

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijrcce.com

Vol. 7, Issue 5, May 2019

Placement Analysis using Data-Mining

Nataraj K B¹, Smitha B A², Madhurya S³, Harshitha S Chavan⁴

Student, Department of Information Science and Engineering, Jain University, Bengaluru, Karnataka, India^{1,3,4}

Assistant Professor, Department of Information Science and Engineering, Jain University, Bengaluru,
Karnataka, India²

ABSTRACT: Nowadays, with the increase in usage of electronic devices (computer, mobiles, etc...) the data collected by them is more abundant and this data can be utilized to fetch some beneficial business predictions and help in making better business decisions. Machine Learning is evolving day by day and there are many algorithms derived to obtain solutions for any of the Big-data related problems. In the paper, the placement related data of a batch of students that who got placed is analyzed and fetching the most frequent attributes of them, by which they got placed. Which helps our peers and juniors to understand, what are the main areas to focus on, so that they get placed in a good company with a high package.

KEYWORDS: Frequent item sets; Regularity; Sliding Window; top-k patterns; Bit-vector; Support; Confidence

I. INTRODUCTION

Data Mining is the Process of extracting similar patterns, from large dataset using Machine Learning Algorithm and Statistics.

Because of the significance of extracting data from the expansive information archives, information mining has turned into a basic part in different fields of human life including business, instruction, medicinal, logical and different considerable information based fields. For example, classification of the things accessible on the online business webpage is a principal issue. A right thing order framework is fundamental for client experience. Thing arrangement can be defined as a directed characterization issue in information mining where the classifications are the objective classes and the highlights are the words making some printed portrayal out of the things. One of the methodologies is to discover bunches at first which are comparative and spot them together in an inert gathering. Presently given another thing, first characterize into a dormant gathering which is called coarse dimension order. At that point, complete a second round of grouping to discover the class to which the thing has a place with. In the ongoing occasions, information mining has a colossal task to carry out in the investigation of human hereditary qualities that finds how a person's DNA succession influences the dangers of creating regular infections, for example, malignant growth and finding extemporized strategies for treating these sicknesses.

Essentially, information mining discovers its uses in cutting edge fields including human rights, reconnaissance and different other business handle, this space of software engineering has been picking up a great deal of fame of late and demonstrates to be one of the proficient, precise types of innovation. In this way, information mining or even AI innovations will keep on winning for good until a choice to coordinate fundamental segments and innovation is found.

In the Paper, firstly all the details of placed students is collected, the data is understood and then decide the attributes of the data by which can analyze and mine useful Patterns from it. We also decide which Associative Algorithm will accurately suit for the data. Later, only the data that is required is fetched for mining from whole dataset and apply algorithm on that data and fetch out the useful patterns from it.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijrcce.com

Vol. 7, Issue 5, May 2019

By using Apriori Algorithm on the dataset for finding the frequent Attribute set, by which the student got cleared the placement. Since, the Paper is mainly focusing on descriptive analysis, it analyses the previously stored placement related data of the batch students and tell what are the major and frequent attributes that helped them to clear the placement, and the Paper results also guide the following peers and juniors through their placements, by telling them what are the main attributes that they need to focus on, before attending the placements.

A detailed explanation of the implementation of data mining principles in the Paper is done in the latter sections.

II. LITERATURE SURVEY

Lately in the data-mining community, there is a lot of interest in mining interesting item sets over a stream of continuous data. since this is applicable to a variety of different areas such as traffic analysis, click stream analysis, call record analysis, etc. all of which involve high data rate with very large volumes, which makes mining these types of data a very challenging and complex task for traditional algorithms based on Apriori, FP-growth and ECLAT.

To overcome the problems faced by the traditional algorithms, [1] [11] frequent-regular itemset mining (FRIM) is adapted for various applications. An itemset that frequently and regularly occurs in a database is identified as an interesting itemset. FRIM has several extensions such as mining itemsets with approximate regularity (periodic) of occurrence, mining (frequent) regular itemsets from incremental database (data stream), mining regular closed itemsets, mining chronic-frequent itemsets, mining high-utility regular itemsets, etc.

[5][6] The top-k frequent-regular itemsets mining over a transactional database (TFRIM) is applied to control the number of generated results. However, the input database TFRIM usually contains a static number of transactions causing TFRIM cannot efficiently discover the itemsets from dynamic database, to improve upon this, TFRIM-DS (Top-k Frequent-Regular Itemsets Miner over Data Streams) is used to discover a set of k itemsets that regularly occur and have highest support in the current data stream.[8][12] The concept of sliding window is applied to scope constant amount of recent transactions where the recent occurrences are considered to be of higher importance than the former occurrences.

[2] An efficient Single Scan Frequent Itemset Mining (SSFIM) for solving FIM problem is by using the generated candidate itemsets from each transaction and stored in a hash table that maintains the support information. Only the non-existing itemsets are inserted when they are generated in the hash table (with initial support value set to 1), while the generation of existing itemsets from a new transaction increases its support value.

[3] Sparse databases can alternatively use a heuristic called EA-SSFIM, which combines the evolutionary procedures and the SSFIM by instantiating the Hadoop clusters using MapReduce programming model to yield MR-SSFIM

[4][9][10] Experiments using SSFIM have showed that it outperforms other accurate FIM algorithms and achieves better scalability compared to the other algorithms. Moreover, EA-SSFIM considerably reduces the runtime compared to the other approaches while maintaining a high accuracy that is very close to SSFIM. Moreover, the results show that MR-SSFIM outperforms existing HPC-based FIM approaches for big and sparse databases. [7] For FIM Single Scan, Two Scan or Multiple Scan approaches can be used. For Big Data, GPU Based Approaches or Cluster based approaches.

III. PROPOSED METHOD

In the Paper, descriptive analysis is done by using the subset approach of Apriori algorithm and the sliding window concept of Top-k Frequent-regular Itemset mining (TFRIM).

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijrcce.com

Vol. 7, Issue 5, May 2019

The flow of the proposed method is briefly described in the figure below.

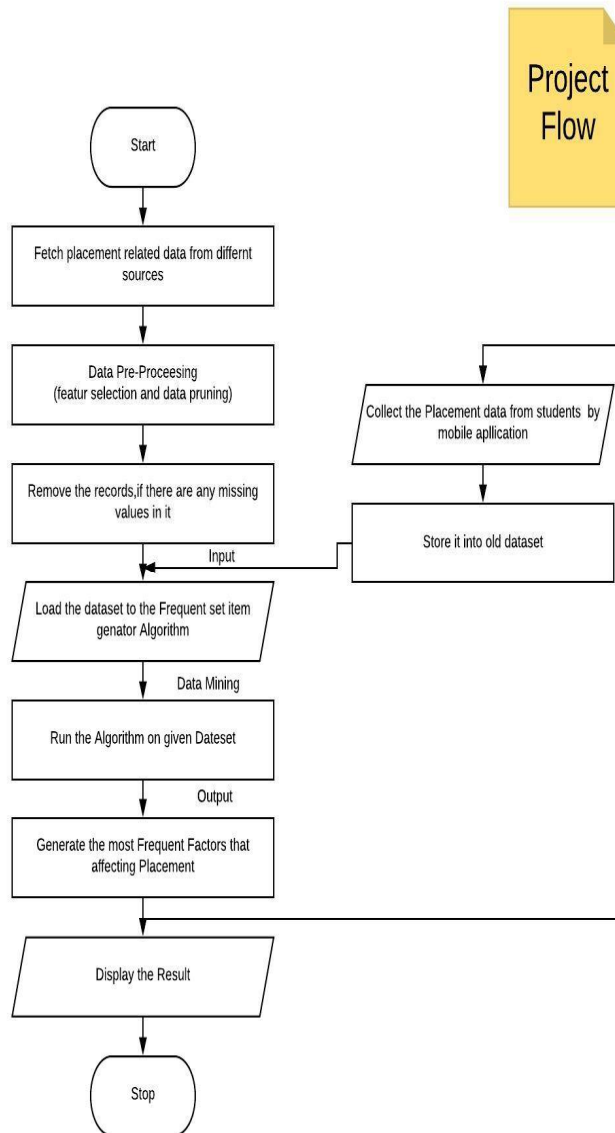


Fig. 1. Project Flow

Firstly, all the placement related details of the batch of students is collected from various sources and translated all the data into one single format and stored in excel sheet. The data included the main attributes the are required for mining. Those attributes are Academic performance, Practical performance, Behavioral performance, Communicational performance and Placed Package. Academic performance and Practical performance are collected by the report cards of that batch. And Behavioral performance and Communication performance are collected by conducting a survey, in which all the lecturers gave their feedback.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijrcce.com

Vol. 7, Issue 5, May 2019

Later, the data is pre-processed where discretization method is used to convert all the continuous valued data into categorical valued data. Then categorizing the values into bad, average, good and excellent depending on the range in which the value is falling.

Later, on the pre-processed data subset approach of apriori algorithm is used and the sliding window concept of TFRIM algorithm. The flow of the apriori algorithm is described in the figure below.

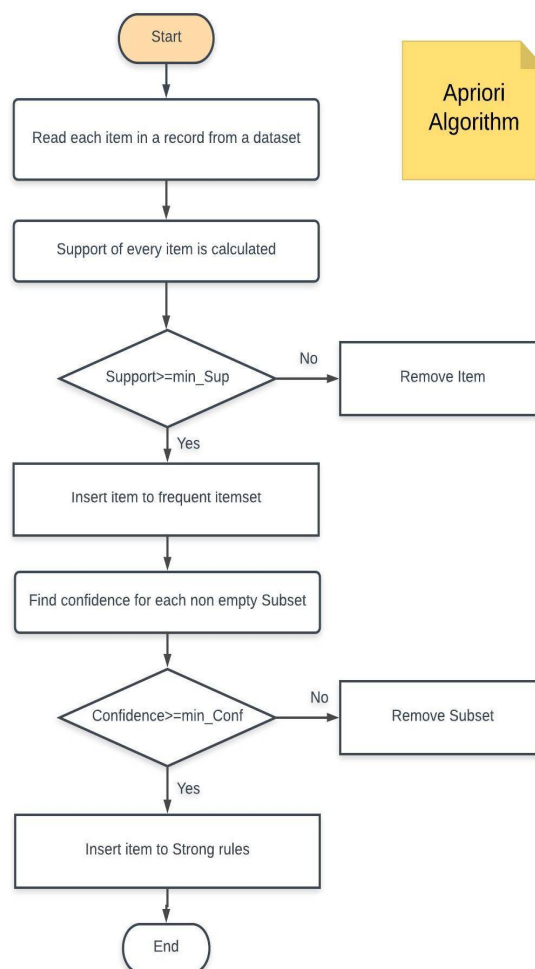


Fig. 2. Apriori Algorithm Flow

Firstly, the support of all the attributes in the data is found out, then the top-k (Here k refers to the size of the sliding window) support attributes is supported and the power set of those attributes is found. And then the top-k item-sets in that power set, whose confidence is more is found out. Thus, concluding the results of the descriptive analysis of the placement data.

IV. CONCLUSION

In the paper, the placement related data of a batch of students is analyzed to understand, what all the major attributes that everyone should focus on, before preparing for interview. Here the most frequent attributes that the batch possessed is found out, which helped them in getting placed.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijrcce.com

Vol. 7, Issue 5, May 2019

REFERENCES

1. Tashinee Mesama and Komate Amphawan on "Mining Top-k Frequent-regular Itemset from Data Streams Based on Sliding Window Technique", Computational Innovation Laboratory, Faculty of Informatics, Burapha University, Chonburi, 20131, Thailand.
2. Youcef Djenouri, Djamel Djenouri, Jerry Chun-Wei Lin (Member, IEEE), And Asma Belhadi on "Frequent Itemset Mining in Big Data with Effective Single Scan Algorithms"
3. Vangipuram Radha Krishna ISBN 978-1-4503-6392-1/18/04 on "Sequential Approach for mining of Temporal Itemsets."
4. "Mining top-k frequent-regular closed patterns," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7882–7894, 2015.
5. G. V. Kumar and V. V. Kumari, "Sliding window technique to mine regular frequent patterns in data streams using vertical format," in *2012 IEEE International Conference on Computational Intelligence and Computing Research*, 2012, pp. 1–4.
6. R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Rec.*, vol. 22, no. 2, pp. 207–216, 1993.
7. M. A. Bhuiyan and M. Al Hasan, "An iterative MapReduce based frequent subgraph mining algorithm," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 608–620, Mar. 2015.
8. K. Wang, L. Tang, J. Han, and J. Liu, "Top down FP-growth for association rule mining," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*, 2002, pp. 334–340.
9. Manish Gupta, Jing Gao, Yizhou Sun, and Jiawei Han, "Integrating community matching and outlier detection for mining evolutionary community outliers", KDD12, In the Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining 2012, pp. 859-867.
10. V.Radhakrishna, P.V.Kumar, V.Janaki, "An Efficient Approach to find Similar Temporal Association Patterns Performing Only Single Database Scan", REVISTA TECNICA DE LA FACULTAD DE INGENIERIA UNIVERSIDAD DEL ZULIA, Volume 39, Issue 1, Feb 2016, pp. 241-255.
11. K. Amphawan and P. Lenca, "Mining top-k frequent/regular patterns based on user-given trade-off between frequency and regularity," in *Advances in Information Technology*, ser. Communications in Computer and Information Science, 2013, vol. 409, pp. 1–12.
12. M. Sreedevi and L. Reddy, "Mining regular closed patterns in transactional databases," in *Proceedings of the 7th International Conference on Intelligent Systems and Control, Jan 4-5, Coimbatore, India*, 2013, pp. 380–383.