



Psychological Impact on Student's Behavior

Vrushali Ambadas Sungar¹

¹Lecturer, Dr.D.Y.Patil School of Engineering School, Polytechnic Lohegaon Pune

ABSTRACT

Now a day's choosing a career out of plenty of options available today has become a difficult task for the students. It is important to consider ones interest, talent, projected growth and sustainability in a particular career before choosing it. It is very commonly seen that many students are rushing towards a specific career path just because some of their relative, friends has done the same or their parents have told them. Also there are few who find the career option very attractive and take it up. And once the student is into that career it results in poor academic record and finally he/she ends up switching to another career option. This causes waste of money and precious time of one's career. Also it demoralizes a candidate as his/her effort results into no success. So it is very important to make the correct choice in the first place itself. These paper survey application of psychological impact on student and also present the analysis result with weka tool. Large amount of data use in video streaming, stock market, bank system, railway reservation so we required to get correct data and find relationship so different data mining technique are developed and used. There are number of popular data mining task in psychological dataset. In these we have use two learners to find error rate on same dataset to achieve these we have use WEKA tool.

Keywords: Data Mining, Weka

1. Introduction

Data mining is a way of presentation of data in correct format which is extracting from raw data. Due to large data base used into day to day life so we need to classify data to our understandability. In some situation, large dataset and dimension of dataset is also large that type of data we will not able to store on cloud. To reduce dimension of dataset, increase the performance of classification we have different algorithms are available in data mining. There are effective tools present in data mining to get correct results. In this paper, we have weka tool to check performance of student based on psychological test. Data mining is also called knowledge discovery process. Knowledge discovery process as following steps:

- **Data cleaning:** from dataset to remove noisy data
- **Data integration:** combine multiple data
- **Data selection:** data relevant to the analysis task are retrieved from the database
- **Data transformation:** data are transforming into correct form by performing summary
- **Pattern evaluation:** Identify the correct pattern
- **Knowledge presentation:** to represent data to use visualization

Data mining tools perform data analysis, data preprocessing, data extraction and data loading it will contributing greatly to business strategies, to present knowledge bases and scientific and medical research.

2. Weka

Weka workbench is a collection of state-of-the-art machine learning algorithm and data preprocessing tools. We can try flexibly new dataset on existing method that describe in weka. These tools provides extensive support for whole process of experimental data mining, including preparing that input data, evaluating learning schemes statistically and visualizing the input data and result of learning. The workbench method for the main data mining problem: regression, classification, clustering, association rule mining and attribute selection. All the input data save in ARFF format. Classification is the main problem in data mining. To reduce large scale data we can apply dataset on learner like support vector machine, regression model, navies method.

In weka, we can apply input dataset on method and analyze its output to learn more about data. Another way is that to apply learned model to newly arrived cases. Third method is to apply several different learners and compare their performance in order to choose one for predication. WEKA is a data mining system developed at the University of Waikato and has become very popular among the academic community working on data mining. We have chosen to develop this system in WEKA as we realize the usefulness of having such a classifier in the WEKA environment [1].

Machine learning learn the different input based upon that output will be given by system. Weka is collection of machine learning algorithm used to solve real world data mining problem. Most of method is written in java and runs on any platform. The algorithm can be applied to dataset or called from your own java code.

The original non-Java version of Weka was a TCL/TK front-end to (mostly third-party) modeling algorithms implemented in other programming languages, plus data preprocessing utilities in C, and a Make file-based system for running machine learning experiments. This original version was primarily designed as a tool for analyzing data from agricultural domains, but the more recent fully Java-based version (Weka 3), for which development started in 1997, is now used in many different application areas, in particular for educational purposes and research. Advantages of Weka include:

- I. Free availability under the GNU General Public License
- II. Portability, since it is fully implemented in the Java programming language and thus runs on almost any modern computing platform
- III. A comprehensive collection of data preprocessing and modeling techniques
- IV. Ease of use due to its graphical user interfaces [2]

In these paper is collected information and survey of result about 100 students on ten attribute it is showing that the on which factor student is better and depend upon total score overall performance of student is generated. There are new rules and relation between selected instances such as evaluation score. Now a day student has lot of stress on their mind so mental ability test will be help them to reduce or to made work load less the following parameters which will help them.

3. METHODS

In machine learning, there are different method is defined. In data mining, classification is main problem. To classify the one instance one or more than cases is known as multi classification problem. Classification is the process to divide into training dataset and testing dataset. In training dataset having know value with actual parameter whereas in testing dataset having to predicate correct value from unknown dataset. The input in the form of concepts instances or attribute. Each instance have special characteristics. There are different types of attributes that is numeric and nominal or categorical. Association rules differs from classification rule in two ways: they cannot predicate class but attribute can and they can predicate more than one attribute's value at a time. These entire instances apply on some data mining methods. To perform successful implementation we require a sound methodology built on best practice. In this study, data mining have following six processes: [3]

- **Problem description:** understand problem statement first then with project goal transforming these goal to problem description and reach to desired result.
- **Understanding Data:** to handle the problem firstly identify data
- **Preparing Data:** to involve data cleaning, transformation and extraction
- **Creating model:** Create model using comparative result
- **Evaluating the models:** to check validity and the utility of the model against each other and goal also.
- **Using the model:** Use the creating model for future decision.

3.1. DATASET

In this study 40 dataset were used which has 10 attribute are numeric type. Dataset have student information Self Awareness, Empathy, Self-Motivation, Mental Stability, Managing Relation, Integrity, Self Development, Value Orientation, Commitment, Altruistic Behavior, Total Score.

Table 1: The list of independent variables used in this study

Sr no	Attribute Name	Type
1	Self Motivation	Numeric
2	Mental Stability	Numeric
3	Managing Relation	Numeric

4	Integrity	Numeric
5	Self Development	Numeric
6	Value Orientation	Numeric
7	Commitment	Numeric
8	Altruistic Behavior	Numeric
9	Sex	String
10	Class	Sting

Table 2. The list of independent variables and values used in this study

Sr no	Attribute Name	Type
1	Sex	Female, Male
2	Class	Class1,Class2,Class3

Table 3: The output variable (Evaluation score) used in the study

Factors	A (Self Awareness)	B (Empathy)	C (Self Motivation)	D (Emotional Stability)	E (Managing Relations)	F (Integrity)	G (Self Development)	H (Value Orientation)	I (Commitment)	J (Altruistic Behavior)
High	11 and above	15 and above	18 and above	11 and above	12 and above	8 and above	6 and above	6 and above	6 and above	6 and above
Normal	4 to 10	7 to 14	9 to 17	4 to 10	5 to 11	4 to 7	2 to 5	2 to 5	2 to 5	2 to 5
Low	3 and Below	6 and below	8 and below	3 and below	4 and below	3 and below	1 and below	1 and below	1 and below	1 and below

3.2. CLASSIFICATION

Classification ion has two parts training dataset and testing dataset that maps knowledge into predefined group and classes. It is also called as supervised learning. It consists of two parts:

3.1. Construction of Models: It is collection of set of predetermined classes. Each attribute is assumed to belong to a predefined class. The set of attribute used for model construction is training set. The model is represented as classification rules, decision trees, or mathematical formulae.

3.2. Usage of Model: for future data classified using this model. In testing model unknown sample is compared with known sample. set is independent of training set, otherwise over-fitting will occur

4. RESULT

4.1. Exploring weka

When you will open weka tool, weka GUI chooser open in front of user. In Weka GUI Chooser start with Explore, Experimenter, Knowledge Flow, Simple CLI. After click on Explore option weka explore window will open. Weka window have again different option. If we want to open file then we will select open file tab. Weka preprocess have ARFF file format.

4.2. ARFF File Format

First all the student dataset is made into excel file format. In next step we will convert excel sheet into arff format using converter.

```
@relation student
@attribute A (Self Awareness) real
@attribute B (Empathy) real
@attribute C (Self Motivation) real
@attribute D (Emotional Stability) real
```

- @attribute E(Managing Relations) real
- @attribute F (Integrity) real
- @attribute G (Self Development) real
- @attribute H (Value Orientation) real
- @attribute I (Commitment) real
- @attribute J (Altruistic Behavior) real
- @attribute Total Score real
- @attribute Sex { F,M }

4.3. Opening dataset

After psychological dataset of student load it shown in figure 1. Weka tool has total six tabs namely Preprocess, Classify, Cluster, Associate, Select Attribute and visualize. Dataset has been loaded successfully we can see the history of dataset with the help of visualize tab which is shown in Figure2

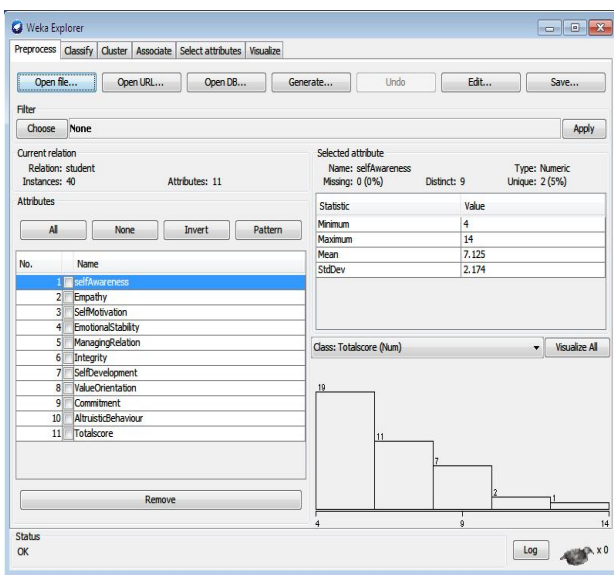


Fig 1: Explorer window

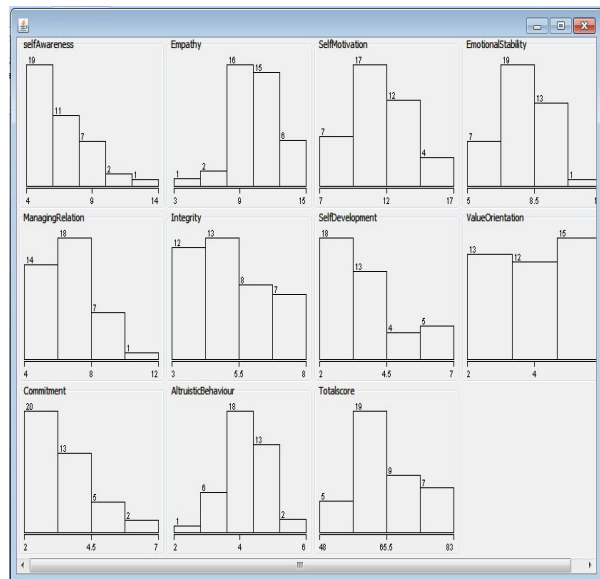


Fig 2: Visualization Window

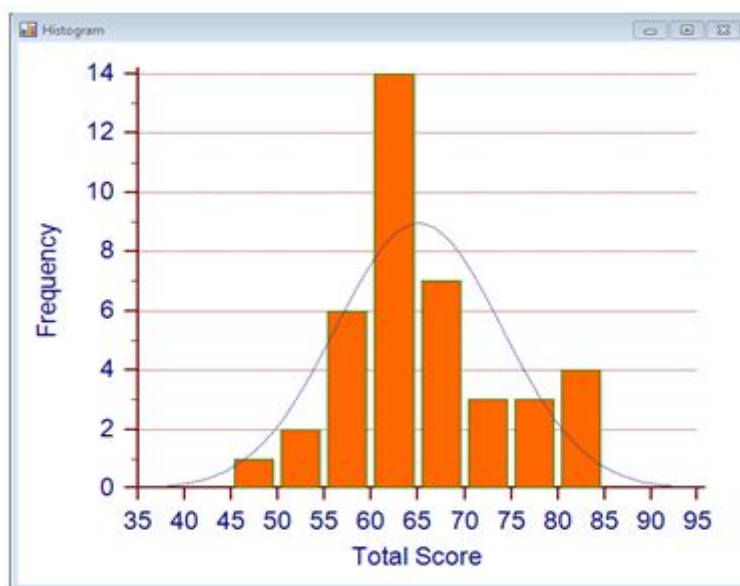


Fig3: Histogram of Total Score

4.4. Choose Classifier

On these dataset we have applied ZeroR is the simplest classification method which relay on class and predicate majority class. It is useful for determining a baseline performance as a benchmark for other classification methods. The following table shows that combined result on training dataset. The following two tables show that difference in mean absolute error and root mean squared error. For testing dataset we have taken 20 instance with 11 attribute.

Table 4: Summary table of ZeorR and Decision Model on training dataset

Classifier Model	Type	Values
ZeroR	Correlation coefficient	0
	Mean absolute error	7.09
	Root mean squared error	8.7931
	Relative absolute error	100 %
	Root relative squared error	100%
Decision Model	Correlation coefficient	0.7182
	Mean absolute error	5.0663
	Root mean squared error	6.1183
	Relative absolute error	71.456 3 %
	Root relative squared error	69.579 9 %

Table 5: Summary table of ZeorR and Decision Tree Model on testing dataset

Classifier Model	Type	Values
ZeroR	Correlation coefficient	0
	Mean absolute error	5.4725
	Root mean squared error	7.3074
	Relative absolute error	100 %
	Root relative squared error	100%
Decision Model	Correlation coefficient	0.703
	Mean absolute error	4.5583
	Root mean squared error	5.2603
	Relative absolute error	83.2953 %
	Root relative squared error	71.9866

5. DISCUSSION

Weka provides to user well platform for using classification, clustering, association rule on their dataset and weka is the platform independent. ZeroR is the simplest classification method which relies on the target and ignores all predictors. Construct a frequency table for the target and select its most frequent value. Whereas decision model is useful for classification and regression. The goal is to create a model that predicts the value of a target variable based on several input variables. In the table 4, the relative absolute error and root relative squared error also is 100% in ZeroR classifier for training dataset. But in decision model these error is only 45% and 52%. So the decision model gives good result as compare to ZeroR.

6. CONCLUSION

Data mining is the non trivial process which extracts knowledge from data. Classification rule and association rule totally opposite to each other. Classification is main problem of data mining. To classify the instances we require classification methods. Here we use ZeroR and Decision Tree classifier for comparison between different type of errors during testing and training dataset. In data mining all other supervised learning available for classification neural network, support vector machine..

References

[1] Shilpa Dhanjibhai Serasiya, Neeraj Chaudhary “Simulation of Various Classifications Results using WEKA”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-3, August 2012

[2] Dr. Sudhir B. Jagtap, Dr. Kodge B. G.” Census Data Mining and Data Analysis using WEKA”, (ICETSTM – 2013) International Conference in “Emerging Trends in Science, Technology and Management-2013, Singapore.



- [3] C. Shearer, "The CRISP-DM model: The new blueprint for data mining" *Journal of Data Warehousing*, (2000). 5: 13-22.
- [4] Bharat Chaudhari, Manan Parikh, "A Comparative Study of clustering algorithms Using weka tools", *International Journal of Application or Innovation in Engineering & Management (IJAIEM) ISSN 2319 - 4847*, Volume 1, Issue 2, October 2012.
- [5] Sunita B Aher, Mr. LOBO L.M.R.J. "Data Mining in Educational System using WEKA", *International Conference on Emerging Technology Trends (ICETT) 2011 Proceedings published by International Journal of Computer Applications® (IJCA)*.
- [6] Roozbeh Razavi-Far, Member, IEEE, Piero Baraldi, and Enrico Zio, Senior Member, IEEE, "Dynamic Weighting Ensembles for Incremental Learning and Diagnosing New Concept Class Faults in Nuclear Power Systems", *IEEE TRANSACTIONS ON NUCLEAR SCIENCE*, VOL. 59, NO. 5, OCTOBER 2012.
- [7] Haibo He, Senior Member, IEEE, Sheng Chen, Student Member, IEEE, Kang Li, Member, IEEE, and Xin Xu, Member, IEEE, "Incremental Learning from Stream Data", *IEEE TRANSACTIONS ON NEURAL NETWORKS*, VOL. 22, NO. 12, DECEMBER 2011 1901.
- [8] Geoffrey Holmes, Richard Kirkby, Bernhard Pfahringer, "A batch -incremental learning for mining data streams"
- [9] Devi Parikh and Robi Polikar, Member, IEEE, "An Ensemble-Based Incremental Learning Approach to Data Fusion", *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS*, VOL. 37, NO. 2, APRIL 2007.
- [10] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 3, pp. 418–435, May/Jun. 1992.
- [11] K. Woods, W. P. J. Kegelmeyer, and K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 405–410, Apr. 1997.
- [12] L. I. Kuncheva, J. C. Bezdek, and R. P. W. Duin, "Decision templates for multiple classifier fusion: An experimental comparison," *Pattern Recognit.*, vol. 34, no. 2, pp. 299–314, 2001.
- [13] *Pattern Anal. Mach. Intell.*, "A theoretical study on six classifier fusion strategies," *IEEE Trans.* vol. 24, no. 2, pp. 281–286, Feb. 2002.
- [14] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Comput.*, vol. 3, no. 1, pp. 79–87, 1991.
- [15] M. J. Jordan and R. A. Jacobs, "Hierarchical mixtures of experts and the EM algorithm," *Neural Comput.*, vol. 6, no. 2, pp. 181–214, 1994.
- [16] H. Drucker, C. Cortes, L. D. Jackel, Y. LeCun, and V. Vapnik, "Boosting and other ensemble methods," *Neural Comput.*, vol. 6, no. 6 pp. 1289–1301, 1994.
- [17] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992
- [18] B. V. Dasarthy and B. V. Sheela, "Composite classifier system design: Concepts and methodology," *Proc. IEEE*, vol. 67, no. 5, pp.