iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** VII **Month of publication:** July 2015

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Association Rule Mining on Language Trend in MNCs

Nutan Dahiya[#1] ,Rachna Dhaka[*2]
[#] *Students, Computer Science & Engineering, Gateway Institute Of Engineering and Technology, Sonipat, Haryana(India)*
[*]*HOD, Computer Science & Engineering, Gateway Institute Of Engineering and Technology, Sonipat,Haryana(India)*

Abstract*— Data mining system discovers patterns and relationships hidden in data, and actually is a part of a larger process called "knowledge discovery" which describes the steps that must be taken to ensure meaningful results. The presented work will focus on implementing different data mining approaches on database which is simulated from different companies. The work has combined two major approaches to provide profiling of companies and language analysis via clustering and association rules. Once the languages will be identified, then association between segments and profiling of companies will be identified. A comparison between the results which will be obtained from using two different algorithms will be done. One algorithm used is Apriori algorithm and the other one is Simple K Means algorithm which is described in my previous paper. The results obtained from these two algorithms will be identified and compared for the purpose of validation of trends of programming language. All the work is done using WEKA tool. This paper is all about association rule mining to analyse languages in different MNCs.
Keywords: language trend, Data mining, Clustering, Segments, Weka tool.

I. INTRODUCTION

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information -information that can be used to increase revenue, cuts costs, or both. Data mining software allows users to analyze data from many different dimensions or angles, categorize it and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Data mining is primarily used today by companies with a strong consumer focus - retail, financial, communication and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics. And, it enables them to determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transactional data. With data mining, student recruitment intelligence system find out point-of-interest records of students for employment to send targeted promotions based on an individual's knowledge. By mining they can find out the required information data to find out the relation between different companies and students according to their intelligence. This project is going to describe the activity related to **recruitment**, its various problems and their solutions. It uses MIS and ERP system for managing various administrator tasks like student data for recruitment.

Management information systems (MIS) are typically computer systems used for managing five primary components:
Hardware, Software, Data (information for decision making), Procedures (design, development and documentation), People (individuals, groups, or organizations)

Academically, the term is commonly used to refer to the study of how individuals, groups, and organizations evaluate, design, implement, manage, and utilize systems to generate information to improve efficiency and effectiveness of decision making, including systems termed decision support systems, expert systems, and executive information systems. Most business schools (or colleges of business administration within universities) have an MIS department, alongside department accounting, finance, management, marketing.

Enterprise resource planning (ERP) is a cross-functional enterprise system driven by an integrated suite of software modules that supports the basic internal business processes of a company. ERP software integrates all facets of an operation, including product planning, development, manufacturing processes, sales and marketing.

ERP gives a company an integrated real-time view of its core business processes such as production, order processing, and inventory management, tied together by ERP applications software and a common database maintained by a database management system

Recruitment refers to the overall process of attracting, selecting and appointing suitable candidates to a one or more jobs

within an organization, either permanent or temporary. Recruitment of candidates is the function preceding the selection, which helps create a pool of prospective employees for the organisation so that the management can select the right candidate for the right job from this pool. The main objective of the recruitment process is to expedite the selection process. There are several recruitment groups that provide employment Opportunities and world class training to students of the Institute in-campus or off-campus in leading organizations/Industry.

All the reputed organizations including MNCs come to the Institute for Campus Recruitment. To achieve its goal, the Placement Cell works towards recognizing the core competencies of students.

A. Significance of the problem

Company recruits the new enrolled students on the basis of company's needs, student grade, technical language and protocols the companies defined for recruitment process. This study can also do a comparative analysis by using different algorithms to validate the result. Unlike the existing system this is an efficient approach (algorithm) which is easy to implement and does the following tasks.

1) Profiling the students according to the Companies in which they placed.
2) It can analyze the language(s) according to which company can recruit.
3) To target the names of companies earlier according to the grades and technical language the students knows.
4) To find out how many students were recruited by which company.
5) To explore the relation of recruitment group with high profiled MNC's.

It also validates the result comes from two different approaches.

II. RESEARCH BACKGROUND

In year 2013, Dorina Kabakchieva, Predicting Student Performance by Using Data Mining Methods for Classification. Data mining methods are often implemented at advanced universities today for analyzing available data and extracting information and knowledge to support decision-making.

In year 2014, Ajay kumar, Swati Singhal, Praveen Dhankher and Anju Gulia, a study was done which works in three phases. First phase, provide profiling of students according to their grades via clustering then, a number of different segments are formed on the basis of type of grades provided to the student by the institution. The students according to their grades are grouped into different clusters according to the company rating. Second phase considers validating the student data using J-48 algorithms. In third phase, a comparison is made between different algorithms (J-48, LMT and Bayes Net algorithms) to improve prediction accuracy of student recruitment datasets.

In year 2015, Praveen Rani, Dr. Rajan Vohra, presents a comprehensive statistical experiment to identify the number of students those are ready for placements and students those are not fulfilling the basic criteria for placement from a large database of all computer engineering students of a college containing their academic record. The design of experiments software named Weka Tool is used for making three clusters of whole database which will categorise the students according to their qualifications.

III. RESEARCH METHODOLOGY

For solving the problem some research techniques and methodologies are used for obtaining the desired result.

A. First of all literatures and research papers were reviewed for getting more information about the problem and knowing which type of work was done by others on this topic and by which method.
B. Then tools required for solving the problem were identified and the best tool was selected from all. Organize filed visits to Engineering colleges that conduct placement drive for reputed companies
C. Determine nature and definition of research problem and work flow of the problem for getting accurate and desired result.
D. Organize the database with useful attributes and populate it then perform data analysis using suitable tool e.g., WEKA in order to generate the result.

IV. CONCEPTUAL FRAMEWORK

Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. An association rule has two parts, an antecedent (if) and a consequent (then). An

antecedent is an item found in the data. A consequent is an item that is found in combination with the antecedent. Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships.

In data mining, association rules are useful for analyzing and predicting customer behavior. They play an important part in shopping basket data analysis, product clustering, catalog design and store layout. To select interesting rules from the set of all possible rules, constraints on various measures of significance and interest can be used. The best-known constraints are minimum thresholds on support and confidence.

Apriori is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions. As is common in association rule mining, given a set of **itemsets**, the algorithm attempts to find subsets which are common to at least a minimum number C of the itemsets. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.

A. Profiling Of Companies

In this problem, all the companies are grouped according to the no. of recruitments they held. The result will in form of different groups that provide information about the number of students and group of companies that recruit these students. The whole data of recruitment process is refined. Then companies are segregated into different groups. A few dataset is taken from an recruitment group and then whole database is simulated to analyze the result.

B. Language Trend Analysis Via Association Rule

From profiling, Companies will found, then by using association rules, link between Companies and Language will be discovered. It also find out the dominant language for each company by using Apriori algorithm. It will generate the different rules according to the company and language of candidate recruited like as:

Rule 1. Company ➔ Language
Rule 2. Language ➔Company

Both rules support each other and having confidence value acc. to the number of entries in database. 0.1 confidence value for each 55 entries. The value of confidence for supporting the rule should lies in between 0 – 1. The support supp(X) of an item set **X** is defined as the proportion of transactions in the data set which contain the item set. The confidence of a rule is defined conf (X➔ **Y**) = supp (**X** U **Y**)/ supp (**X**) Then choose the strongest and more confident link among all links and define that link as a rule.

V. RESULTS AND DISCUSSION

A database of 448 records/entries are created and simulated for the purpose of analysing dominant language in the companies.

A. Company Wise Count Of Each Language

A company may recruit not only for one language. So, it represents the various language in each company.

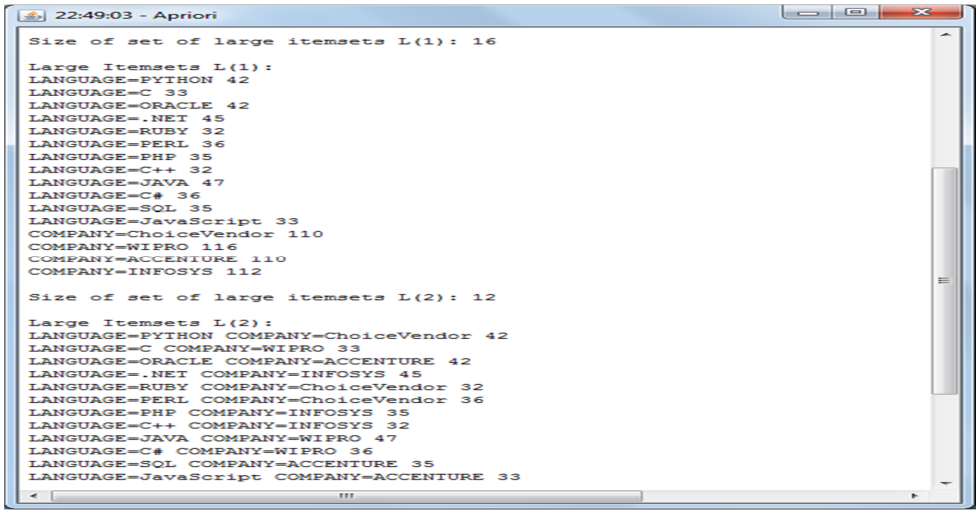


Fig. 1 Company wise count of each Language

There are total 8 rules supporting each other according to their dominant category.

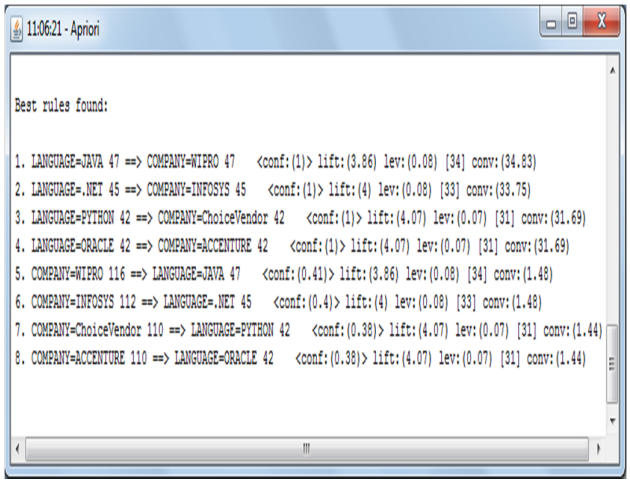


Fig.2 Association Rule

B. Discovering Travel Pattern Via Association Rule

Segments	All Languages	Count Of Languaes	Dominant Language	Rule Formed
ChoiceVendor	Perl Ruby Python	36 32 42	Python	ChoiceVendor ➔ Python
Wipro	Java C C#	47 33 36	Java	Wipro ➔ Java
Infosys	PHP .Net C++	35 45 32	.Net	Infosys ➔ .Net
Accenture	Oracle Java Script SQL	42 33 35	Oracle	Accenture ➔ Oracle

Table 1 Language Trend via Association Rule

C. Validation Of Language Trends

In this step we compare the results obtained by two different techniques i.e. Clustering approach and association rule. So now we compare the dominant language received from analyzing the database of companies using clustering and association. We have already discussed and showed the result of clustering method in our previous paper whereas this paper gives us the results of Association rule.

After getting results from two different techniques, the result will be analyzed to see whether the results are same and supporting each other or not. If results will same then rules generated by association rule will valid and provide the correct language trend.

\

Volume 3 Issue VII, July 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

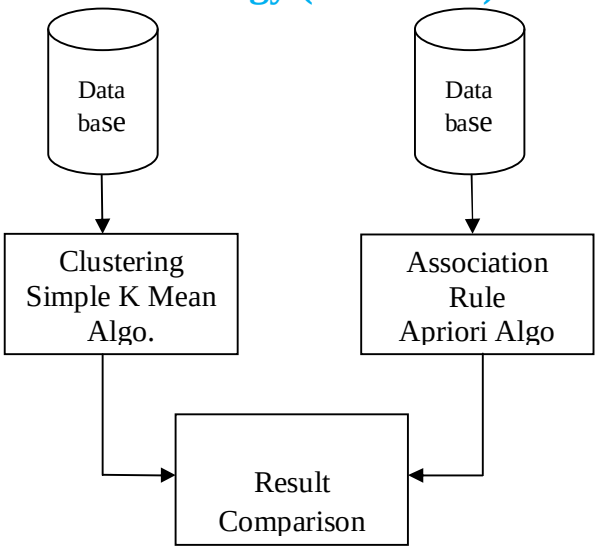


Table 2 Validation of Results

Company	Dominant Language Via Clustering	Dominant Language Via Association Rule
ChoiceVendor	Pyhton	Pyhton
Wipro	Java	Java
Infosys	.Net	.Net
Accenture	Oracle	Oracle

Table 3 Analysis of Results

VI. CONCLUSION

The main motive of this thesis is to figure out which IT company uses which language to develop its own programs which will be used by other companies in outside world for daily operations. This also helps students in getting their dream job in their desired companies by knowing those companies' requirements.

Lots of students pass each year and prepare for exams to get hired by MNCs or other reputed companies. And for that purpose they should be knowing which company wants their future employees to have specific knowledge about their preferred language. This thesis does that work for students and provides them a specific language to work on and gain skills in that. Thus it makes the task very easy for students and for MNCs as well. MNCs will have those aspiring candidates who know quite well about that specific company's need they want to get in and as a result better candidates hired by company who have good amount of knowledge.

VII. FUTURE WORK

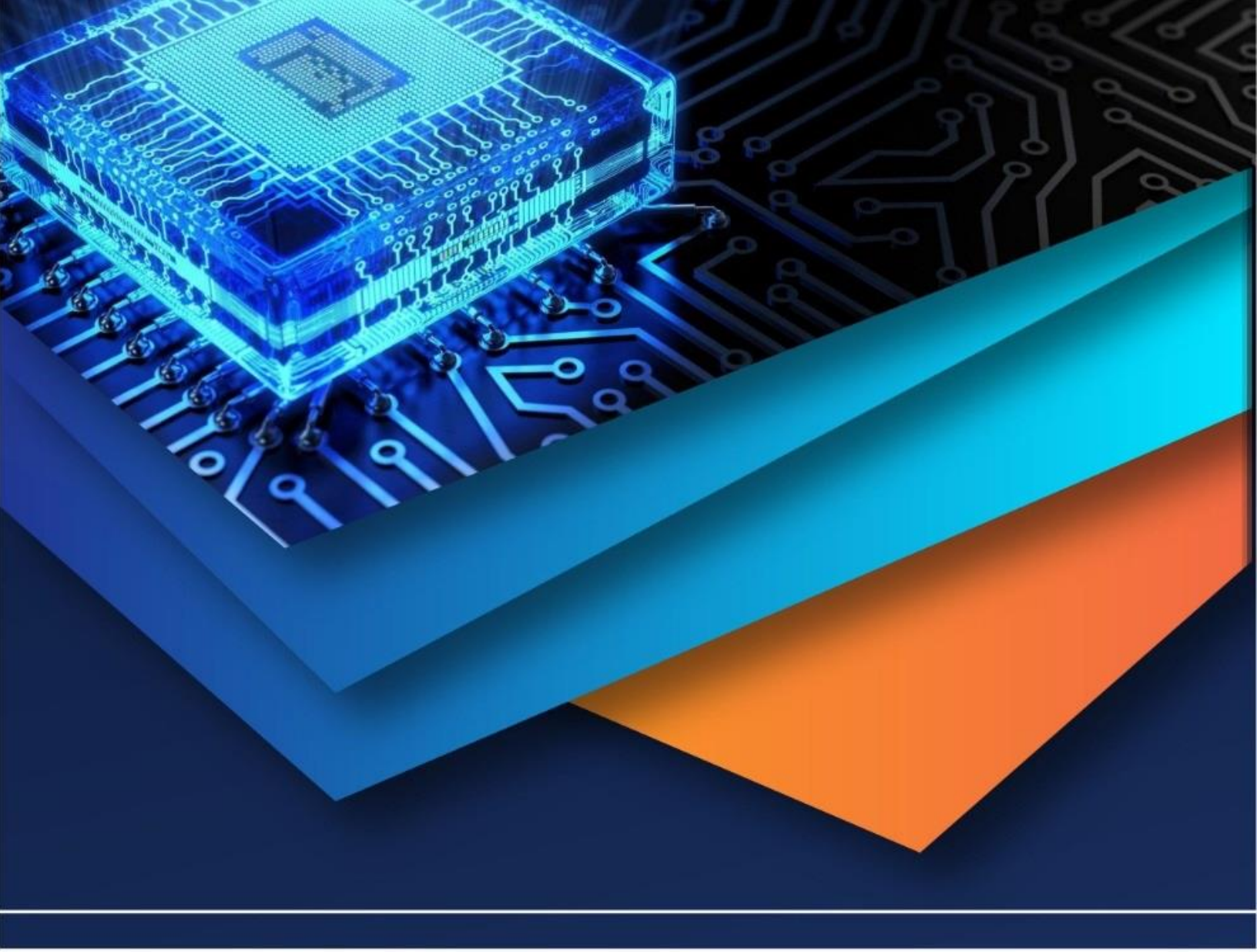
Any decent research is always focused on the future scope and improvements that can be carried out seeing the demand of time. When this work will be applied to the problems real world faces it will provide the desired and accurate results as this thesis is made very calculated and after thoroughly analysis of simulated database. The results suggest that when real data from companies will be analyzed students can firmly execute their plans to get knowledge in that particular programming language to get their dream job. So is with companies they will hire better and most suitable bunch of candidates for them.

VIII. ACKNOWLEDGEMENT

Author would like to thank to their head Ms. Rachna, HOD Computer Science & Engineering, Gateway Institute Of Engineering and Technology, Sonipat, Haryan for her continuous and valuable support during this whole project.

REFERENCES

[1].http://mckinseyonsociety.com/education-to-employment/report/
[2].http://www.naukrihub.com/recruitment/importance-of-recruitment.html
[3].http://www.uni-weimar.de/medien/webis/teaching/lecturenotes/machine- learning/unit-en-decision-trees-algorithms.pdf
[4].http://www.cs.princeton.edu/courses/archive/spr07/cos424/papers/mitchell- dectrees.pdf
[5]. http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf
[6]. http://www.d.umn.edu/~padhy005/Chapter5.html
[7]. http://www.cs.ccsu.edu/~markov/ccsu_courses/DataMining-7.html
[8]. http://www.cs.umd.edu/~samir/498/10Algorithms-08.pdf
[9]. http://wiki.pentaho.com/display/DATAMINING/Classifiers
[10].http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.html
[11]. http:// www.ijcsit.com/.../ijcsit2015060371.pdf
[12]. http:// www.ijcsms.com/journals/volume%2520.pdf



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY