

International Journal of Information Technology, Research and Applications (IJITRA)

Justin Rajasekeran (2022). A Study on Students' Performance in compliance with Industry requirements using Data Mining Techniques. International Journal of Information Technology, Research and Applications, 1(3), 1-5.

ISSN: 2583 5343

DOI: 10.5281/zenodo.7384567

The online version of this article can be found at:
<https://www.ijitra.com/index.php/ijitra/issue/archive>

Published by:
PRISMA Publications

IJITRA is an Open Access publication. It may be read, copied, and distributed free of charge according to the conditions of the Creative Commons Attribution 4.0 International license.

International Journal of Information Technology, Research and Applications (IJITRA) is a journal that publishes articles which contribute new theoretical results in all the areas of Computer Science, Communication Network and Information Technology. Research paper and articles on Big Data, Machine Learning, IOT, Blockchain, Network Security, Optical Integrated Circuits, and Artificial Intelligence are in prime position.



<https://www.prismapublications.com/>

A Study on Students' Performance in compliance with Industry requirements using Data Mining Techniques

Justin Rajasekeran

Department of Information Technology, University of Technology and Applied Sciences, Shinas, Oman

Article Info

Article history:

From the proceedings of 2nd Symposium on Information Technology and Applied Mathematics – SITAM 2019, 6th November 2019.

Host: University of Technology and Applied Sciences, Shinas, Sultanate of Oman.

Keywords:

Academic performance prediction
Decision making
Data Mining
Cluster

ABSTRACT

Education is one of the best aspects of our life which makes the student to learn the value of education. It makes the student technically sound enough with the confident to face the world. For the academic institution and the industry play a vital role in each student's education. To meet the industry standard, it is important to identify student's performance becomes challenging due to high value of educational databases. So, it is necessary to analyze and monitor the student performance in order to meet industry requirements. This paper identifies suitable method for predicting the performance of the students in Oman using data mining techniques.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Justin Rajasekeran
Information Technology Department,
University of Technology and Applied Sciences, Shinas,
Oman
Email: justin.rajasekeran@shct.edu.om

1. INTRODUCTION

In the increasing competition of the educational society it is most important issues to identify the students' performance in their exit level of the institutions especially using data mining techniques. This is the field of educational data mining research areas. It has need in this decade. The reason behind this research work is, there is always a gap between industry and the educational institutions. Due to this students' performance range is shows in a different curve. By identifying their lack of performance helps the student community to learn and improve the skills needed for the industry sectors. The available resources and infrastructure may not be sufficient to meet the industry standards [1].

The data collected from the difference sources needed a proper method of extracting knowledge from large repositories for decision making [2]. By this the quality of the education may be improved for outgoing student community in various educational institutional.

The quality education is one of the most important key factor for any institution. To enhance the quality of the education by improving decision making procedures with the useful knowledge. The quality criteria can be set as course assessment, teachers' assessment and the student assessment. The new research field concerned with this method to explore the data from educational settings to use better understand the student community [3]. This learning is called Educational Data Mining[EDM]. This EDM is considered as technology to provide the behavior of lecturer, and student. This paper tried to establish the capability of data mining in identifying and extracting high volume data from the available resources. There are some techniques like clustering analysis, regression analysis, statistical classification which helps to experiment and evaluate the process in the educational institution.



Fig-1 Data Mining Techniques

Data Mining

Data Mining is also referred as KDD means Knowledge Discovery in Database used to extract knowledge from the huge data. There are data mining techniques are used to discover patterns and relationships in decision making. Figure 2 shows the steps involved extracting knowledge from huge amount of data.

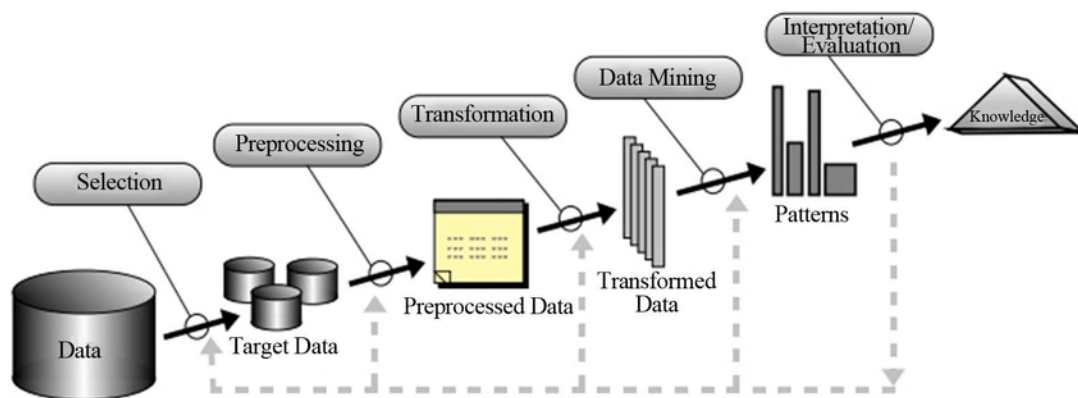


Figure 2. Process of extracting knowledge from data

There are various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, are used for knowledge discovery from databases [4].

Classification

Classification is the most common data mining technique which employs a set of pre-classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network-based classification algorithms. This classification process involves learning and classification.

Clustering

Clustering could be an identification of similar classes of objects. This technique helps to identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes.

Predication

Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. Today unfortunately, many real-world problems are not simply prediction. For example, sale rate, stock prices, and product failure rates are all very difficult to predict because there is a decency of complex interactions [5]. The same model types can often be used for both regression and classification. For example, the CART

(Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical response variables) and regression trees (to forecast continuous response variables) [6].

Data Mining Process

In this study, we make two contributions. One is text mining methods to extract data from job descriptions to understand the job market. Second, a study using sample data to support this paper. Generally, job descriptions are given by the industry employers, but it is not in standard format or well structured. There may be useful and not useful information, therefore important information need to be extracted and cluster the useful information such as required skill set.

The text mining methodology may be used to understand the industry standard. It is the process of extract data from the huge volume of unstructured text data. It also performs to identify concepts, patterns and other attributes. This process further helps to identify attributes which comes frequently like job titles and job descriptions. K-means clustering may be used over the extracted attributes to differentiate the different available jobs in the industry [7].

A study happened nearly 100s of job postings from a popular job web site. This study identified different jobs available for different level of expertise of undergraduate students. This study may help students, institution and industry to get insights.

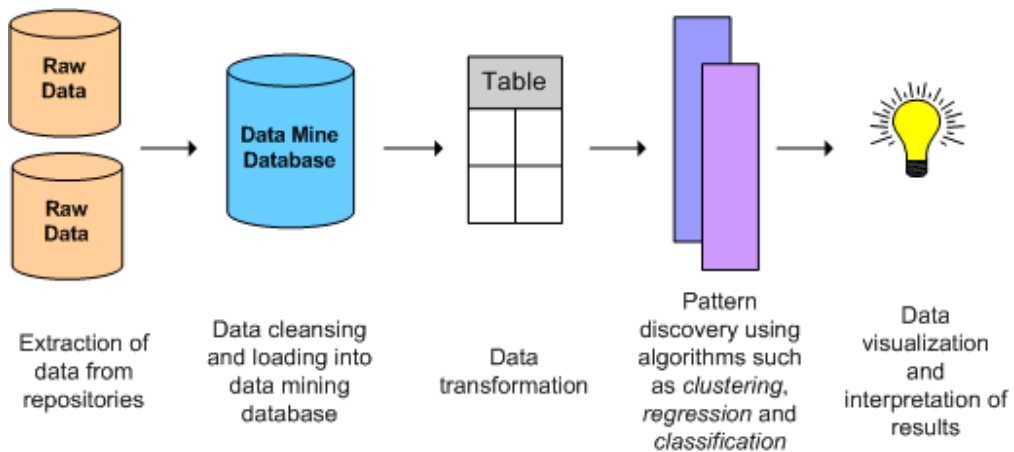


Figure 3. Clustering process

Methodology

There are two datasets are consider from the masters computers degree from a private university for the year 2018. The most common job postings are identified for 2013 and 2018 to analyse the trend. Each record contains job title and job description.

Job Description

[Send me Jobs like this](#) |   

We are looking for PHP Developers apart from several other positions.

Desired Profile:

Proficient in core PHP, MySQL, jQuery

Sound knowledge and experience on any MVC frameworks.

Basic knowledge of HTML/ CSS

Excellent communication and interpersonal skills.

Quality obsessive, team player, quick learner, problem solver

Roles & Responsibility

- Develop PHP/ MySQL applications using MVC Frameworks
- Integrate front- end screens in web application and make front- end functional using jQuery
- Collaborates with fellow developers, team lead, UI developers and QA to produce high quality web applications.
- Candidate who belongs from indore or near by cities can apply

Fig 4. Sample Job Description

The above figure contains the following information:

Technical skills: HTML, PHP, MySQL, jQuery, MVC frameworks

Soft skills: team player, quick learner, problem solver, communication and interpersonal skills

Table 1 Skill Information

Attribute	Freq. in 2013	Freq. in 2014	Freq. in 2015	Freq. in 2016	Freq. in 2017	Freq. in 2018
IT Admin	56%	58%	70%	66%	71%	77%
Project Manager	76%	81%	80%	85%	85%	86%
SAP consultant	45%	50%	54%	60%	65%	78%
Technical Support Executive	65%	72%	77%	81%	85%	89%
Architect	54%	58%	60%	63%	66%	70%
ERP Executive	60%	63%	69%	74%	78%	82%
Business Analyst	48%	50%	59%	67%	74%	84%
.NET Developer	70%	69%	74%	84%	88%	90%
Software Engineer	77%	59%	67%	74%	84%	92%
Web Developer	56%	61%	67%	74%	78%	89%

By using clustering, the different types of job within the discipline is identified. The Latent Semantic Analysis will be used to identify the job descriptions represented as a job vector. The “i” value of the job vector is equal to the inverse document frequency of the set of possible words. Then run K-means clustering on the vector value and report the keyword from each cluster as representatives. The above table express the ratio of the job titles in the specified year and also job description also varies in every year [7].

The study shows the job title frequently asked in the mentioned year. Though the technology changes more technical skills are required. On the other hand, many soft skills are more frequent which the students have to increase their skills for the competitive job.

The below table shows the students’ performance when they appear to job interview.

Table 2: Students performance when attend interview

	OCGPA (4)	Technical skills (10)	Soft Skills (10)	Result (10)
Student 1	3.65	4.5	4	4.05
Student 2	3.1	5.76	3.76	5.46
Student 3	3.67	6.2	3.9	4.59

Table 3: Sample Large clustered data

Label	Attribute in Cluster Centroid	2013	2014	2015	2016	2017	2018
Software Development, Testing	Java, Test, SQL, Tool, Server,	56%	58%	70%	66%	71%	77%
Web Development	HTML, SQL, Web, Asp, Javascript, Server, Java, XML, Database	76%	81%	80%	85%	85%	86%

System Development	C++, Sort, Expens, Java, GUI	45%	50%	54%	60%	65%	78%
System Admin	Network, Hardware, Troubleshoot, Install, User, Configure	65%	72%	77%	81%	85%	89%
Embedded system	Multimedia, Debug, Embed, Device, C++	54%	58%	60%	63%	66%	70%

Conclusion


This study presented a text mining methodology to extract the raw data and compare, cluster the different job descriptions. This method identifies required skills for the industry in this competitive world. This study considered 500 undergraduate dataset and analyzed. The student and the institutions may get insight about the industry expected skills which is in high demand.

This study suggest that all students acquire the basic knowledge of technical skills and strong soft skills in order to meet the industry standard. The institution may be considered the attributes which need high influence in the student performance to meet the industry standard for further action for improvement.

REFERENCES

- [1] Buckley, F. & Harary, F. (1990). Distance in graphs, Addison – Wesley, Longman.
- [2] Chartrand, G. & Zhang, P. (2004). Distance in graphs-taking the long view, AKCE J. Graphs Combin.1-13.
- [3] Trinajstic, N.(1983). Chemical Graph Theory (CRC Press, Boca Raton, FL).
- [4] Gutman, I. Ruscic, B, Trinajstic N, & Wilcox C.F. (1975). Graph theory and molecular orbitals. XII Acyclic Polyenes, J. Chem. Phys. 62, 3399-3405.
- [5] Gutman, I. &Trinajstic, N.(1972). Graph theory and molecular orbitals. Total φ –electron energy of alternant hydrocarbons, Chem. Phys. Lett. 17(4), 535 – 538.
- [6] Dobrynin, A.A. & Kochetova, A.A. (1994). Degree distance of a graph: A degree analogue of the Wiener index, J. Chem.inf. Comput. Sci 34,1082 – 1086.
- [7] Gutman, I.(1994) Selected properties of the Schultz molecular topological index, J.Chem. Inf.Comput.Sci. 34,1087-1089.
- [8] Schultz, H.P.(1989). Topological organic chemistry 1. Graph Theory and Topological Indics of Alkanes, J. Chem. Inf. Comput. Sci 29,227 – 228.
- [9] Chen, S. & Guo, Z. (2010). A lower bound on the degree distance in a tree, Int. J. Contem. Math. Sci. 5(13),649 - 652.

BIOGRAPHIES OF AUTHORS

	<p>Justin Rajasekaran, received the Master degree in Computer Application from Madurai Kamaraj University, Tamil Nadu, Chennai, in 1998. He is currently working as a Lecturer at University of Technology and Applied Sciences-Shinas, Sultanate of Oman. His research interest includes data mining, big data analytics, block chain technology, Data Privacy & Security, Software Quality and Testing, and Information Security. He can be contacted at email: Justin.rajasekaran@shct.edu.om</p>
---	---