ISSN:0975-9646

Predicting Educational Performance of a Student Failure and Dropout by using Data mining Techniques

U.Harilatha[1], N.Sudhakaryadav[2]M.Tech

[1]*PG Student,* [2]*Assistant Professor*
[1 2]*Department of Computer Science and Engineering,*
Madanapalle Institute of Technology and Science, JNTUA

***Abstract*- There are numerous data or text mining methods have exists projected for the mining functional pattern during the manuscript documents and the quality of the extracted features is the key issue to the text mining. The quality evidence for the existing text mining schemes owed to great of surroundings, expression and noises. Though, the data mining works with effectively, the update discover patterns having issues on the open research especially in the text mining as domain. Data mining methods are applied to predict college failure and dropout of the student. This project is used for real data on college students for prediction of failure and dropout. It implements white-box classification methods, like decision trees and induction rules. Decision tree could be a decision support tool that represented as like graph or a model of decision. It consists of nodes, in which the internal nodes are denoted as test on attributes. Attribute is nothing but real data of student that collected from college in middle or educational activity. A path from root to leaf is represents classification rules and it consists of 3 forms of nodes, which includes decision node, probability node and finish node. It can be used in verdict examination. Using this method, try to boost their correctness for computing the students may not pass or dropout by first; with all accessible characteristics next and then choosing best attributes. Attribute selection is done by Java programing language. Hence the data processing tool mainly works in prediction and classification of knowledge. Java programing language supports much normal data processing task information pre-processing, clustering, classification and have choice of information is rebalanced victimization price responsive classification that is Naive Bayes rule. The naive classifier works based on Bayes rule of probability and it accepts all attributes that contained in dataset, it takes some samples for creating classification. The outcomes are compared and also the models with the best results are exposed.**

***Keywords* : data mining, educational performance, classification.**

I. INTRODUCTION

There is a rapid growth in the computer and network technologies in recent years. In this technology, numeric data's also made available in the current time and it show the fast growth in this field. This type of technologies is simple to gather and provisions in a huge quantity of unstructured or semi-structured text or data's are present in form of webpage's, HTML/XML archives, emails, and text files. And these copy information can be an idea with the great level text types of databases, it becomes significant to expand disciplined tackle to determine exciting knowledge or news from such data warehouses. There are numerous functions such as business management and market investigation; it can be benefits with knowledge and information extracted from a huge amount of data or text. Data mining is therefore a necessary step in gathering of information and discovery in large vast data warehouse.

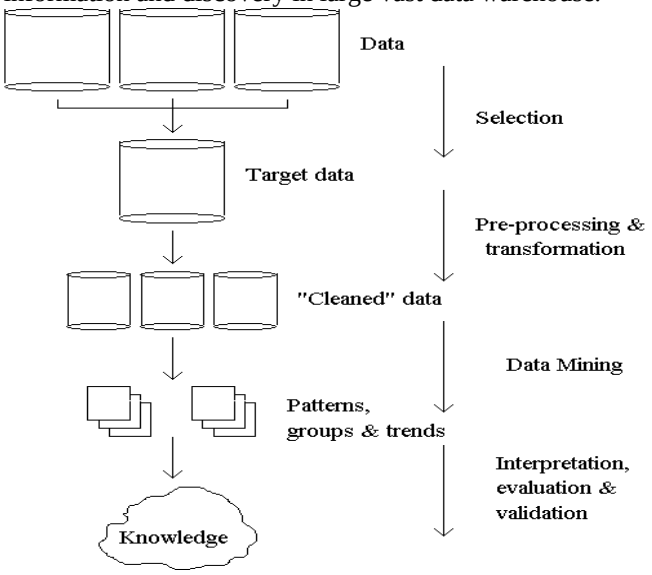

Figure1: Sample Flow Diagram of Data Mining.

Our contribution in this paper, we develop an efficient data mining technique to Predicting educational performance of a student failure and dropout from the school and colleges. In this paper we propose, it implements white-box classification methods, like decision trees and induction rules. Decision tree could be a decision support tool that represented as like graph or a model of decision. It consists of nodes, in which the internal nodes are denoted as test on attributes. Attribute is nothing but real data of student that collected from college in middle or educational activity. A path from root to leaf is represents classification rules and it consists of 3 forms of nodes, which includes decision node, probability node and finish node. It can be used in verdict examination. Using this method, try to boost their correctness for computing the students may not pass or dropout by first; with all accessible characteristics next and then choosing best attributes. Attribute selection is done by Java programing

language. Hence the data processing tool mainly works in prediction and classification of knowledge. Java programing language supports much normal data processing task information pre-processing, clustering, classification and have choice of information is rebalanced victimization price responsive classification that is Naive Bayes rule. The naive classifier works based on Bayes rule of probability and it accepts all attributes that contained in dataset, it takes some samples for creating classification.

II. LITERATURE REVIEW

In this section, we briefly discuss the works which is similar techniques as our approach but serve for different purposes.

Loretta Auvi, Anthony Don, Ben Shneiderman, Elena Zheleva, Catherine Plaisan, Machon Gregory, Tanya Clement, and Sureyya Tarkan in this paper the author proposed about the Feature Lens, visualize a text or data compilation at numerous stages of granularity and facilitate the consumers to discover interesting text or data patterns in the data warehouse. The current accomplishment focuses on everyday entry sets of n-grams, as they incarcerate the replication of accurate or comparable terminology in the compilation. Users can locate meaningful co-occurrences of data patterns or text by envisaging them within and transversely documents in the text collection in the databases. This also consents the users to recognize the sequential progression of tradition such as goes up and down or sudden appearance of text prototypes. The boundary could be worn to discover other copy features as fine.

Ah-Hwee Tan proposed data or text mining, It is also known as text data mining or knowledge discovery. From textual databases refers to the procedure of removing interest and significant model or knowledge from copy documents. There is a fast growth in the computer and network technologies in recent years. In this technology, numeric data's also made available in the current time and it show the fast growth in this field. This critique challenges to shack some lights to the query. A text mining structure involves two components: unstructured text documents transform into intermediate form by using text refining and knowledge sanitization that deduces patterns or knowledge from the intermediate form. In conclusion, we emphasize the upcoming challenges of text mining and the opportunities it offers.

M. Rajman, and R. Besancon, proposed the common framework of knowledge discovery, This type of technologies is simple to gather and provisions in a huge quantity of unstructured or semi-structured text or data's are present in form of webpage's, HTML/XML archives, emails, and text files. And these copy information can be an idea with the great level text types of databases, it becomes significant to expand disciplined tackle to determine exciting knowledge or news from such data warehouses.

In Collaborative data publishing more providers. It desire to calculate an identify view of their data. Without disclosing any responsive and personal information. In a single data source setting the problem of inferring information from recognized data have been generally studied. A data recipient, i.e., an attacker, for example, P0, attempts to infer background knowledge, BK and additional information about data records using the published data, t*. We regard as the shared data publishing setting among level partitioned data diagonally numerous data providers, Each contributing a separation of records Ti. Managing and analyzing a huge number of low-level alerts is very difficult and wearing for system administrators. To decrease the no.of alerts. And create more comprehensible the alert correlation methods has been proposed. Alert correlation methods are different in terms of their correctness, presentation and adaptiveness. In this, we are presenting an repeated annotation advance that initially aligns the data units on a result page into different groups. Such that the data in the one collection is having the equal semantic. Now, we can understand it from special aspects. And then aggregate the different comments to predict an last explanation label for each set.

In text categorization (TC) the grouping of bigram and unigrams was preferred for text indexing and evaluated on a selection of attribute estimation function. For Web document management a phrase-based text representation was also proposed. Thinking about the difficulty of the information centre location, This document explores the personality of knowledge based agents to find out data removal opportunities. Within these active databases. The purpose is to main focus on the most important problems, at the same time, produce a number of flexibility. In which undesired metrics are suitable in features of their negative and positive impact analysis from an power view the likely reason was that a phrase-based method had "lower text frequency for conditions and lower stability of project " as mentioned .

The term-based ontology removal methods are also providing various view for text representations. For e.g, hierarchical clustering was used to find out hyponymy and synonymy relatives among keywords. And also, in arrange to improve the performance of term-based ontology mining, the pattern evolution technique was introduced.

For web databases annotation process there are many existing number of techniques proposed. mainly work has focused on considering the information recipient as an attacker and a single data provider setting. A huge body of text assumes limited background knowledge of the attacker, and defines privacy by considering specific types of attacks using relaxed adversarial notion. Representative philosophy include ldiversity, k-anonymity, and t-closeness. Some of recent works have studied perturbation techniques under these syntactic privacy notions and modelled the instance level background knowledge as corruption.

Recently, a new concept-based these information units are prearranged dynamically into result pages for individual browsing and converted into machine process able unit and assigned meaningful labels. The encoding of information units requires lot of human efforts to annotate data units manually. Thus, lack in scalability. To overcome this, automatic assigning of data units within the SRRs is required. Annotated in different aspects and aggregated to predict a final label. Finally, a wrapper is constructed.

Wrappers are commonly used as translators which annotate new result pages from the same web database. This automatic annotation approach is highly effective and more scalable.

III. PROPOSED SYSTEM

Data mining methods are applied to predict college failure and dropout of the student. This project is used for real data on college students for prediction of failure and dropout. It implements white-box classification methods, like decision trees and induction rules. Decision tree could be a decision support tool that represented as like graph or a model of decision. It consists of nodes, in which the internal nodes are denoted as test on attributes. Attribute is nothing but real data of student that collected from college in middle or educational activity. A path from root to leaf is represents classification rules and it consists of 3 forms of nodes, which includes decision node, probability node and finish node. It can be used in verdict examination. Using this method, try to boost their correctness for computing the students may not pass or dropout by first; with all accessible characteristics next and then choosing best attributes. Attribute selection is done by Java programing language. Hence the data processing tool mainly works in prediction and classification of knowledge. Java programing language supports much normal data processing task information pre-processing, clustering, classification and have choice of information is rebalanced victimization price responsive classification that is Naive Bayes rule. The naive classifier works based on Bayes rule of probability and it accepts all attributes that contained in dataset, it takes some samples for creating classification. The models with the best results are also exposed by comparing the outcomes.

IV. EXPERIMENTAL RESULTS

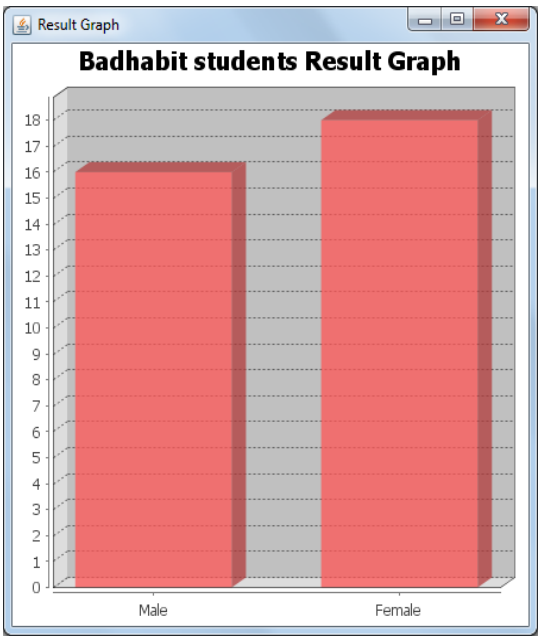

Figure2:Badhabit students result graph

Above diagram shows graph which specifying the mined data graphically from college information by applying navie bayes algprithm. Here we can observe that students who are with bad habits by different bading male and female based on given data.

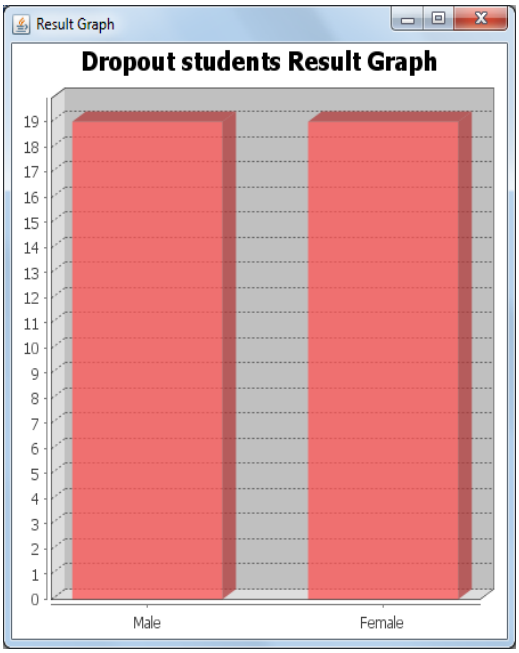

Figure3:Dropout students result graph

Above diagram shows graph which specifying the mined data graphically from college information by applying navie bayes algprithm. Here we can observe that students dropouts by male and female differentiation based on given input data.

CONCLUSION

By observing above results we can conclude that, the envisage student stoppage at college can be a complicated task not merely because it is a multifactor difficulty but also because the available data is usually imbalanced. We proposed effective technique in this paper for to predict educational performance of a student failure and dropout from the colleges based on attribute is nothing but real data of student that collected from college in middle or educational activity. A path from root to leaf is represents classification rules and it consists of 3 forms of nodes, which includes decision node, probability node and finish node. It can be used in verdict examination. Using this method, try to boost their correctness for computing the students may not pass or dropout by first; with all accessible characteristics next and then choosing best attributes. Attribute selection is done by Java programing language. Hence the data processing tool mainly works in prediction and classification of knowledge. Java programing language supports much normal data processing task information pre-processing, clustering, classification and have choice of information is rebalanced victimization price responsive classification that is Naive Bayes rule. Our proposed approach works efficiently when compared to other previously approached schemes.

REFERENCES

[1] A. Salguero, C. Roldán, and F. Araque, "Factors influencing university drop out rates,"Comput. Educ., vol. 53, no. 3, pp. 563–574, 2009.

[2] L. A. Alvares Aldaco, "Comportamiento de la deserción y reprobación en el colegio de bachilleres del estado de baja california: Caso plantel ensenada," inProc. 10th Congr. Nat. Invest. Educ., 2009, pp. 1–12.

[3] S. Ventura and C. Romero, "Educational data mining: A survey from1995 to 2005,"Expert Syst. Appl., vol. 33, no. 1, pp. 135–146, 2007.

[4] N. V. Kalyankar and M. N. Quadril, "Drop out feature of student data for academic performance using decision tree techniques," Global J. Comput. Sci. Technol., vol. 10, pp. 2–5, Feb. 2010.

[5] S. Kotsiantis, K. Patriarcheas, and M. Xenos, "A combinational incremental ensemble of classifiers as a technique for predicting students'performance in distance education,"Knowl. Based Syst., vol. 23, no. 6, pp. 529–535, Aug. 2010.

[6] C. Romero and S. Ventura, "Educational data mining: A review of thestate of the art,"IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40,no. 6, pp. 601–618, Nov. 2010.

[7] S. Kotsiantis, "Educational data mining: A case study for predicting dropout—prone students," Int. J. Know. Eng. Soft Data Paradigms, vol. 1, no. 2, pp. 101–111, 2009.

[8] A. Dapena-Janeiro, J. Más-Estellés, A. Valderruten-Vidal, R. Satorre-Cuerda, R. Alcover-Arándiga, F. Llopis-Pascual, R. Mayo-Gual, T. Rojo-Guillén, M. Bermejo-Llopis, E. Menasalvas-Ruiz, J. Gutiérrez-Serrano, E. Tovar-Caro, and J. García-Almiñana, "Rendimiento académico de los estudios de informática en algunos centros españoles," in Proc. 15th Jornadas Enseñanza Univ. Inf., Barcelona, Rep. Conf., 2009, pp. 5–12.

[9] A. Parker, "A study of variables that predict dropout from distance education, "Int. J. Educ. Technol., vol. 1, no. 2, pp. 1–11, 1999.

[10] I. Giannoukos, I. Lykourentzou, V. Nikolopoulos, V. Loumos, and G. Mpardis, "Dropout prediction in e-learning courses through the combination of machine learning techniques,"Comput. Educ., vol. 53, no. 3, pp. 950–965, 2009.

[11] A. León and E. Espíndola, "La deserción escolar en américa latina: Un Tema prioritario para la agenda regional,"Revista Iberoamer. Educ., vol. 1, no. 30, pp. 39–62, 2002.

[12] T. Aluja, "La minería de datos, entre la estadística y la inteligen-cia artificial," Quaderns d'Estadística Invest. Operat., vol. 25, no. 3, pp. 479–498, 2001.

[13] M. M. Hernández, "Causas del fracaso escolar," in Proc. 13th Congr. Soc. Española Med. Adolescente, 2002, pp. 1–5.

[14] I. H. Witten and F. Eibe,Data Mining, Practical Machine Learning Tools and Techniques, 2nd ed. San Mateo, CA, USA: Morgan Kaufman, 2005.

[15] M. A. Hall and G. Holmes, "Benchmarking attribute selection techniques for data mining," Dept. Comput. Sci., Univ. Waikato, Hamilton, New Zealand, Tech. Rep. 00/10, Jul. 2002.