Literature survey LSM

Team Mendit: An ML based employee and employer mediating web portal.

Name	Register number
Manan Sharma	18BCI0132
Rakshith Sachdev	18BCI0109
Ayush Rana	18BCE2305
Hrithik Ahuja	18BCE2154
Rohan Allen	18BCI0247

Jyoti V Gautam et al. proposed that scheduling problem has been an active area of research in computing systems since their inception. The Apache Hadoop framework has emerged as most widely adopted framework for distributed data processing because of open source and allowing use of commodity hardware. Job scheduling has become an important factor to achieve high performance in Hadoop cluster. Several scheduling algorithms have been developed for Hadoop-MapReduce model which vary widely in design and behaviour, handling different issues such as locality of data, user share fairness and resource awareness. This paper highlights fundamental issues in job scheduling, presents classification of Hadoop schedulers, and discusses presented survey of existing scheduling algorithm. Moreover, paper also discusses features, advantages, and limitations of the scheduling algorithms. This paper also discusses about how various resource monitoring tools or frameworks help in achieving better result from MapReduce. It also discusses customized MapReduce frameworks used for improving the performance. This paper would be useful to beginners and researchers for understanding the state-of-the-art on scheduling in Big data processing [1].

Dorothea Calmels proposed that the job sequencing and tool switching problem is a combinatorial optimisation problem that appears in various industries, mainly in the manufacturing sector. Although tool

switching is only part of a much larger decision-making process in manufacturing systems, it has a major impact on the overall performance of the system by affecting the total set-up time as well as the machine and tool utilisation. Over the past few decades, various approaches have been applied to tool switching problems. This study provides a comprehensive review of the literature on the job sequencing and tool switching problem. Studies are identified and compared through a structured literature review and classified according to a new framework which allows the identification of problem perspectives and solution methods. The results show that current research does not fully exploit real-world situations and studies are often limited to a single-machine and uniform tool size and set-up time. Finally, this literature review summarises the current research results and provides directions for future research [2].

S. Clyde, Jianping Zhang and Chih-Chung Yao proposed in their article that it is challenging to automate job classification due to the many dynamic classes and features, concept drip, uncertainty, and noisy data. This paper presents a solution to this problem by combining an incremental learning subsystem with a job classifier. The authors describe our design and implementation using a complete object-oriented approach known as Object Oriented Systems Modelling (OSM) that uses a smooth mapping to most object-oriented programming languages. The software being written in C++ is performing better than other learning/classification algorithms in experimentation. The production version is performing very accurately. [3]

F. M. Javed Mehedi Shamrat et al. proposed that jobseekers tend to register themselves in many job portals and find various job circulars through different sources. Since all job circulars are posted, it is easy to find one that interests them. In addition, these job portals collect information on those who are applying for jobs through these portals. When a new job advertisement is posted on the portal, job seekers receive an email notification. Using a decision tree algorithm, email notification becomes more relevant and personalized for job-seekers.

This decision tree is used to analyse job seekers' fields of interest, expected salaries, and experience to make decisions on which job circular should be advertised to which job seekers. It makes the recruitment process more engaging for jobseekers and helps them relate to the notification. [4]

L. Xu et al. proposed that there has been substantial attention paid to resource prediction in cloud computing jobs (e.g., CPU/memory usage). Existing models predict jobs utilizing regression techniques based on job historical data under the impractical assumption that the new jobs will be of the same class as the historical ones. This problem can be overcome by taking into consideration the job category for effective resource prediction. A number of existing works on job classification ignore or use a very naive idea of resource consumption variability, leading to unsatisfactory classification accuracy and slow running time. In this paper, the authors introduce a new and efficient job classification approach, called Bejo. Taking inspiration from the textual classification methods which classify documents according to distribution of the textual words, Bejo treats each job as a document. This distribution of "resource words" describes and categorizes a job based on how each snapshot data is assigned to a certain resource word during collection. In order to address the unique challenges of high noise and tight time budget of cloud job classification, an E1 norm minimization formulation is used to assign each resource snapshot to a resource word. A comprehensive dataset is collected for jobs classification and resource consumption prediction on cloud platforms, and Bejo outperforms state-of-the-art algorithms in quality and efficiency. Experimental work also indicates that by adding an additional class of job to the existing regression method, the relative error on resource consumption can be dramatically reduced. [5]

Habiba, Sultana Umme et al. proposed that lately, because of headway in present day innovation and social correspondence, publicizing new position posts has become basic issue in the current world. In this way, counterfeit work posting forecast task will be an extraordinary worry for all. Like numerous other grouping undertakings, counterfeit

occupation presenting expectation leaves a part of difficulties to confront. This paper proposed to utilize distinctive information mining strategies and grouping calculation like KNN, choice tree, uphold vector machine, innocent bayes classifier, arbitrary woodland classifier, multilayer perceptron and profound neural organization to anticipate a task post on the off chance that it is genuine or deceitful. We have tested on Employment Scam Aegean Dataset (EMSCAD) containing 18000 examples. Profound neural organization as a classifier, performs incredible for this grouping task. We have utilized three thick layers for this profound neural organization classifier. The prepared classifier shows around 98% grouping exactness (DNN) to foresee a deceitful occupation post.[6]

S. Deshmukh et al. proposed that Hadoop MapReduce is perhaps the most mainstream freely accessible systems for enormous information preparing in the cloud climate. Hadoop accommodates the requirements of a wide assortment of potential clients yet doesn't give a way to upgrade the scheduler for singular clients. This is the thing that we have endeavoured to do. The MapReduce planning framework comprises of a single Job Tracker per bunch and various Task Trackers. One of the essential obligations of the work tracker is to plan all the client occupations which makes upgrading the planning by the Job Tracker an intriguing issue. For an improved booking structure, we have actualized booking which likewise takes into account the real asset necessities of the work, as restricted to depending totally on number of free Map and Lessen openings. In this paper, an endeavor is made to make a scheduler that can learn and adjust to any conceivable application. The scheduler orders the errands to be appointed into two classes, schedulable and non-schedulable. This cycle weeds out positions that are probably not going to run on a hub utilizing a cycle that is computationally modest. The experimentation result identifies the work which will over-burden a specific hub and demonstrate the equivalent to the scheduler. Hence, in the event that it is exceptionally impossible that a task will effectively run on a specific hub, that work will be delegated non-schedulable and not be considered for execution

on that hub. This will keep occupations from executing somewhat, at that point coming up short also, being rescheduled.[7]

A. D. Hartanto et al. proposed that the human asset staff in an organization is an individual in charge of finding new laborers. To get a certified new labour force, a human asset staff should be particular toward the machine regarding capacity and character. This investigation gives an elective point of view to a human asset in getting one's character information through their tweets on a Twitter account. This investigation utilizes the Naive Bayes Classifier calculation with W-IDF (Weighted-Inverse Document Frequency) weighting to order the character of enlisted people into one of Plate's character speculations, in particular Dominance, Influence, Dauntlessness, and Compliance. By utilizing preparing information and test information upwards of 120 individual Twitter records and marking of words that have been confirmed by analysts, gotten character circulation. The order of the tweet information is Predominance 90 records, Influence 10 records, Steadiness 8 records and Compliance 12 records. Assessment of the precision level of 36.67%.[8]

Ikudo, Akina & Lane et al. proposed that portraying individuals' occupations is significant for both arrangement and examination. Nonetheless, as huge scope regulatory records are progressively being utilized to portray work market movement, it will get essential to discover new mechanized ways to deal with depicting occupations. We apply new AI strategies to new wellsprings of information and research the capability of utilizing calculations to order occupations. We secure that position titles are both characteristically uproarious and conflicting across associations, however a subset of them can be alloted algorithmically, with little effect on precision.[9]

These days, the main objective of students in universities has become acquiring knowledge and specific job skills. Knowledge relating to a specific situation is extremely essential to make informed decisions and

therefore, knowledge management in any organization is quintessential in staying one step ahead of the competition in today's tough and globalized world. In this paper, the authors have come up with the design of an online recruitment system, that enables employers to post their job ads and allows job seekers to interact with the aforementioned employers to secure jobs. This portal captures job requirements based on the industry's requirements .[10]

The present study utilizes VirtualBox virtual environment technology to develop the personal and compact size of multi-node big data VM platform with Spark and Hadoop cluster that can effectively replicate and provides an environment for developers to easily design and implement Spark and Hadoop Map/Reduce programming. By using the multi-node Hadoop VM system, developers can conduct Map/Reduce programing completely the same as that in the real multi-node Hadoop cluster. To demonstrate its capability and applicability, this study performs the benchmark by using the big data VM platform and a physical Multi-Node Hadoop Cluster. Based on the standard WordCount benchmarking, the computing time of the physical multi-node Hadoop cluster is 3.7 times faster than that of VM Hadoop cluster. The benchmark results show that the big data VM platform is an ideal platform for the portal and Map/Reduce programming, Spark programming and testing purposes, and the physical Hadoop cluster is the most appropriate for production runs. In addition, the big data VM platform contains a web portal development module designed to support applications that implement big data computing services for the engineering and science users. Such applications are inherently complex, potentially accessing data from a variety of sources and distributing applications to a variety of clients. This portal development module can act as multiple roles in many projects such as personal portals, small business portals, enterprise portals, educational portal, infrastructure portal, and other types of

portals. Finally, the big data VM platform, in term of a big data development platform, is ready for users to download.[11]

With a purpose to broaden the research on online consumer behavior, w.r.t. online employment search sources viz. commercial job web sites (job portals), this paper investigates the influence (relation) of identified factors on job seekers, interacting / transacting with Naukri.com.,as online job seeking avenue. Through univariate, bivariate and multivariate analysis carried out through factor analysisby using IBM SPSS AMOS the paper found out the relationships (strength and direction) among the variables under study. Results of the study may be useful for Naukri.com to successfully deliver offerings in the market place, to appreciate and correct perceptions of job seekers interacting / transacting with naukri.com, to discover job seekers' attitudes by understanding how job seekers psychologically engage with the portal, how job seekers acquire, evaluate and use job search information as well as job search experiences and how job seekers make decisions to spend their available resources (time, money, effort) while employing the Naukri.com as job seeking avenue.[12]

This paper proposes and highlights the need of an online job board system for colleges and its effectiveness in bridging the gap between college students and career opportunities. Traditionally, employment websites have been used in HR management for finding candidates and in recruitment. This work is based on a job portal built for one of the leading engineering campuses of Nepal, Pulchowk Campus, which is a variation of such job boards designed specifically to serve the students of the Campus. With services like job recommendations to students based on their skills, and candidate filtering to assist companies in candidate matching, the system is expected to be of use for both students for exploring jobs, and companies to find potential candidates suited for the job [13].

This article focuses on creating an approach to find similar job positions based on an existing job position in a system for evaluating competencies. The proposed method uses data mining methods to find similar job positions on a job portal based on input data. The input data is the job position title and the list of competencies. The proposed approach then accesses the relevant job positions in a job portal and compares them with the entered inputs. The output of the approach is a list of the most similar job positions in the job portal, including their description. Then HR manager should use this output for completing the assigned job position and its competencies. The proposed approach is then experimentally verified [14].

The purpose of this paper is to learn more about demand for competences is crucial for revealing the complex relationship between employee selection, different strands of education and training and labor market regulation. – Employer skills requirements even for low- and medium-skilled jobs are highly specific. Formal education requirements are higher than they "should" be. No detectable "basic package" of general cognitive skills for low- and medium-skilled jobs was found. Employer demand focusses on non-cognitive skills and specific cognitive skills. Specificity of skill requirement across different sectors or occupation groups differs vastly between different types of low- and medium-skilled jobs and is linked to the interactive nature of the job, not to the qualifications or the experience required [15].

Lately, Morocco is on the road to become the hub of the automotive industry in Africa. By the encouragement of the Moroccan Government, foreign investments such as Renault Group expand their manufacturing industry in Morocco. This sector continues to grow with the new factory of PSA Peugeot Citroen Group. In fact, such new investments promise to create new jobs opportunities, however with the lack of a study that analyses this sector, the needs of the automotive job

market are still ambiguous. In this paper, we collect job ads from 22 Moroccan job portals from January 2017 until January 2018. We use several machine learning and data analysis techniques to extract and mine the needs of the automotive industry. We also use Random Forest algorithm to predict the salary from job offers using the requirements of job ads. Findings of this paper show that the automotive sector hires graduates not only in the manufacturing area but also in the service and trade area. For the most demanded occupations, we find Technician, Operator and Machinist. We also find that the most required skills in this industry are: mechanics and logistics. The embedded programming languages required in the manufacturing area are Matlab and Assembly and for design software Catia and Solidworks. Findings of this paper will be beneficial for educational institutions to update and add new branches to fulfil the demand of this sector. Moreover, it will help boost the student's motivation with the salary prediction system [16].

Shanti Dahal et al. proposed that public health programmes work to improve the health and well-being of communities by identifying the risks affecting people at different stages of life and finding best ways to minimize them. This article attempts to map the public health jobs available in India and contribute towards providing new insights in recruitment strategies for public health professionals (PHPs). A total of 427 public health jobs advertised in recruitment portals, newspapers and websites of organizations during 2012–2015 were analysed for title, qualification, location and job disciplines. Higher qualification in public health and/or social sciences is much preferred followed by MBBS and management. The largest group of vacancies consisted of programme management responsibilities followed by profiles in RMNCH. Delhi, Orissa, Madhya Pradesh and Bihar had the highest number of jobs [17].

Uma Pavan Kumar Kethavarapu and Dr.S.Saraswathi proposed that the basis of our research is to construct a job recommendation system to the job seekers by collecting the job portals data. Due to huge amounts of the data in job portals the employers are facing difficulty in the identification of right candidate for the required skill and experience.

The job seekers are also facing the problem of getting the suitability of the job based on their skill and experience. The knowledge acquisition based on the requirements is very difficult in case of huge amounts of the data sources. In fact classical development of domain ontology is typically entirely based on strong human participation. It does not adequately fit new applications requirements, because they need a more dynamic ontology and the possibility to manage a considerable quantity of concepts that human cannot achieve alone. The main focus of our work is to generate a job recommendation system with the details of job by taking account into the data posted in the web sites and data from the job seekers by the creation of dynamic ontology. We strongly believe that our system will give the best outcome in case of suitable job recommendation for both employers and job seekers without spending much time. To achieve this first we have extracted the data from various web pages and stored the collected data into .csv files. In the second stage the stored input files are used by the similarity measure and ontology creation module by generating the corresponding Web Ontology Language (.owl) file. The third stage is creating the ontology with the generated. owl by using protégé tool [18].

References

[1] Jyoti V Gautam et al., "A Survey on Job Scheduling Algorithms in Big Data Processing".

[2] Dorothea Calmels, "REVIEW ARTICLE The job sequencing and tool switching problem: state-of-the-art literature review, classification, and trends", International Journal of Production Research, 2019 Vol. 57, Nos. 15–16, 5005–5025, https://doi.org/10.1080/00207543.2018.1505057.

[3] S. Clyde, Jianping Zhang and Chih-Chung Yao, "An object-oriented implementation of an adaptive classification of job openings," Proceedings the 11th Conference on Artificial Intelligence for Applications, Los Angeles, CA, USA, 1995, pp. 9-16, doi: 10.1109/CAIA.1995.378795.

[4] F. M. Javed Mehedi Shamrat, Z. Tasnim, P. Ghosh, A. Majumder and M. Z. Hasan, "Personalization of Job Circular Announcement to Applicants Using Decision Tree Classification Algorithm," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-5, doi: 10.1109/INOCON50539.2020.9298253.

[5] L. Xu, J. Cao, Y. Wang, L. Yang and J. Li, "Bejo: Behavior Based Job Classification for Resource Consumption Prediction in the Cloud," 2014 IEEE 6th International Conference on Cloud Computing Technology and Science, Singapore, 2014, pp. 10-17, doi: 10.1109/CloudCom.2014.48.

[6] Habiba, Sultana Umme; Islam, Md. Khairul; Tasnim, Farzana. In: 2021 2nd International Conference on Robotics, Electrical and Signal Processing **Techniques** (ICREST) Robotics, Electrical and Signal Processing **Techniques** (ICREST), 2021 2nd International Conference on. :543-546 Jan, 2021

[7] S. Deshmukh, J. V. Aghav and R. Chakravarthy, "Job Classification for MapReduce Scheduler in Heterogeneous Environment," 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies, Pune, India, 2013, pp. 26-29, doi: 10.1109/CUBE.2013.15.

[8] A. D. Hartanto, E. Utami, S. Adi and H. S. Hudnanto, "Job Seeker Profile Classification of Twitter Data Using the Naïve Bayes Classifier Algorithm Based on the DISC Method," 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 2019, pp. 533-536, doi: 10.1109/ICITISEE48480.2019.9003963.

[9] Ikudo, Akina & Lane, Julia & Staudt, Joseph & Weinberg, Bruce. (2019). Occupational classifications: A machine learning approach. Journal of Economic and Social Measurement. 44. 1-31. 10.3233/JEM-190463.

[10] V. K. Sehgal, A. Jagtiani, M. Shah, A. Sharma, A. Jaiswal and D. Mehta, "Job Portal - A Web Application for Geographically Distributed Multiple Clients," 2013 1st International Conference on Artificial Intelligence, Modelling and Simulation, Kota Kinabalu, Malaysia, 2013, pp. 199-204, doi: 10.1109/AIMS.2013.38.

[11] C. Wu, W. Chang, W. Tsai, F. Lin, C. Lee and C. Yang, "Multi-node Big Data VM Platform and Job Submission Portlet," 2017 5th Intl Conf on Applied Computing and Information Technology/4th Intl Conf on Computational Science/Intelligence and Applied Informatics/2nd Intl Conf on Big Data, Cloud Computing, Data Science (ACIT-CSII-BCD), Hamamatsu, Japan, 2017, pp. 228-233, doi: 10.1109/ACIT-CSII-BCD.2017.72.

[12] Dr. Umesh B. Patwardhan, "Perception of Job Seekers Towards Monster.Com as an Online Job Seeking Avenue", in Journal for Contemporary Research in Management, 2016. Available: https://pdfs.semanticscholar.org/d34b/81bb385c808f2f016c6e5d6aac 695e2b185c.pdf.

[13] V. Yadav, U. Gewali, S. Khatri, S. R. Rauniyar and A. Shakya, "Smart Job Recruitment Automation: Bridging Industry and University," *2019 Artificial Intelligence for Transforming Business and Society (AITB)*, Kathmandu, Nepal, 2019, pp. 1-6, doi: 10.1109/AITB48515.2019.8947445.

[14] O. Pektor, B. Walek and I. Martinik, "A Data Mining Approach for Finding Similar Job Positions to Upgrade a Job Position in a System for Evaluating Competencies," 2020 21th International Carpathian Control Conference (ICCC), High Tatras, Slovakia, 2020, pp. 1-5, doi: 10.1109/ICCC49264.2020.9257298.

[15] Miroslav Beblavý et al., "The surprisingly exclusive nature of medium- and low-skilled jobs Evidence from a Slovak job portal", https://www.emerald.com/insight/publication/issn/0048-3486.

[16] I. Makdoun, G. Mezzour, K. M. Carley and I. Kassou, "Analyzing the Needs of the Automotive Job Market in Morocco," 2018 13th International Conference on Computer Science & Education (ICCSE), Colombo, 2018, pp. 1-6, doi: 10.1109/ICCSE.2018.8468854.

[17] Shanti Dahal et al., "Mapping of Public Health Jobs in India—Where Can the Public Health Graduates Be Employed?", Journal of Health Management 20(1) 73–83 © 2018 Indian Institute of Health Management Research SAGE Publications sagepub.in/home.nav DOI: 10.1177/0972063417747725.

[18] Uma Pavan Kumar Kethavarapu and Dr.S.Saraswathi, "Concept Based Dynamic Ontology Creation for Job Recommendation System", International Conference on Computational Modeling and Security (CMS 2016).