

Model of Recommendation System to Enhance the Employability: Machine Learning Approach

Ms. Priyanka Shah*

S. K. Patel Institute of Management
and Computer Studies, Kadi Sarva
Vishwavidyalaya University,
Gandhinagar, Gujarat, India
iampriyashah@gmail.com

Dr. Paresh V. Virpariya

Department of Computer Science,
Sardar Patel University, Vallabh
Vidyanagar, Gujarat, India
pvvirpariya@yahoo.com

Dr. Hardik B. Pandit

Faculty of Computer Science and
Applications, Charotar University of
Science & Technology, Changa,
Gujarat, India
hardik00@gmail.com

ABSTRACT

In the recent era, various types of data are available on the Web in various forms such as text, images, audio, and video. Extracting information of interest from this web content is one of the applications of Web Content Mining. Computer Science is one of the most demanding fields where students can get the best recruitment in the IT (Information Technology) industry through job portals. Still, there is a gap of skills between academia and industry. One of the prominent factors is non-competent syllabus. This Research Paper introduces the Industry Institute Interaction Recommendation System-IIIRS model of a Web Mining technique. It is based on a recommendation system to enhance the employability of students, by bridging the gap between the institution's curriculum and the required skills set at the industry end. The proposed model includes application of data extraction, data pre-processing, classification, and clustering. The model presents the state of the art skills required by the software industry, and gives recommendations about required updates in the syllabus of academic institute. The model serves the purpose of both industry and academia by bridging the gap of skill sets due to incompetent syllabus.

CCS CONCEPTS

• **Computing methodologies** → Machine learning; Machine learning approaches.

KEYWORDS

Classification, Clustering, WebContentMining, WebScraping, RecommendationSystem

ACM Reference Format:

Ms. Priyanka Shah*, Dr. Paresh V. Virpariya, and Dr. Hardik B. Pandit. 2022. Model of Recommendation System to Enhance the Employability: Machine Learning Approach. In *International Conference on Information Management & Machine Intelligence (ICIMMI 2022)*, December 23, 24, 2022, Jaipur, India. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3590837.3590949>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICIMMI 2022, December 23, 24, 2022, Jaipur, India

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9993-7/22/12...\$15.00
<https://doi.org/10.1145/3590837.3590949>

1 INTRODUCTION

After postgraduate, most students are ready for employment in the private or public sector. According to the eighth edition of the India Skills Report (ISR)-2021, only 47.38% in2019, 46.21% in2020 and 45.9% of graduates in 2021are employable [1]. Various factors are responsible for unemployment, like technological changes, lack of skill-based education, lack of communication skills, and outdated syllabus. Academia needs to have updated course content that will be suitable for industry requirements [2].

In the era of digitization, every university must establish sustainability while adhering to the wider platform. The internet is the only source. Through this, one can fetch the data from the academic course outlines. Each university has a different format for outlining the course content. Job Portal is a platform where the industry can post their jobs on the web. Job Portal is an easily accessible, faster, reliable, and accurate process [3].

The World Wide Web consists of various documents that contain text, images, audio, or video data. A recommendation system is an information retrieval system based on a predictive web content mining algorithm. Henceforth, it is possible to suggest preferable products or services. For instance, it applies to which product to purchase or sell, what service to be used, which location to be explored, what career opportunity can be focused on, and others.

Web mining refers to the process of analyzing large datasets to extract useful information available on the web. Web content mining is one of the techniques of web mining, a process of extracting useful information from online sources like pictures, audio, video, text, etc. The Web content mining process does not involve data extraction or collection processes [4]. The process of extracting data from a large dataset to uncover insights by structuring it into a convenient format is achieved by a process called web scraping. As the data is being scrapped from multiple job portals, all the data needs to be processed to identify and filter out into useful data. A dataset of the required attributes of job domains needs to be created for the system. Web scraping is the process of extracting data from web sources and structuring it into a more convenient format. It is important to perform web scraping before data preprocessing and analysis [5]. The main aim of this paper is to introduce the model that gives recommendations to academics for regenerating the curriculum syllabus in the area to meet the requirements of the industry.

2 LITERATURE REVIEW

As a part of literature review, various models have been studied and analyzed.

Chuck Litecky et al. [6] have extracted 2,44,460 job ads from job portals using hierarchical agglomerative clustering, k-means cluster analysis Methodology. Researcher analyzed the data and found 20 job types and clusters of similar skill sets for specific job definitions.

NurulEzzaAsyikin Mohamed Almi et al. [7] tried to find the gap between the industry's requirements and the readiness of graduate students in the software engineering field using Sampling Method. Researcher finds graduate students' skills by taking inter-views of 20 students. From these interview questions, Researcher concludes various reasons like lack of confidence, readiness, knowledge about the field, etc are responsible for this gap.

Deepali V Musale et al. [8] have developed an android application that pro-vides suitable job recommendations to the students by matching their profiles with the company's profile using semantic matching and tree-based knowledge matching. This application also sends notifications of a matched job profile to students by sending SMS.

K.ThamaraiSelvi, R.Ramya [9] researchers have analyzed the attitudes in the direction of job seekers towards employment Using KNIME tools. The researcher studied the essentiality of the e-recruitment process. This paper has analyzed the expectations and satisfactions of job seekers from different job portals. It includes Data Integration, Data Selection, Data Cleaning, Data Transformation, Data Mining, Pattern Evaluation and knowledge, Presentation, Decision.

YogeshBharambe et al. [10] has done classification analysis using different classification algorithms to predict the employability status of the students using Data Mining with intelligent learning algorithm. Evaluation and eligibility tests have been taken to find out the students' skill sets, including strengths and weaknesses. These students' skill sets are mapped with job profiles, suggesting all companies for which students are eligible and can apply using data mining techniques.

Jaichandran R et al. [11] has developed a recommendation system for students that recommend which skills are required by industry using Text Mining and Natural Language Processing. The researcher used a collaborative filtering algorithm to match the ex-act profile from a resume with the required skill set of the industry [11].

3 OBJECTIVE

The main objective of this model is to map the gap between the learning skills including in course content and the required skills of industries. It is the fundamental responsibility of the educational institutions to create such curricula that will be responsive to the requirements of the industry. The following are some objectives that would be assisted by the proposed model:

- Identify current required skills in Computer Science field of IT Industries from job advertisements.
- Discover current minimum required skills in specific job domain of Computer Science field.
- Identify the domain of Information Technology which has maximum requirement in Industry.
- Recognize the learning skills that must be integrated in the academic course outline.

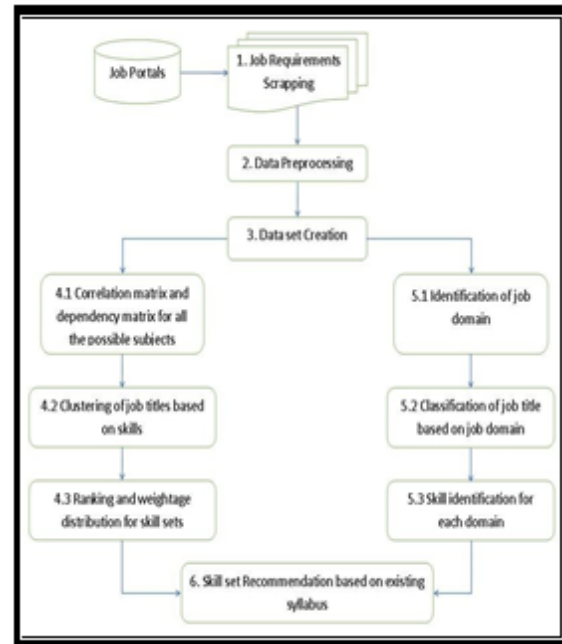


Figure 1: IIIRS Model

4 METHODOLOGY

The proposed logical framework methodology of the IIIRS model is divided into seven stages as shown in figure 1

IIIRS Model shown in figure 1 works in following steps

- Step 4.1: Job Requirement Scrapping
- Step 4.2: Data Preprocessing
- Step 4.3: Data set Creation
- Step 4.4: Create a correlation matrix and a dependency matrix for all the possible subjects
- Step 4.5: clustering of job titles based on skills
- Step 4.6: Ranking and Weightage distribution for skills set
- Step 4.7: Identification of the Job Domain
- Step 4.8: Classification of job titles based on job domains
- Step 4.9: Skill Identification for each domain
- Step 4.10: Skill based Recommendations based on existing system

4.1 JOB REQUIREMENT SCRAPPING

4.1.1 Selecting Job Portals for IIIRS implementation. There are various job portal web- sites like Moster.com, Naukari.com, Internshala.com, etc., that allow employers to post job requirements in different domains. Individual modules have been created for each job portal according to their website structure in Python. Any web-site can have two types of structure: dynamic content delivery and static content delivery. In dynamic content delivery, websites fetch the data through their internal API and dynamically place the data into the HTML component. In dynamic content delivery, the API network request is being monitored and the network request is being spoofed in a Python script to get the data.

4.1.2 Identify the attributes from Job Portals. At this stage, identify and select only those attributes from the job domain which are requisite for the recommendation model. Identify the attributes from Job Portals like Job Title, Salary, Experience, Education, Key Skills, Location, JD, Source, JD_DATE, JB_UNIQID.

4.1.3 Scrape Job Details based on identified attributes. Web scraping is the process of extracting data from web sources and structuring it into a more convenient format. Web scraping needs to be performed before data preprocessing and analysis. At this stage, a scraper code was developed in Python and 46 keywords were defined for IT jobs in Job Portals.

4.2 DATA PREPROCESSING

Data preprocessing involves various steps:

4.2.1 Interpolation. The extracted data may not be cleaned and may contain missing values or duplicate values. So it is necessary to apply different techniques to clean the data and also find out missing values.

4.2.2 Feature Extraction. All the extracted data is not useful for mining. So, select only those attributes that are required for mining. Performed Data Preprocessing on the acquired dataset – created UDF for cleaning datasets by removing missing values, duplicate values, new lines, tab, and from the strings and other anomalies.

4.3 DATASET CREATION

Various websites from the relevant URL addresses have been searched, and the Portable Document Files have been downloaded using a scraping tool. Scraping tools are implemented using Python. The scraped document, which forms the initial data set, is retrieved in an excel format.

4.4 CREATE A CORRELATION MATRIX AND A DEPENDENCY MATRIX FOR ALL THE POSSIBLE SUBJECTS

At this stage, Using Principal Component Analysis (PCA), a correlation matrix will be created that shows which skill is dependent on which subject.

4.5 CLUSTERING OF JOB TITLES BASED ON SKILLS

At this stage, decoding a cluster algorithm depending upon the Eigen value, a cluster of all the job titles will be created where that particular skill occurred.

4.6 RANKING AND WEIGHTAGE DISTRIBUTION FOR SKILLS SETS

At this stage, rankings will be created to show which job skill occurred how many times in all the scraped data and upon which subject it depended.

4.7 IDENTIFICATION OF THE JOB DOMAIN

At this stage, the domain of the subject will be determined weight for the bias function according to the skills required by the industry and the learning outcomes of the subject, thus the related job domain of the subject can be specified.

4.8 CLASSIFICATION OF JOB TITLES BASED ON JOB DOMAINS

At this stage, the job titles will be classified for multiple jobs belonging to the same skill set but having different names using classification algorithm in the relation with ROC (Receiver Operating Characteristic) curve. Those jobs will be classified in the same domain.

4.9 SKILL IDENTIFICATION FOR EACH DOMAIN

On the basis of the job description, the total number of skills required for the job will be identified and it'll be determined that which subject provides the required skills.

4.10 SKILL SET RECOMMENDATION BASED ON EXISTING SYLLABUS

Using Recommendation algorithm, the skill will be searched in the job dataset to see how frequently it occurs in jobs. And the same process will be repeated in a loop for each skill offered by the subjects. The skill will be stored based on their frequency in the job domain. The subject providing the skill, with a high frequency, will be recommended to academic institutes to be added to the syllabus in order to provide students with skill-based education suitable for industry requirements.

5 CONCLUSION

The proposed IIIRS model helps to bridge the gap of skill sets between academia and industry. The application domain for this recommendation model will be an academic platform where students and academicians are always looking to learn about future and current trends of the job market so that they can select the course to integrate into their syllabus. This recommendation model would be helpful together to uplift the employment potential ratio. This recommendation model also enables us to shape the future of the students and proportionally increase the Employment ratio.

REFERENCES

- [1] M. P. Ang, "Evaluating The Responsiveness Of Bs Is Curriculum To Industry Skill Requirements Aileen L. Rillon, Mba, Mim A Research Report Submitted To The Ateneo De Naga University Research Council October 2017," No. October 2017.
- [2] "CareerAndSkillsRecommendationsUsingDataMiningTechnique: Matching Right People For Right Profession, In Pakistani Context," Vfast Trans. Softw. Eng., Vol. 7, No. 1, Pp. 93–101, 2018, Doi: 10.21015/Vtse.V13i3.510.
- [3] M. Pejic-Bach, T. Bertoncel, M. Meško, And Ž. Krstić, "Text Mining Of Industry 4.0 Job Advertisements," Int. J. Inf. Manage., Vol. 50, No. July, Pp. 416–431, 2020, Doi: 10.1016/J.Ijinfomgt.2019.07.014.
- [4] F. Johnson And S. Kumar Gupta, "Web Content Mining Techniques: A Survey," Int. J. Comput. Appl., Vol. 47, No. 11, Pp. 44–50, 2012, Doi: 10.5120/7236-0266.
- [5] N. K. Kaur, "Topic Detection And Tracking Using Web Mining 1," Vol. 02, No. 01, Pp. 5–10, 2015.
- [6] C. Litecky, A. Aken, A. Ahmad, And H. J. Nelson, "Skill Sets Mining For Computing Jobs," IEEE Softw., Pp. 78–85, 2008.

- [7] N. E. A. M. Almi, N. A. Rahman, D. Purusothaman, And S. Sulaiman, "SoftwareEngineering Education: The Gap Between Industry's Requirements And Graduates' Readiness," Isc. 2011 - 2011 IeeeSymp, Comput. Informatics, Pp. 542-547,2011,Doi:10.1109/Isci.2011.5958974.
- [8] D.VMusale,M.K.Nagpure,K.S.Patil,AndR.F.Sayyed,"JobRecommendationSystem Using Profile Matching And Web-Crawling," Int. J. Adv. Sci. Res. Eng. Trends, Vol.1, No.2,Pp.29-34, 2016.
- [9] K. T. Selvi And R. Ramya, "An Impact On Electronic-Recruitment And Its Perception Towards Job Portal Function Through Search Engines Among Job Seekers UsingKnimeData MiningTool," No. 1,Pp.10-18,2016.
- [10] Y. Bharambe, N. More, M. Mulchandani, R. Shankarmani, And S. G. Shinde, "AssessingEmployabilityOfStudentsUsingDataMiningTechniques,"2017Int.Conf.Adv. Comput. Commun. Informatics, Icacci 2017, Vol. 2017-January, Pp. 2110-2114,2017, Doi:10.1109/Icacci.2017.8126157.
- [11] R. Jaichandran, S. Leelavathy, K.S. Raja, P. Kumar, S.K. Majhi, And V.S. Thakur, "Job Portal Resume Evaluation System Using Text Mining And Natural Language Processing,"Eur.J. Mol.Clin.Med.,Vol.7,No.4,Pp.2800-2807,2020.