

Prediction of Attributes in TNSET preparation using Machine Learning

Subashini Chandrasekar
Department of Chemistry,
Rajalakshmi institute of technology,
Poonamlee,
Chennai,
Tamil Nadu, India.
Subashini335@gmail.com

Kala Raja Mohan
Department of Mathematics,
Vel Tech Rangarajan Dr. Sagunthala R&D
Institute of Science and Technology,
Avadi, Chennai – 600 062,
Tamil Nadu, India.
kalamohan24@yahoo.co.in

ABSTRACT

Nowadays, clearance of competitive examinations is the major challenge faced by many of the educators. The method of selecting toppers, based on the rank or grade or percentage of marks in an examination is the task involved in such competitive examination. This paper aims at bringing out an analysis on determining the salient factors involved in getting through TNSET in mathematics using the concept of machine learning.

Key Words: TNSET - Preparation - Confusion Matrix - Machine Learning

I. INTRODUCTION

Selection of relevant candidates in any field based on their knowledge, aptitude, skill and other factors is made through suitable assessment. The assessment conducted, lists out the marks or grades. The required number of candidates are chosen from the prepared list and given the credit of clearing the exam. This procedure of selecting the candidates is referred as competitive exams.

Competitive exams are conducted in all fields with various purpose. Few exams like NEET are conducted with the purpose of providing admission for higher studies. Few exams like CSIR NET aim at giving certification to people to get through it for using it as a tool for jobs prescribed. These types of exams are conducted either with objective or descriptive type of questions. Assessment of questions are done based on the number of questions with the specified time management technique.

Yiyu Yao and Yan Zhao (2008) discussed about attribute reduction using rough topology [1]. J.N. Nasiri and M.Mashinchi (2009) made a study on decision tables [2]. M. Lellis Thivagar et al (2012) discussed about medical events [3]. Tao Yan and Chongzhao Han (2014) proposed a rough topology method of attribute selection using conditional entropy [4]. Qinghua Zhang et al (2016) made a survey on applications of rough set topology [5]. Nirmala Rebecca Paul (2016) made a medical diagnosis analysis using rough topology [6]. Pooja Chaturveri et al (2017) discussed about the various concepts involved in rough set topology [7]. Kanchana.M and Rekha.S (2020) discussed about Corona virus diagnosis [8]. G. Gomathi et al (2016, 2021) has made a study fuzzy rough set using its properties with local closed sets and local compactness [9].

Shixia Liu et al in 2017 discussed various machine learning methods [10]. Yue Liu et al in 2017 made their study on designing models using machine learning [11]. KM Jyoti Rani in 2020 performed a predictive analysis using machine learning [12]. Mahmood Safaei et al in 2021 performed various methods on machine learning to predict obesity [13]. Umarani et al in 2021 proposed sentimental analysis applying machine learning [14].

In this chapter, TNSET, the competitive exam conducted for academicians which provides a key to enter jobs at Government and Private colleges is taken into consideration. TNSET is the state level examination conducted by Tamil Nadu Government to post graduate holders in testing the subject knowledge of the academicians. In clearing the test in mathematics, has become a big dream for many mathematicians. The various aspects required in the preparation of the examination are discussed. From that, the important factors which play a major role in clearing the examination are identified.

II. STANDARD DEFINITIONS

- Information System: The data collected with respect to the attributes, from various persons referred as objects, are arranged in the form of a matrix. The matrix is called as information system.
- Machine Learning: A field of study in which the computers get the ability to learn without programming technique is machine learning.
- MATLAB: It refers to matrix laboratory. Arrays and matrices are the basic that operates on MATLAB. It is based on a programming system used to analyze data.
- True Positive (TP): It refers to the correct positive prediction.
- False Positive (FP): It refers to the incorrect positive prediction.
- True Negative (TN): It refers to the correct negative prediction.
- False Negative (FN): It refers to the incorrect negative prediction.
- Confusion Matrix: The performance of a classification algorithm is visualized and summarized with the help of confusion matrix. It is represented in tabulated form, summarizing the number of positive and negative predictions. The format of confusion matrix is given in Table 1 below.

Table 1 : Confusion Matrix representation

CONFUSION MATRIX		PREDICTED	
		N	P
ACTUAL	N	TN	FP
	P	FN	TP

Table 1 : Confusion Matrix representation

- Accuracy: It refers to the ratio of representing the number of all correctly predicted values to the total values. Higher the value of accuracy, higher the predicted value is true.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

- Precision: It refers to the number of actually positive values from all the positively predicted class values, Precision also shows a good result, when its value is higher.

$$Precision = \frac{TP}{TP + FP}$$

- Recall: It refers to the number of correctly predicted value, from all the positive classes. Higher the recall values denotes higher the predicted value is true.

$$Recall = \frac{TP}{TP + FN}$$

- F1-Score: It refers to the harmonic mean of recall and precision. It take values between 0 and 1.

$$F1\ Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

III. Factors involved in preparing for TNSET:

There are various factors behind clearing of the mentioned examination in mathematics. Few of them are listed below with higher weightage are discussed in this section.

- The candidate learnt through any private coaching centres
- The candidate preparing for CSIR NET exam too
- The candidate is employed
- The candidate faced stress while preparing for the examination?

IV. METHODOLOGY

With a motive to predict on the factors involved in TNSET preparation in mathematics, it is proposed to conduct a survey on general public, in identifying the various activities related to the attributes taken for consideration as stated above. Google form with the list of following questionnaires has been circulated to collect data for the analysis. All the questions are given in objective type questions with answer as 'yes' or 'no'. The responses related to the survey questionnaire are recorded from around respondents from various age groups.

MATLAB:

The proposed methodology is then analyzed using machine learning concept with MATLAB. In order to make the analysis, factors involved in the analysis such as confusion matrix, accuracy, precision, recall and F - score are calculated using MATLAB. The major part of the MATLAB coding is given below for reference in Table 2.

Table 2: MATLAB coding for obesity prediction analysis

```
confusion_matrix = confusionmat(true_labels, predicted_Q1);
true_positive = confusion_matrix(1, 1);
false_positive = confusion_matrix(2, 1);
false_negative = confusion_matrix(1, 2);
true_negative = confusion_matrix(2, 2);
accuracy = (true_positive + true_negative) / sum(confusion_matrix(:));
precision = true_positive / (true_positive + false_positive);
recall = true_positive / (true_positive + false_negative);
f1_score = 2 * (precision * recall) / (precision + recall);
```

The responses obtained from decision attribute is compared with the responses received for all other attributes and the factors specified above for the analysis are calculated. The output obtained after the execution of the coding is given below in Table 3.

Table 3: Output obtained from MATLAB coding

```
Enter the true values
[1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1]
true_labels =
1 1 1 1 0 1 0 0 1 1 1 0 1 1 0 1 0 1 0 1 0 1 1 1
Enter the predicted values of Question 1
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
predicted_Q1 =
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Enter the predicted values Question 2
[1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0]
predicted_Q2 =
1 1 1 0 1 0 1 0 1 1 0 1 0 0 1 0 1 1 0 1 0 0 1 0 0
Enter the predicted values of Question 3
[1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1]
predicted_Q3 =
1 1 1 1 1 1 1 1 0 1 0 1 0 1 1 0 1 0 1 0 1 1 1 0 1
Enter the predicted values of Question 4
[0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0]
predicted_Q4 =
0 1 0 0 1 0 1 0 0 1 0 1 1 0 1 0 0 0 0 1 0 1 1 0 0
Decision on Question 1
Confusion Matrix:
8 0
17 0
Accuracy:
0.3200
Precision:
0.3200
Recall:
1
F1 Score:
0.4848
Decision on Question 2
Confusion Matrix:
2 6
10 7
Accuracy:
```

0.3600
Precision:
0.1667
Recall:
0.2500
F1 Score:
0.2000
Decision on Question 3
Confusion Matrix:
1 7
6 11
Accuracy:
0.4800
Precision:
0.1429
Recall:
0.1250
F1 Score:
0.1333
Decision on Question 1
Confusion Matrix:
2 6
13 4
Accuracy:
0.2400
Precision:
0.1333
Recall:
0.2500
F1 Score:
0.1739

The values obtained from the output are listed below in Table 4 for easy reference.

Table 4: Values obtained in Tabular form

Attribute	Confusion Matrix	Accuracy	Precision	Recall	F1 Score
Q 1	$\begin{pmatrix} 8 & 0 \\ 17 & 0 \end{pmatrix}$	0.3200	0.3200	1	0.4848
Q 2	$\begin{pmatrix} 2 & 6 \\ 10 & 7 \end{pmatrix}$	0.3600	0.1667	0.2500	0.2000
Q 3	$\begin{pmatrix} 1 & 7 \\ 6 & 11 \end{pmatrix}$	0.4800	0.1429	0.1250	0.1333
Q 4	$\begin{pmatrix} 2 & 6 \\ 13 & 4 \end{pmatrix}$	0.2400	0.1333	0.2500	0.1739

Depending on the attributes specified, the change in the values of attributes are found to have their individual values. The graphical representation of the values obtained is presented below in Figure 1 given below.

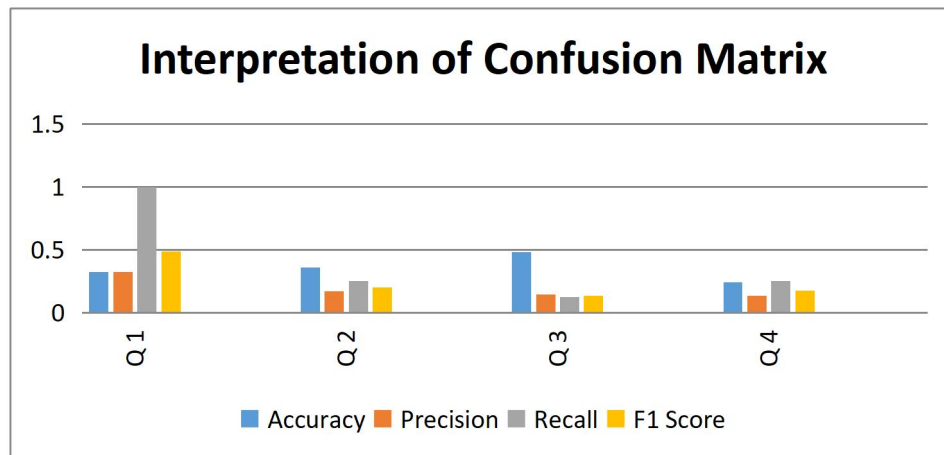


Figure 1: Interpretation of Confusion Matrix

V. Conclusion:

Predictive analysis on various attributes related to TNSET is carried out, MATLAB software has been used in carrying out the analysis. The values involved in the analysis such as accuracy, precision, recall and F1-score are calculated using confusion matrix obtained. The attribute with highest weightage is identified. The attributes in which more care to be taken to clear the examination are identified.

REREENCES

- [1] Yiyu Yao and Yan Zhao, "Attribute reduction in decision- theoretic rough set models", Information Sciences, Vol. 178, 2008, pp. 3356-3373.
- [2] J.N. Nasiri and M.Mashinchi, "Rough Set and Data Analysis in Decision Tables", Journal of Uncertain Systems, Vol. 3, 2009, pp. 232-240.
- [3] M. Lellis Thivagar, Carmel Richard and Nirmala Rebecca Paul, "Mathemcatical Innovations of a Modern Topology in Medical Events", Journal of Information Science, Vol. 2, 2012, pp. 33-36.
- [4] Tao Yan and Chongzhao Han, "A Normal Approach of Rough Conditional Entropy-Based Attribute Selection for Incomplete Decision System", Mathematical Problems in Engineering, 2014, pp.1-15.
- [5] Qinghua Zhang, Qin Xie and Guoyin Wang, "A survey on rough set theory and its applications", CAAI Transactions on Intelligence Technology, Vol. 1, 2016, pp. 323-333.
- [6] Nirmala Rebecca Paul, "Rough Topology Based Decision Making in Medical Diagnosis", International Journal of Mathematics Trends and Technology, Vol. 18, 2016, pp. 40-43.
- [7] Pooja Chaturveri, A. K. Daniel and Kumari Khusboo, "Concept of Rough Set Theory and its Applications in Decision Making Processes", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, 2017, pp. 43-46.
- [8] Kanchana.M and Rekha.S, "Decision Making Using Rough Topology and Indiscernibility for Corna Virus Diagnosis", International Journal of Scientific Research in Science, Engineering and Technology, Vol. 7, 2020, pp. 31-33.
- [9] G. Gomathi, R. Narmada Devi and D. Sophia Ponmalar , "A study on Properties of Fuzzy Rough Topological Spaces", International Journal of Advanced Research in Management, Architecture, Technology and Engineering, Vol. 7, 2021, pp. 13-1
- [10] Shixia Liu, Xiting Wang, Mengchen Liu, Jun Zhu, "Towards better analysis of machine learning models: A visual analytics perspective", Visual Informatics 2017, Vol. 1, 2017, 48-56.
- [11] Yue Liu, Tianlu Zhao, Wangwei Ju, Siqi Shi, "Materials discovery and design using machine learning", J Materiomics, Vol. 3, 2017, 159-177.
- [12] KM Jyoti Rani, "Diabetes Prediction using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Vol.6-4, 2020, 3083-3087.
- [13] Mahmood Safaei, Elankovan A. Sundararajan, Maha Driss, Wadii Boulila, Azrulhizam Shapi, "A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity", Computers in Biology and Medicine 2021, Vol. 136, 2021, 1-17.
- [14] V. Umarani, A. Julian, J. Deepa, "Sentiment Analysis using various Machine Learning and Deep

Learning Techniques”, Journal of the Nigerian Society of Physical Science, Vol. 3, 2021, 385–394