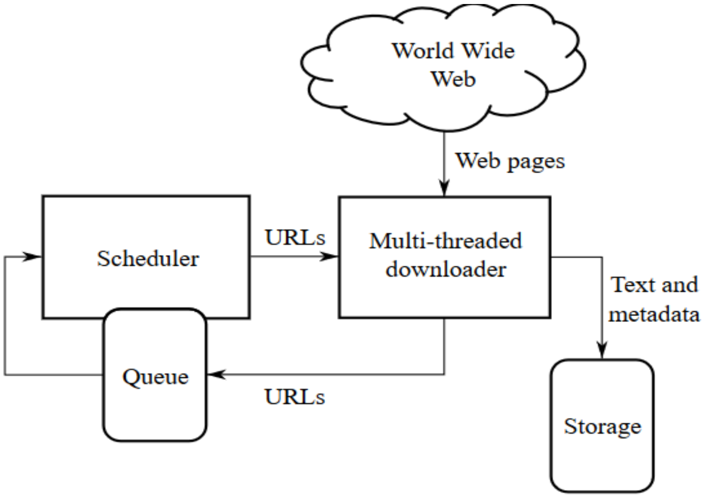


Figure 1.1: Architecture of Web Crawling.

Clustering: Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters).

Artificial Intelligence: Artificial intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for using the information), reasoning (using rules to reach approximate or definite conclusions) and self-correction.

Chapter 2: Literature Review

In this research, a system is proposed to make an intelligent business model that will automate the skill.jobs website and make it more efficient. The system gathers information automatically and gives decision assessing the data from the database to make the system more efficient. Moreover the system gives a visual presentation of the data in the database more better and easier analysis. Supervised and unsupervised machine learning algorithms are proposed in the system to make the data analysis more effective. Amongst supervised learning algorithms, decision tree algorithm is proposed. Likewise, amongst unsupervised learning algorithms, kmeans clustering algorithm is proposed. The decision tree makes the decision of suitable candidates from the database compared to the job circular. The clustering algorithm gives a visual presentation of data in the database. Implementing the proposed algorithms, the skill.jobs website will be more effective in finding jobs for the jobseekers and the job seekers will get more enriched information from the website.

2.1. Article Searching Procedure

I used a systematic searching procedure to identify all of the available articles that discuss the business intelligence model, specially related with web crawling, using machine learning techniques. In my systematic procedure, I search three keywords from Science Direct databases in order to access the article. I used the keywords, “Intelligence business model with data science”, “web crawling to grab the web data” and “intelligence business model using machine learning” to find journal articles published in English Language between years 2012 to 2019.

2.2. Previous works

In recent years some research paper has been published, where researchers have shown how to collect data using web crawler technology, classify group of data using unsupervised algorithm and visualize data with proper manner.

2.2.1. Web crawler

A mobile web crawler is an automated computer program, which transfers itself to web servers in an attempt to download information and contents. Dynamically changing nature of web requires that mobile web crawlers must be able to intelligently decide about new pages and the changes in already crawled pages. This ability of mobile web crawler allows minimizing the consumption of resources [1]. This paper aims to provide an efficient crawling mechanism for implementing a mobile web crawler, which intelligently decides about page changes to reduce overall load on web resources and helps search engines to increase web crawling speed and expand their reach in indexing.

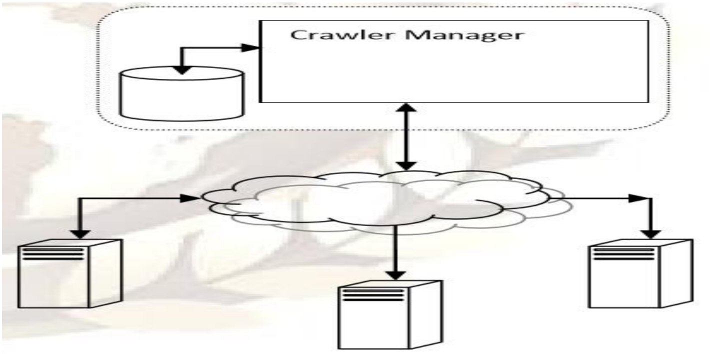


Figure 2.1: Mobile Web Crawler Mechanism.

In this paper, we will discuss some recent techniques for crawling web pages belonging to specific topics. We discuss the following classes of techniques: (1) Intelligent Crawling Methods. (2) Collaborative Crawling Methods. We will also discusses some creative ways of combining different kinds of linkage- and user-centered methods in order to improve the effectiveness of the crawl [2].

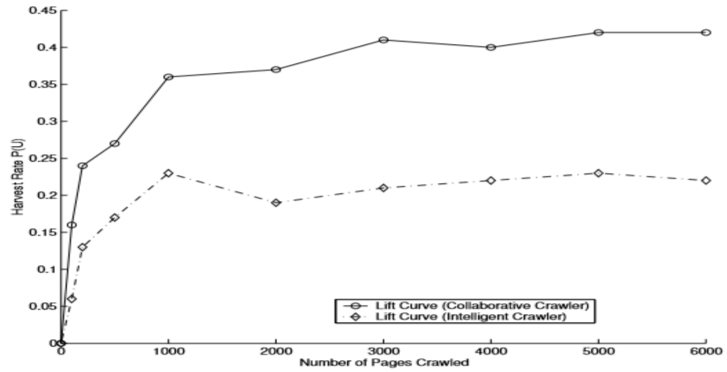


Figure 2.2: Performance of Collaborative and Intelligent Crawler (Predicate is category "SPORTS")

It is an essential method for collecting data on, and keeping in touch with the rapidly increasing Internet. This Paper briefly reviews the concepts of web crawler, its architecture and its various types [3].

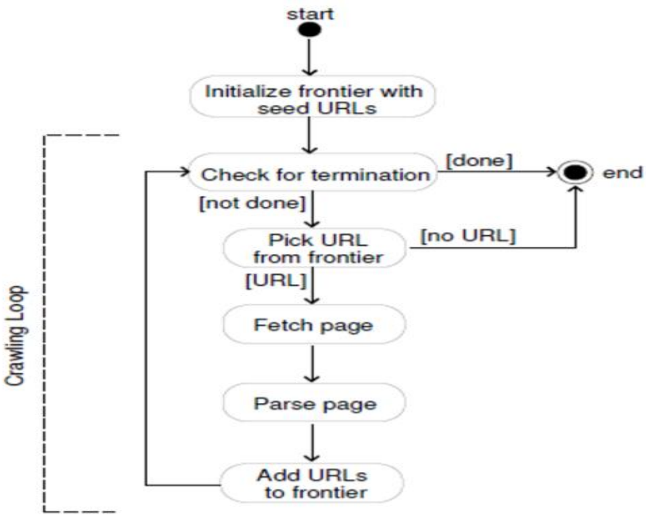


Figure 2.3: Flow of a basic crawler.

2.2.2. Machine Learning Algorithms

In this paper, the concept of data mining was summarized and its significance towards its methodologies was illustrated. This paper also conducts a formal review of the area of rule extraction from ANN and GA [4].

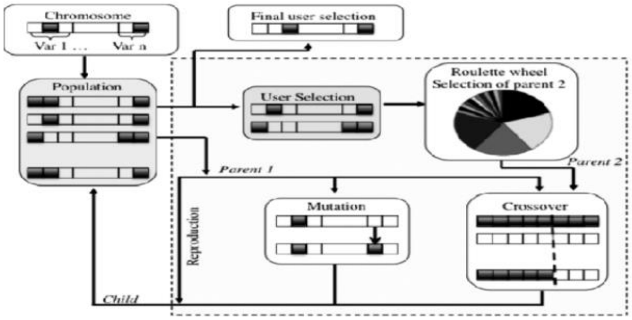


Figure 2.4: Structural view of Genetic Algorithm.

This paper presents the top 10 data mining algorithms identified by the IEEE International Conference on Data Mining (ICDM) in December 2006: C4.5, k-Means, SVM, Apriori, EM, PageRank, AdaBoost, KNN, Naive Bayes, and CART. With each algorithm, we provide a description of the algorithm, discuss the impact of the algorithm, and review current and further research on the algorithm. These 10 algorithms cover classification, clustering, statistical learning, association analysis, and link mining, which are all among the most important topics in data mining research and development [5].

This paper summarizes an approach to synthesizing decision trees that has been used in a variety of systems, and it describes one such system, ID3, in detail. Results from recent studies show ways in which the methodology can be modified to deal with information that is noisy and/or incomplete [6]. A reported shortcoming of the basic algorithm is discussed and two means of overcoming it are compared.

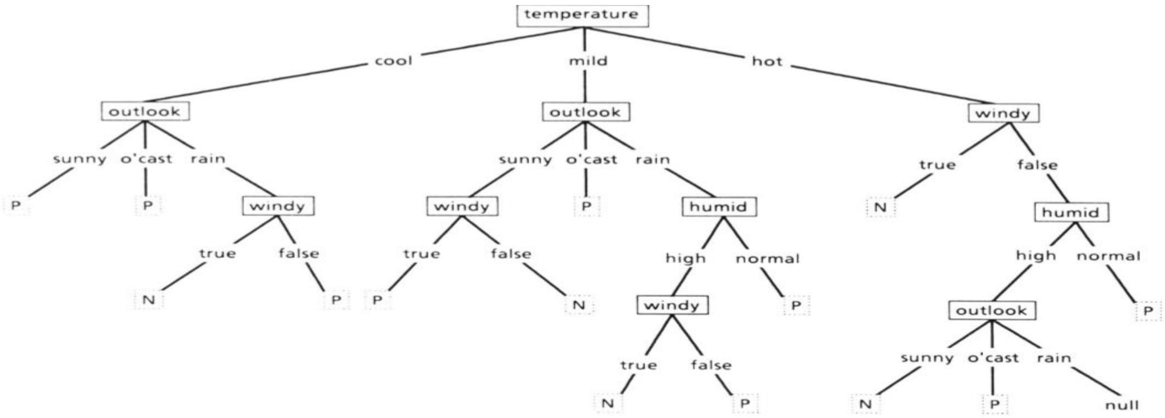


Figure 2.5: A complex decision tree.

Big data refers to datasets that are not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques. This paper aims to analyze some of the different analytics methods and tools which can be applied to big data, as well as the opportunities provided by the application of big data analytics in various decision domains [7].

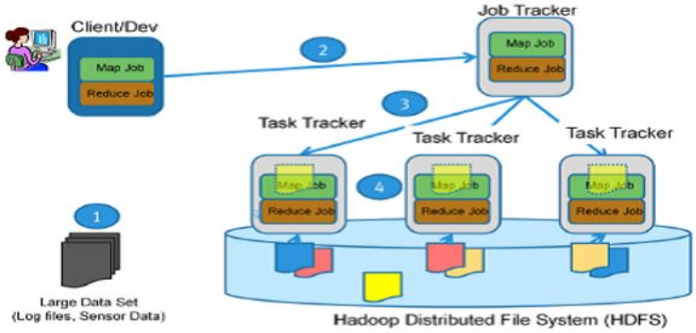


Figure 2.6: MapReduce and HDFS.

Decision tree is a classification technique in which a model is created that anticipates the value of target variable depends on input values. ID3 and C4.5 are commonly used decision tree algorithms. These algorithms are based on Hunt’s algorithm. Goal of this study is to provide review of these decision tree algorithms. At first we present concept of Data Mining, Classification and Decision Tree. Then we present ID3 and C4.5 algorithms and we will make comparison of these two algorithms [8].

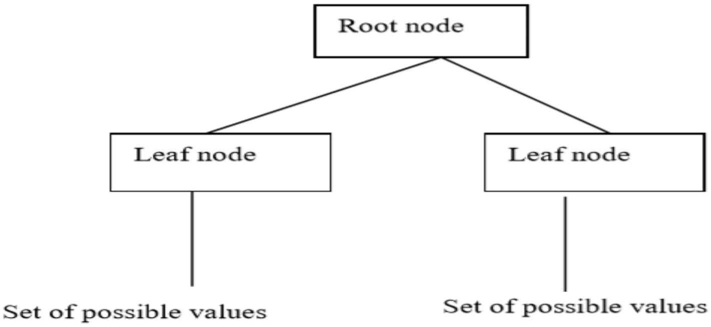


Figure 2.7: Decision tree induction.

In this paper, we propose a novel kNN type method for classification that is aimed at overcoming these shortcomings. Our method constructs a kNN model for the data, which replaces the data to serve as the basis of classification. The value of k is automatically determined, is varied for different data, and is optimal in terms of classification accuracy. The construction of the model reduces the dependency on k and makes classification faster [9].

In this algorithm the neighbor samples are automatically determined using clustering techniques. After partitioning the train set, the labels of cluster centers are determined. For specifying the class label of a new test sample, the class label of the nearest cluster prototype is used. Computationally, the NC method is faster than KNN, K2 times. Also the clustering techniques lead to find the best number of neighbors based on the nature of feature space [10].

This paper presents a KNN text categorization method based on shared nearest neighbor, effectively combining the BM25 similarity calculation method and the Neighborhood Information of samples.

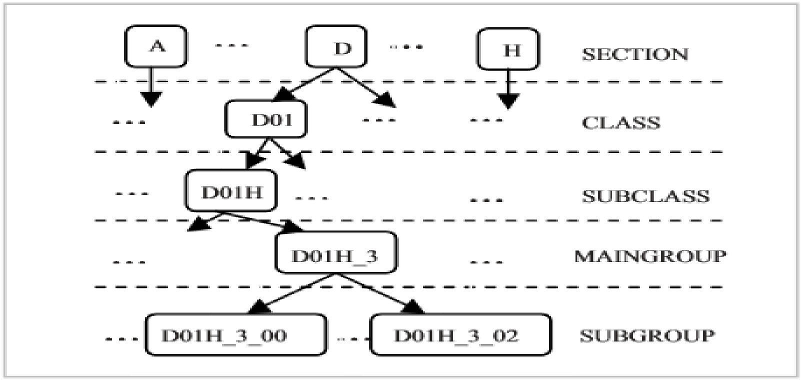


Figure 2.8: Patented multi-level tree.

2.3. Research Gap from Previous Works

According to use research knowledge, an intelligence business model using machine learning approaches is proposed and implemented. For developing our proposed system, we have used python programming with its default library, develop algorithm, and web crawling technology, machine leaning algorithms. We are developed business intellectual model that will optimize the process of job seeking, skill development for job seekers and for human resource department (HRD) to hire candidates.

2.4. Summary

Studying the works done previously, it is possible to decide which algorithm and technology is best suited to implement the proposed system. Besides, the most effective and efficient system can be implemented from the sea of technology available. It is also seen that such a system has not been yet implemented.

Chapter 3: Research Methodology

Skill.jobs is a job portal that contains the data of thousands of job seekers and job circulars from different companies. The system gathers the job seekers' data from the profiles created by the job seekers. At the same time, the job circulars are gathered after the companies provide the job circular information to skill.jobs themselves. However if a company fails to provide their latest job circular information to skill.jobs, the web portal will not get the information in any other way. Furthermore, when a job is posted, every person registered in the system gets a notification which can be a bother. Notifications should be sent only to relevant users. At the same time, as users are looking for jobs, it is also important to arrange training session based on the types of jobs most in demand in market. Therefore, a system is proposed that will make the portal intelligent and more efficient.

3.1. Proposed System

The primary target is to look at the exhibition of different Machine Learning calculations so as to perform with the arrangement of skill.jobs. This proposed module incorporated with three-stage, for example, the Clusters' comparative sort of quest for new employment stage (CSK) is a method for realizing the interest is to make a visual chart demonstrating groups of comparative sorts of occupation looked by the activity searchers in the site of aptitude. employments, the email notices send stage (ENS) is dependable to send email notices to the activity searchers when an occupation roundabout is posted in the site of skill. Jobs, separate the activity roundabout stage (EJC) is the best approach to remove the activity round post from the vocation segment of every one of the organization's site. The outcome demonstrates the fruitful bunching of comparative pursuit of employment, email notice sent to explicit individuals and concentrates the data from the web. The business intelligence model can diminish the exertion of the skill. Jobs worker's and create appropriate graphitic results for the particular class. In this way, I chose to gather the example dataset of the skill.jobs and work on it by the canny plan of action that I proposed. On the off chance that the model effectively executed, it very well may be turned over for skill.jobs and in particular, it can diminish the human exertion and increment efficiency. To free up knowledge workers from repetitive tasks that a machine can do as well or even better, you need a combination of technologies to collect and process the data (e.g., IoT and cloud applications), artificial intelligence and Machine Learning. We are working on topics that actually a business intellectual model that will optimize the process of job seeking, skill development for job seekers and for human resource department (HRD) to hire candidates. Within three phase we are accomplished the whole process.

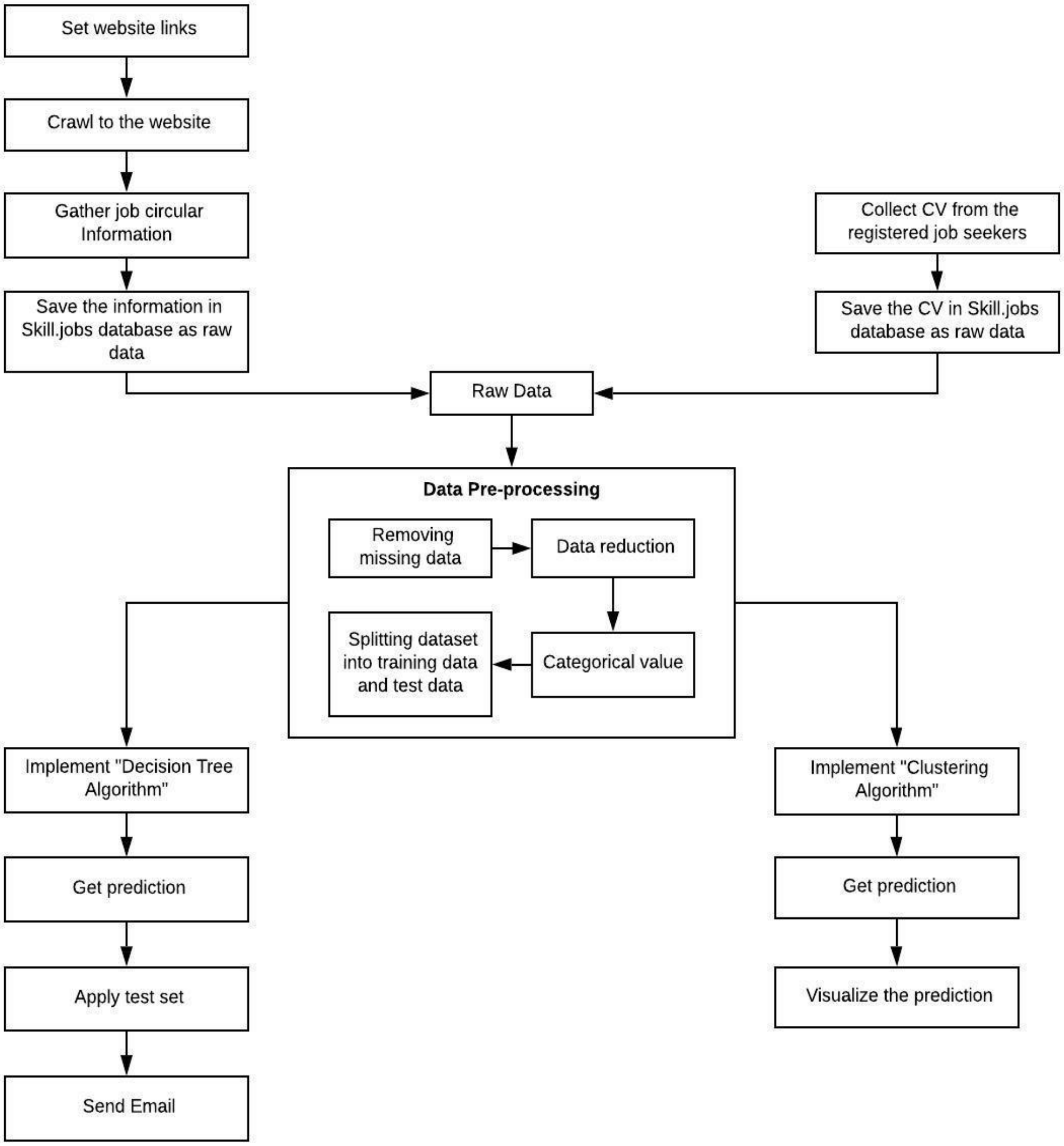

Set website links
Crawl to the website
Gather job circular Information
Save the information in Skill.jobs database as raw data
Collect CV from the registered job seekers
Save the CV in Skill.jobs database as raw data
Raw Data
Data Pre-processing
Removing missing data
Data reduction
Categorical value
Splitting dataset into training data and test data
Implement "Decision Tree Algorithm"
Get prediction
Apply test set
Send Email
Implement "Clustering Algorithm"
Get prediction
Visualize the prediction

Figure 3.1: System Architecture.

3.2. System Overview

Our proposed system contains three (3) phases. Those are given below:

1. Extract the job circular phase (EJC)
2. Clusters similar kind of job search phase (CSK)
3. Email notifications send phase (ENS)

3.3 Implementation

The first phase is to extract the job circular phase (EJC). In this phase, the job circular will be automatically gathered and stored into the database of the system using a web crawler. The career link of different websites will be set in the web crawler. The web crawler will visit the links provided and read the latest job circular from the websites and write it into the database of the system so that even without companies proving the job circular themselves, skill.jobs will always stay updated. Secondly the cluster similar kind of job search phase (CSK) uses the unsupervised K-means algorithm. For this phase first the job title of the job seekers are retrieved. At the same time, from the web site, when any user searches for a job in the search bar, the search history is recorded in the database. This record is retrieved as well. Merging the two dataset, we get a total data of the jobs that are currently in demand by the job seekers. This jobs are sorted in different categories based on their kinds and plotted in a scatter graph. Using the K-means algorithm, clusters are created in the graph. The clusters are analyzed to identify which kind of jobs are most in demand and skill enhancement programs can be arranged to help the maximum number of people to find themselves better jobs. That last phase of the system is the email notifications send phase (ENS). The database of skill.jobs contains both the detailed information of job seekers and job circulars of different companies. Based on the data, the system suggests which job seekers will be suitable for a certain job. This decision is made with the help of a decision tree algorithm. The decision tree uses three attributes from the job seekers' data and three attributes from the job circular data. These attributes are job title, years of experience and expected salary of job seekers and job position, required years of experience and offered salary of job circulars. The algorithm uses job title as the root node and sets conditions to make a decision. If all the conditions of job circular matches any job seeker for the database, those job seekers receive an email notification of the job from skill.jobs. It is so that everyone receives notifications that are relevant making the system more efficient.

3.3.1 Extract the job circular phase (EJC)

In this phase, a web crawler was implemented. A web crawler that is also called a web spider is a program that browses the web in a methodical manner to gather information. Web crawlers are used to gather data or copy the pages of any website it visits. But most importantly a web crawler is used to gather some specific data from a web site.

The skill.jobs website post job circular for companies that are looking for recruitments. In order to post the circular in the website, the system has to extract the job circular post from the career section of each of the company's website.

Using web crawling, the job circular information will be extracted from the websites which will be saved as parameters in the database of skill.jobs. In the web crawler that is implemented, a number of URLs of career section of the companies' website were set in a queue. The crawler gets a URLs from the queue and visits the webpages. From the web pages it extract the posts of the job circulars published by the companies, than copy and save the information in the database of skill.jobs.

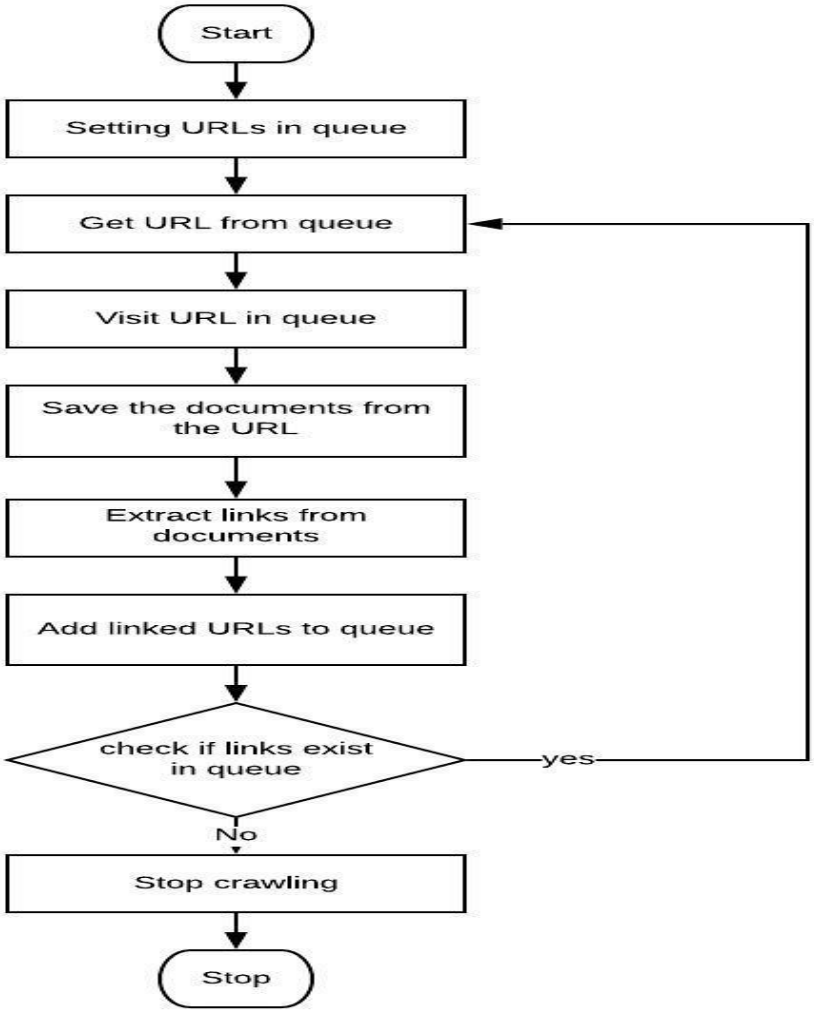


Figure 3.2: Web Crawling Action Architecture.

3.3.2 Clusters similar kind of job search phase (CSK)

In this phase, the K-means algorithm is implemented. K-means algorithm is an unsupervised algorithm that takes a number of data points and group them into a k number of clusters. Here k denote the number of clusters, i.e. if k = 3, the number of clusters received in the end will be 3 clusters. In k-means algorithm, data points a plotted across a scatter graph and k number of clusters are set. k number of centroid will be formed in the graph. A computation will be done to the number of iteration set to find the data points nearest to the centroids based on Euclidean distance. After the maximum iteration the clusters of data points around the centroids will be the final clusters.

In the database of skill.jobs, the list of job titles are available. Is the list of jobs, candidate are doing or are interested in doing. At the same time, candidates search of the type of job they are willing to do in the search bar. As a result the search record makes another list of jobs that are in demand among the job seekers. Merging the two lists, a dataset can be obtained that contains the data of jobs, job seeker are interested in. Using this data, skill.jobs can identify the field of jobs people are most interested in and arrange for training sessions so that job seekers can become more professional and skilled in those particular fields and get better jobs.

K-means algorithm makes a cluster of the similar types of jobs from the dataset. In order to do that first the data of jobs positions is retrieved from the database. At the same time, the search record from the search box of the website is stored into the database and merged with the data from dataset. These data are categorized into different types and is vectorized as x-vector and y-vector. These values are used as x-axis and y-axis for the scatter graph of k-means. The number of clusters, k is set as the number of categories, and the number of iteration is set as required. After plotting the data point in the scatter graph, the distance between the points and centroids is calculated and reassigned up to the maximum number of iteration. Finally a scatter graph with the required number of clusters are shown. From the graph, the cluster that are highly dense are the type of jobs, job seekers are most interested in and workshops can be arrange to enhance their skills as those jobs are most demanded.

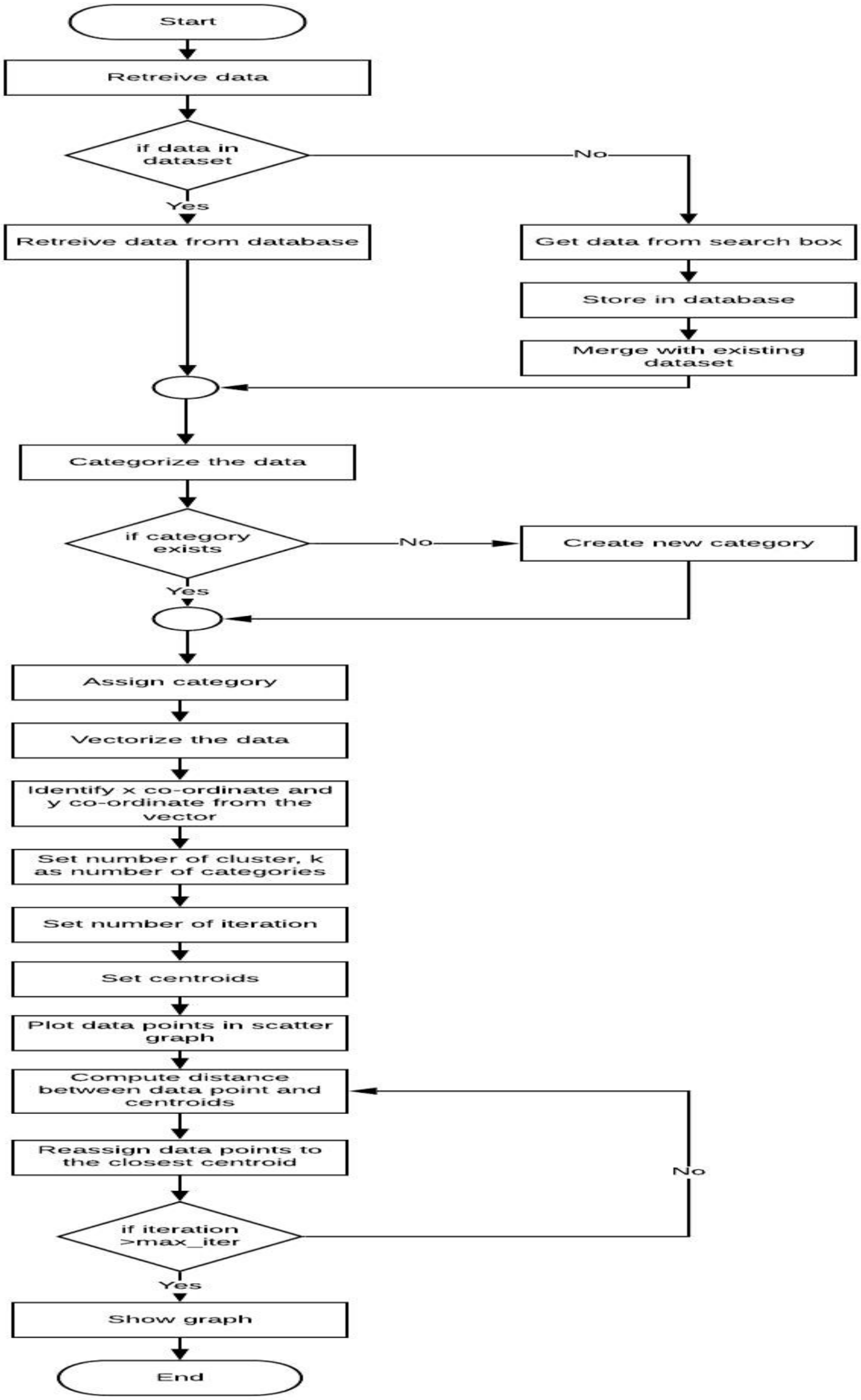

Start
Retreive data
if data in dataset
No
Yes
Retreive data from database
Get data from search box
Store in database
Merge with existing dataset
Categorize the data
if category exists
No
Create new category
Yes
Assign category
Vectorize the data
Identify x co-ordinate and y co-ordinate from the vector
Set number of cluster, k as number of categories
Set number of iteration
Set centroids
Plot data points in scatter graph
Compute distance between data point and centroids
Reassign data points to the closest centroid
No
if iteration >max_iter
Yes
Show graph
End

Figure 3.3: Data Clustering With K-means Algorithm Architecture.

3.3.3. Email Notifications Send Phase (ENS)

In this part, a decision tree algorithm is implemented. A decision tree is a supervised learning algorithm that can be used for classification. A decision tree is commonly used to solve a problem or to make a decision using a tree like structure and hence the name 'decision tree'. Each node represents an attribute and the best attribute is placed in the root node. Branch nodes help make decisions and the final decision is made in the leaf node. The feature values of the attributes are categorical values. A decision tree checks a certain condition or value of an attribute and decides which direction of the branch the next move should be.

In the database of skill.jobs, there is a list of job circulars present. At the same time, there is a list of job seekers with their detailed information present. Using the two sets of data, with the help of a decision tree, it is possible to match a number of job seekers who will be fit for a certain job circular. Those job seekers can be notified about the job circular via email so that they do not miss the opportunity to apply for the job. Besides if the job circular is advertised to every job seeker, irrespective of their field of interest, the advertisement will lose its importance. Therefore email notifications must only be send to the job seekers for whom the job circular is meaningful.

The decision tree algorithm gives a list of candidates who are suitable for a certain job. This decision is made based on three attributes. These are, job position, experience and expected salary. For that, first the list of the job seekers with their detailed information is to be retrieved. At the same time, the job circular for which candidates are needed to be retrieved. All the data must be preprocessed before implementing the decision tree such as, string values must be converted to categorical values for which the job position in both job seekers' dataset and job circular dataset are categorized and vectorized. At the same time all null values are removed. After preprocessing and cleaning the datasets, the decision tree is implemented. In the decision tree, the job position is considered as the root node. Since, if the field of interest doesn't match, it will not matter if the person has years of experience in another job field and has suitable expectation of salary. For all the candidates, first the decision tree will check if their job position matches. If it does, it will check if the years of experience of the candidate is greater or equal to the years of experience required for the job. If the condition matches, it will check if the expected salary of the candidate is less than or equal to the offered salary. When all the conditions will match, the decision tree will give an output of a list of email address of the candidate who are suitable for the job so that they can be notified via email. Using the email address provided by the job seekers, skill.jobs will send an email notification to the job seeker that a job that matches their profile has been circulated and they can apply for that job.

This process will reduce unnecessary email notification for the job seekers and everyone will get relevant emails only.

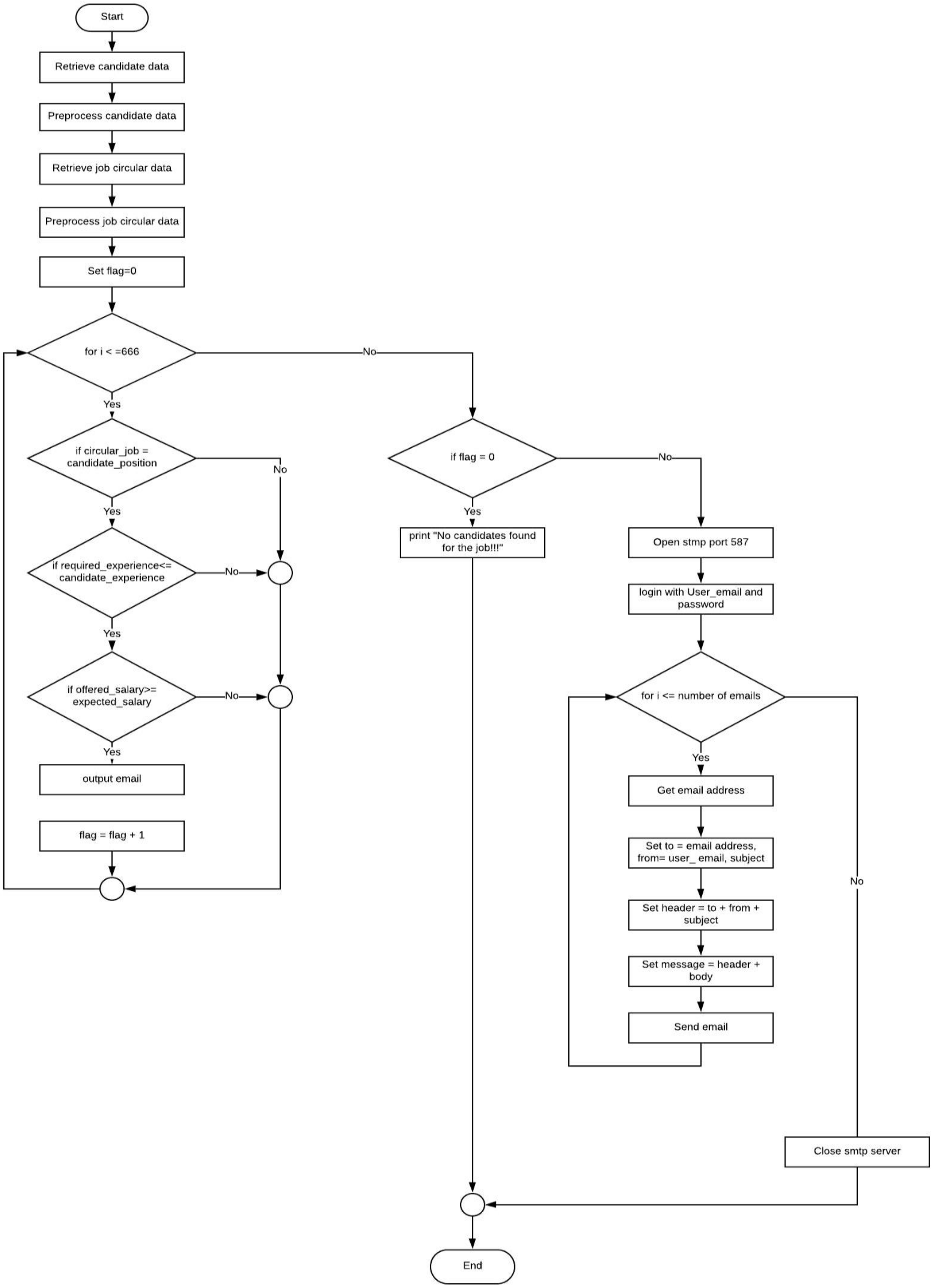

Start
Retrieve candidate data
Preprocess candidate data
Retrieve job circular data
Preprocess job circular data
Set flag=0
for i < =666
No
Yes
if circular_job = candidate_position
No
Yes
if required_experience<= candidate_experience
No
Yes
if offered_salary>= expected_salary
No
Yes
output email
flag = flag + 1
if flag = 0
No
Yes
print "No candidates found for the job!!!"
Open stmp port 587
login with User_email and password
for i <= number of emails
Yes
Get email address
Set to = email address, from= user_ email, subject
Set header = to + from + subject
Set message = header + body
Send email
No
Close smtp server
End

Figure 3.4: Email Sending Using Decision Tree Algorithm Architecture

3.4 Summary

The implementation of the three phases gives a total system that is intelligent enough to gather data on its own from the web. Using the data, the system can decide the candidates who are suitable for a job based on their qualifications. Furthermore, the system can also tell the user which kind of jobs are more needed and requires improvement in job seekers qualifications.

Chapter 4: Results and Discussion

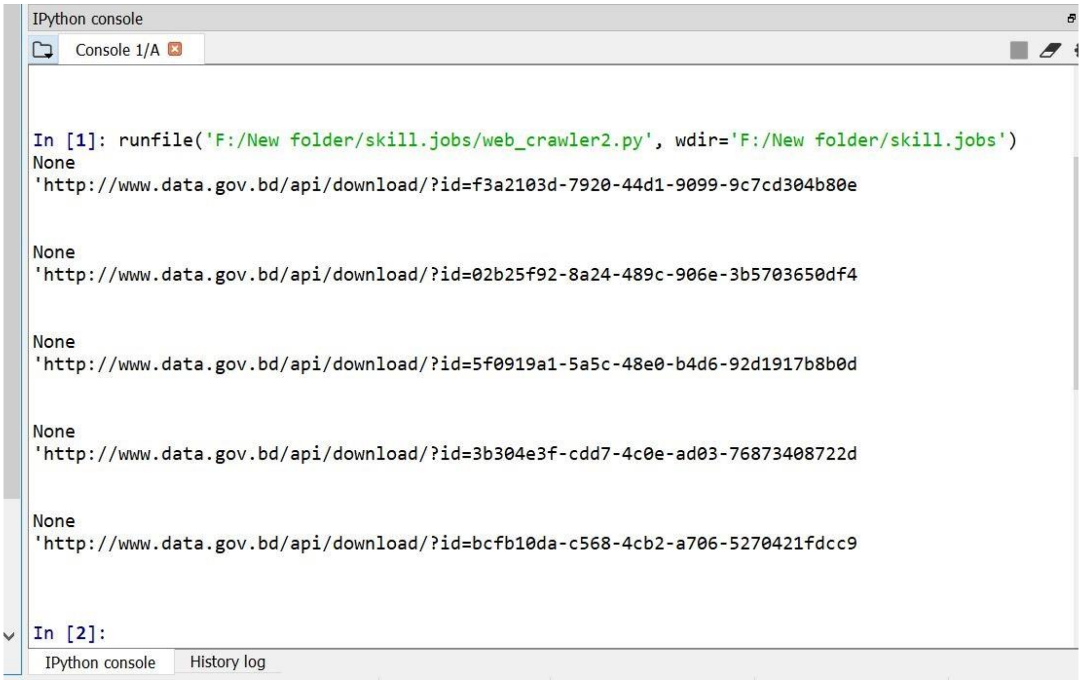
This experiment depended on a greatly rich thought. By setting up an algorithm to keep running out of sight of the relative direction learning purpose, we could accomplish uncertain measures of watching time. This enabled us to endeavor a more profound hunt than would have been conceivable in a period apportioned circumstance. With an exceptionally restricted spending we assembled and introduced a beneficiary, spectrometer and control programming, all of which have performed honorably.

4.1. Extracted data using web crawler

A HTTP link of a website is set in the web crawler using which the web crawler goes in the website and extracts the required data from the website's source, copy it and stores it for further use. Such data is extracted using the implemented web crawler in order to gather data for the system.

4.1.1. Extracted HTTP links

In figure 4.1 we can see, in the HTML file, the link did not contain any text so there is no title and the variable contained nothing and 'None' is printed out. In the next line we see a HTTP link that the web crawler extracted from the webpage. All the links in the page are extracted using the developed web crawler.

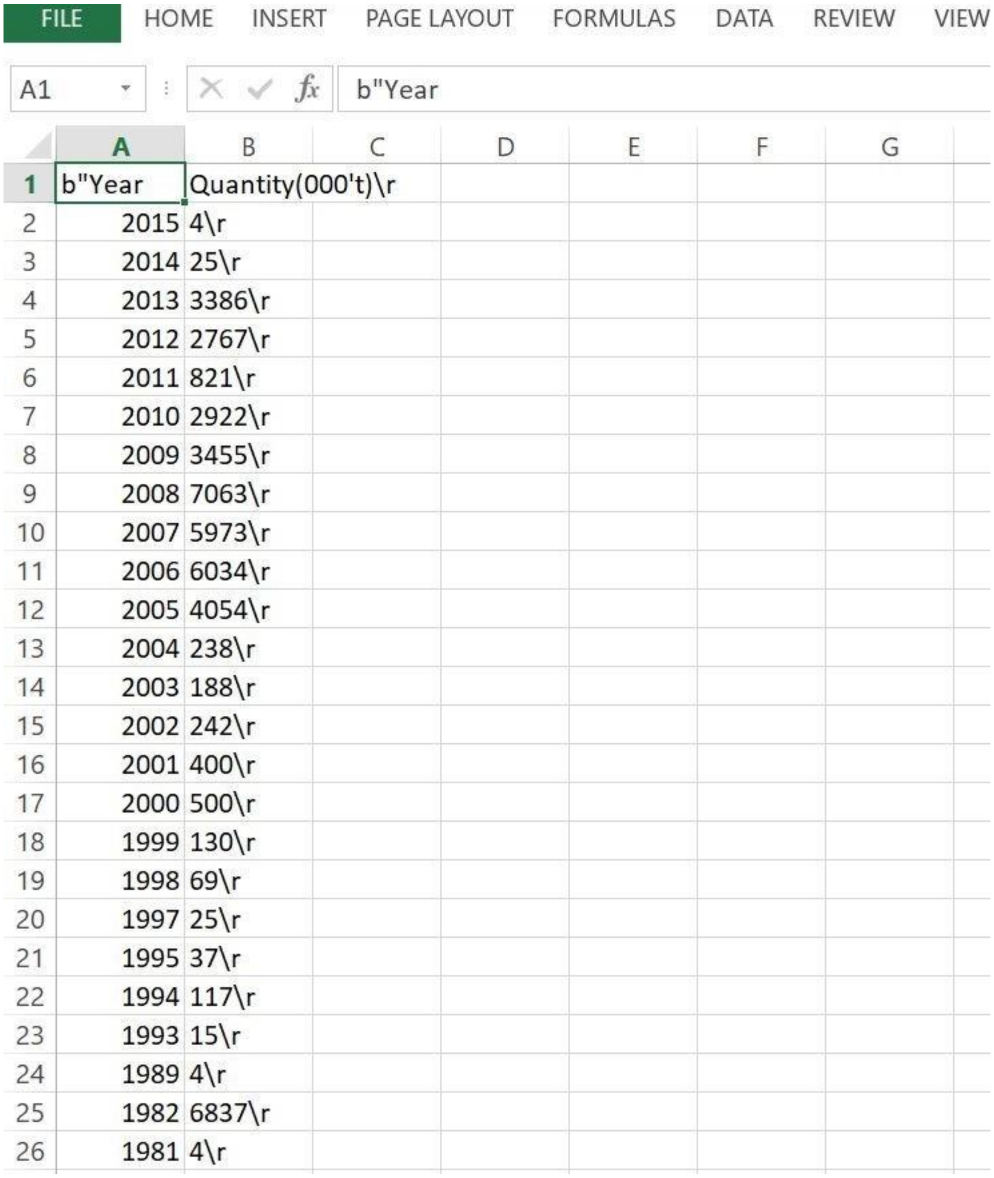


```
In [1]: runfile('F:/New folder/skill.jobs/web_crawler2.py', wdir='F:/New folder/skill.jobs')
None
'http://www.data.gov.bd/api/download/?id=f3a2103d-7920-44d1-9099-9c7cd304b80e


None
'http://www.data.gov.bd/api/download/?id=02b25f92-8a24-489c-906e-3b5703650df4


None
'http://www.data.gov.bd/api/download/?id=5f0919a1-5a5c-48e0-b4d6-92d1917b8b0d


None
'http://www.data.gov.bd/api/download/?id=3b304e3f-cdd7-4c0e-ad03-76873408722d


None
'http://www.data.gov.bd/api/download/?id=bcfb10da-c568-4cb2-a706-5270421fdcc9



In [2]:
```

Figure 4.1: Hyperlinks of Files as Output of the Web Crawler

4.1.2. Extracted data using web crawler

In figure 4.2, the dataset that was present in the webpage whose link was extracted using the web crawler is being copied and stored. Using this stored information, further processing in the other phases will be done.



	A	B
1	b"Year	Quantity(000't)\r
2	2015	4\r
3	2014	25\r
4	2013	3386\r
5	2012	2767\r
6	2011	821\r
7	2010	2922\r
8	2009	3455\r
9	2008	7063\r
10	2007	5973\r
11	2006	6034\r
12	2005	4054\r
13	2004	238\r
14	2003	188\r
15	2002	242\r
16	2001	400\r
17	2000	500\r
18	1999	130\r
19	1998	69\r
20	1997	25\r
21	1995	37\r
22	1994	117\r
23	1993	15\r
24	1989	4\r
25	1982	6837\r
26	1981	4\r

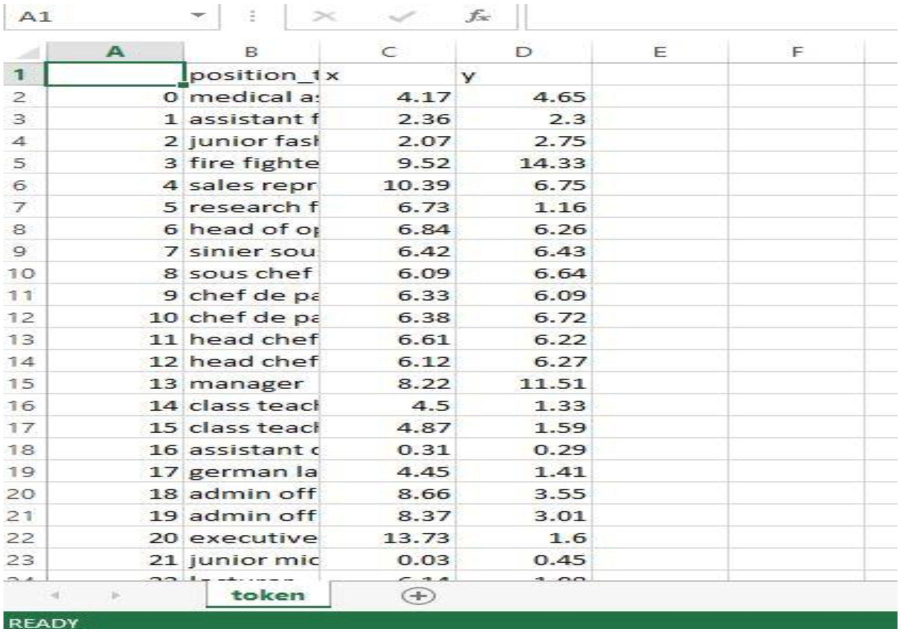
Figure 4.2: The Data of the Downloaded File

4.2. Clusters of similar jobs

To make a cluster of similar types of jobs first the data has to be preprocessed in order to make in suitable to implement any algorithm. After preprocessing, the kmeans clustering algorithm is implemented that gives a plotted graph. This plotted graph shows the clusters of job. In indicates, the most dense the cluster is, the more demanded the type of job is.

4.2.1 Pre-processed data for clustering

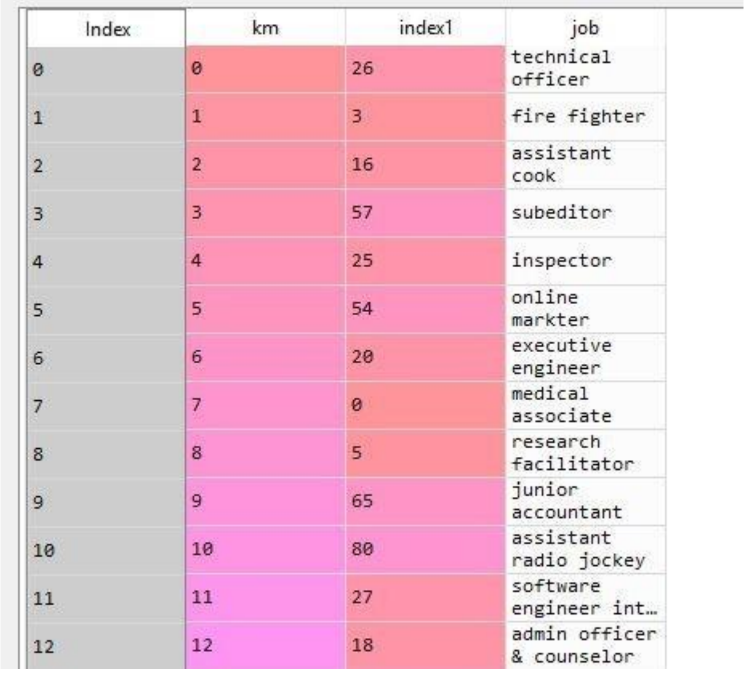
In figure 4.3, a new dataset is created using the “position_title” that contains a row of all the job positions of the job seekers registered into the skill.jobs website. After categorizing each job into different categories, vector values, x-vector and y-vector is set.



	A	B	C	D
1		position_t	x	y
2	0	medical a	4.17	4.65
3	1	assistant f	2.36	2.3
4	2	junior fasl	2.07	2.75
5	3	fire fighte	9.52	14.33
6	4	sales repr	10.39	6.75
7	5	research f	6.73	1.16
8	6	head of op	6.84	6.26
9	7	sinier sou	6.42	6.43
10	8	sous chef	6.09	6.64
11	9	chef de pa	6.33	6.09
12	10	chef de pa	6.38	6.72
13	11	head chef	6.61	6.22
14	12	head chef	6.12	6.27
15	13	manager	8.22	11.51
16	14	class teacl	4.5	1.33
17	15	class teacl	4.87	1.59
18	16	assistant c	0.31	0.29
19	17	german la	4.45	1.41
20	18	admin off	8.66	3.55
21	19	admin off	8.37	3.01
22	20	executive	13.73	1.6
23	21	junior mic	0.03	0.45

Figure 4.3: Dataset After Pre-processing For Clustering

In figure 4.4, the dataframe contains the values of “y_kmeans” in the “km” column with the respective index number of the “position_title” for which the “y_kmeans” vector was set. matching the index, the “position_title is add in the dataframe in the “job” column.



Index	km	index1	job
0	0	26	technical officer
1	1	3	fire fighter
2	2	16	assistant cook
3	3	57	subeditor
4	4	25	inspector
5	5	54	online markter
6	6	20	executive engineer
7	7	0	medical associate
8	8	5	research facilitator
9	9	65	junior accountant
10	10	80	assistant radio jockey
11	11	27	software engineer int…
12	12	18	admin officer & counselor

Figure 4.4: Dataframe Containing Labels of Each Cluster

4.2.2 Clustered data

The final scatter graph (figure 4.5) is shown that indicates 29 different clusters in along with a color code legend. Analyzing the density of each cluster, in can be identified which job is in most demand among the job seekers and training sessions or skill enhancement programs can be arranged in order to prepare or improve the skills of the job seekers so that they can find better jobs.

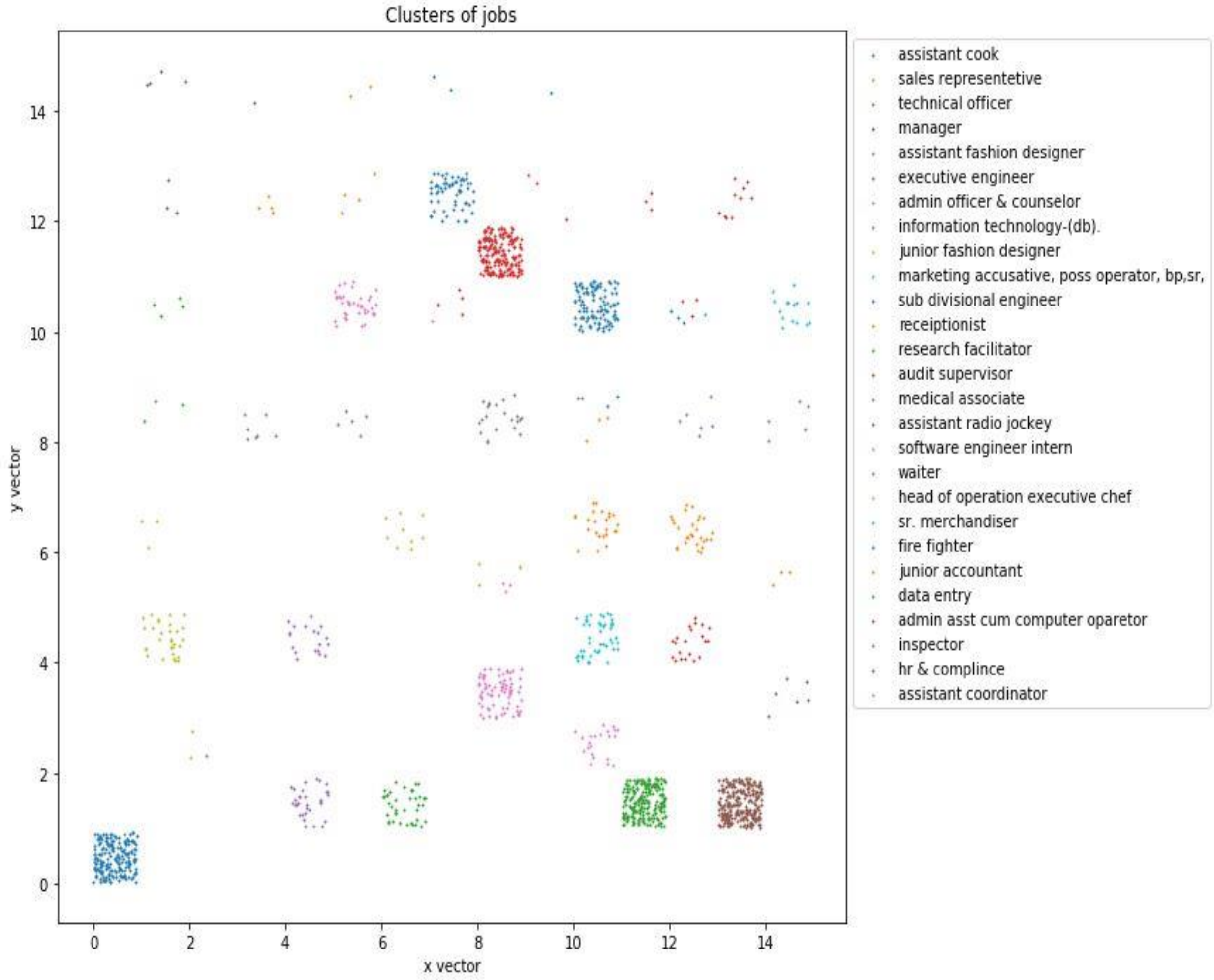


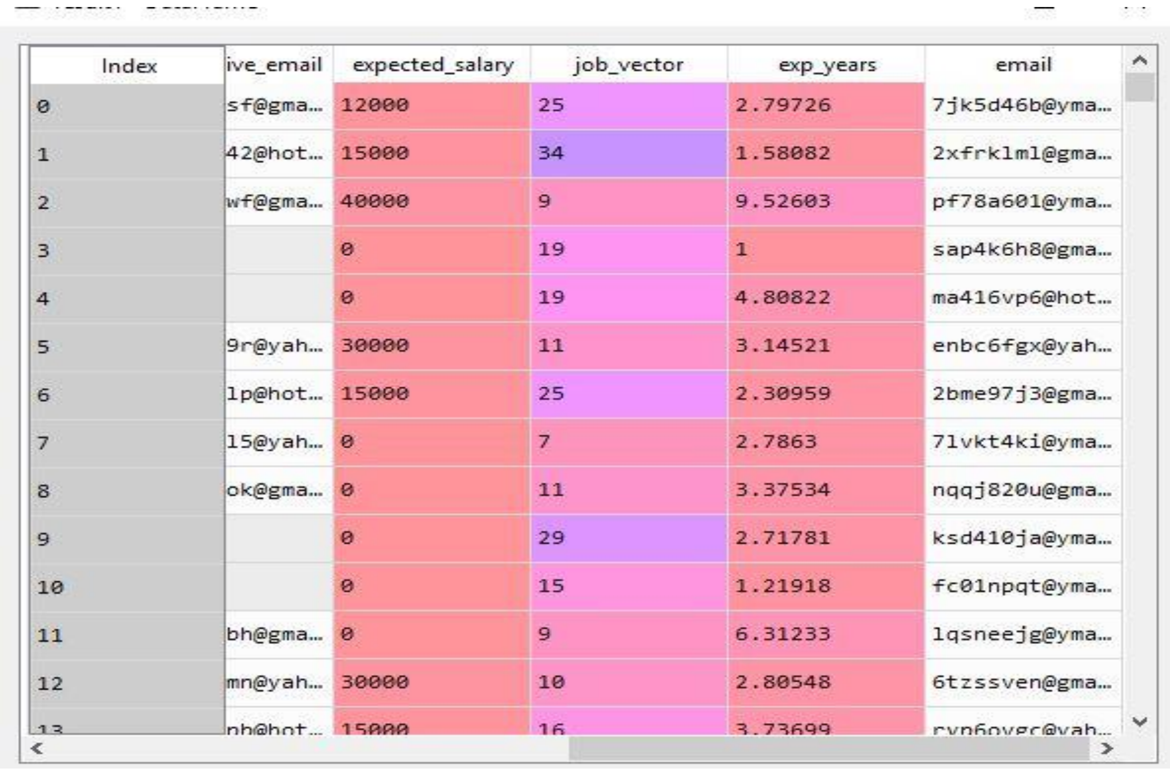
Figure 4.5: K-means Cluster of Types of Jobs

4.3. Email Notifications

The phase of sending email notification consists of three parts. Forst the raw data must be cleansed and preprocessed to make it suitable for implementing a classification algorithm. Once the data is preprocessed the decision tree algorithm is implemented. The decision tree gives as a decision that says which job seekers are suitable for a job and their email addresses are found out in order to send them notifications. Finally email notifications are send out to the addresses available from the database of the system.

4.3.1 Preprocessed data for decision making

In figure 4.6, after preprocessing the dataset, this is the final data frame received which is suitable for implementing the decision tree.

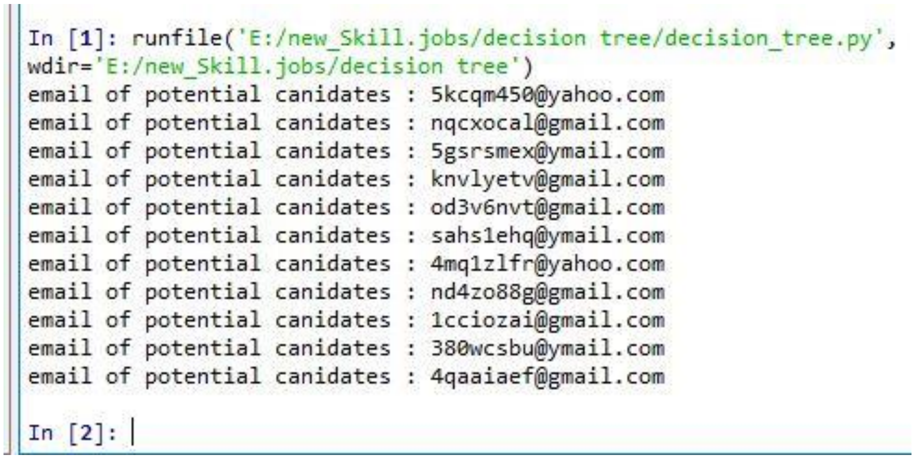


Index	ive_email	expected_salary	job_vector	exp_years	email
0	sf@gma…	12000	25	2.79726	7jk5d46b@yma…
1	42@hot…	15000	34	1.58082	2xfrklml@gma…
2	wf@gma…	40000	9	9.52603	pf78a601@yma…
3		0	19	1	sap4k6h8@gma…
4		0	19	4.80822	ma416vp6@hot…
5	9r@yah…	30000	11	3.14521	enbc6fgx@yah…
6	1p@hot…	15000	25	2.30959	2bme97j3@gma…
7	15@yah…	0	7	2.7863	71vkt4ki@yma…
8	ok@gma…	0	11	3.37534	nqqj820u@gma…
9		0	29	2.71781	ksd410ja@yma…
10		0	15	1.21918	fc01npqt@yma…
11	bh@gma…	0	9	6.31233	1qsneejg@yma…
12	mn@yah…	30000	10	2.80548	6tzssven@gma…
13	nh@hot…	15000	16	3.73699	ryn6ovgc@yah…

Figure 4.6: Dataframe After Jobseekers’ Data is Pre-processed

4.3.2 Decision making

In figure 4.7, if suitable candidates are found for any job, a list of email of the candidates will be shown as output.

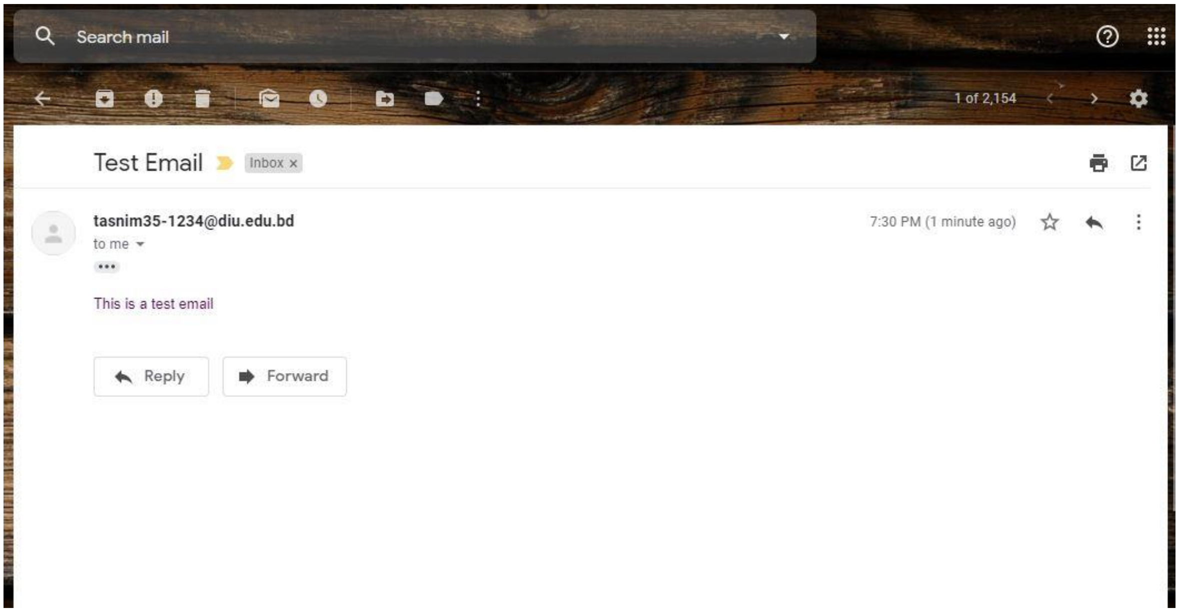


```
In [1]: runfile('E:/new_Skill.jobs/decision tree/decision_tree.py',
wdir='E:/new_Skill.jobs/decision tree')
email of potential canidates : 5kcqm450@yahoo.com
email of potential canidates : nqcxocal@gmail.com
email of potential canidates : 5gsrsmex@ymail.com
email of potential canidates : knvlyetv@gmail.com
email of potential canidates : od3v6nvt@gmail.com
email of potential canidates : sahs1ehq@ymail.com
email of potential canidates : 4mq1zlfr@yahoo.com
email of potential canidates : nd4zo88g@gmail.com
email of potential canidates : 1cciozai@gmail.com
email of potential canidates : 380wcsbu@ymail.com
email of potential canidates : 4qaaiaef@gmail.com

In [2]:
```

Figure 4.7: Output of Decision Tree

4.3.3 Sending Email

In figure 4.8, this is the test email, send from the user to one of the email address is the list



Test Email

tasnim35-1234@diu.edu.bd

to me

7:30 PM (1 minute ago)

This is a test email

Figure 4.8: The Email Sent

4.4. Summary

From the implemented system, it is seen that the output gives as exactly the results that were expected from the proposed system. Here we get the data from websites on the internet without them being provided manual. The automation of extracting the data makes the database of the system more enriched and efficient. At the same time, jobseekers are receiving more relevant notifications that helps them look for job faster and more effective. Furthermore, the jobseekers are now getting access to information of what kinds of jobs have more competition in market so that they can enhance their skills with workshops or training sessions enabling the chance of giving a better job more easily.

Chapter 5: Conclusion and Recommendations

The proposed system contains various phases that helps the website of skill.jobs to perform more efficiently and automate the system. The admin useres of the system and at the same the job seekers and HRD of dffernt company will experience a much more advanced and improved website with the implemetation of the proposed sytem. Since it is seen that the expected results are derived from the proposed phase, if the phase are implemented in the system, we will get a complete experience.

5.1. Findings and contributions

The system is proposed for skill.jobs which is one the biggest job sites in the country. The system contains three different phases. Each phase is responsible to improve the efficiency of the system. First of all, extract the job circular phase (EJC) is responsible for extracting the job circulars from different company's website automatically using a web crawler. The web crawler extracts the job circular from the company website's career section and store it in the database of skill.jobs. Next is the cluster similar kind of job search phase (CSK). This phase is tasked to give a graph containing the cluster of same kind of jobs from the data in the database of skill.jobs and the data gathered from the search history in the site by users, using the K-means algorithm. Analyzing the density of the cluster, skill enhancement training programs can be organized to help jobseekers get better jobs. Finally there is the email notifications send phase (ENS). In this phase a decision tree is implemented to make decision about to whom email notification of a certain job circular should be send. This phase uses the data stored in the database, provided by the jobseekers in order to match the job seekers' job position, salary and experience to find a suitable job from all the job circulars in the skill.jobs database. When job circular is matched with a suitable job seeker, the job seeker receives an email allowing him to know about the job circular. This phases together makes an efficient system that automatically gathers job circulars and allow job seekers to know if any job matches their requirement which helps job seekers to apply for jobs more efficiently. At the same time, analyzing the demand job seekers, training programs can be arranged that helps enhance skills to get better jobs.

5.2. Recommendations for Future Work

Machine learning and data science both are most important topics. This model will be helpful or tremendous discoveries for the skill.jobs organization. The business scholarly model can be diminish the exertion of the skill.jobs representative's and produce legitimate graphitic result for the particular classification. Along these lines, I chose to gather the example dataset of the skill.jobs and work on it by the savvy plan of action that I proposed. On the off chance that the model effectively actualized, it very well may be turn over for skill.jobs and above all it can lessen the human exertion and increment efficiency. The entire concepts can be changed into a scholarly model by various learning calculations (Fully Automated, Artificial intelligence).

In our coming exploration, we are hopeful to apply Artificial Intelligence and fully automated system approach in “Intelligence business model for skill.jobs” with merge of current developed model.

References

[1] Dr. Gulshan Ahuja, "An Efficient Mechanism for Navigating Web Using Mobile Web Crawler." International Journal of Engineering Research and Applications (IJERA) Vol. 3, Issue 2, 2013.

[2] Charu C. Aggarwal," On Learning Strategies for Topic Specific Web Crawling", IBM T. J. Watson Research Center, (2005).

[3] Trupti V. Udapure, Ravindra D. Kale and Rajesh C. Dharmik, "Study of Web Crawler and its Different Types" IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727Volume 16, Issue 1, Ver. VI, 2014.

[4] Nikita Jain and Vishal Srivastava, "DATA MINING TECHNIQUES: A SURVEY PAPER", IJRET: International Journal of Research in Engineering and Technology Volume: 02 Issue: 11, 2013.

[5] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand and Dan Steinberg, "Top 10 algorithms in data mining", Springer-Verlag London Limited 2007.

[6] Dr. S.Vijayarani1 and Ms. S.Sharmila2," RESEARCH IN BIG DATA – AN OVERVIEW", Informatics Engineering, an International Journal (IEIJ), Vol.4, No.3, September 2016.

[7] Himani Sharma and Sunil Kumar," A Survey on Decision Tree Algorithms of Classification in Data Mining",International Journal of Science and Research (IJSR) ,April 2016

[8] Davinder Kaur, Rajeev Bedi and Dr. Sunil Kumar Gupta, "REVIEW OF DECISION TREE DATA MINING ALGORITHMS:ID3 AND C4.5", Proceedings of International Conference on Information Technology and Computer ScienceJuly 11-12, 2015.

[9] D. P. Acharjya and Kauser Ahmed P, "A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools", (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 7, No. 2, 2016.

[10] Szil´ard Vajda1 and K.C. Santosh ,"A fast k-nearest neighbor classifier using unsupervised clustering", Communications in Computer and Information Science, April 2017

[11] M A Syakur, 2B K Khotimah, 3E M S Rochman and B D Satoto, "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster", IOP Conf. Series: Materials Science and Engineering 336 (2018) 012017.

[12] Cepy Slamet, Ali Rahman, Muhammad Ali Ramdhani, and Wahyudin Darmalaksana, "Clustering the Verses of the Holy Qur'an using K-Means Algorithm" , Asian Journal of Information Technology 15(24): 5159-5162, 2016.

[13] Kumar, M., Bhatia, R., Rattan, D. (2017) "A survey of Web crawlers for information retrieval." Wiley Interdiscip. Rev. Data Min. Knowl. Discov. e1218. doi:10.1002/widm.1218

[14] Eshan Sherkat, Julien Velcin and Evangelos E. Milios , "Fast and simple deterministic seeding of Kmeans for text document clustering", 9th International conference of the CLEF Association, CLEF 2018.

[15] Zhao, F., Zhou, J., Nie, C., Huang, H., & Jin, H. "SmartCrawler: A Two-stage Crawler for Efficiently Harvesting Deep-Web Interfaces." IEEE transactions on services computing, 9(4), 608-620

[16] Fong, P.K. and Weber-Jhanke, J.H,"Privacy Preserving Decision Tree Learning using Unrealized Data Sets",IEEE Transactionson knowledge and Data Engineering,2012 Vol.24,No.2, February 2012, pp. 353-364.

[17] Kabra, R.R. and Bichkar, R.S.,"Performance Prediction of Engineering Students using Decision Tree", International Journal of Computer Applications, Vol.36,No.11, December 2011, pp. 8-12.

[18] Karaolis, M.A. &Moutiris, J.A, "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining with Decision Trees", IEEE Transactions on Information Technology in Biomedicine,Vol.14, No.3, May 2010, pp. 559-566.

[19] Kesavraj, G. and Sukumaran, S., "A Study on Classification Technique in Data Mining", 4th ICCNT-2013.

[20] Sautikar, A.V., Bhujada, V., Bhagat,P.&Khaparde, A.," A Review paper on Various Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.4,Issue 4, April 2014, pp. 98-101.

[21] Li, L. & Zhang, X. (2010), "Study of Data Mining Algorithm based on Decision Tree", International Conference on Computer Design and Applications (ICCDA 2010), Vol.1, pp. 155-158.

[22] Yi-Yang, G. and Man-ping, R. , "Data Mining and Analysis of Our Agriculture based on the Decision Tree", ISECS International Colloquium on Computing, Communication, Control and management, 2009, pp. 134-138.

[23] Zhang, X.F. and Fan, L.," A Decision Tree Approach for Traffic accident Analysis of Saskatchewan Highways", 26th IEEE Canadian Conference of Electrical and Computer Engineering(CCECE) 2013.

[24] Zhang, T., Fulk, G.D. & Tang, W.,"Using Decision Tree to Measure Activities in People with stroke", 35th Annual International Conference of the IEEE EMBS, July 13, pp.6337-6340.

[25] Suknovic, .M, Delibasic, B., Jovanovic, M., Vukecevic, M., Obradovic, Z.,"Reusable components in decision tree induction algorithm",Comp Stat Februaury 2011.

[26] Connor, M., Kumar, P.: Fast construction of k-nearest neighbor graphs for point clouds. IEEE Transactions on Visualization and Computer Graphics **16**(4), 599{608 (2010)

[27] Gou, J., Du, L., Zhang, Y., Xiaong, T.: A new distance-weighted k-nearest neighbor classifier. Journal of Information and Computational Science **9**(6), 1429{1436 (2012)

[28] Jain, A.K.: Data clustering: 50 years beyond k-means. Pattern Recogn. Lett. **31**(8), 651{666 (2010)

[29] Junaidi, A., Vajda, S., Fink, G.A.: Lampung - a new handwritten character benchmark: Database, labeling and recognition. In: International Workshop on Multilingual OCR (MOCR), pp. 105{112. ACM, Beijing, China (2011)

[30] Lifshits, Y., Zhang, S.: Combinatorial algorithms for nearest neighbors, near duplicates and small-world design. In: SODA, pp. 318{326 (2009)

[31] Vajda, S., Junaidi, A., Fink, G.A.: A semi-supervised ensemble learning approach for character labeling with minimal human effort. In: ICDAR, pp. 259{263 (2011)

[32] Qing-yun Dai, Chun-ping Zhang and Hao Wu ,"Research of Decision Tree Classification Algorithm in Data Mining", International Journal of Database Theory and Application Vol.9, No.5 (2016), pp.1-8

[33] Yan-yan SONG and Ying LU "Decision tree methods: applications for classification and prediction", Shanghai Arch Psychiatry. 2015 Apr 25; 27(2): 130–135

[34] A Pranav and S Chauhan , "Efficient Focused Web Crawling Approach for Search Engine," Int. J. Kalol Inst. Technol. Res. Canter 2015

[35] Brin, S., Page, L. (2012) "Reprint of: The anatomy of a large-scale hypertextual web search engine." Comput. Networks. 56 (18):3825–3833. doi:10.1016/j.comnet.2012.10.007

[36] Kumar, M., Bhatia, R., Rattan, D. (2017) "A survey of Web crawlers for information retrieval." Wiley Interdiscip. Rev. Data Min.Knowl. Discov. e1218. doi:10.1002/widm.1218

[37] Kumar M, Bhatia R, Ohri A, Kohli A. "Design of focused crawler for information retrieval of Indian origin Academicians." In Advances in Computing, Communication, & Automation (ICACCA)(Spring), International Conference on 2016 Apr 8, IEEE:1-6

[38] Zhao, F., Zhou, J., Nie, C., Huang, H., & Jin, H. "SmartCrawler: A Two-stage Crawler for Efficiently Harvesting Deep-Web Interfaces." IEEE transactions on services computing, 9(4), 608-620. (2016)

[39] Priyatam PN, Vaddepally SR, Varma V. "Domain specific search in indian languages." In Proceedings of the first workshop on Information and knowledge management for developing region 2012 Nov 2, ACM: 23-30.

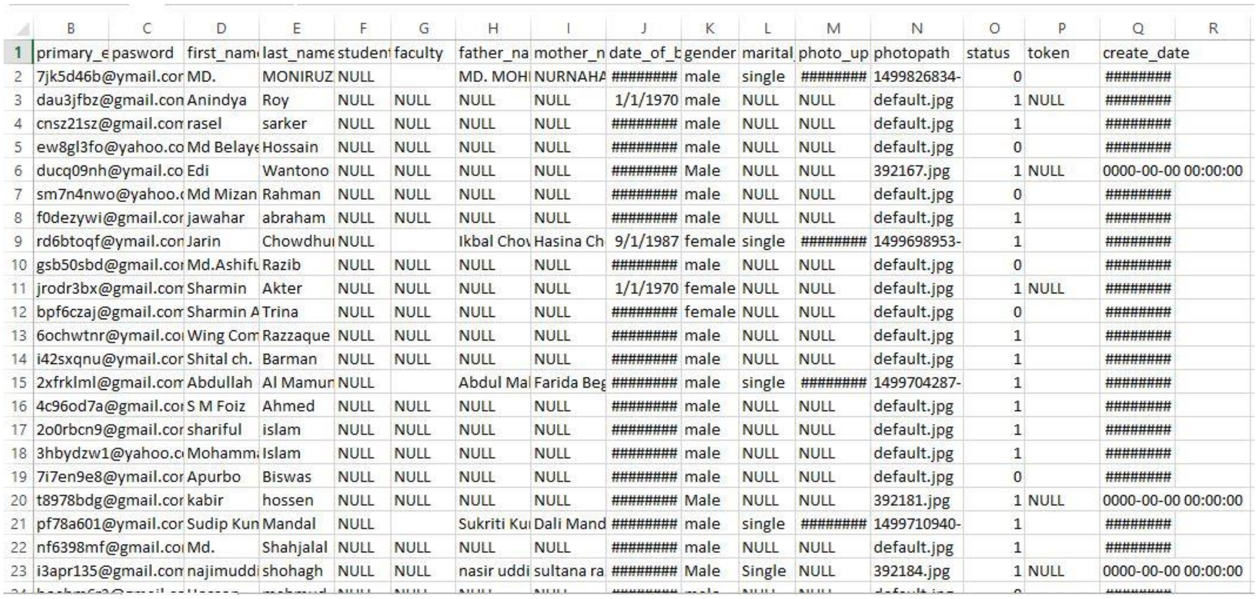
[40] Guojun, Zheng, et al. "Design and application of intelligent dynamic crawler for web data mining," Automation (YAC), 2017 32nd Youth Academic Annual Conference of Chinese Association of. IEEE, 2017.

[41] D. Debraj, and D. Payel "STUDY OF DEEP WEB AND A NEW FORM BASED CRAWLING TECHNIQUE," International Journal of Computer Engineering & Technology (IJCET) Vol. 7, pp. 36-44, Jan-Feb 2016.

[42] S. Saranya, B. S. E. Zoraida, and P. Victor Paul, "A Study on Competent Crawling Algorithm (CCA) for Web Search to Enhance Efficiency of Information Retrieval," Artificial Intelligence and Evolutionary Algorithms in Engineering Systems. Springer, New Delhi, 2015.

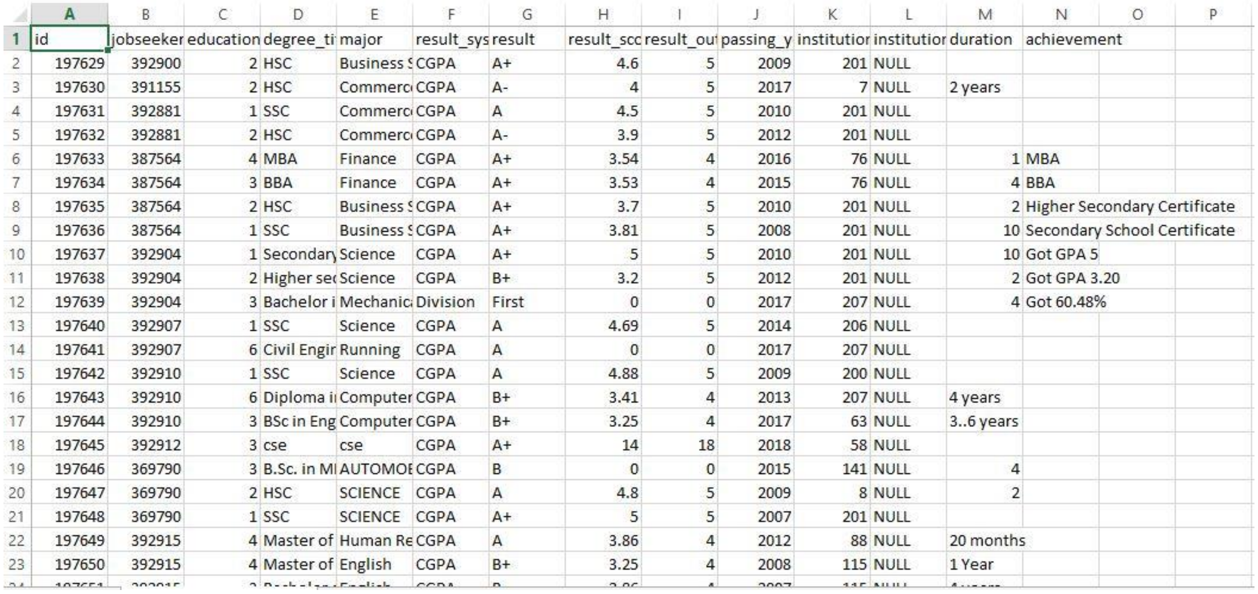
APPENDIX

Appendix A: jb_jobseeker dataset



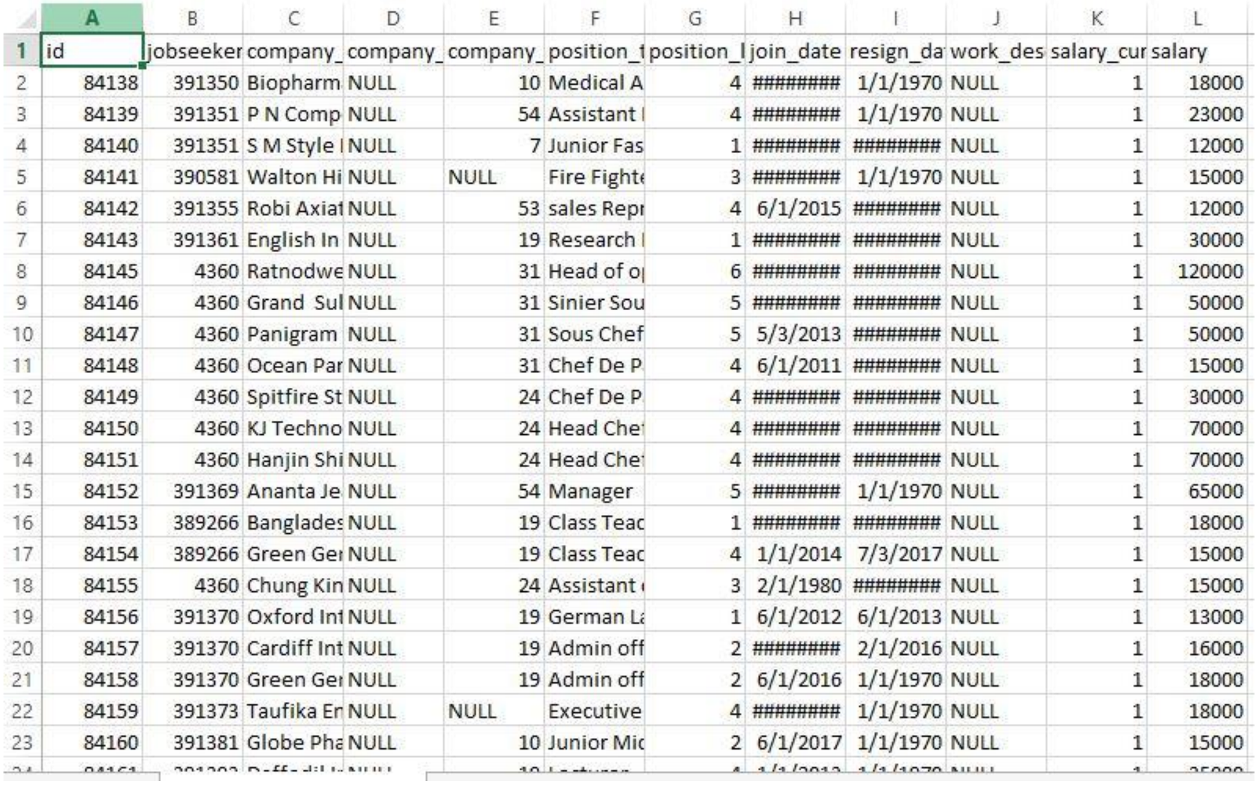
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	primary_e	pasword	first_nam	last_name	student	faculty	father_na	mother_n	date_of_b	gender	marital	photo_up	photopath	status	token	create_date	
2	7jk5d46b@ymail.cor		MD.	MONIRUZ	NULL		MD. MOHI	NURNAHA	########	male	single	########	1499826834-	0		########	
3	dau3jfbz@gmail.con		Anindya	Roy	NULL	NULL	NULL	NULL	1/1/1970	male	NULL	NULL	default.jpg	1	NULL	########	
4	cnsz21sz@gmail.com		rasel	sarker	NULL	NULL	NULL	NULL	########	male	NULL	NULL	default.jpg	1		########	
5	ew8gl3fo@yahoo.co		Md Belaye	Hossain	NULL	NULL	NULL	NULL	########	male	NULL	NULL	default.jpg	0		########	
6	ducq09nh@ymail.co		Edi	Wantono	NULL	NULL	NULL	NULL	########	Male	NULL	NULL	392167.jpg	1	NULL	0000-00-00 00:00:00	
7	sm7n4nwo@yahoo.c		Md Mizan	Rahman	NULL	NULL	NULL	NULL	########	male	NULL	NULL	default.jpg	0		########	
8	f0dezywi@gmail.cor		jawahar	abraham	NULL	NULL	NULL	NULL	########	male	NULL	NULL	default.jpg	1		########	
9	rd6btoqf@ymail.con		Jarin	Chowdhu	NULL		Ikbal Chov	Hasina Ch	9/1/1987	female	single	########	1499698953-	1		########	
10	gsb50sbd@gmail.co		Md.Ashifu	Razib	NULL	NULL	NULL	NULL	########	male	NULL	NULL	default.jpg	0		########	
11	jrodr3bx@gmail.con		Sharmin	Akter	NULL	NULL	NULL	NULL	1/1/1970	female	NULL	NULL	default.jpg	1	NULL	########	
12	bpf6czaj@gmail.com		Sharmin A	Trina	NULL	NULL	NULL	NULL	########	female	NULL	NULL	default.jpg	0		########	
13	6ochwtnr@ymail.co		Wing Com	Razzaque	NULL	NULL	NULL	NULL	########	male	NULL	NULL	default.jpg	1		########	
14	i42sxqnu@ymail.cor		Shital ch.	Barman	NULL	NULL	NULL	NULL	########	male	NULL	NULL	default.jpg	1		########	
15	2xfrklml@gmail.com		Abdullah	Al Mamur	NULL		Abdul Ma	Farida Beg	########	male	single	########	1499704287-	1		########	
16	4c96od7a@gmail.co		S M Foiz	Ahmed	NULL	NULL	NULL	NULL	########	male	NULL	NULL	default.jpg	1		########	
17	2o0rbcn9@gmail.cor		shariful	islam	NULL	NULL	NULL	NULL	########	male	NULL	NULL	default.jpg	1		########	
18	3hbydzw1@yahoo.c		Mohamm	Islam	NULL	NULL	NULL	NULL	########	male	NULL	NULL	default.jpg	1		########	
19	7i7en9e8@ymail.cor		Apurbo	Biswas	NULL	NULL	NULL	NULL	########	male	NULL	NULL	default.jpg	0		########	
20	t8978bdg@gmail.cor		kabir	hossen	NULL	NULL	NULL	NULL	########	Male	NULL	NULL	392181.jpg	1	NULL	0000-00-00 00:00:00	
21	pf78a601@ymail.cor		Sudip Kun	Mandal	NULL		Sukriti Ku	Dali Mand	########	male	single	########	1499710940-	1		########	
22	nf6398mf@gmail.co		Md.	Shahjalal	NULL	NULL	NULL	NULL	########	male	NULL	NULL	default.jpg	1		########	
23	i3apr135@gmail.com		najimuddi	shohagh	NULL	NULL	nasir uddi	sultana ra	########	Male	Single	NULL	392184.jpg	1	NULL	0000-00-00 00:00:00	

Appendix B: jb_jobseeker_education dataset



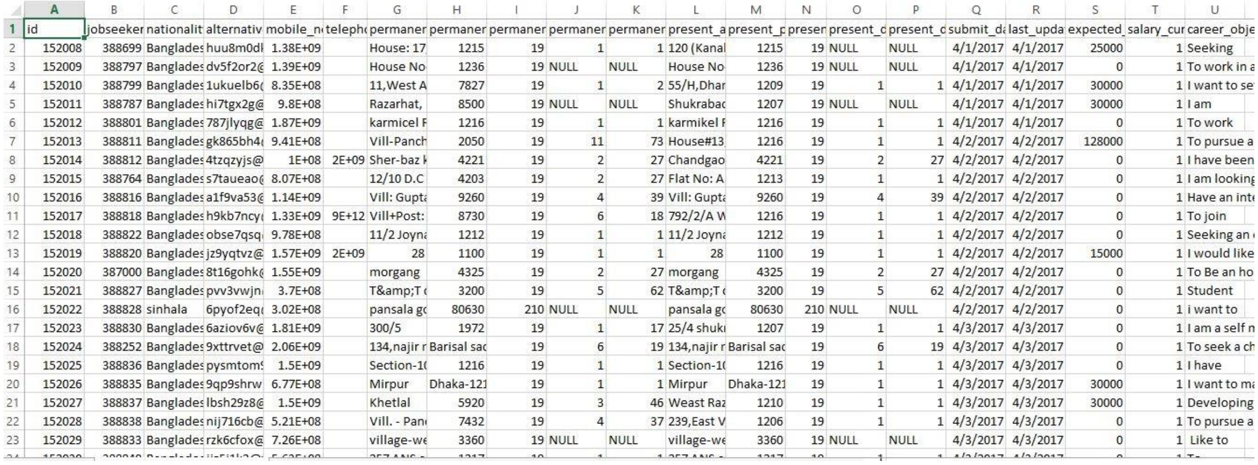
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	id	jobseeker	education	degree_ti	major	result_sys	result	result_scc	result_ou	passing_y	institutior	institutior	duration	achievement		
2	197629	392900	2	HSC	Business S	CGPA	A+	4.6	5	2009	201	NULL				
3	197630	391155	2	HSC	Commerc	CGPA	A-	4	5	2017	7	NULL	2 years			
4	197631	392881	1	SSC	Commerc	CGPA	A	4.5	5	2010	201	NULL				
5	197632	392881	2	HSC	Commerc	CGPA	A-	3.9	5	2012	201	NULL				
6	197633	387564	4	MBA	Finance	CGPA	A+	3.54	4	2016	76	NULL	1	MBA		
7	197634	387564	3	BBA	Finance	CGPA	A+	3.53	4	2015	76	NULL	4	BBA		
8	197635	387564	2	HSC	Business S	CGPA	A+	3.7	5	2010	201	NULL	2	Higher Secondary Certificate		
9	197636	387564	1	SSC	Business S	CGPA	A+	3.81	5	2008	201	NULL	10	Secondary School Certificate		
10	197637	392904	1	Secondary	Science	CGPA	A+	5	5	2010	201	NULL	10	Got GPA 5		
11	197638	392904	2	Higher sec	Science	CGPA	B+	3.2	5	2012	201	NULL	2	Got GPA 3.20		
12	197639	392904	3	Bachelor i	Mechanic	Division	First	0	0	2017	207	NULL	4	Got 60.48%		
13	197640	392907	1	SSC	Science	CGPA	A	4.69	5	2014	206	NULL				
14	197641	392907	6	Civil Engir	Running	CGPA	A	0	0	2017	207	NULL				
15	197642	392910	1	SSC	Science	CGPA	A	4.88	5	2009	200	NULL				
16	197643	392910	6	Diploma i	Computer	CGPA	B+	3.41	4	2013	207	NULL	4 years			
17	197644	392910	3	BSc in Eng	Computer	CGPA	B+	3.25	4	2017	63	NULL	3..6 years			
18	197645	392912	3	cse	cse	CGPA	A+	14	18	2018	58	NULL				
19	197646	369790	3	B.Sc. in M	AUTOMOE	CGPA	B	0	0	2015	141	NULL	4			
20	197647	369790	2	HSC	SCIENCE	CGPA	A	4.8	5	2009	8	NULL	2			
21	197648	369790	1	SSC	SCIENCE	CGPA	A+	5	5	2007	201	NULL				
22	197649	392915	4	Master of	Human Re	CGPA	A	3.86	4	2012	88	NULL	20 months			
23	197650	392915	4	Master of	English	CGPA	B+	3.25	4	2008	115	NULL	1 Year			

Appendix C: jb_jobseeker_experience dataset



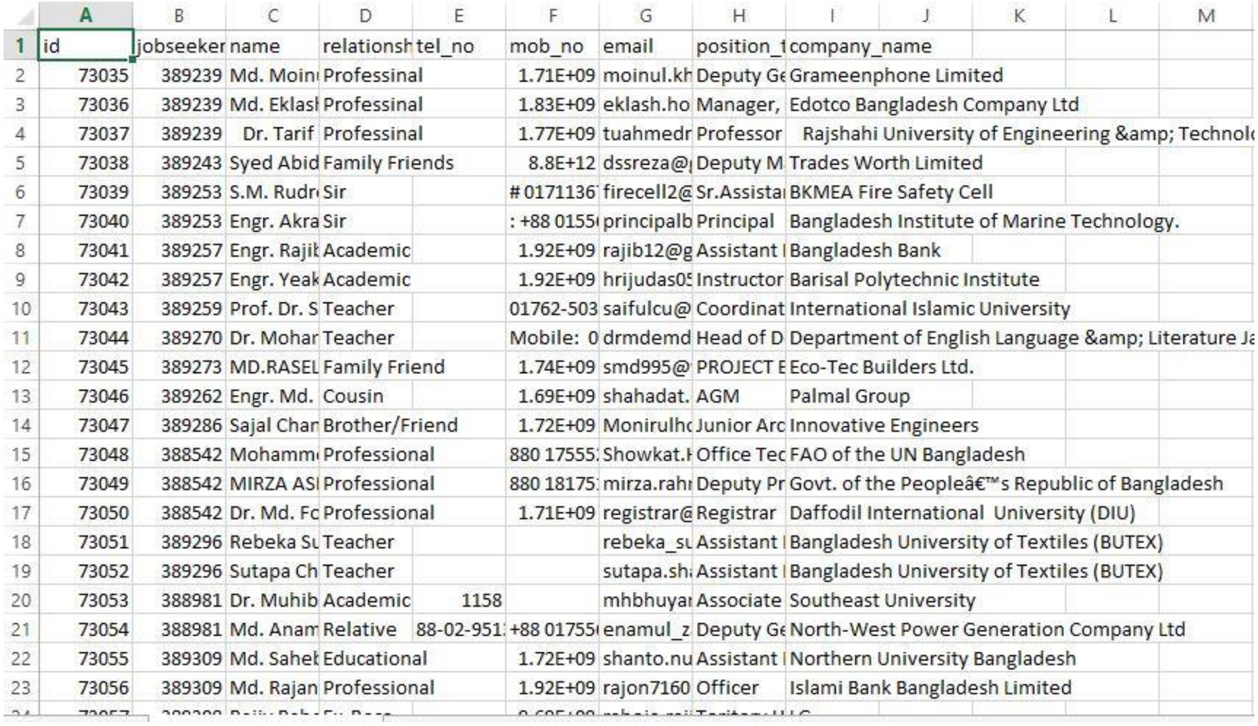
	A	B	C	D	E	F	G	H	I	J	K	L
1	id	jobseeker	company_	company_	company_	position_t	position_l	join_date	resign_da	work_des	salary_cur	salary
2	84138	391350	Biopharm	NULL	10	Medical A	4	########	1/1/1970	NULL	1	18000
3	84139	391351	P N Comp	NULL	54	Assistant	4	########	1/1/1970	NULL	1	23000
4	84140	391351	S M Style I	NULL	7	Junior Fas	1	########	########	NULL	1	12000
5	84141	390581	Walton Hi	NULL	NULL	Fire Fighte	3	########	1/1/1970	NULL	1	15000
6	84142	391355	Robi Axiat	NULL	53	sales Repr	4	6/1/2015	########	NULL	1	12000
7	84143	391361	English In	NULL	19	Research I	1	########	########	NULL	1	30000
8	84145	4360	Ratnodwe	NULL	31	Head of o	6	########	########	NULL	1	120000
9	84146	4360	Grand Sul	NULL	31	Sinier Sou	5	########	########	NULL	1	50000
10	84147	4360	Panigram	NULL	31	Sous Chef	5	5/3/2013	########	NULL	1	50000
11	84148	4360	Ocean Par	NULL	31	Chef De P	4	6/1/2011	########	NULL	1	15000
12	84149	4360	Spitfire St	NULL	24	Chef De P	4	########	########	NULL	1	30000
13	84150	4360	KJ Techno	NULL	24	Head Che	4	########	########	NULL	1	70000
14	84151	4360	Hanjin Shi	NULL	24	Head Che	4	########	########	NULL	1	70000
15	84152	391369	Ananta Je	NULL	54	Manager	5	########	1/1/1970	NULL	1	65000
16	84153	389266	Banglades	NULL	19	Class Teac	1	########	########	NULL	1	18000
17	84154	389266	Green Ger	NULL	19	Class Teac	4	1/1/2014	7/3/2017	NULL	1	15000
18	84155	4360	Chung Kin	NULL	24	Assistant	3	2/1/1980	########	NULL	1	15000
19	84156	391370	Oxford Int	NULL	19	German La	1	6/1/2012	6/1/2013	NULL	1	13000
20	84157	391370	Cardiff Int	NULL	19	Admin off	2	########	2/1/2016	NULL	1	16000
21	84158	391370	Green Ger	NULL	19	Admin off	2	6/1/2016	1/1/1970	NULL	1	18000
22	84159	391373	Taufika En	NULL	NULL	Executive	4	########	1/1/1970	NULL	1	18000
23	84160	391381	Globe Pha	NULL	10	Junior Mic	2	6/1/2017	1/1/1970	NULL	1	15000

Appendix D: jb_jobseeker_personal_info



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	id	jobseeker	nationalit	alternativ	mobile_n	teleph	permaner	permaner	permaner	permaner	permaner	present_a	present_p	presen	present_d	present_d	submit_da	last_upda	expected_	salary_cur	career_obje
2	152008	388699	Banglades	huu8m0dl	1.38E+09		House: 17	1215	19	1	1	120 (Kanal	1215	19	NULL	NULL	4/1/2017	4/1/2017	25000	1	Seeking
3	152009	388797	Banglades	dv5f2or2@	1.39E+09		House No	1236	19	NULL	NULL	House No	1236	19	NULL	NULL	4/1/2017	4/1/2017	0	1	To work in a
4	152010	388799	Banglades	1ukuelb6	8.35E+08		11,West A	7827	19	1	2	55/H,Dhar	1209	19	1	1	4/1/2017	4/1/2017	30000	1	I want to se
5	152011	388787	Banglades	hi7tgx2g@	9.8E+08		Razarhat,	8500	19	NULL	NULL	Shukrabad	1207	19	NULL	NULL	4/1/2017	4/1/2017	30000	1	I am
6	152012	388801	Banglades	787jlyqg@	1.87E+09		karmicel F	1216	19	1	1	karmikel F	1216	19	1	1	4/1/2017	4/1/2017	0	1	To work
7	152013	388811	Banglades	gk865bh4	9.41E+08		Vill-Panch	2050	19	11	73	House#13	1216	19	1	1	4/2/2017	4/2/2017	128000	1	To pursue a
8	152014	388812	Banglades	4tzqzyjs@	1E+08	2E+09	Sher-baz k	4221	19	2	27	Chandgao	4221	19	2	27	4/2/2017	4/2/2017	0	1	I have been
9	152015	388764	Banglades	s7taueao	8.07E+08		12/10 D.C	4203	19	2	27	Flat No: A	1213	19	1	1	4/2/2017	4/2/2017	0	1	I am looking
10	152016	388816	Banglades	a1f9va53@	1.14E+09		Vill: Gupta	9260	19	4	39	Vill: Gupta	9260	19	4	39	4/2/2017	4/2/2017	0	1	Have an inte
11	152017	388818	Banglades	h9kb7ncy	1.33E+09	9E+12	Vill+Post:	8730	19	6	18	792/2/A W	1216	19	1	1	4/2/2017	4/2/2017	0	1	To join
12	152018	388822	Banglades	obse7qsq	9.78E+08		11/2 Joyna	1212	19	1	1	11/2 Joyna	1212	19	1	1	4/2/2017	4/2/2017	0	1	Seeking an
13	152019	388820	Banglades	jz9yqtvz@	1.57E+09	2E+09	28	1100	19	1	1	28	1100	19	1	1	4/2/2017	4/2/2017	15000	1	I would like
14	152020	387000	Banglades	8t16gohk	1.55E+09		morgang	4325	19	2	27	morgang	4325	19	2	27	4/2/2017	4/2/2017	0	1	To Be an ho
15	152021	388827	Banglades	pvv3vwjn	3.7E+08		T&amp;T	3200	19	5	62	T&amp;T	3200	19	5	62	4/2/2017	4/2/2017	0	1	Student
16	152022	388828	sinhala	6pyof2eq	3.02E+08		pansala gc	80630	210	NULL	NULL	pansala gc	80630	210	NULL	NULL	4/2/2017	4/2/2017	0	1	i want to
17	152023	388830	Banglades	6aziov6v@	1.81E+09		300/5	1972	19	1	17	25/4 shuk	1207	19	1	1	4/3/2017	4/3/2017	0	1	I am a self n
18	152024	388252	Banglades	9xttrvet@	2.06E+09		134,najir r	Barisal sad	19	6	19	134,najir r	Barisal sad	19	6	19	4/3/2017	4/3/2017	0	1	To seek a ch
19	152025	388836	Banglades	pysmtom	1.5E+09		Section-1	1216	19	1	1	Section-1	1216	19	1	1	4/3/2017	4/3/2017	0	1	I have
20	152026	388835	Banglades	9qp9shrw	6.77E+08		Mirpur	Dhaka-121	19	1	1	Mirpur	Dhaka-121	19	1	1	4/3/2017	4/3/2017	30000	1	I want to ma
21	152027	388837	Banglades	lbsh29z8@	1.5E+09		Khetlal	5920	19	3	46	Weast Raz	1210	19	1	1	4/3/2017	4/3/2017	30000	1	Developing
22	152028	388838	Banglades	nij716cb@	5.21E+08		Vill. - Pan	7432	19	4	37	239,East V	1206	19	1	1	4/3/2017	4/3/2017	0	1	To pursue a
23	152029	388833	Banglades	rzk6cfox@	7.26E+08		village-we	3360	19	NULL	NULL	village-we	3360	19	NULL	NULL	4/3/2017	4/3/2017	0	1	Like to

Appendix E: jb_jobseeker_reference



	A	B	C	D	E	F	G	H	I
1	id	jobseeker	name	relationsh	tel_no	mob_no	email	position_t	company_name
2	73035	389239	Md. Moin	Professinal		1.71E+09	moinul.kh	Deputy Ge	Grameenphone Limited
3	73036	389239	Md. Eklasl	Professinal		1.83E+09	eklash.ho	Manager,	Edotco Bangladesh Company Ltd
4	73037	389239	Dr. Tarif	Professinal		1.77E+09	tuahmedr	Professor	Rajshahi University of Engineering &amp; Technol
5	73038	389243	Syed Abid	Family Friends		8.8E+12	dssreza@	Deputy M	Trades Worth Limited
6	73039	389253	S.M. Rudr	Sir		# 0171136	firecell2@	Sr.Assista	BKMEA Fire Safety Cell
7	73040	389253	Engr. Akra	Sir		: +88 0155	principalb	Principal	Bangladesh Institute of Marine Technology.
8	73041	389257	Engr. Rajib	Academic		1.92E+09	rajib12@g	Assistant	Bangladesh Bank
9	73042	389257	Engr. Yeak	Academic		1.92E+09	hrijudas05	Instructor	Barisal Polytechnic Institute
10	73043	389259	Prof. Dr. S	Teacher		01762-503	saifulcu@	Coordinat	International Islamic University
11	73044	389270	Dr. Mohar	Teacher		Mobile: 0	drmdemd	Head of D	Department of English Language &amp; Literature Ja
12	73045	389273	MD.RASEL	Family Friend		1.74E+09	smd995@	PROJECT E	Eco-Tec Builders Ltd.
13	73046	389262	Engr. Md.	Cousin		1.69E+09	shahadat.	AGM	Palmal Group
14	73047	389286	Sajal Char	Brother/Friend		1.72E+09	Monirulhc	Junior Arc	Innovative Engineers
15	73048	388542	Mohamm	Professional		880 17555	Showkat.H	Office Tec	FAO of the UN Bangladesh
16	73049	388542	MIRZA ASI	Professional		880 18175	mirza.rahı	Deputy Pr	Govt. of the Peopleâ€™s Republic of Bangladesh
17	73050	388542	Dr. Md. Fc	Professional		1.71E+09	registrar@	Registrar	Daffodil International University (DIU)
18	73051	389296	Rebeka Su	Teacher			rebeka_su	Assistant	Bangladesh University of Textiles (BUTEX)
19	73052	389296	Sutapa Ch	Teacher			sutapa.sh	Assistant	Bangladesh University of Textiles (BUTEX)
20	73053	388981	Dr. Muhib	Academic	1158		mhbhuyar	Associate	Southeast University
21	73054	388981	Md. Anam	Relative	88-02-951	+88 01755	enamul_z	Deputy Ge	North-West Power Generation Company Ltd
22	73055	389309	Md. Sahel	Educational		1.72E+09	shanto.nu	Assistant	Northern University Bangladesh
23	73056	389309	Md. Rajan	Professional		1.92E+09	rajon7160	Officer	Islami Bank Bangladesh Limited