ISSN 2320 - 2602
Volume 11 No. 7, July 2022
International Journal of Advances in Computer Science and Technology
Available Online at http://www.warse.org/IJACST/static/pdf/file/ijacst011172022.pdf
https://doi.org/10.30534/ijacst/2022/011172022

Decision making Assessment model for University admission

Dr.Nikhat Khan
Institute of Company Secretaries of India
nikhatk2004@yahoo.com

Received Date : June 04, 2022 Accepted Date : June 28, 2022 Published Date : July 07, 2022

ABSTRACT

Educational data mining has the potential to enhance the decision making by providing services to academicians. The students are confused when they aspire to get admission in selected colleges. Depending upon the university ranking the students get admission based on certain parameters. This paper presents a decision making assessment model that relies on combination of Distillation of Data through Human Judgment and fuzzy inference systems (FISs) to accommodate imprecise data using fuzzy rules after applying data attribute reduction to overcome uncertainty and complexity in the assessment model. This model was implemented using Weka and Matlab software.

Key words: Data Mining Techniques, Data Preprocessing, Education Data Mining (EDM), Fuzzy Inference System.

1. INTRODUCTION

The present day world has undergone a remarkable transformation in the past few decades and the education system has also accordingly changed a lot. Students aspire to get admissions in reputed colleges. Students are targeting universities leaving them confused about their coursework selection. In such a situation they need an assessment model that is helpful to fulfil their requirements. For enhancing teaching and learning Educational data mining should be included and there should be strong collaboration across research, commercial, and educational sector For this educational data has be collected , stored ,mined and then analyzed.

Machine learning research has yielded techniques for knowledge discovery and data mining that extracts novel and potentially useful information from large amounts of unstructured data. These techniques find patterns in data and then helpful in building predictive models that probabilistically is helpful in predicting an outcome. Applications of these developed models can then be used in computing analytics and decision making over large datasets. The Educational data mining process transforms raw data obtained from educational systems into potentially useful information that could have a remarkable impact on educational research and practice. This process is similar to other application areas of data mining like manufacturing, banking , business, genetics, medicine, etc. because it is based on similar steps as the general process of data mining: pre-processing, data mining and post-processing.

Educational data mining makes use of variety of methods to analyze data. Supervised model is a machine learning technique which provides inference of prediction models from training instances in which the target attribute values are known. Unsupervised model machine learning technique infer models from training instances in which the target attribute values are not known. Parameter estimation is a statistical techniques for inferring parameters of probabilistic models from the available data. Relationship mining is comprised of the identification of relationships between variables –these relationships may be associative, sequential, correlational or causal in nature. Distillation of data through human judgment comprises of representing data in intelligible ways using statistics, visualization methods and interactive information interfaces.

2. LITERATURE SURVEY

Fuzzy Sets and Numbers

Zadeh was the first who introduced the fuzzy logic concept to deal with vagueness, uncertainty or imprecision in problems [1]. Fuzzy set theory, which is based on fuzzy logic, assigns a degree of membership to a particular object or a variable in a given set [1] [8].

Definition 1 (Fuzzy set): Fuzzy set A is defined in terms of a universal set X by a membership function that assigns a value $\mu A(x)$ in the interval [0,1], i.e., $A: X \rightarrow [0,1]$ to each element $x \in X$ [1][8].

Definition 2 (Fuzzy number): A fuzzy set A in $\mathbb{R}$ satisfies the following conditions [2] [8].:

- A is normal.
- $A\alpha$ is a closed interval for every $\alpha \in (0,1]$.
- The support of A is bounded.

Definition 3 (Linguistic variable): A linguistic variable is a variable whose values are words or sentences in a natural or artificial language [3] [8]. A linguistic variable is characterized by the variables X, V, W, and M where:

- X is the name of the linguistic variable (e.g., success in admission).
- V is the set of linguistic values which X can take (e.g., {good, bad, very bad}).
- W is the physical domain from which X, the linguistic variable takes its crisp values (e.g., [yes, no]).
- M is a semantic rule that relates each linguistic value in V having a fuzzy set in W.

Definition 4 (FIS): An FIS has three parts: Fuzzification, a fuzzy inference engine, and defuzzification. The Fuzzy inference deals with the process of mapping from a given input to an output using the logic of fuzzy. This mapping further provides a basis through which decision making happens.. The fuzzy inference engine uses fuzzy logic operations to generate an output by considering the logical relationships between the input variables. The defuzzification process converts the fuzzy output set into crisp values [4].

There are several inference methods; however, the most commonly used methods in the fuzzy community are Mamdani [5] and Takagi and Sugeno [6].

3. PROBLEM DESCRIPTION AND ANALYSIS

This section describes the problem domain and data that is required to be investigated in this research.

Educational data mining involves developing new methods to discover knowledge from educational and academic database and it can be used further for decision making in educational and academic systems. It is useful in many different areas including identifying priority learning needs for different groups of students ,identifying at risk students, increasing graduation rates, effectively assessing institutional performance, maximizing campus resources, helpful in prediction of university admission and optimizing subject curriculum renewal

3.1 DATA PREPARATION AND MINING

The data set is obtained from website Kaggle.com. This research uses past educational data that is gained from Kaggle, a website that provides many kinds of datasets for machine learning and data scientists. It contains various factors attributed towards picking the right university. This dataset is created for prediction of Graduate Admissions from an Indian perspective [10]. It contains data of 100 different students. Data set is classified into 9 different parameters which are considered important during the application for Masters. Those parameters are: GRE scores, TOEFL scores, university rating, statement of purpose, letter of recommendation, CGPA, research paper, chance of admit.[7]. This research uses past educational data that is gained from Kaggle, a website that provides many kinds of datasets for machine learning and data scientists This raw data is then cleaned, transformed, and represented in some visualizations charts by using Weka.

Weka is an open-source data mining and visualization framework. Weka was developed at the University of Waikato, New Zealand [9]. Figure1 shows the user interface of Weka. This paper at initial stage uses Weka as a tool for data visualization and mining.

Figure axis labels are often a source of confusion. Use words rather than symbols. As an example, write the quantity “Magnetization,” or “Magnetization *M*,” not just “*M*.” Put units in parentheses. Do not label axes only with units. As in Figure 1, for example, write “Magnetization (A/m)” or “Magnetization ($A \cdot m^{-1}$),” not just “A/m.” Do not label axes with a ratio of quantities and units. For example, write “Temperature (K),” not “Temperature/K.”

Multipliers can be especially confusing. Write “Magnetization (kA/m)” or “Magnetization (10^3 A/m).” Do not write “Magnetization (A/m) × 1000” because the reader would not know whether the top axis label in Figure 1 meant 16000 A/m or 0.016 A/m. Figure labels should be legible, approximately 8 to 12 point type.

Figure 1: Weka Interface

3.1 Assessment model

In this paper we are intending to develop an assessment model by developing a fuzzy based prediction system in which the rules are framed by Distillation of Data through Human Judgment and also by finding the important attributes through best search method of attribute selection in Weka as shown in figure 2 below:

1) *Attribute selection using Attribute Evaluator Technique*

The raw machine learning data consists of a mixture of attributes. Out of these attributes some are relevant and helpful in making predictions. Through Weka it is possible to select those features in the available data that are most useful or most

relevant for the problem through an automated way. In machine learning this process is called feature selection. Further feature selection is divided into two parts:

- Attribute Evaluator
- Search Method.

The attribute evaluator is the technique by which each attribute in the available dataset known as a feature is evaluated in the context of the output variable .The search method is the technique through which different combinations of attributes in the dataset are taken in order to arrive on a short listing the chosen features.

Some Attribute Evaluator techniques require the use of specific Search Methods. For example, the CorrelationAttributeEval technique can only be used with a Ranker Search Method that evaluates each attribute and lists the results in a rank order. In this paper the attribute evaluator using best search method is used. It is observed that LOA, CGPA and Research are the important parameters that we have arrived at. Hence these parameters will form important rule base in our fuzzy inference system

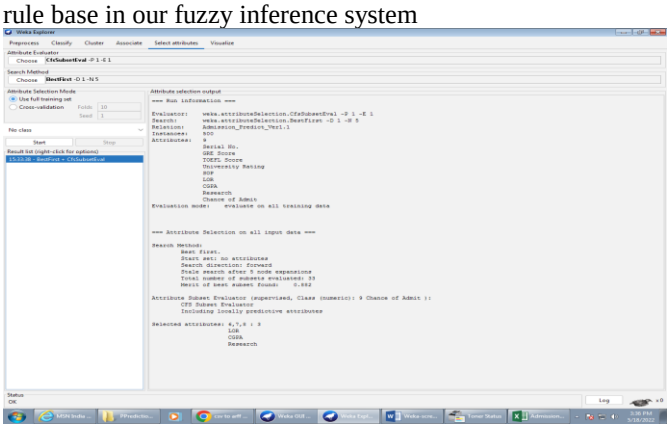


Figure 2: Attribute selection using Attribute Evaluator Technique using Weka tool.

2) Rule generation using classifier function of Weka tool.

In this we have used the M5-Rules Algorithm method of classification as depicted in Figure 3. The data modeling is done in WEKA environment. The dataset consists of 500 instances of past educational data.

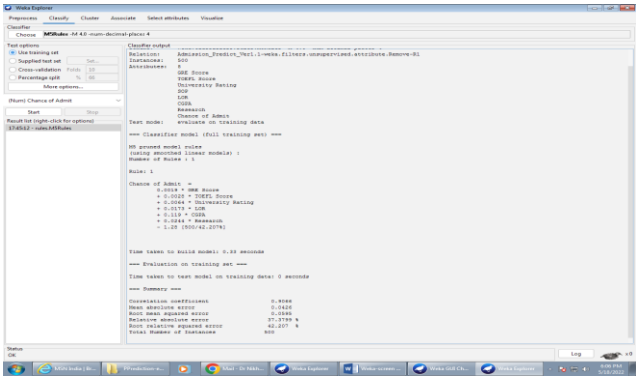
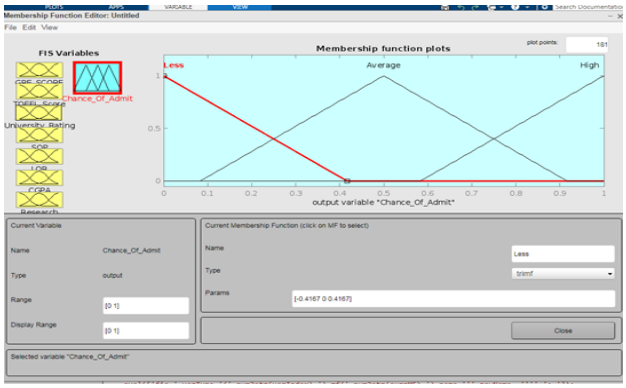


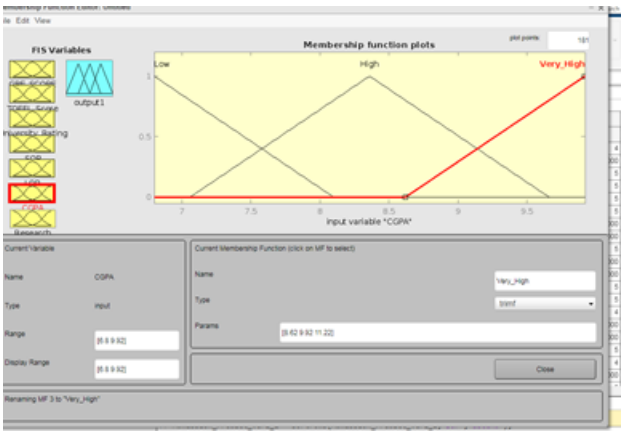
Figure 3: Rule generation using classifier function of Weka tool.

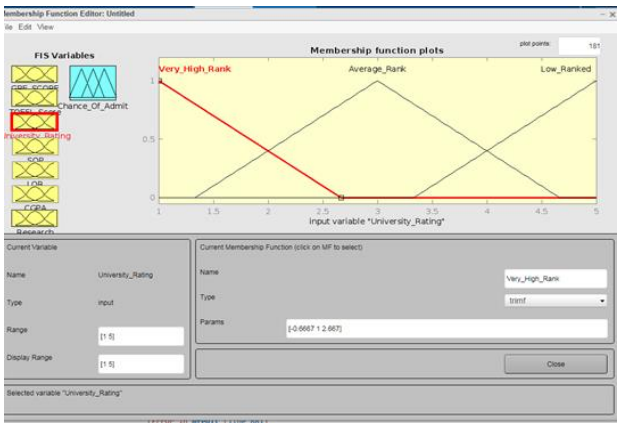
3) Developing the Fuzzy Inference System using Matlab tool.

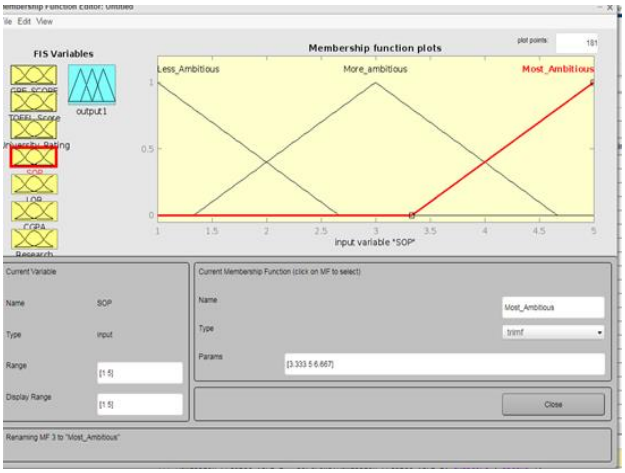
Subsequent to preprocessing of data one of the ways of predicting the admission is by developing a Fuzzy expert system which helps to classify large set of student data set .Fuzzy provide a good platform for making conditions or rules and we can easily implement this technique by using fuzzy inference system. These rules arise from various perspectives belonging to varying, overlapping or conflicting objectives. For instance some may give higher weightage to research work while others may not consider that parameter as important. Such decisions are based on experience, human intuition rather than precision of the data.

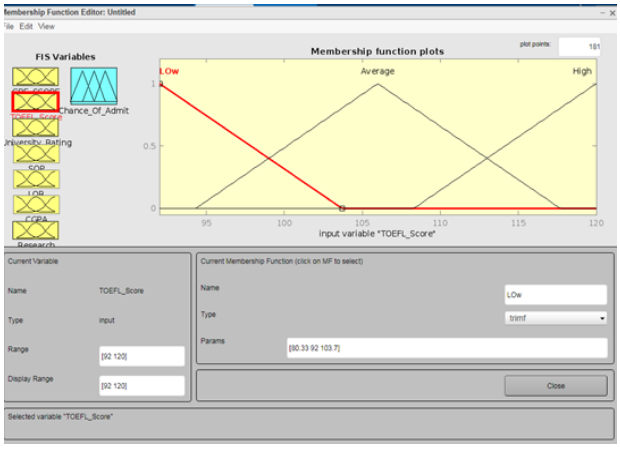
Distillation of Data for Human Judgment: This is an important area within educational data mining. Humans can make inferences about data in many cases, when things are beyond the scope of completely automated data mining methods. Data is distilled for human judgment in educational data mining for two key purposes: identification and classification. In our case we are using the same for framing rules of the fuzzy inference system. In addition to this the rules are framed based on the automated data mining as depicted in figure 2. When data is distilled for identification, data is displayed in ways that enable a human being to easily identify well-known patterns that are nonetheless difficult to formally express. In a fuzzy set theory the basic idea is that an element belongs to a fuzzy set with a certain degree of membership. The fuzzy set is a set with fuzzy boundaries. Because of which a proposition is neither True nor False, but may be partly true or false. This degree of partial truth is usually taken as a real number in the interval [0, 1]. As an example, the preferences for the CGPA score could be described in the Prediction system in fuzzy terms as ‘Low’, ‘High’ or ‘VeryHigh’ (as shown in Fig. 4). A fuzzy set A of universe X defined by a function ìA(x) is known as the Membership Function (MF): ìA(x) : X [0, 1]. Where ìA(x) = 1 if x is totally in A, ìA(x) = 0 if x is not in A and 0 < ìA(x) < 1 if x is partly in A.

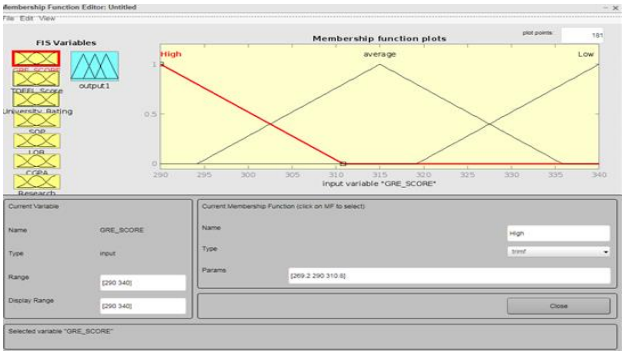


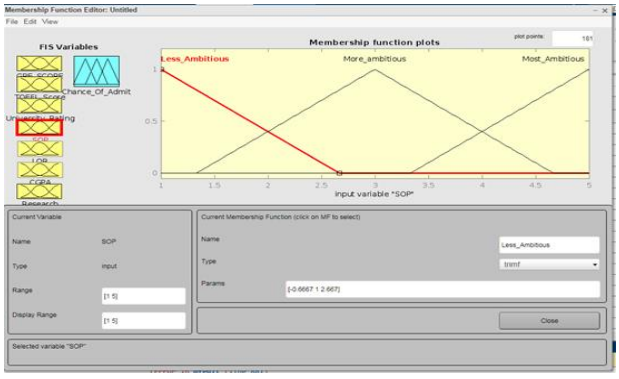

Membership function plots
FIS Variables
Low
High
Very_High
input variable "CGPA"

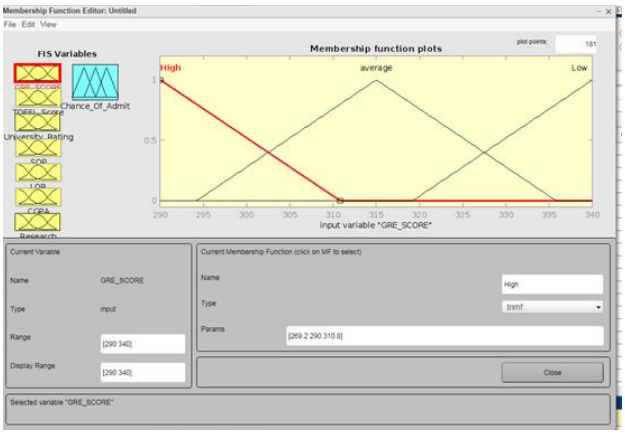

Membership function plots
FIS Variables
Very_High_Rank
Average_Rank
Low_Ranked
input variable "University_Rating"


Membership function plots
FIS Variables
Less_Ambitious
More_ambitious
Most_Ambitious
input variable "SOP"


Membership function plots
FIS Variables
LOw
Average
High
input variable "TOEFL_Score"


Membership function plots
FIS Variables
High
average
Low
input variable "GRE_SCORE"


Membership function plots
FIS Variables
Less_Ambitious
More_ambitious
Most_Ambitious
input variable "SOP"


Membership function plots
FIS Variables
High
average
Low
input variable "GRE_SCORE"

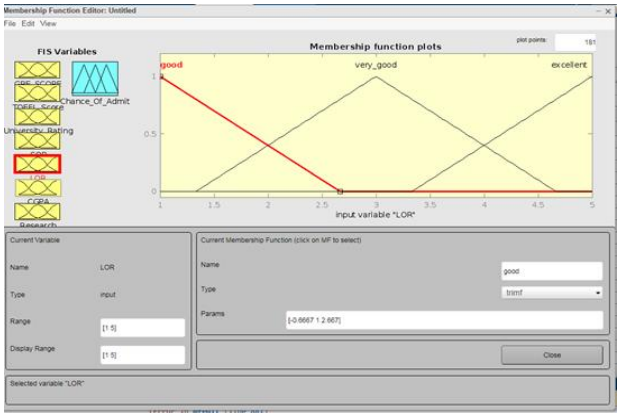

Membership function plots
FIS Variables
good
very_good
excellent
input variable "LOR"

Figure 4: Fuzzification using Matlab tool.

4) Defuzzification using Matlab tool.

Defuzzification is the method in which fuzzy quantity is converted into a precise quantity. The main important commonly used defuzzification method is that the center of area method (COA), which is also commonly referred as centroid method. This method determines the central area of fuzzy set and returns the corresponding crisp value. Defuzzification methods are used to convert the fuzzy values to their corresponding crisp values. There are five defuzzification methods. They are Centroid, Smallest of Maximum (SOM), Middle of Maximum (MOM), bisector and Largest of Maximum (LOM).

The defuzzification method used in the proposed work is centroid method as depicted in Fig 5.In this it returns the middle of area under the curve.

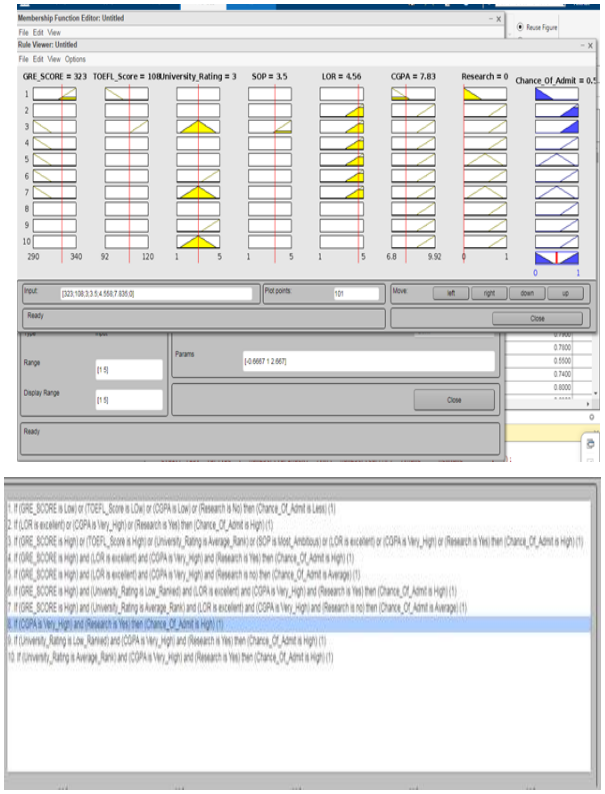


Figure 5: Defuzzification using Matlab tool.

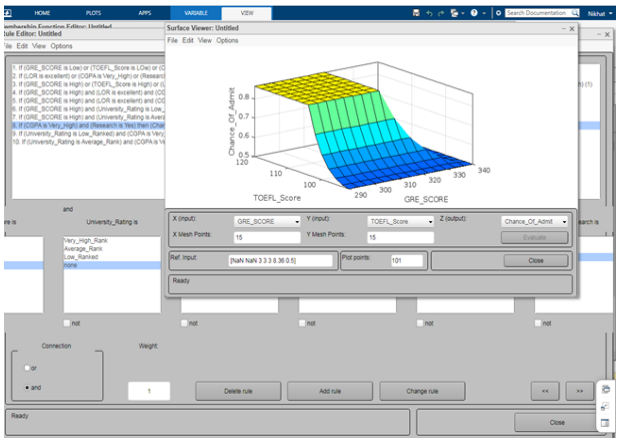


Figure 6: Defuzzification sample using Matlab tool.

The system has been tested by educational experts and one of the tested values has been considered which is depicted in Figure 6

If you wish, you may write in the first person singular or plural and use the active voice ("I observed that ..." or "We observed that ..." instead of "It was observed that ..."). Remember to check spelling. If your native language is not English, please get a native English-speaking colleague to proofread your paper.

4. CONCLUSION

The decision is made by the student using the assessment model tested by educational experts. Modelling through combination of mining techniques for framing fuzzy rules along with distillation of data using human judgment and further the imprecise nature of preferences of getting admission in suitable college renders fuzzy inference system an excellent tool for modeling those preferences. This helps in lessening the computational cost substantially when the candidate applies in the university in a more realistic way. Further neuro fuzzy system could also be explored for designing systems.

REFERENCES

1. L. A. Zadeh, "Fuzzy sets," Information and Control, vol. 8, no. 3, pp. 338-353, 1965.
2. M. Naderpour, J. Lu, and G. Zhang, "An intelligent situation awareness support system for safety-critical environments," Decision Support Systems, vol. 59, pp. 325-340, 2014.
3. L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," Information Sciences, vol. 8, no. 3, pp. 199-249, //, 1975.
4. A. S. Markowski, M. S. Mannan, A. Kotynia et al., "Application of fuzzy logic to explosion risk assessment," Journal of Loss Prevention in the Process Industries, vol. 24, no. 6, pp. 780-790, 2011.
5. E. H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic synthesis," IEEE Transactions on Computers, vol. C- 26, no. 12, pp. 1182-1191, 1977.
6. S. Nazir, S. Colombo, and D. Manca, "Testing and analyzing different training methods for industrial operators: an experimental approach," Computer Aided Chemical Engineering, K. Andrzej and T. Ilkka, eds., pp. 667-672: Elsevier, 2013.
7. Dr.Nikhat Khan " Study of techniques and applications of data mining in Education, International Journal of Application or Innovation in Engineering and Management.
8. Jay Daniel, Mohsen Naderpour, Member, IEEE, Chin-Teng Lin, Fellow, IEEE A Fuzzy Multilayer Assessment Method for EFQM https://ieeexplore.ieee.org/document/8481412
9. https://www.primescholars.com/articles/forex-data-analysis-using-weka-92645.html
10. https://www.kaggle.com/mohansacharya/graduate-admissions/tasks?taskId=459