Prediction of Student Admission using Fuzzy based Education Data Mining

Dr. Nikhat Khan

Institute of Company Secretaries of India
nikhatk2004[at]yahoo.com

Abstract: *Educational data mining is used to increase the learning outcome of the students. Due to the increased use of technology in educational systems large amount of data is collected which makes it imperative to use EDM for improving the teaching and process of learning. This paper surveys the relevant studies in the educational field and includes the data and methodologies used in those studies. In this paper the fuzzy systems have been used for solving the problem of student aspiring to study in certain university. The student has to apply in universities which are a costly affair so to solve that problem and help students in decision making of applying in the shortlisted universities this fuzzy inference system has been designed and modeled.*

Keywords: Fuzzy Inference System, Education Data Mining (EDM), Data Preprocessing, Data Mining Techniques

1. Introduction

Educational data mining (EDM) is a method for extracting useful information that is potentially useful for an organization or for the students. EDM is useful in many different areas including identifying priority learning needs for different groups of students, effectively assessing institutional performance, identifying at - risk students, increasing graduation rates, maximizing campus resources, and optimizing renewal of subject curriculum. Educational data mining can be used to attract, maintain and retain the students to achieve the profitability of University. Analyzing students' data is crucial for discovering, detecting and understanding which instructional practices are effective. Data mining techniques are used to extract useful knowledge from raw data. The extracted knowledge is valuable and significantly affects the decision maker [1].

In the analytically mature organization, analytics is a way of life and data is considered as a very valuable asset. The data can be used for solving various problem and doing the analysis of the same using various data techniques which helps in decision making. If there is no mathematical model then fuzzy logic can be used for solving a problem.

The fuzzy system requires linguistic rules instead of learning examples as prior knowledge. In Fuzzy system the input and output variables have to be described linguistically. If the knowledge is incomplete, wrong or contradictory, then the fuzzy system has to be updated. Since there is no formal approach for the same, the updating is performed in a heuristic way. In this paper the systematic study of introduction to fuzzy logic is done and then further the combination is applied on educational data to achieve better results.

Similarly Students aspiring to get admission in selected colleges are in a state of confusion when they need to decide to apply in universities. Depending upon the ranking of the universities the students get admission based on certain parameters like higher score in GRE, TOEFFEL or research experience. The Student scoring high on these parameters can aim for highly ranked university.

2. Basics of Fuzzy Inference System

The Fuzzy Logic is a concept introduced in 1965 by Lofti A. Zadeh in his research paper "Fuzzy Sets" [2]. . He is considered as the father of Fuzzy Logic. Fuzzy method is used for modeling the uncertainty in human reasoning. This is one of the techniques used in data mining which gives approximate result rather than exact result. The limitations of the crisp set are removed by the invention of fuzzy set. The Fuzzy logic provides the way to represent linguistic variable and subjective attributes of the real world in computing. Fuzzy provides an uncertain behavior and this uncertainty in fuzzy logic is represented by stochastic uncertainty and lexical uncertainty [7].

In fuzzy sets, each elements is mapped to [0, 1] by membership function. $\mu A: X \rightarrow [0, 1]$ Where [0, 1] means real numbers between 0 and 1 (including 0 and 1). In Fuzzy logic the fuzziness is best characterized by its membership function. Alternatively the membership function represents the degree of truth in fuzzy logic. In fuzzy logic the natural language expression IF antecedent THEN consequent is referred to as the Fuzzy IF - THEN rule base. Fuzzy Inference System is the basic unit of a fuzzy logic system which has decision making as its primary work. It uses the "IF…THEN" rules along with connectors "OR" or "AND" for framing essential decision rules. In the Fuzzy Inference system the output will always be a fuzzy set irrespective of input which could either be fuzzy or crisp. There would be defuzzification unit with FIS to convert fuzzy variables into crisp variables. The basic Fuzzy Inference System comprises of rule - base having IF - THEN rules, database which defines fuzzy sets membership function that are used in fuzzy rules. There is decision making unit which performs operation on rules. The conversion of crisp quantities into fuzzy quantities. Is done by Fuzzification Interface Unit. Finally the conversion of fuzzy quantities into crisp quantities is done by Defuzzification Interface Unit.
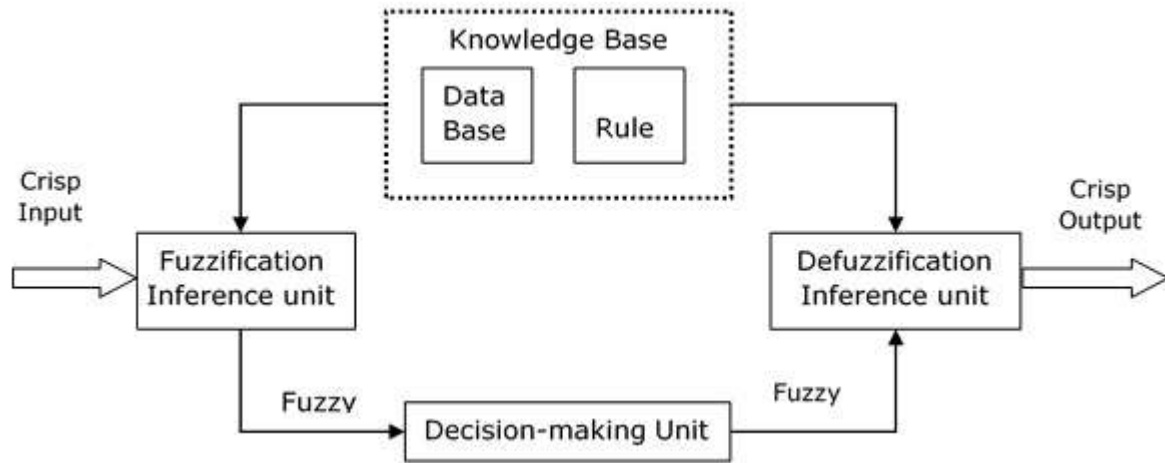
Figure 1: Diagram of basic Fuzzy Inference System

When knowledge is represented in the form of IF - THEN rules then Fuzzy systems can be applied to problems. The system is integrated with a Problem - specific a priori knowledge. Only heuristics are used in Fuzzy Systems. Training pattern set and system modeling are not required. This process requires the knowledge of experts. When the process of tuning happens, it is required to add, remove, or change a rule, or sometimes can change the weight of a rule.

3. DataSET

The data set which is obtained from website Kaggle. com [3]. It contains various factors attributed towards picking the right university. This dataset is created for prediction of Graduate Admissions from an Indian perspective. 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.

4. Methodology and Implementation

The primary step to building a model for our prediction system is choosing the right dataset. In this case for predictions, a dataset was chosen which contains all the important attributes that would affect the chances of admit. This is followed by data cleaning where the missing values are handled present in various fields. Once the data is ready to be analyzed, various tools and libraries are used to visualize the data and perform analysis. Fig [1] [2] [3] [4] [5] [6] [7] [8] [9] [10]. This includes visualizing bar graphs and the correlation matrix. Further we apply fuzzy logic on the processed data.

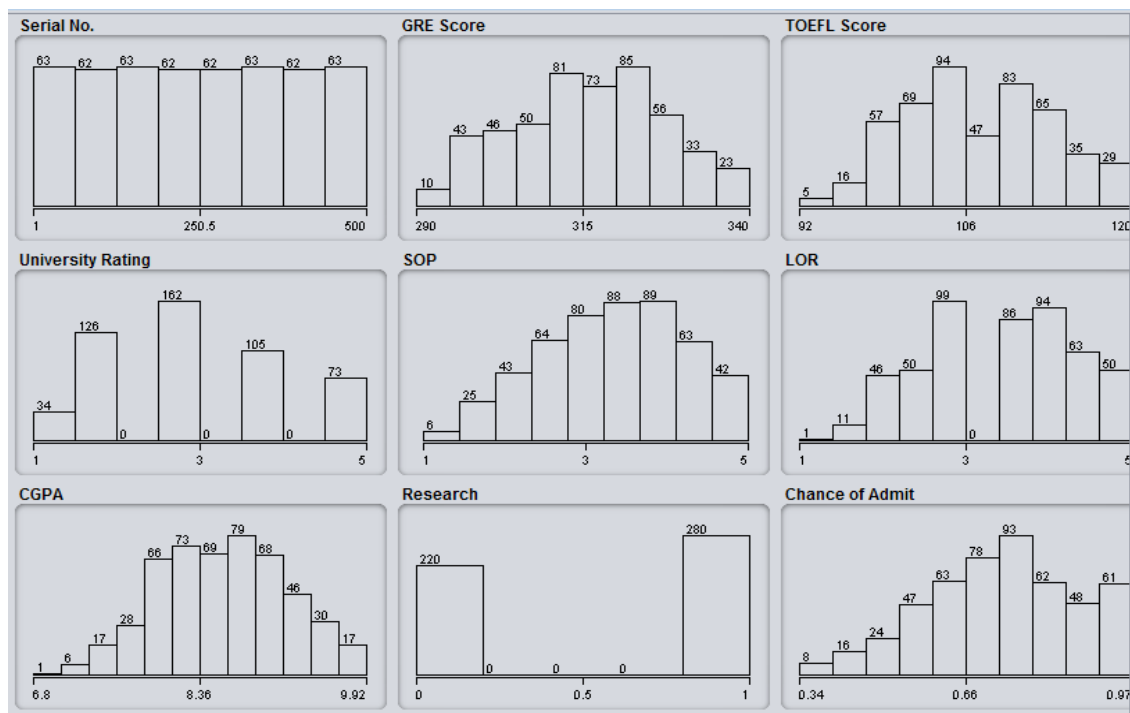The data available is processed using weka data mining tool as illustrated in Figures below.

Figure 2: data visualized using weka tool

Further the various important data is pre - processed using the Weka Explorer.
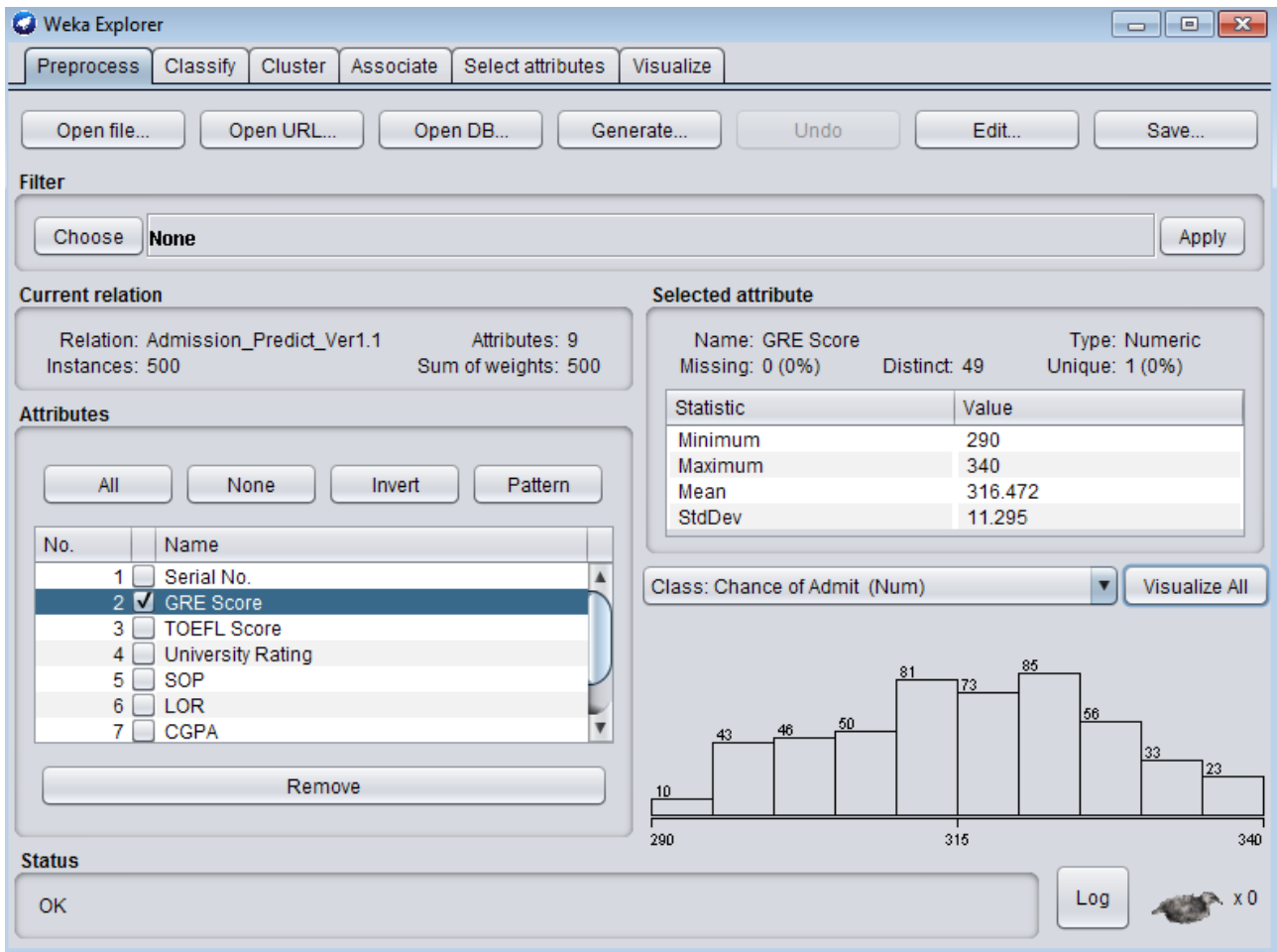

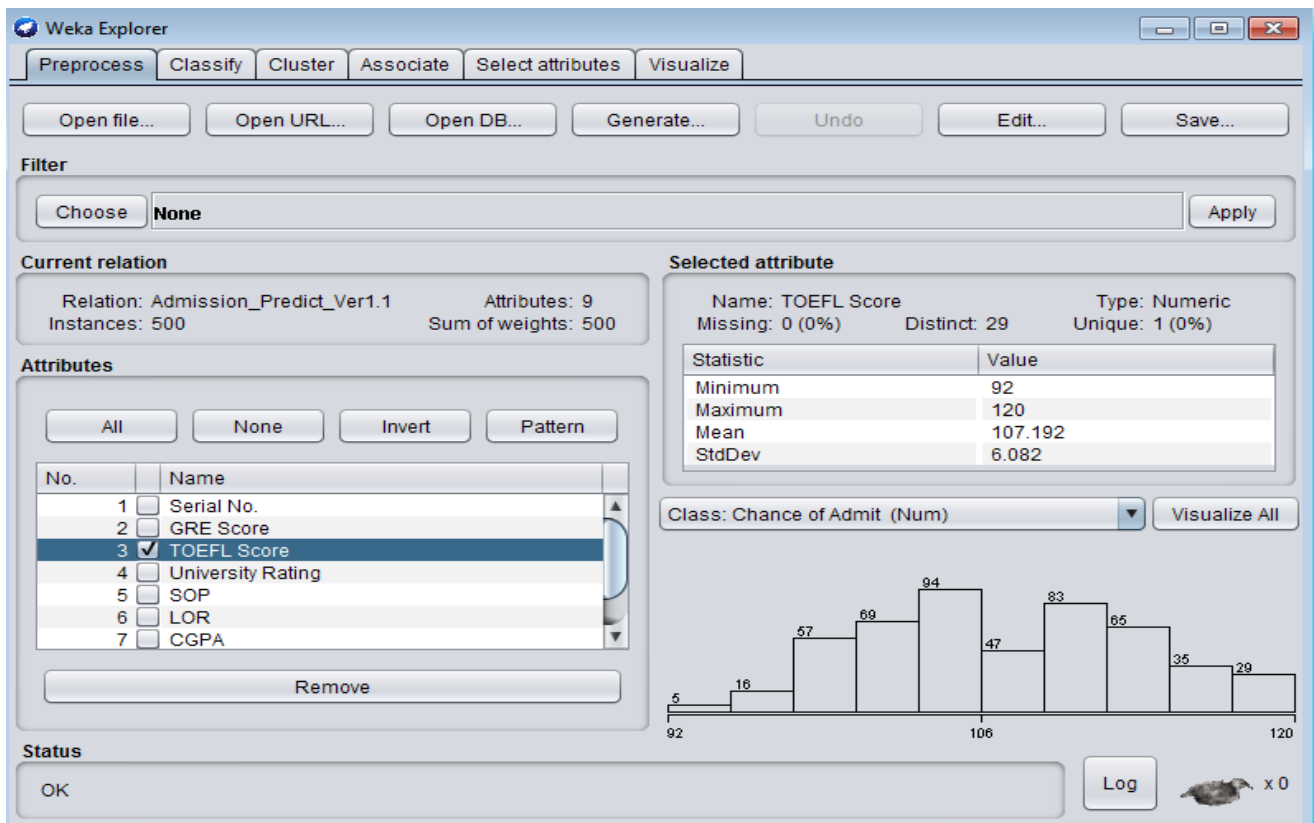**Figure 3:** Preprocessing of data having attribute GRE Score using weka tool.

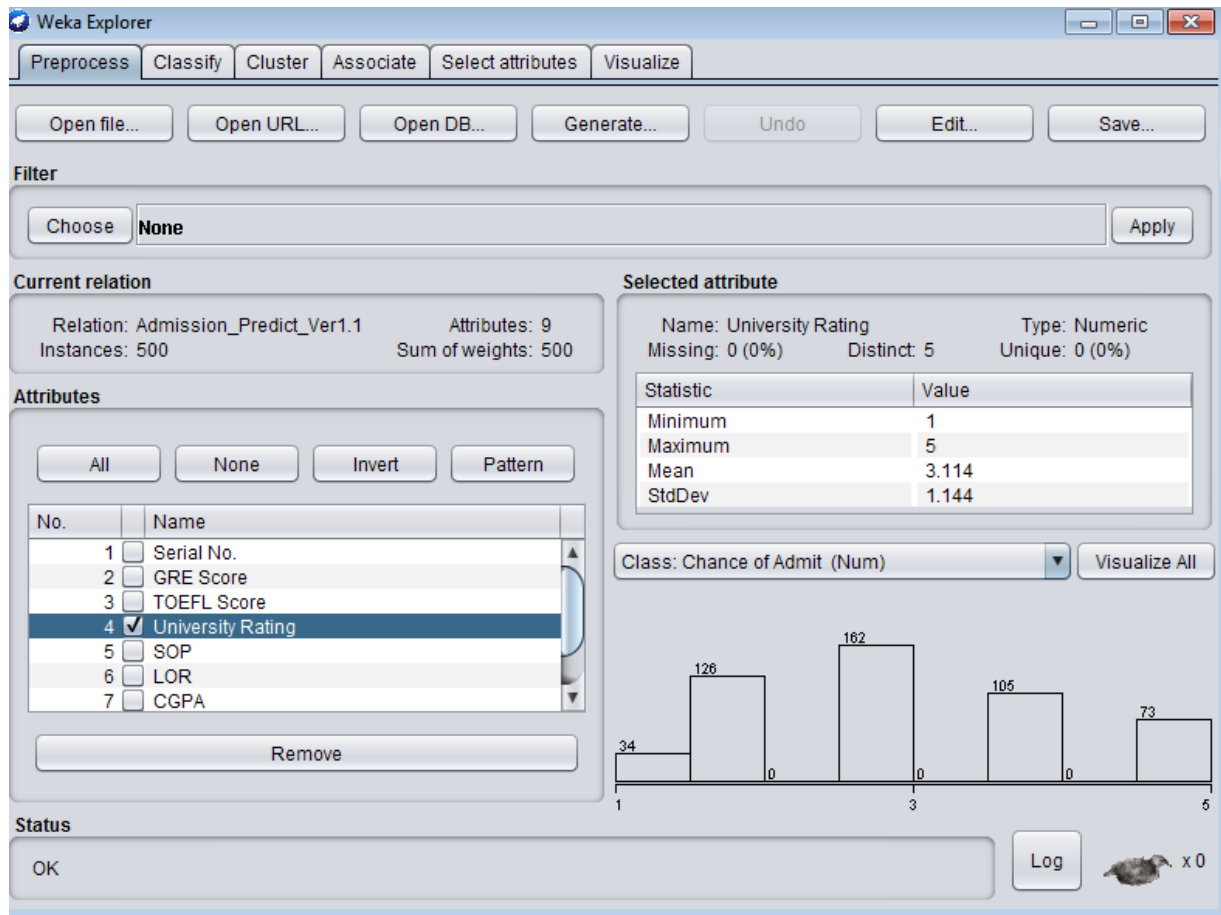**Figure 4:** Preprocessing of data having attribute TOEFL Score using weka tool.

Figure 5: Preprocessing of data having attribute University Rating using weka tool.
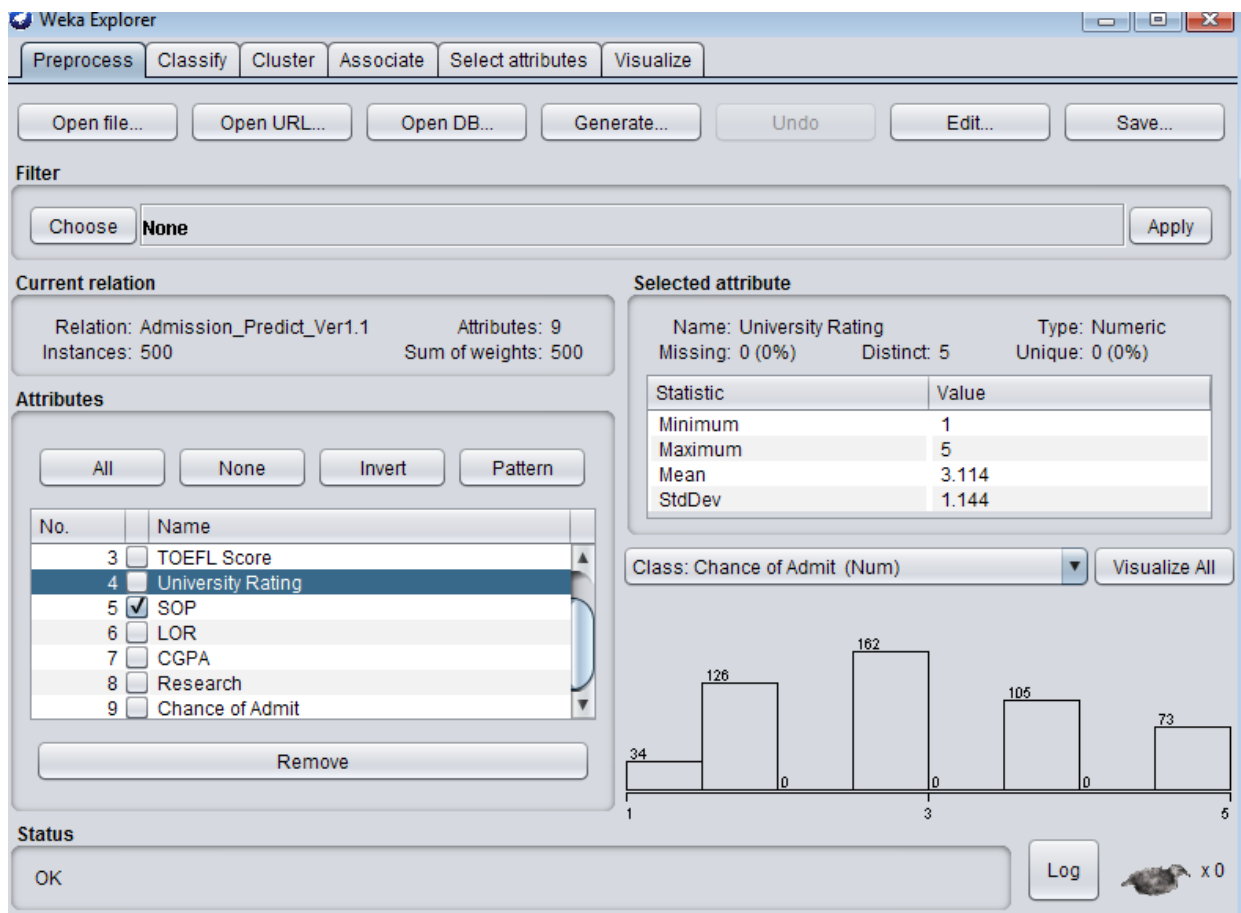

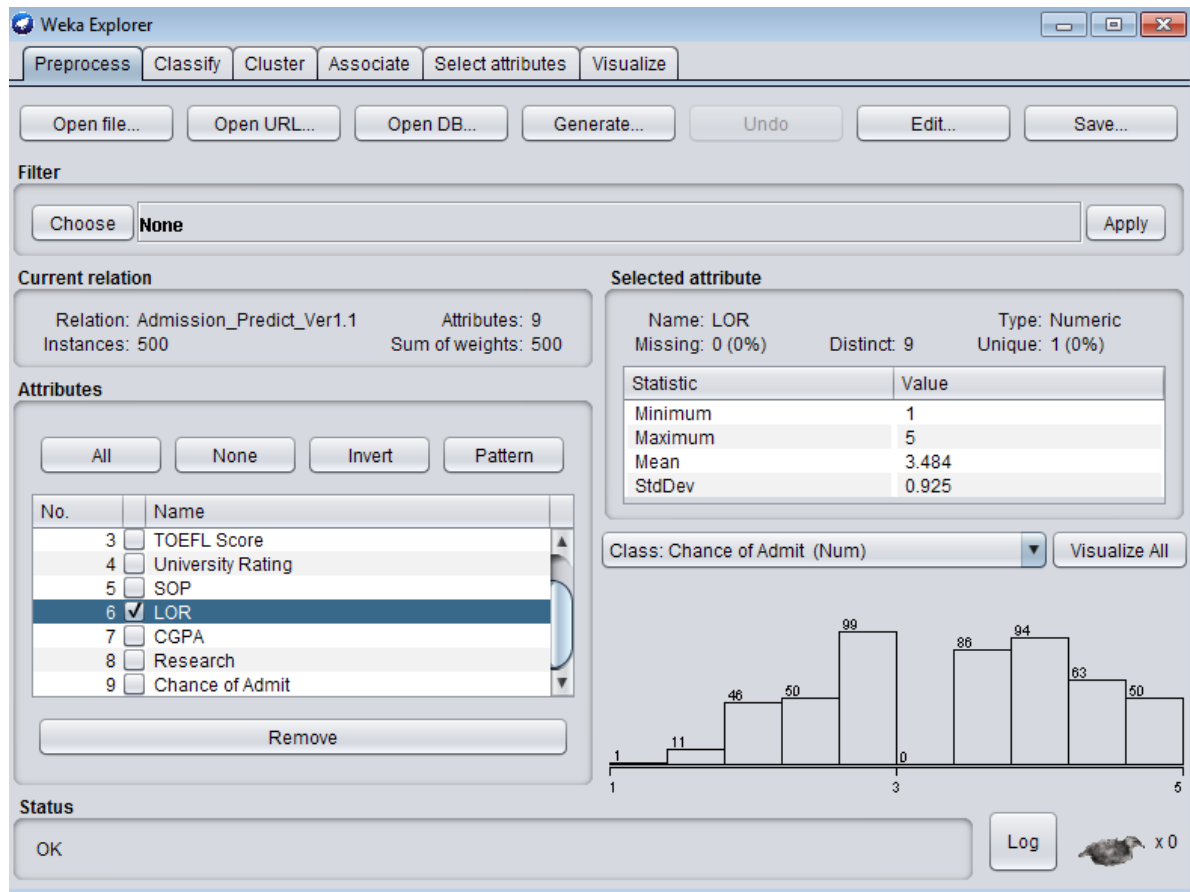**Figure 6:** Preprocessing of data having attribute SOP using weka tool.

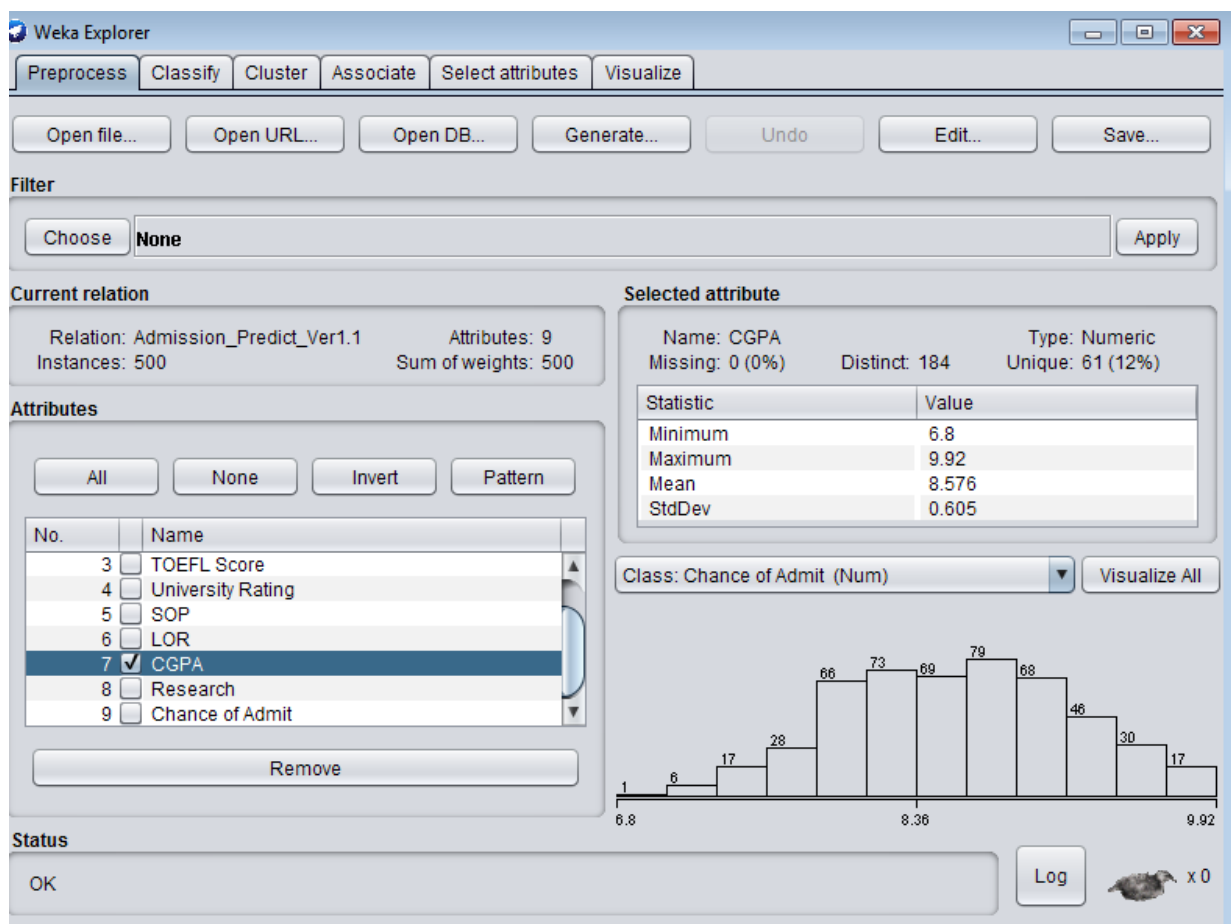**Figure 7:** Preprocessing of data having attribute LOR using weka tool.

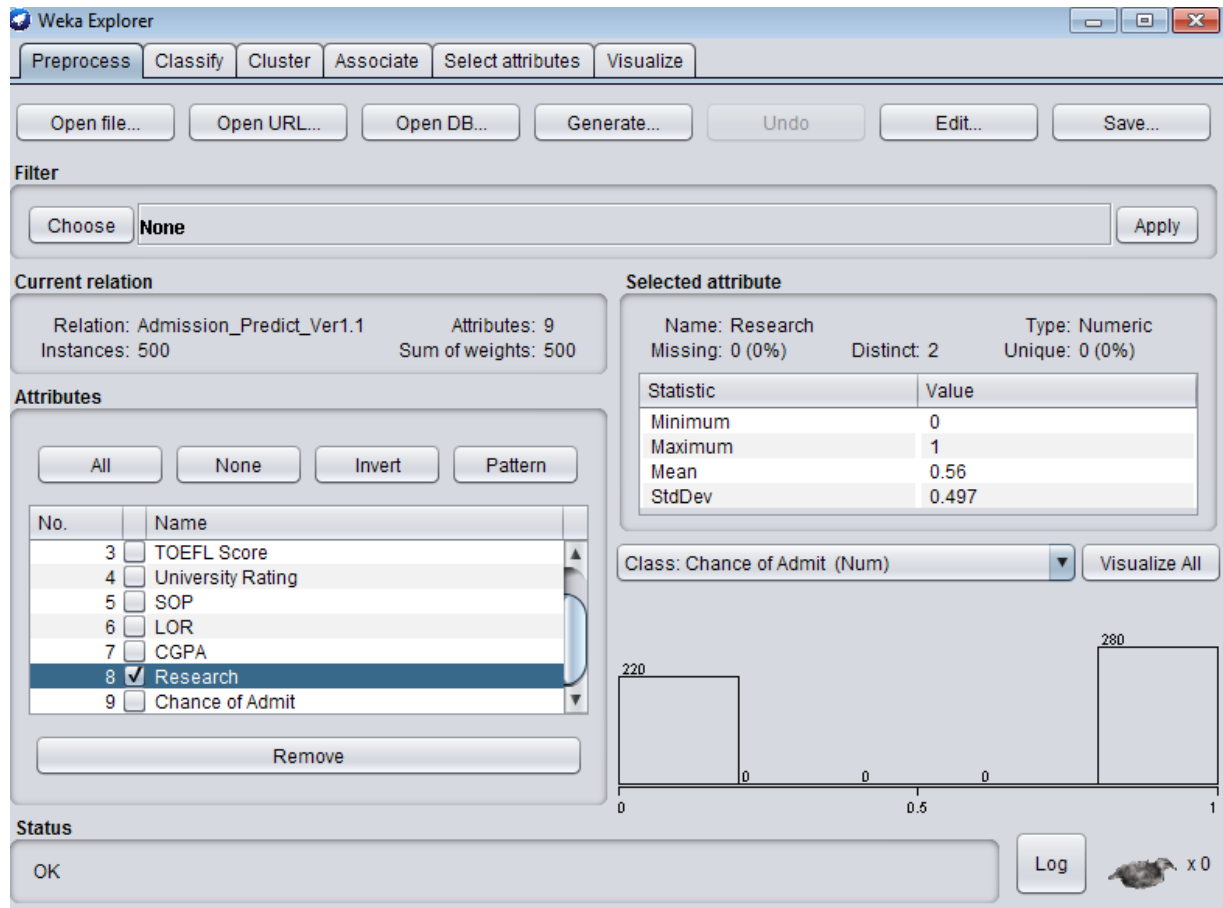**Figure 8:** Preprocessing of data having attribute CGPA using weka tool.

Figure 9: Preprocessing of data having attribute Research using weka tool.
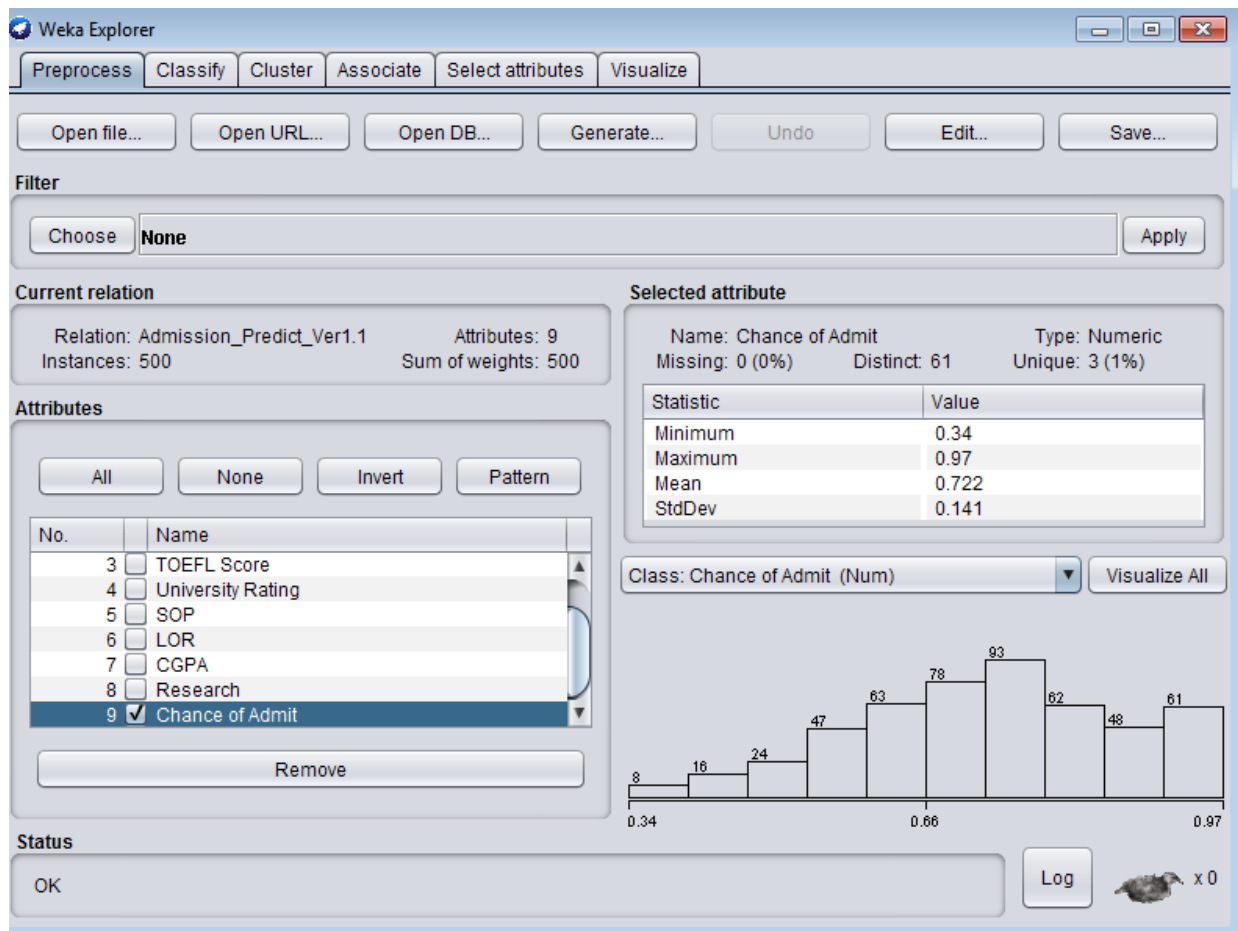

Figure 10: Preprocessing of data having attribute Chance of Admit using weka tool.

Developing the Fuzzy System using preprocessed data

Subsequent to preprocessing of data one of the ways of predicting the admission is by developing a fitness function comprising of weighed sums of factors impacting the decision of admission. The weights would represent the preferences provided by the institutes. These preferences 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.

The inherently imprecise nature of preferences of getting admission in suitable college renders fuzzy technology an excellent candidate for modeling those preferences. This will lessen the computational cost substantially when the student applies in the university in a more realistic way. A university admission prediction system using fuzzy inference system is beneficial for students in determining their chances of acceptance to a specific university. The fuzzy inference system could be designed based on the data related to previous applicants to various universities and this could be modelled and tested based on their admit or reject status.

The acceptance or reject status is an outcome of merging of rules in fuzzy knowledge based systems, by virtue of fuzzy sets. The basic idea of fuzzy set theory is that an element belongs to a fuzzy set with a certain degree of membership. The fuzzy set is simply a set with fuzzy boundaries. Hence, a proposition is neither True nor False, but may be partly true or false to some degree. This degree 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 'Very Low', 'Low', 'medium' High or 'VeryHigh' (as shown in Fig.9). 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. One of the foremost requirements in application of FL is the determination of MFs. It may involve knowledge of one or more experts. The approach of machine learning can also be used to automatically derive fuzzy sets and their MFs. The typical MFs used in fuzzy knowledge - based systems are the triangular and trapezoidal functions as those provide an adequate representation of the expert knowledge and significantly simplify the computational process [4]. We envisage fuzzy preferences taking the form of significance parameters and preference parameters. A significance parameter (SP) tells 'how important' certain aspect/criteria is for the overall fitness of the layout. [5] Whereas, a preference parameter (PP) tells 'how much' of certain aspect/criteria should be incorporated in the inference system. For instance in the fuzzy preference parameters in Prediction system is the rating of college while the parameters like CGPA, GRE scores, TOEFL scores etc. The Fuzzy inference system would accept fuzzy and/or crisp preferences and transform them in crisp weights, using the fuzzy rules present in the knowledge base, for the use in some fitness evaluation function. It is observed that certain parameters could have significant interaction with one another affecting more than one value of crisp weights used in evaluation phase.

Fuzzy inference is the process of mapping a given input to an output using the theory of fuzzy sets. There are various fuzzy inference mechanisms, the Mamdani - style inference method is one of the most prevalent technique for capturing experts' knowledge allowing for more intuitive and linguistic description of expertise involving four steps: Fuzzification of inputs, rule evaluation, aggregation of rule outputs, and finally defuzzification [4], [6].

5. Results and Conclusion

Many students apply to universities to get admission the colleges of their choice. Most of them have no prior knowledge, sufficient resources and lack awareness and hence apply in the wrong universities which created further problems to them. They waste their precious time and money. To overcome this problem this fuzzy inference system could be a useful tool to take proper decision for applying in the college. This will help the students to take decision based on the fuzzy inference prediction tool.

References

[1] Dr. Nikhat Khan " Study of techniques and applications of data mining in Education , " International Journal of Application or Innovation in Engineering and Management.

[2] L. A. Zadeh. ―Fuzzy sets‖. Inf. & Contr., 8, pp.338–353, 1965.

[3] **Mohan S Acharya, Asfia Armaan, Aneeta S Antony: A Comparison of Regression Models for Prediction of Graduate Admissions, IEEE International Conference on Computational Intelligence in Data Science 2019**

[4] Negnevitsky M., 2002. Artificial Intelligence: A Guide to Intelligent System,. Pearson Education, Sydney

[5] https: //www.researchgate. net/publication/220901925_Fuzzy_Inferencing_in_the _Web_Page_Layout_Design

[6] Nikolopoulos C., 1997. Expert Systems: An Introduction to First and Second Generation and Hybrid Knowledge Based Systems, Marcel Dekker Inc., New York

[7] Dr. Nikhat Khan and Dr. F. Z. Khan, "Data Mining for Fuzzy Decision system in Banking, " CiiT International Journal of Data Mining Knowledge Engineering, no. January 2013.

Author Profile

Dr. Nikhat Khan is a Visionary IT Professional with more than 20 years of experience in the IT sector spanning multiple domains and worked in various Government, private and public sector organizations. Adept at working in fast - paced environments and picking up new skills on the go. Providing effective leadership in implementing innovative and technical solutions, planning strategies to drive comprehensive growth. Enthusiastic team

player, with a focus on meeting organizational goals and surpassing expectations. She has visited US, UK and German countries for managing and executing IT projects for renowned international clients. She has worked in multiple national and international organizations. Currently she is working as a Director in the Institute of Company Secretaries of India. She has received MCA degree from D. A. V. V, Indore in the year 1998 and PhD in Computers from Maulana Azad National Institute of Technology, Bhopal (MANIT). She is eager to know about the advances in business intelligence, data mining and fuzzy logic. She has number of publications in National and International Journals.