Project Report

On

Job Recommendation based on job profile clustering and job seeker behavior

Submitted for partial fulfillment of requirement for the degree of

BACHELOR OF ENGINEERING

(Computer Science and Engineering)

Submitted By

Vaibhav Sanjayrao Dharmik

Sanket Rambhau Gulhane

Devendra Babasaheb Talhande

Girish Sudhakar Sawsundar

Under the Guidance of

Dr. Vaishali H. Deshmukh

Department of Computer Science & Engineering,

Prof. Ram Meghe Institute of Technology & Research, Badnera

2021-2022

Department of Computer Science & Engineering
Prof. Ram Meghe Institute of Technology & Research, Badnera
2021 - 2022

CERTIFICATE

This is to certify that the Project (8KS07) entitled

Job Recommendation based on job profile clustering and job seeker behavior

is a bonafide work and it is submitted to the

Sant Gadge Baba Amravati University, Amravati

By

Vaibhav Sanjayrao Dharmik
Sanket Rambhau Gulhane
Devendra Babasaheb Talhande
Girish Sudhakar Sawsundar

in the partial fulfillment of the requirement for the degree of ***Bachelor of Engineering*** *in* ***Computer Science & Engineering****, during the academic year* ***2021-2022*** *under my guidance.*

Dr. Vaishali H. Deshmukh
Guide
Department of Computer Sci.& Engg.
Prof. Ram Meghe Institute Of Technology & Research, Badnera

Dr. G. R. Bamnote
Head,
Department of Computer Sci.& Engg.
Prof. Ram Meghe Institute Of Technology & Research, Badnera

External Examiner

_____

ACKNOWLEGEMENT

With great pleasure we hereby acknowledge the help given to us by various individuals throughout the project. This Project itself is an acknowledgement to the inspiration, drive and technical assistance contributed by many individuals. This project would have never seen the light of this day without the help and guidance we have received.

We would like to express our profound thanks to Dr. V. H. Deshmukh for her guidance and constant supervision as well as for providing necessary information regarding the project & also for her support in completing the project. We would also thank the faculties of the Department of Computer Science & Engineering, for their kind co-operation and encouragement which help us in completion of this project. We owe an incalculable debt to all staffs of the Department of Computer Science & Engineering for their direct and indirect help.

Our thanks and appreciations also go to our colleague in developing the project and people who have willingly helped us out with their abilities.

We extend our heartfelt thanks to our parents, friends and well-wishers for their support and timely help. Last but not the least; we thank the God Almighty for guiding us in every step of the way.

Vaibhav S. Dharmik _____

Sanket R. Gulhane _____

Devendra B. Talhande _____

Girish S. Sawsundar _____

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SCREENSHOTS

ABSTRACT

Many students search for summer jobs during the vacation, but there are always too many choices. We need to find a way to help people choose a best summer job. We constructed a three-tier system to comprehensively illustrate the factors that high school / Jr. college / University students need to consider when looking for a summer job from the criteria of comfort, salary, personal gain, and matching degree. Under each criterion lie several sub-criteria (which are discussed later in detail).

In this project we proposed an online suitable job recommendation system based on user's profile and preferences. We proposed job recommendation model using **K-means Clustering** algorithm which takes job offers as an input and create clusters based on similar attributes. The matching job cluster will be recommended to the users as per their profiles.

CHAPTER 1

INTRODUCTION

The highly competitive and dynamic nature of the job market as well as personal preferences and goals lead individuals to change their jobs at some point in their lives. Moving to a new job, however, is not an easy decision, which may depend on many factors, such as salary, job description, and geographical location. Making successful job transitions is essential for a successful professional career. In this work, we build an automated system that can recommend jobs to people based on their past job histories in order to facilitate the process of selecting a new job. We believe that such a system can successfully exploit the job transitions performed by other employees. That is, we propose recommending jobs to people based on inference from the job transition patterns observed in the past. These patterns may involve features extracted from the business profiles of employees (e.g., years of experience, educational degree, job title), the profiles of institutions1 (e.g., industry, type, size), and the job transitions themselves (e.g., frequency of transitions between jobs, average time spent in a job).

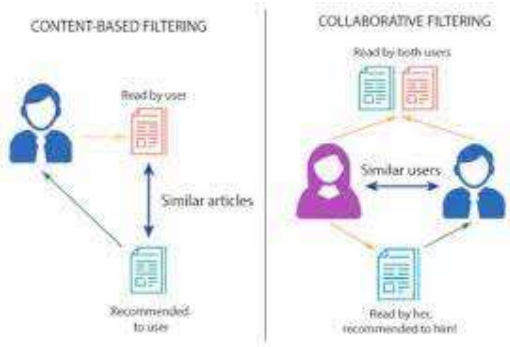


Fig. 1. Content based filtering and collaborative filtering recommendation

The framework we propose is based on K means clustering. Given an employee's past job history, the objective of the learning model is to accurately predict the next institution that the employee will move to. The predicted institution can then be recommended to the employee as the next step in his/her career. To evaluate our

framework, we use a large sample of job transitions extracted from the publicly available employee profiles in the Web. From this sample, we extract a number of features that we use to train and test our machine learning model. The results of our experiments demonstrate that the transition of an employee to an institution can be quite accurately predicted, significantly improving over a baseline predictor that always predicts the most frequent institution in the data. Our results indicate that the most important feature in predicting a job transition is the current institution of the employee.

1.1 Problem Definition

In existing job recommendation sites, job offers which are matching with user's profile used to recommend user. To improve this recommendation, we proposed machine learning algorithms which will find user's preferences and recommend job offers as per the preferences, profile and requirements of the recruiters.

1.2 Problem Objectives

- To implement K means clustering algorithm
- To develop an online job recommendation application for job seekers
- To implement recommendation model

CHAPTER 2
LITERATURE SURVEY

According to Narendra Nathan that the identification of the parameters of interest is achieved by a priori specification of the distribution of wage-vacancies, as required by the applicants.

The divided the labor flow into different age groups and gender groups. For the young male population group, business cycle, government expenditure for education and reward of formal education are found to be the significant determinant factors and for the females of the same age group, business cycle and reward from formal education are the significant determinants factors.

Along with widespread Internet access, job vacancies can not only be seen through print media, but can also be accessed online. Until now, there have been many websites and applications providing job vacancies. There are many backgrounds of applicants, while companies also have different specifications in job vacancies. Therefore, an application is needed to help applicants find job vacancies that match their abilities and fields of interest and that can recommend specific job vacancies for undergraduates from universities.

Previous research related to applications such as the Puspasari research, where the use of applications can increase efficiency and data information on special processes that are displayed. In other research related to applications such as in Pramono's research, application design with concepts and interfaces can provide alternative information solutions for users.

Data Clustering is one of the Data Mining methods that is unsupervised. There are two types of data clustering that are often used in the data grouping process, namely hierarchical (hierarchical) data clustering and non-hierarchical (non-hierarchical) data clustering. K-Means is a non-hierarchical data clustering method that attempts to partition existing data into one or more clusters/groups.

This method partitions data into clusters so that data with the same characteristics are grouped into the same cluster and data with different characteristics

are grouped into other groups. The purpose of this data clustering is to minimize the objective function set in the clustering process, which generally tries to minimize variations within a cluster and maximize variations between clusters. The benefits of Clustering are as Object Identification (Recognition) for example in the field of Image Processing, Computer Vision or robot vision. In addition, it is a Decision Support System and Data Mining such as market segmentation, area mapping, marketing management etc.

CHAPTER 3
SYSTEM ANALYSIS

3.1 PROBLEM STATEMENT:

The highly competitive and dynamic nature of the job market as well as personal preferences and goals lead individuals to change their jobs at some point in their lives. Moving to a new job, however, is not an easy decision, which may depend on many factors, such as salary, job description, and geographical location. Making successful job transitions is essential for a successful professional career. In this

3.2 SYSTEM REQUIREMENTS:

Our project has some software requirements as follows:

3.2.1 Eclipse Software:

Eclipse is an integrated development environment (IDE) used in computer programming. It contains a base workspace and an extensible plug-in system for customizing the environment. We used this software to develop our project because of its compatibility with java programming language. The Eclipse SDK includes the Eclipse Java development tools (JDT), offering an IDE with a built-in Java incremental compiler and a full model of the Java source files. This allows for advanced refactoring techniques and code analysis. The IDE also makes use of a workspace, in this case a set of metadata over a flat file space allowing external file modifications as long as the corresponding workspace resource is refreshed afterward.

Eclipse implements the graphical control elements of the Java toolkit called Standard Widget Toolkit (SWT), whereas most Java applications use the Java standard Abstract Window Toolkit (AWT) or Swing. Eclipse's user interface also uses an intermediate graphical user interface layer called JFace, which simplifies the construction of applications based on SWT.

3.2.2 Apache Tomcat Server:

Apache Tomcat (called "Tomcat" for short) is an open-source implementation of the Java Servlet, JavaServer Pages, Java Expression Language and WebSocket technologies. Tomcat provides a "pure Java" HTTP web server environment in which Java code can run. We have used Tomcat 4.x which was released with Catalina (a servlet container), Coyote (an HTTP connector) and Jasper (a JSP engine). As our project is a web application for detecting duplicate images in user's own created database so we used this because it has also added user— as well as system-based web applications enhancement to add support for deployment across the variety of environments. It also tries to manage sessions as well as applications across the network.

3.2.3 MySQL

MySQL is an open-source relational database management system (RDBMS). A relational database organizes data into one or more data tables in which data types may be related to each other; these relations help structure the data. SQL is a language programmers use to create, modify and extract data from the relational database, as well as control user access to the database. In addition to relational databases and SQL, an RDBMS like MySQL works with an operating system to implement a relational database in a computer's storage system, manages users, allows for network access and facilitates testing database integrity and creation of backups.

3.3 TECHNOLOGIES INVOLVED

3.3.1 Jakarta Server Pages (JSP):

Jakarta Server pages is one of the original java web technology which is being widely used to create dynamic web pages that can connect to java backend. It is built on top of the Java Servlet specification. JSP may be viewed as a high-level abstraction of Java servlets. JSPs are translated into servlets at runtime, therefore JSP is a Servlet; each JSP servlet is cached and re-used until the original JSP is modified. Jakarta Server Pages can be used independently or as the view component of a server-side model–view–controller design, normally with JavaBeans as the model and Java servlets (or a framework such as Apache Struts) as the controller. JSP allows Java code and certain predefined actions to be interleaved with static web markup content,

such as HTML. The resulting page is compiled and executed on the server to deliver a document. The compiled pages, as well as any dependent Java libraries, contain Java byte code rather than machine code. Like any other .jar or Java program, code must be executed within a Java virtual machine (JVM) that interacts with the server's host operating system to provide an abstract, platform-neutral environment. JSPs are usually used to deliver HTML and XML documents, but through the use of OutputStream, they can deliver other types of data as well. The Web container creates JSP implicit objects like request, response, session, application, config, page, pageContext, out and exception. JSP Engine creates these objects during translation phase. Architecturally, JSP may be viewed as a high-level abstraction of Java servlets. JSPs are translated into servlets at runtime, therefore JSP is a Servlet; each JSP servlet is cached and re-used until the original JSP is modified. JSP can be used independently or as the view component of a server-side modelâ€“viewâ€“controller design, normally with JavaBeans as the model and Java servlets as the controller.

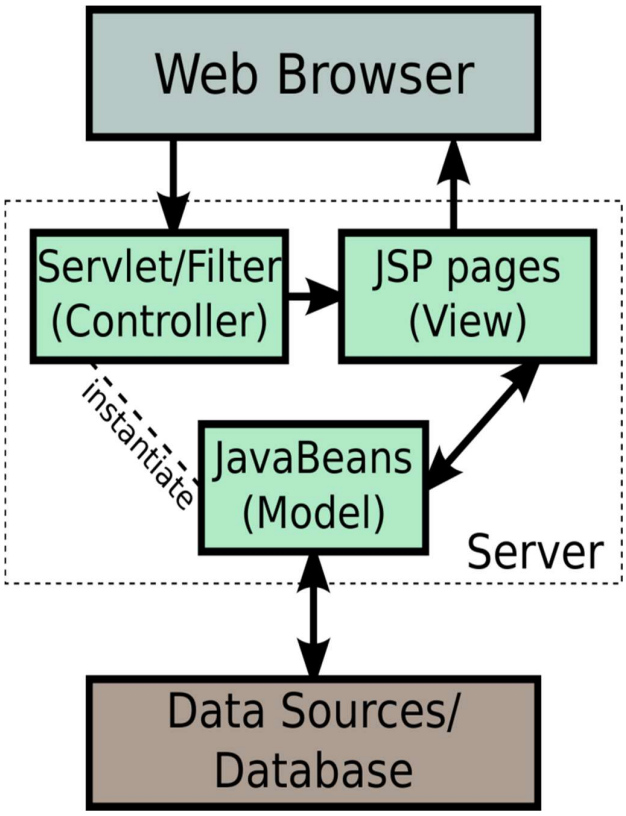


Fig. 2. JSP model 2 architecture

JSP technology is the extension to Servlet technology. The main features of JSP technology are as follows:

- ❖ A language for developing JSP pages, which are text-based documents that describe how to process a request and construct a response
- ❖ An expression language for accessing server-side objects
- ❖ Mechanisms for defining extensions to the JSP language

Servlets provide URL mapping and request handling capabilities in your Java web applications. Request handling is the bread and butter of Java web application development. In order to respond to requests from the network, a Java web application must first determine what code will respond to the request URL, then marshal a response. Every technology stack has a way of accomplishing request-response handling. In Java, we use servlets (and the Java Servlet API) for this purpose. Think of a servlet as a tiny server whose job is to accept requests and issue responses.

3.3.2 Bootstrap:

Bootstrap is the most popular open-source framework full of useful and common classes to use in any project. It helps to develop responsive and mobile-first websites faster and easier. It is known for its faster and effortless responsive web development assistance, Bootstrap web design methodology utilize HTML and CMS based templates for user interface components like forms, navigations, alerts, buttons, typography in addition to optional JavaScript extensions. Bootstrap is a web framework that focuses on simplifying the development of informative web pages (as opposed to web apps). The primary purpose of adding it to a web project is to apply Bootstrap's choices of color, size, font and layout to that project. As such, the primary factor is whether the developers in charge find those choices to their liking. Once added to a project, Bootstrap provides basic style definitions for all HTML elements. The result is a uniform appearance for prose, tables and form elements across web browsers. In addition, developers can take advantage of CSS classes defined in Bootstrap to further customize the appearance of their contents. For example, Bootstrap has provisioned for light-and dark-colored tables, page headings, more prominent pull quotes, and text with a highlight. Bootstrap also comes with several JavaScript components in the form of jQuery plugins. They provide additional user interface elements such as dialog boxes, tooltips, and carousels. Each Bootstrap

component consists of an HTML structure, CSS declarations, and in some cases accompanying JavaScript code. They also extend the functionality of some existing interface elements, including for example an auto-complete function for input fields. The most prominent components of Bootstrap are its layout components, as they affect an entire web page. The basic layout component is called "Container", as every other element in the page is placed in it. Developers can choose between a fixed-width container and a fluid-width container. While the latter always fills the width of the web page, the former uses one of the four predefined fixed widths, depending on the size of the screen showing the page:

Smaller than 576 pixels
576–768 pixels
768–992 pixels
992–1200 pixels Larger than
1200 pixels

Once a container is in place, other Bootstrap layout components implement a CSS Flexbox layout through defining rows and columns. A precompiled version of Bootstrap is available in the form of one CSS file and three JavaScript files that can be readily added to any project. The raw form of Bootstrap, however, enables developers to implement further customization and size optimizations. This raw form is modular, meaning that the developer can remove unneeded components, apply a theme and modify the uncompelled Sass files.

3.3.3 JavaScript

JavaScript is one of the core technologies of the WWW (World Wide Web). It enables interactive web pages and is an essential part of web applications. It has application programming interfaces (APIs) for working with text, dates, regular expressions, standard data structures, and the Document Object Model (DOM). Almost all the websites and web browser uses JavaScript engines to execute client side page behavior. JavaScript engines were originally used only in web browsers, but they are now embedded in some servers, usually via Node.js. They are also embedded in a variety of applications created with frameworks such as Electron and Cordova.

3.3.4 Spring MVC

Spring's Web MVC framework is designed around a `DispatcherServlet` that dispatches requests to handlers, with configurable handler mappings, view resolution, locale and theme resolution as well as support for upload files. The default handler is a very simple `Controller` interface, just offering a `ModelAndView handleRequest(request, response)` method. This can already be used for application controllers, but you will prefer the included implementation hierarchy, consisting of, for example `AbstractController`, `AbstractCommandController` and `SimpleFormController`. Application controllers will typically be subclasses of those. Note that you can choose an appropriate base class: if you don't have a form, you don't need a form controller. This is a major difference to Struts. Spring Web MVC allows you to use any object as a command or form object - there is no need to implement a framework-specific interface or base class. Spring's data binding is highly flexible: for example, it treats type mismatches as validation errors that can be evaluated by the application, not as system errors. All this means that you don't need to duplicate your business objects' properties as simple, untyped strings in your form objects just to be able to handle invalid submissions, or to convert the Strings properly. Instead, it is often preferable to bind directly to your business objects. This is another major difference to Struts which is built around required base classes such as `Action` and `ActionForm`.

Compared to WebWork, Spring has more differentiated object roles. It supports the notion of a `Controller`, an optional command or form object, and a model that gets passed to the view. The model will normally include the command or form object but also arbitrary reference data; instead, a WebWork `Action` combines all those roles into one single object. WebWork does allow you to use existing business objects as part of your form, but only by making them bean properties of the respective `Action` class. Finally, the same `Action` instance that handles the request is used for evaluation and form population in the view. Thus, reference data needs to be modeled as bean properties of the `Action` too. These are (arguably) too many roles for one object.

Spring's view resolution is extremely flexible. A `Controller` implementation can even write a view directly to the response (by returning `null` for the `ModelAndView`). In the normal case, a `ModelAndView` instance consists of a view name and a model `Map`, which contains bean names and corresponding objects (like a command or form, containing reference data). View name resolution is highly configurable, either via bean names, via a properties file, or via your own `ViewResolver` implementation. The fact that the model (the M in MVC) is based on the `Map` interface allows for the complete abstraction of the view technology. Any renderer can be integrated directly, whether JSP, Velocity, or any other rendering technology. The model `Map` is simply transformed into an appropriate format, such as JSP request attributes or a Velocity template model.

CHAPTER 4
SYSTEM ARCHITECTURE

4.1 K-MEANS CLUSTERING:

K-means clustering is most widely used clustering algorithm which is used in many areas such as information retrieval, computer vision and pattern recognition. K-means clustering assigns n data points into k clusters so that similar data points can be grouped together. It is an iterative method which assigns each point to the cluster whose centroid is the nearest. Then it again calculates the centroid of these groups by taking its average. The algorithm 1 shows the basic approach of K-means clustering.

1: An initial clustering is created by choosing k random centroids from the dataset.

2: For each data point, calculate the distance from all centroids, and assign its membership to the nearest centroid.

3: Recalculate the new cluster centroids by the average of all data points that areassigned to the clusters.

4: Repeat step 2 until convergence.

Algorithm 1 K-Means Clustering:

The working of Algorithm 1 can be explained clearly with the help of an example, which is shown on Figure 2.

Figure 2 shows the graphical representation for working of K-means algorithm. In the first step there are two sets of objects. Then the centroids of both sets are determined. According to the centroid again the clusters are formed which gave the different clusters of dataset. This process repeats until the best clusters are achieved. There are abundant tools available for data mining. Some of them are Rapid Miner, R, Knime, Own Code, Weka Or Pentaho, Statistica, Sas Or Sas Enterprise Miner, Orange, Tanagra, And Matlab.

A Review ON K-means DATA Clustering APPROACH

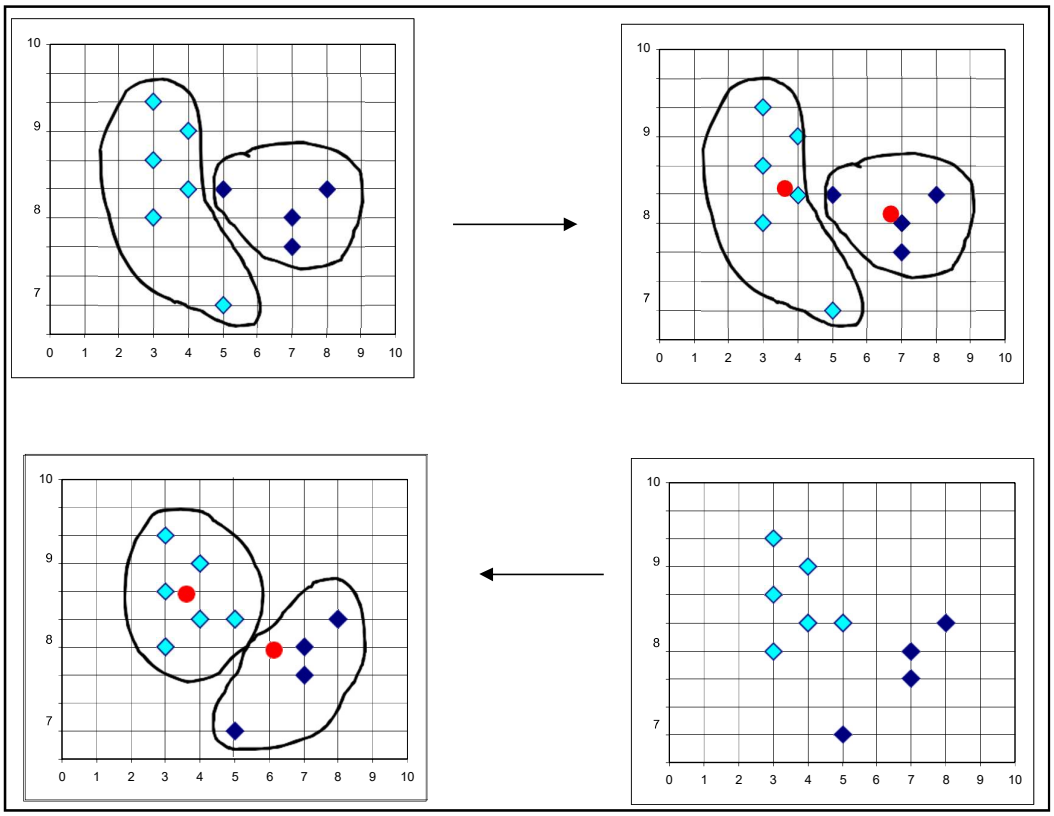


Fig 3 Working of K-means clustering algorithm

4.2 LIMITATIONS:

K-means clustering has some of the limitations which need to get overcome. Several people got multiple limitations while working on their research with K-means algorithm. Some of the common limitations are discussed below.

4.2.1 Outlier:

It has been observed by several researchers that, when the data contains outliers there will be a variation in the result that means no stable result from different executions on the same data. Outliers are such objects they present in dataset but do not result in the clusters formed. Outliers can also increase the sum of squared error within clusters. Hence it is very important to remove outliers from the dataset. Outliers

can be removed by applying preprocessing techniques on original dataset.

4.2.2 Number of Clusters:

Determining the number of clusters in advance is always been a challenging task for K-means clustering approach. It is beneficial to determine the correct number of clusters in the beginning. It has been observed that sometimes the number of clusters are assigned according to the number of classes present in the dataset. Still it is an issue that on what basis the number of clusters should be assigned [49].

4.2.3 Empty Clusters:

If no points are allocated to a cluster during the assignment step, then the empty clusters occur. It was an earlier problem with the traditional K-means clustering algorithm.

4.2.4 Non-globular Shapes and Sizes:

With the K-means clustering algorithm if the clusters are of different size, different densities and non-globular shapes, then the results are not optimal. There is always an issue with the convex shapes of clusters formed.

4.3 Applications:

There are diverse applications of clustering techniques in the fields of finance, health care, telecommunication, scientific, World Wide Web, etc. Some of the applications are discussed below.

4.3.1 Clustering Algorithm in Identifying Cancerous Data:

Clustering algorithm can be used in identifying the cancerous data record within a dataset. Different people tried on this application by assigning labels to known samples of datasets as cancerous and non-cancerous. Then randomly the data samples are mixed together and different clustering algorithms were applied. The result of clustering has been analyzed to know the correctly clustered samples. Accuracy of clustering is calculated easily as the labels of samples were known initially.

4.3.2 Clustering Algorithm in Search Engines:

Clustering algorithm plays an important role in the functioning of search engines. Hence it will act as a backbone to search engines. Search engines try to group similar kind of objects into one cluster and dissimilar objects into other. The performance of the search engines depends on the working of the clustering techniques. The chances of getting the required results on the front page are more if the clustering technique is better.

4.3.3 Clustering Algorithm in Academics:

Students' academic progress monitoring has been a vital issue for academic society of higher learning. With clustering technique this issue can be managed easily. Based on the scores obtained by the students they are grouped into different clusters, where each cluster shows the different level of performance. By calculating the number of students' in each cluster we can determine the average performance of a class all together…

4.3.4 Clustering Algorithm in Wireless Sensor Network Based Application:

Clustering Algorithm can be used efficiently in Wireless Sensor Network's based application. It can be used in landmine detection. Clustering algorithm plays a role of finding the cluster heads which collects all the data in its respective cluster.

CHAPTER 5
IMPLIMENTATION AND RESULT

5.1 IMPLEMENTATION

5.1.1 Database Connectivity to JAVA Application:

To connect Java application with the MySQL database, we need to follow 5 following steps.

- Driver class: The driver class for the MySQL database is "*com.mysql.jdbc.Driver"*.
- Connection URL: The connection URL for the MySQL database is "*jdbc:mysql://localhost:8080/jobsdb?user=root&&password=root"* where jdbc is the API, MySQL is the database, localhost is the server name on which MySQL is running, we may also use IP address, 8080 is the port number and *jobsdb* is the database name.
- Username: The default username for the MySQL database is *root*.
- Password: It is the password given by the user at the time of installing the MySQL database. In project, we are going to use root as the password.

To connect to MySQL from Java, we have used the JDBC driver from MySQL. The MySQL JDBC driver is called MySQL Connector/J. JDBC provides an abstraction layer between Java applications and database servers, so that an application's code does not need to be altered in order for it to communicate with multiple database formats. Rather than connecting to the database directly, the applications send requests to the JDBC API, which in turn communicates with the specified database through a driver that converts the API calls into the proper dialect for the database to understand.

5.1.2 Web Scrapping:

Web scraping refers to the process of extracting a significant amount of information from a website using scripts or programs. Such scripts or programs allow one to extract data from a website, store it and present it as designed by the creator.

We will extract the information which was entered by the user from webpage to create user's own database for handling the user's data and to suggest jobs. The time taken to extract information from a particular source is significantly reduced as compared to manually copying and pasting the data if we use web scraping.

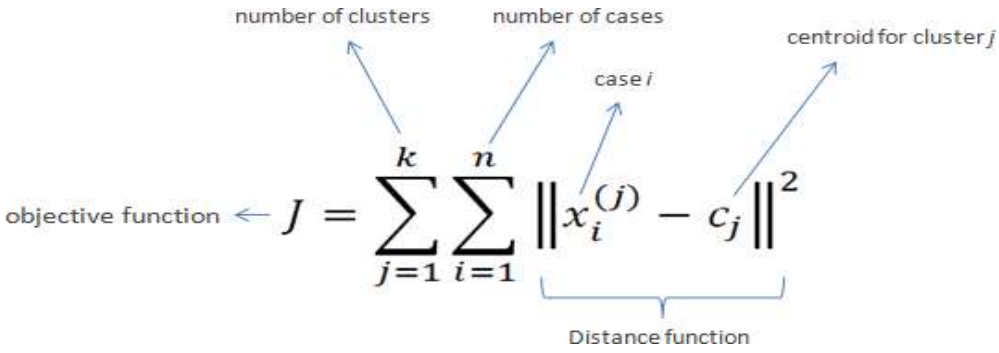
5.1.3 User's Data / Job's Comparison:

After successful creation of user database (After registration) or after user login, user's data gets compared with jobs by means of clustering (K-means Algorithms gets triggered).

The Application compares user's data like percentage of SSC, HSC, Graduation, etc. With the requirements of the company (jobs profile). The application will compare the values and list out the jobs which have the value near to the value of a current user's data. And the matched jobs get displayed to the user interface.

5.1.4 K-Means Clustering Algorithm for Segmentation:

K-Means clustering algorithm is an unsupervised algorithm and it is used to segment the interest area from the background. It clusters, or partitions the given data into K-clusters or parts based on the K-centroids. The algorithm is used when you have unlabeled data (i.e. data without defined categories or groups). The goal is to find certain groups based on some kind of similarity in the data with the number of groups represented by K. The objective of K-Means clustering is to minimize the sum of squared distances between all points and the cluster center.



- **Steps in K-Means algorithm:**

1. Choose the number of clusters K.
2. Select at random K points, the centroids (not necessarily from your dataset).
3. Assign each data point to the closest centroid → that forms K clusters.
4. Compute and place the new centroid of each cluster.
5. Reassign each data point to the new closest centroid. If any reassignment took place, go to step 4, otherwise, the model is ready.

- **Key Features of k-means Clustering Algorithm:**

1. It is very smooth in terms of interpretation and resolution.
2. For a large number of variables present in the dataset, K-means operates quicker than Hierarchical clustering.
3. While re-determining the cluster center, an instance can modify the cluster.
4. K-means reforms compact clusters.
5. It can work on unlabeled numerical data.

- **Limitations with K-means:**

1. Sometimes, it is quite tough to forecast the number of clusters, or the value of k.
2. The output is highly influenced by original input, for example, the number of clusters.
3. An array of data substantially hits the concluding outcomes.
4. In some cases, clusters show complex spatial views, then executing clustering is not a good choice.
5. Also, rescaling is sometimes conscious, it can't be done by normalization or standardization of data points, the output gets changed entirely

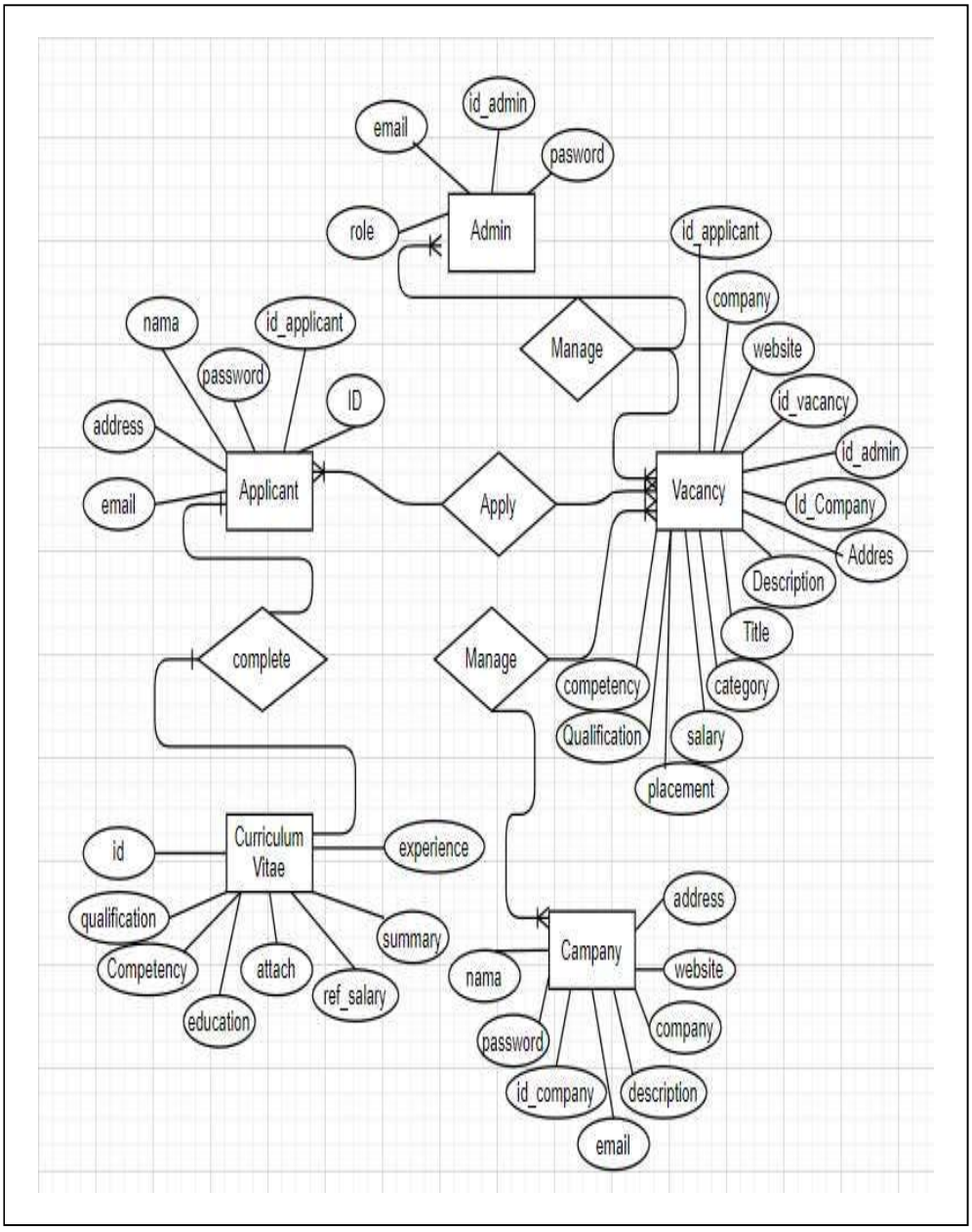
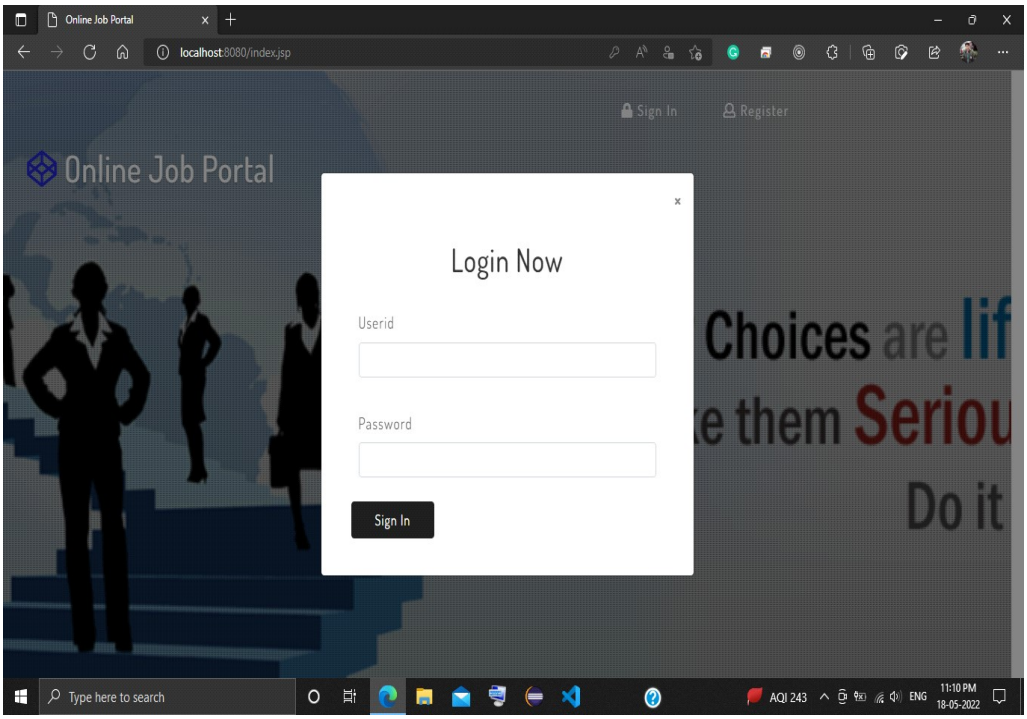


Fig. 4. Entity Relationship Diagram

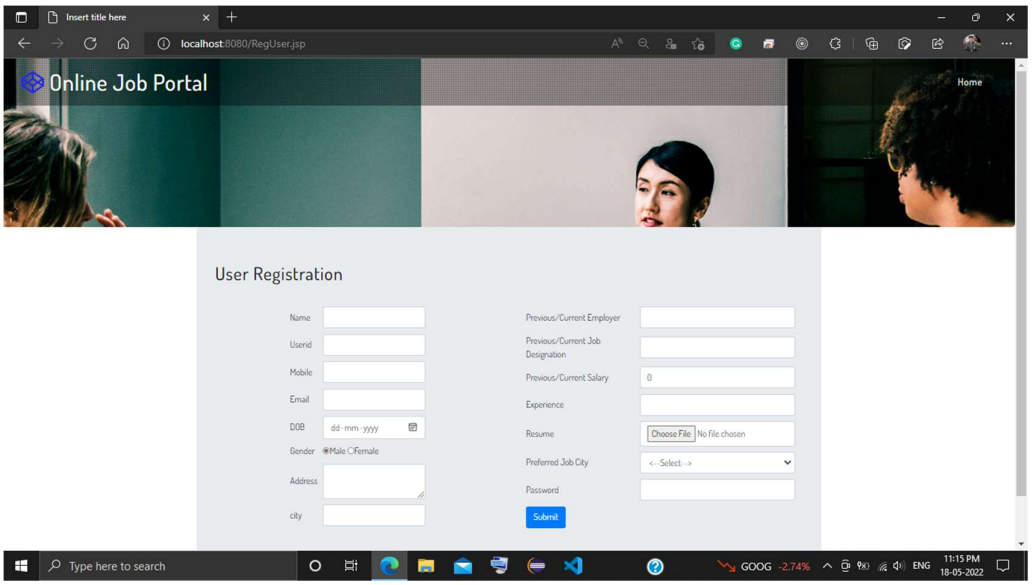
5.2 RESULT

5.2.1 Clustering of Jobs:

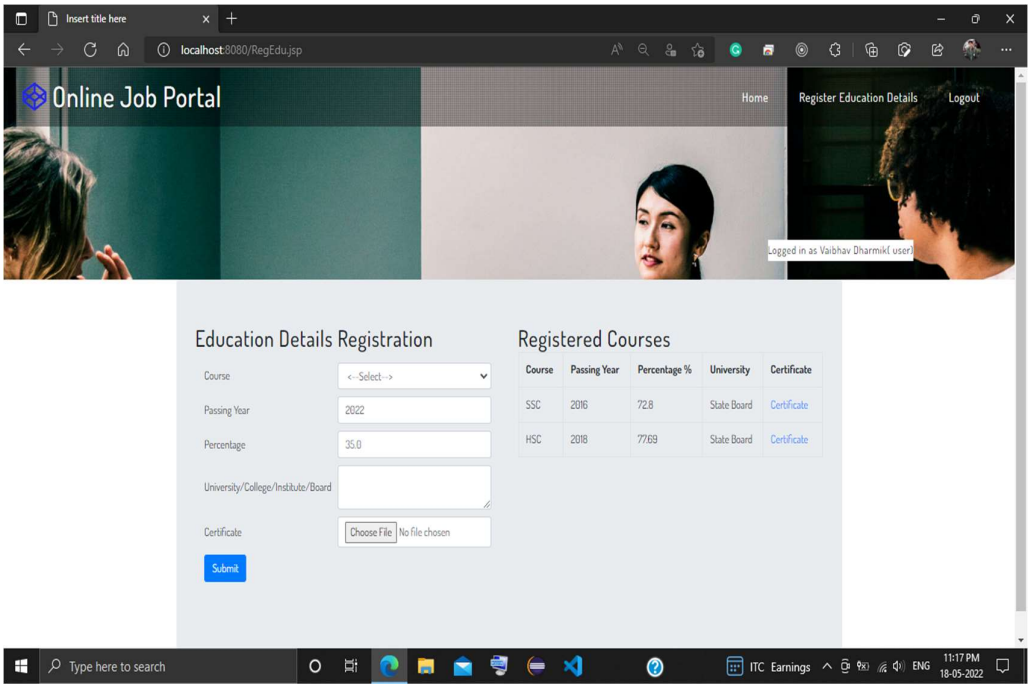
Screenshot 1: Login (User and Admin)



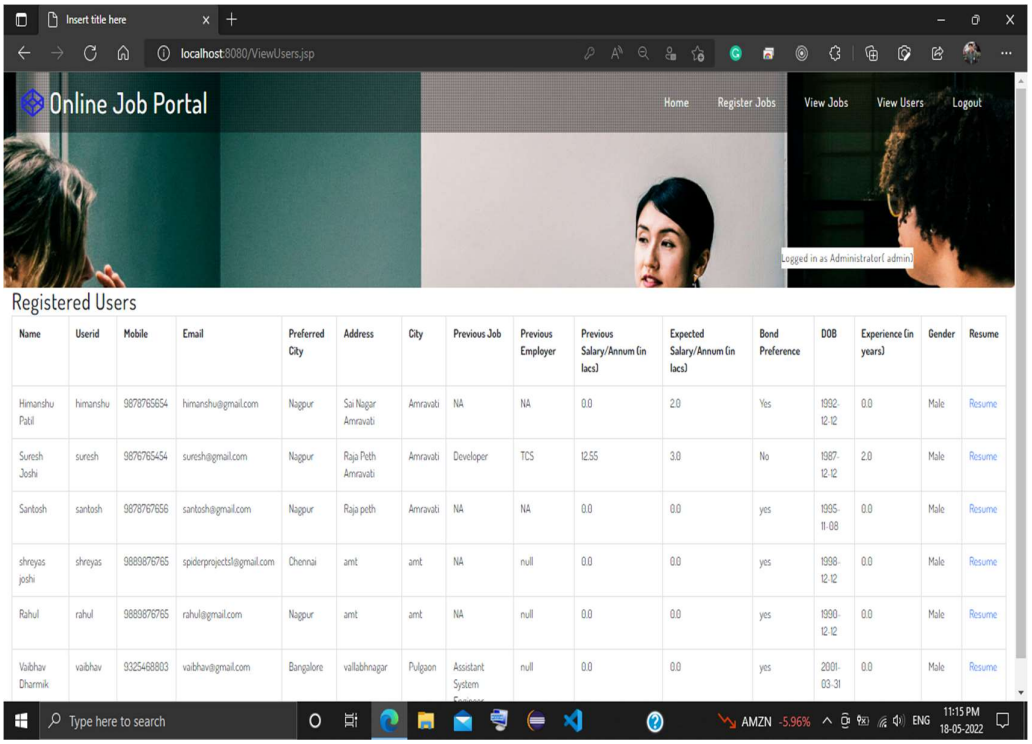
Screenshot 2: User Registration



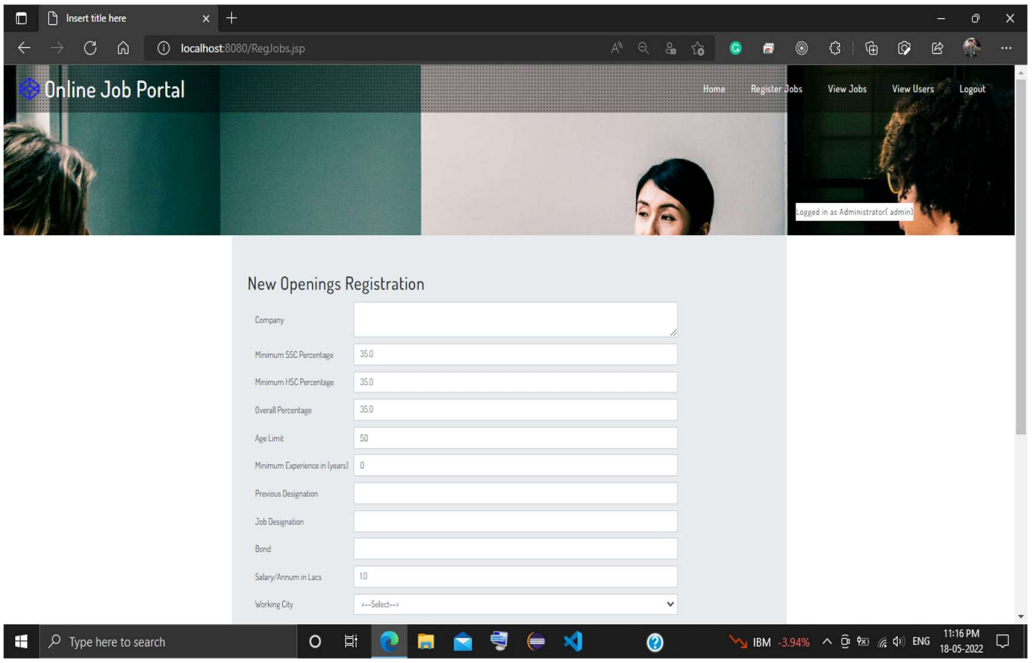
Screenshot 3: User Educational Details



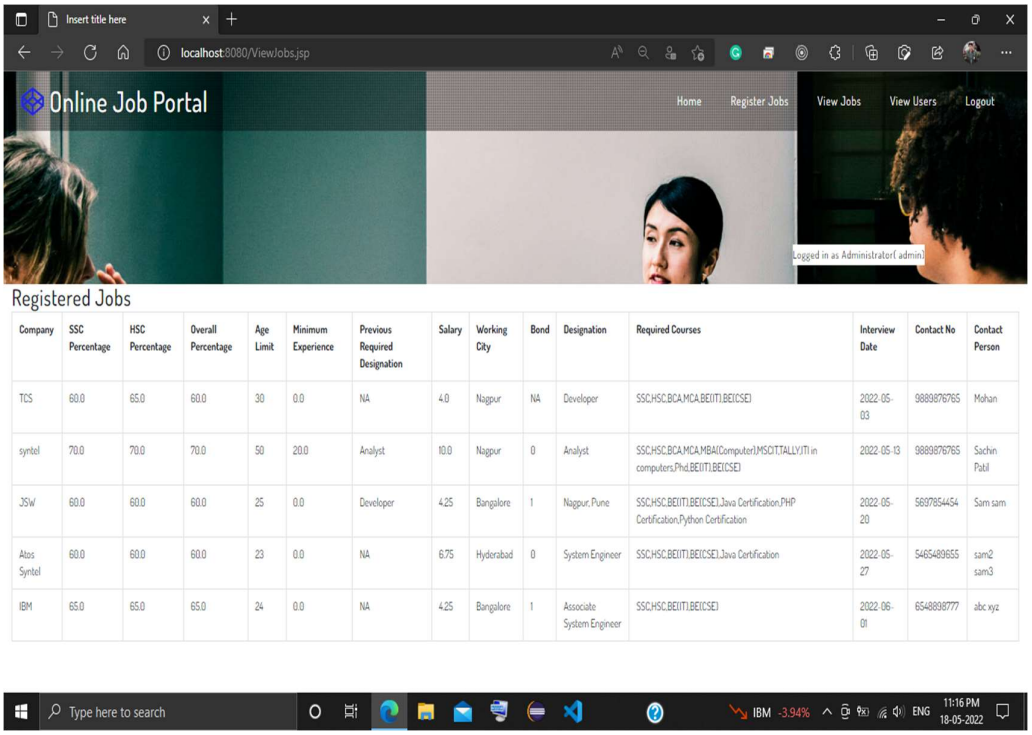
Screenshot 4: Admin Registered Users



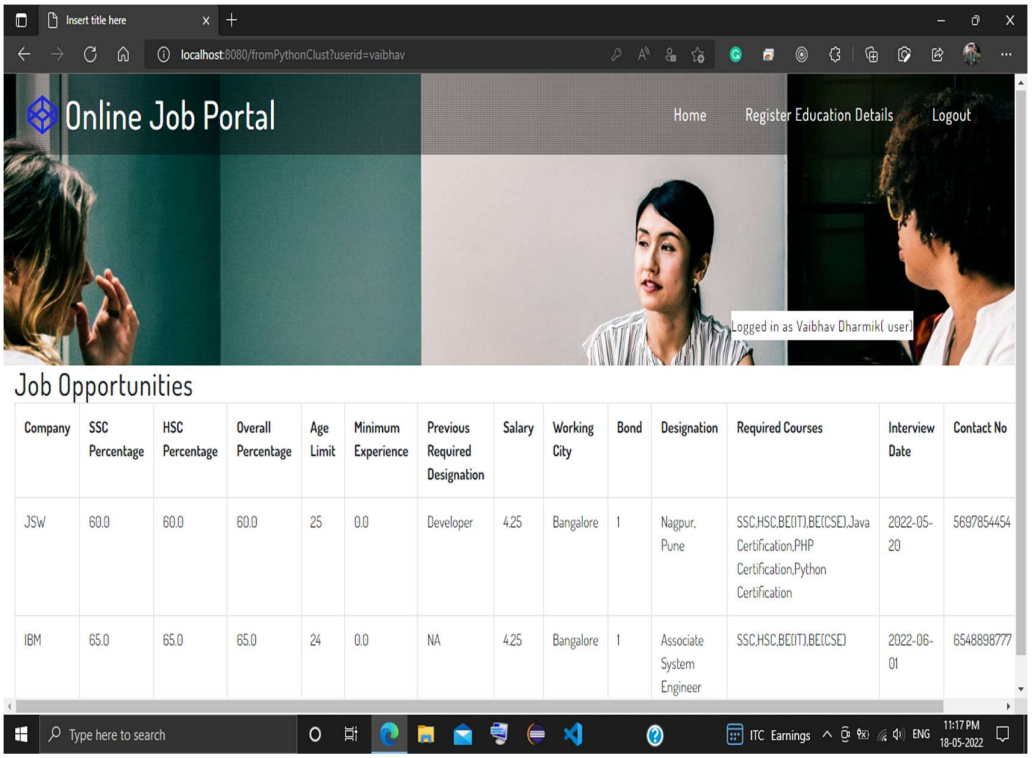
Screenshot 5: Admin Jobs Registration



Screenshot 6: Admin Registered Jobs



Screenshot 7: Recommended Jobs to User



CHAPTER 6
CONCLUSION AND FUTURE SCOPE

6.1 CONCLUSION

In this Project, we presented a job recommender model aiming to extract meaningful data from job postings using data-clustering methods. As a result, job offers are divided into job clusters based on their common features and job offers are matched to job seekers according to their interactions.

6.1 FUTURE SCOPE

Our future Work will focus on training and evaluating our model using Word2vec method and k -means clustering algorithms used to capture and represent the context of job profiles. Subsequently, it will be easy to match set of job offers to a given job seeker based on its past interactions toward specific job offers. The dataset that will be used is built from scraping job search websites.

REFERENCES

[1] A. Kumar, R. Sinha, V. Bhattacherjee, D. S. Verma, S. Singh, "modeling using K-means clustering algorithm", *IEEE 2012, 1st international conference on recent advances in information technology(RAIT).*

[2] A. K. Jain, prof. S. Maheshwari, "survey of recent clustering techniques in data mining", *international journal of computer science and management research*, vol 1 issue 1 Aug 2012.

[3] A. Saurabh, A. Naik, "Wireless sensor network based adaptive landmine detection algorithm, " *2011 3rd International Conference on Electronics Computer Technology (ICECT),* vol.1, no., pp.220, 224, 8-10 April 2011.

[4] B. Biggio, I. Pillai, S. R. Bulò, D. Ariu, M. Pelillo, F. Roli," is data clustering in adversial settings secure?", *Proceedings of the 2013 ACM workshop on Artificial intelligence and security*, Pages 87-98.

[5] B. K. Mishra, N. R. Nayak, A. Rath, S. Swain, "far efficient K-means clustering algorithm", *Proceedings of the 2012 ACM's International Conference on Advances in Computing, Communications and Informatics,* Pages 106-110.

[6] B. K Tripathy, A. Ghosh, G.K. Panda, "Kernel based K-means clustering using rough set, " *2012 International Conference on Computer Communication and Informatics (ICCCI),* vol., no., pp.1, 5, 10-12 Jan. 2012.

[7] D. Jimenez, E. Ferretti, V. Vidal, P. Rosso, and C. F. Enguix, "The influence of semantics in IR using LSI and K-means clustering techniques", *ACM Proceedings of the 1st international symposium on Information and communication technologies*, Pages 279-284.

[8] D. Napoleon, P.G. Lakshmi, "An efficient K-Means clustering algorithm for reducing time complexity using uniform distribution data points, " *Trendz in Information Sciences & Computing (TISC), IEEE 2010*, vol., no., pp.42, 45, 17-19 Dec. 2010.

[9] D. Sisodia, L. Singh, S. Sisodia, K. Saxena, "Clustering Techniques: A Brief Survey of Different Clustering Algorithms", *International Journal*

of Latest Trends in Engineering and Technology (IJLTET), Vol. 1 Issue 3 September 2012.

[10] D. S. Modha, W. S. Spangler*, "Feature Weighting in k-Means Clustering*", *ACM Journal of Machine Learning*, Volume 52 Issue 3, September 2003, Pages 217 – 237.

[11] D. T. Pham, S. S. Dimov, and C. D. Nguyen, "Selection of K in K-means clustering", *Proc. IMechE Vol. 219 Part C: J. Mechanical Engineering Science, IMechE 2005.*

[12] E. A. Khadem, E. F. Nezhad, M. Sharifi, "Data Mining: Methods & Utilities",
*Researcher*2013; 5(12):47-59. (ISSN: 1553-9865).

[13] E. Casper, C. C. Hung, E. Jung, and M. Yang, "**A quantum-modeled K-means clustering algorithm for multi-band image segmentation**", *Proceedings of the 2012 ACM Research in Applied Computation*, Pages 158- 163.

[14] E. Kijsipongse, S. U-ruekolan, "Dynamic load balancing on GPU clusters for large-scale K-Means clustering, " *2012 IEEE International Joint Conference on Computer Science and Software Engineering (JCSSE),* vol., no., pp.346, 350, May 30 2012-June 1 2012.

[15] G. D. Fatta, F. Blasa, S. Cafiero, G. Fortino, "Epidemic K-Means Clustering, " *2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW),* vol., no., pp.151, 158, 11-11 Dec. 2011.

[16] H.K. Yogish, G.T. Raju, "Clustering of Preprocessed Web Usage Data Using ART1 Neural Network and Comparative Analysis of ART1, K-Means and SOM Clustering Techniques, " *2013 5th IEEE International Conference on Computational Intelligence and Communication Networks (CICN),* vol., no., pp.322, 326, 27-29 Sept. 2013.

[17] H.T Dashti, T Simas, R.A Ribeiro, A Assadi, A Moitinho, "MK-means-Modified K-means clustering algorithm, " *The 2010 IEEE International Joint Conference on Neural Networks (IJCNN),* vol., no., pp.1, 6, 18-23 July 2010.

[18] H. Xiong; J. Wu; J. Chen, "K-Means Clustering Versus Validation Measures: A Data-Distribution Perspective, " *IEEE Transactions on*

Systems, Man, and Cybernetics, Part B: Cybernetics, vol.39, no.2, pp.318, 331, April 2009.

[19] H. Xiuchang, SU Wei, "An Improved K-means Clustering Algorithm", *JOURNAL OF NETWORKS*, VOL. 9, NO. 1, JANUARY 2014.

[20] J. A. Silva, E. R. Faria, R. C. Barros, E. R. Hruschka, A. C. P. L. F. d. Carvalho, J. Gama, "Data stream clustering: A survey", *ACM Computing Surveys (CSUR),* Volume 46 Issue 1, October 2013, Article No. 13.

[21] J. Feng; Z. Lu; P. Yang; X. Xu, "A K-means clustering algorithm based on the maximum triangle rule, " *2012 IEEE International Conference on Mechatronics and Automation (ICMA),* vol., no., pp.1456, 1461, 5-8 Aug. 2012.

[22] J. Wang; J. Wang; Q. Ke; G. Zeng; S. Li, "Fast approximate k-means via cluster closures, " *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* vol., no., pp.3037, 3044, 16-21 June 2012.

[23] J. Wang; X. Su, "An improved K-Means clustering algorithm, " *2011 IEEE 3rd International Conference on Communication Software and Networks (ICCSN),* vol., no., pp.44, 46, 27-29 May 2011.

[24] J. Xue, X. Liu, "A K-nearest Based Clustering Algorithm by P Systems with Active Membranes*", Journal of Software*, Vol 9, No 3 (2014), 716-725, Mar 2014.

[25] J. Zhu; H. Wang, "An improved K-means clustering algorithm, " *2010 The 2nd IEEE International Conference on Information Management and Engineering (ICIME),* vol., no., pp.190, 192, 16-18 April 2010.

[26] J. MacQueen, "Some methods for classification and analysis of multivariate observations." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* Volume 1: Statistics, 281--297, University of California Press, Berkeley, Calif., 1967.

[27] K. Alsabti, S. Ranka, V. Singh, "An efficient k-means clustering algorithm", *1998 Proceedings of IPPS/SPDP Workshop on High Performance Data Mining.(date location).*

[28] K. Akkaya, F. Senel, B. McLaughlan, "Clustering of Wireless Sensor and Actor Networks based on Sensor Distribution and Inter-actor

Connectivity" *ACM Journal of Parallel and Distributed Computing,* volume 69, Issue 6, June, 2009, Pages 573-587.

[29] K. Honda, R. Nonoguchi, A. Notsu, H. Ichihashi, "PCA-guided k-Means clustering with incomplete data, " *2011 IEEE International Conference on Fuzzy Systems (FUZZ),* vol., no., pp.1710, 1714, 27-30 June 2011.

[30] K. Hornik, I. Feinerer, M. Kober, C. Buchta, "Spherical k-Means Clustering",
Journal of Statistical Software, September 2012, Volume 50, Issue 10.

[31] K. Ichikawa, S. Morishita, "A simple but powerful heuristic method for accelerating k-means clustering of large-scale data in life science, " *2013 IEEE Transactions on Computational Biology and Bioinformatics,* vol.PP, no.99, pp.1, 1.

[32] Y. Cheng, Y. Xie, K. Zhang, A. Agrawal, A. Choudhary, "CluChunk: clustering large scale user-generated content incorporating chunklet information", *ACM Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems*, Programming Models and Applications, Pages 12-19.

[33] Y.K. Lam, P. W. M. Tsang, C.S. Leung, "PSO-based K-Means clustering with enhanced cluster matching for gene expression data", *Neural Computing and Applications*, June 2013, Volume 22, Issue 7-8, pp 1349-1355.

[34] Y.M. Cheung, "k*-Means: A new generalized *k*-means clustering algorithm", *Pattern Recognition Letters, Volume 24, Issue 15, November 2003, Pages 2883-2893.*

[35] Y. Sun; G. Liu; K. Xu, "A k-Means-Based Projected Clustering Algorithm, " *2010 Third International Joint Conference on Computational Science and Optimization (CSO),* vol.1, no., pp.466, 470, 28-31 May 2010.

[36] Y. S. Patil, M.B. Vaidya, "A Technical Survey on cluster analysis in data mining", *International Journal of Emerging Technology and Advanced Engineering*, ISSN 2250-2459, Volume 2, Issue 9, September 2012.