ASSESSMENT OF PERFORMANCE OF ENGINEERING STUDENTS USING EDUCATIONAL DATA MINING

[1]Hema Malini B H, [2] Dr. L Suresh,
[1]BMS Institute of Technology & Management,[2] Cambridge Institute of Technology

Abstract-The number of engineering colleges in South India is considerably high. Since engineering courses are most sought by the parents, the competition amongst institutions is also more. The institution with better ranking and credentials get better cut off ranking students. The present study is on educational data mining where the academic results of students for one course is obtained through questionnaire, the internal assessment and university results are obtained and a study is made by applying machine learning algorithms. For present study, Naïve Bayes algorithm is applied which gives 100%.

Keywords—Naive Bayes, Machine learning, Educational Data mining.

Introduction

In the present situation, there is lot of competence in engineering education. Average enrolment in engineering colleges as per statistics is 70%. Many of the institutes stay as mediocre institutes. Only a few can survive the competition offered by the global market. The expectations of parents and students are high. The productivity of the institution has to go high considerably to meet the demands of stakeholders. The faculty has to make an impact in class by implementing various teaching learning techniques to reach all categories of students. Presently, since outcome based education is in place, it is the responsibility of the faculty to check out whether the deliverable in class was reachable by all categories of students, the class toppers, the average performers and slow learners. Also, the challenge for every faculty is to make the students attend the classes voluntarily. In this regard, an effort is made to take the opinion of students on all the topics covered in class and the university results are collected to check the outcome. A study is made and outcome is obtained.

Some Machine Learning Techniques

Supervised Learning

1. Decision Trees: A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance-event outcomes, resource costs, and utility.

From a commercial decision point of view, a decision tree is the least number of yes/no questions that have to be asked, to assess the probability of making a precise decision, most of the time. The method allows you to approach the problem in a structured and methodical way to arrive at a logical conclusion [7].

2. **Naive Bayes Classification**: Naive Bayes classifiers are a group of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independent assumptions between the features[7].

The featured image is the equation-

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)}$$

Where

P(A|B) - posterior probability,

P(B|A) - likelihood,

P(A) - class prior probability, and

P(B) - predictor prior probability.

Some of real world examples are:

- Check a slice of text expressing positive or negative emotions?
- To mark an email as spam or not spam
- Classify a news article about technology, politics, or sports
- Used for face recognition software.

Unsupervised Learning

Clustering Algorithms: Clustering is the job of grouping a set of objects such that objects in the same group (*cluster*) are more similar to each other than to those in other groups [8].

Every clustering algorithm is different, and here are a couple of them:

- Centroid-based algorithms
- Density-based algorithms

- Connectivity-based algorithms
- Dimensionality Reduction
- Probabilistic
- Neural networks / Deep Learning

Literature Survey

ManolisChalaris et al. [1] have considered the educational data from Technological Educational Institute of Athens. They have distributed the questionnaires to students and have collected the data for evaluation. The authors study the competencies of data mining in the perspective of Higher Education.

HemaMalini B. H et al. [2] have considered numerous factors of faculty which affects the quality of the students produced to the universal market. The authors considered the personal and professional credentials of five potential faculties. Also, the result for the courses they have handled for four consecutive years is collected. A work is made to map the faculty credentials with the result produced. The conclusions are drawn based on the performance of students.

PoojaThakar et al. [3] have done a comprehensive survey, a travelogue (2002-2014) towards educational data mining and its scope in the future.

Kamal Bunkar et al. [4] have used classification technique in data mining, to support in improving the quality of the higher educational system by assessing the student data and have used the main attributes which affect the student performance in courses. The authors have obtained the data from first year students of Vikram University, Ujjain of course B.A. A system that enables the use of the generated rules is made which allows students to foresee the final grade in a course to be studied.

Tjioe Marvin Christian et al. [5] have used students' education data, personal data, admission data, and academic data. One of data mining methods, NBTree classification technique, was adopted to foresee the students' performance. Numerous experiments were implemented to discover a prediction model for students' performance. The class labels of performance ofstudents were students' status in study, graduates predicates, and length of study. Research was conducted with two-level classification, the faculty level and the university level. The resultant model showed that certain attributes had significant impact over students' performance.

AshishDutt et al. [6] provide around thirty years of (1983_2016) systematic literature review on clustering algorithms and its usability and applicability in the context of EDM. Future visions are drawn based on the literature reviewed, and opportunities for further research are identified.

Present Work

In thepresent work, a data collection is done from 30 students by preparing a questionnaire about all the topics coveredin the course Automata Theory and Computability taught for the students of Department of Computer Science and Engineering in BMS Institute of Technology & Management, Bengaluru. The input file is CSV file. The academic score in the internal assessment test and university marks is also collected. The tool used for data analysis is WEKA. Naïve Bayesan algorithm is applied on the dataset obtained. So, the number of instances is 30 and attributes is 20.

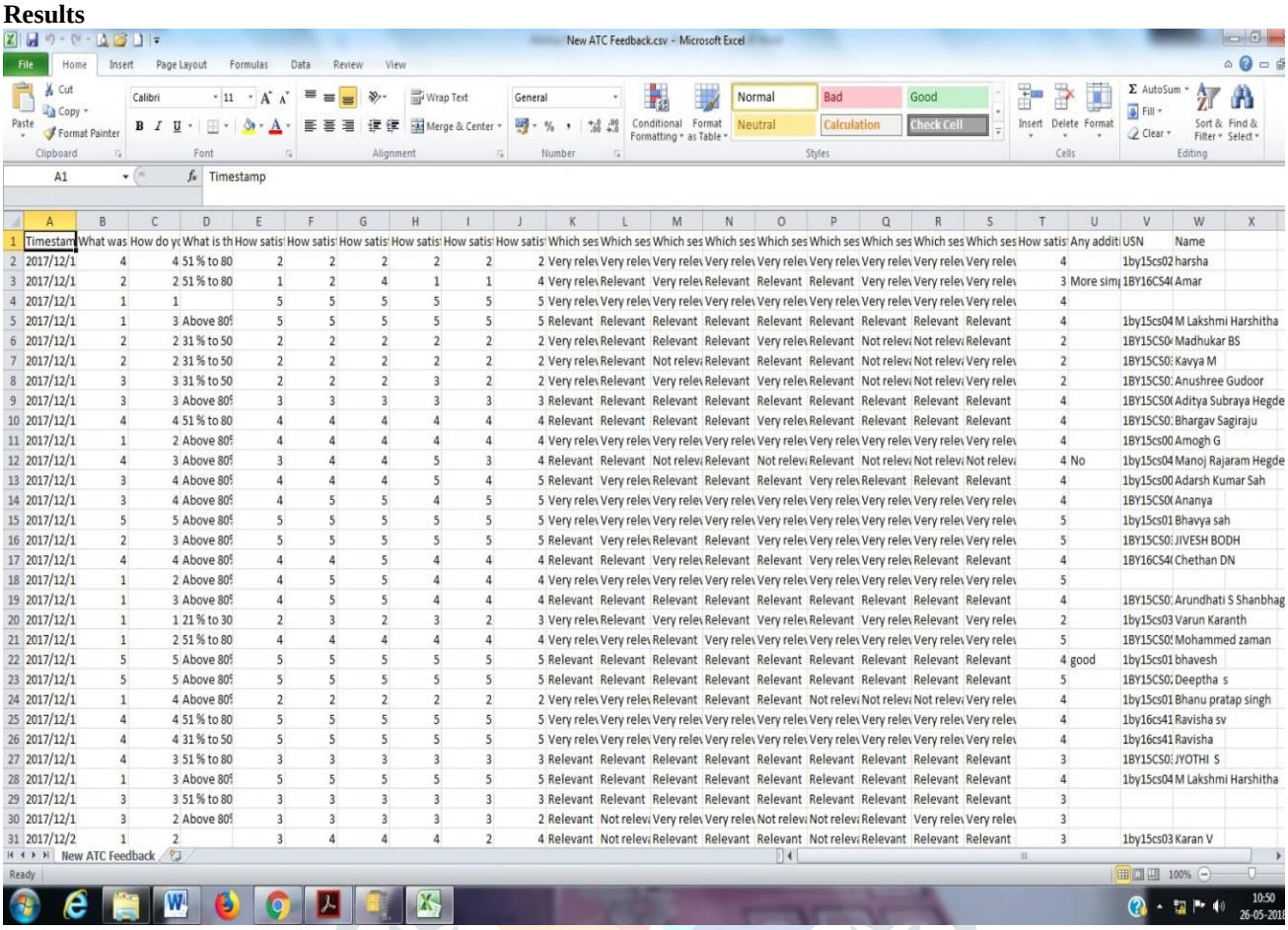
Methodology

The Google form was created with questionnaire and the same was communicated. The survey was conducted. A total of 30 students took up the survey. There were 20questions to be answered. So, the instances were 30 and attributes were 20. The result obtained was downloaded in .csv file. The same was converted to .arff file using WEKA tool. The data preprocessing is done, using WEKA. The confusion matrix and analysis is obtained. Naïve Bayes algorithm is applied on the data set.

The following questions were posed to the students:

1. What was the level of your prior knowledge in Automata Theory?
2. How do you rate the complexity of the subject?
3. What is the percentage of coverage of syllabus in class?
4. How satisfied were you with the Course/Subject? [Your Level of understanding the subject]
5. How satisfied were you with the Course/Subject? [Teaching ability of the faculty]
6. How satisfied were you with the Course/Subject? [Communication of Faculty]
7. How satisfied were you with the Course/Subject? [Learnability of this subject]
8. How satisfied were you with the Course/Subject? [Innovation in the teaching techniques]
9. How satisfied were you with the Course/Subject? [Interactive sessions held for the course]
10. Which sessions did you find most relevant? [DFSM, NFSM]
11. Which sessions did you find most relevant? [Regular Expressions]
12. Which sessions did you find most relevant? [Pumping Lemma]
13. Which sessions did you find most relevant? [Push Down Automata]
14. Which sessions did you find most relevant? [Simplification of Grammar]
15. Which sessions did you find most relevant? [Turing Machines]
16. Which sessions did you find most relevant? [CFG]
17. Which sessions did you find most relevant? [Derivations (LMD, RMD), Parse Trees]
18. Which sessions did you find most relevant? [Ambiguous Grammar]
19. How satisfied were you with the session content?
20. Any additional comments regarding the sessions or overall course?

Results



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Timestam	What was	How do yo	What is th	How satis	How satis	How satis	How satis	How satis	How satis	Which ses	Which ses	Which ses	Which ses	Which ses	Which ses	Which ses	Which ses	Which ses	How satis	Any additi	USN	Name
2	2017/12/1	4	4	51 % to 80	2	2	2	2	2	2	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	4		1by15cs02	harsha
3	2017/12/1	2	2	51 % to 80	1	2	4	1	1	4	Very relev	Relevant	Very relev	Relevant	Relevant	Relevant	Very relev	Very relev	Very relev	3	More simp	1BY16CS4(	Amar
4	2017/12/1	1	1		5	5	5	5	5	5	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	4			
5	2017/12/1	1	3	Above 80%	5	5	5	5	5	5	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	4		1by15cs04	M Lakshmi Harshitha
6	2017/12/1	2	2	31 % to 50	2	2	2	2	2	2	Very relev	Relevant	Relevant	Relevant	Very relev	Relevant	Not relev	Not relev	Relevant	2		1BY15CS0	Madhukar BS
7	2017/12/1	2	2	31 % to 50	2	2	2	2	2	2	Very relev	Relevant	Not relev	Relevant	Relevant	Relevant	Not relev	Not relev	Very relev	2		1BY15CS0	Kavya M
8	2017/12/1	3	3	31 % to 50	2	2	2	3	2	2	Very relev	Relevant	Very relev	Relevant	Very relev	Relevant	Not relev	Not relev	Very relev	2		1BY15CS0	Anushree Gudoor
9	2017/12/1	3	3	Above 80%	3	3	3	3	3	3	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	4		1BY15CS0	Aditya Subraya Hegde
10	2017/12/1	4	4	51 % to 80	4	4	4	4	4	4	Relevant	Relevant	Relevant	Relevant	Very relev	Relevant	Relevant	Relevant	Relevant	4		1BY15CS0	Bhargav Sagiraju
11	2017/12/1	1	2	Above 80%	4	4	4	4	4	4	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	4		1BY15cs00	Amogh G
12	2017/12/1	4	3	Above 80%	3	4	4	5	3	4	Relevant	Relevant	Not relev	Relevant	Not relev	Relevant	Not relev	Not relev	Not relev	4	No	1by15cs04	Manoj Rajaram Hegde
13	2017/12/1	3	4	Above 80%	4	4	4	5	4	5	Relevant	Very relev	Relevant	Relevant	Relevant	Very relev	Relevant	Relevant	Relevant	4		1by15cs00	Adarsh Kumar Sah
14	2017/12/1	3	4	Above 80%	4	5	5	4	5	5	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	4		1BY15CS0	Ananya
15	2017/12/1	5	5	Above 80%	5	5	5	5	5	5	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	5		1by15cs01	Bhavya sah
16	2017/12/1	2	3	Above 80%	5	5	5	5	5	5	Relevant	Very relev	Relevant	Relevant	Very relev	Very relev	Very relev	Very relev	Very relev	5		1BY15CS0	JIVESH BODH
17	2017/12/1	4	4	Above 80%	4	4	5	4	4	4	Relevant	Relevant	Very relev	Relevant	Relevant	Very relev	Very relev	Relevant	Relevant	4		1BY16CS4(	Chethan DN
18	2017/12/1	1	2	Above 80%	4	5	5	4	4	4	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	5			
19	2017/12/1	1	3	Above 80%	4	5	5	4	4	4	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	4		1BY15CS0	Arundhati S Shanbhag
20	2017/12/1	1	1	21 % to 30	2	3	2	3	2	3	Very relev	Relevant	Very relev	Relevant	Very relev	Relevant	Very relev	Relevant	Very relev	2		1by15cs03	Varun Karanth
21	2017/12/1	1	2	51 % to 80	4	4	4	4	4	4	Very relev	Very relev	Relevant	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	5		1BY15CS0	Mohammed zaman
22	2017/12/1	5	5	Above 80%	5	5	5	5	5	5	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	4	good	1by15cs01	bhavesh
23	2017/12/1	5	5	Above 80%	5	5	5	5	5	5	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	5		1BY15CS0	Deeptha s
24	2017/12/1	1	4	Above 80%	2	2	2	2	2	2	Very relev	Very relev	Relevant	Relevant	Relevant	Not relev	Not relev	Not relev	Very relev	4		1by15cs01	Bhanu pratap singh
25	2017/12/1	4	4	51 % to 80	5	5	5	5	5	5	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	4		1by16cs41	Ravisha sv
26	2017/12/1	4	4	31 % to 50	5	5	5	5	5	5	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	Very relev	4		1by16cs41	Ravisha
27	2017/12/1	4	3	51 % to 80	3	3	3	3	3	3	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	3		1BY15CS0	JYOTHI S
28	2017/12/1	1	3	Above 80%	5	5	5	5	5	5	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	4		1by15cs04	M Lakshmi Harshitha
29	2017/12/1	3	3	51 % to 80	3	3	3	3	3	3	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	Relevant	3			
30	2017/12/1	3	2	Above 80%	3	3	3	3	3	2	Relevant	Not relev	Very relev	Very relev	Not relev	Not relev	Relevant	Very relev	Very relev	3			
31	2017/12/2	1	2		3	4	4	4	2	4	Relevant	Not relev	Relevant	Relevant	Relevant	Not relev	Relevant	Relevant	Relevant	3		1by15cs03	Karan V

Fig 1: Input File to Weka: Responses to Questionnaire

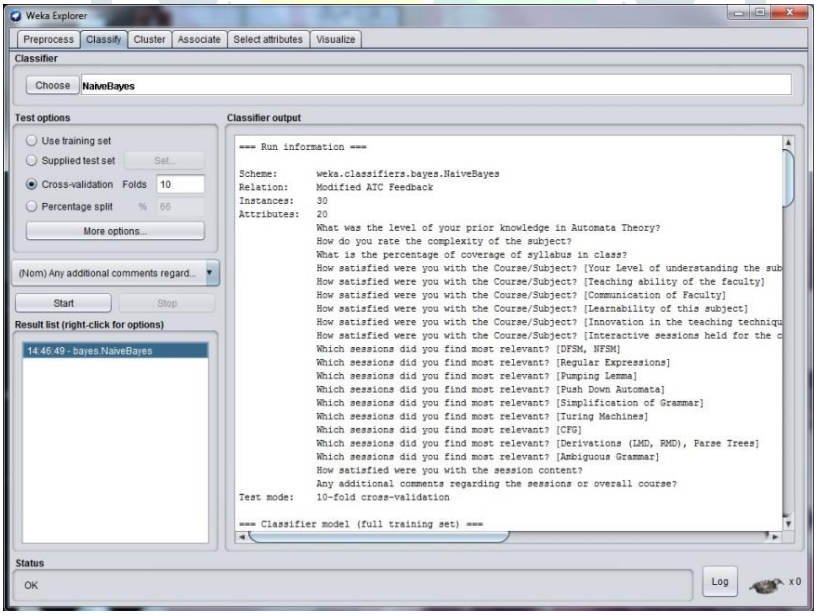


Fig 2: Naïve Bayes Run Information

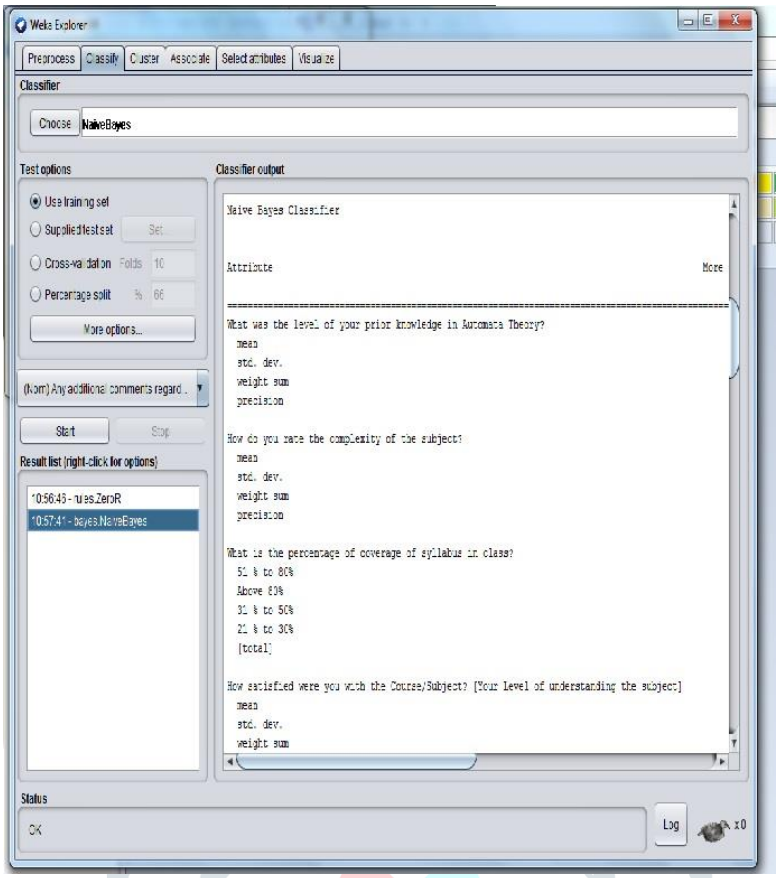


Fig 3: Naïve Bayes Attributes

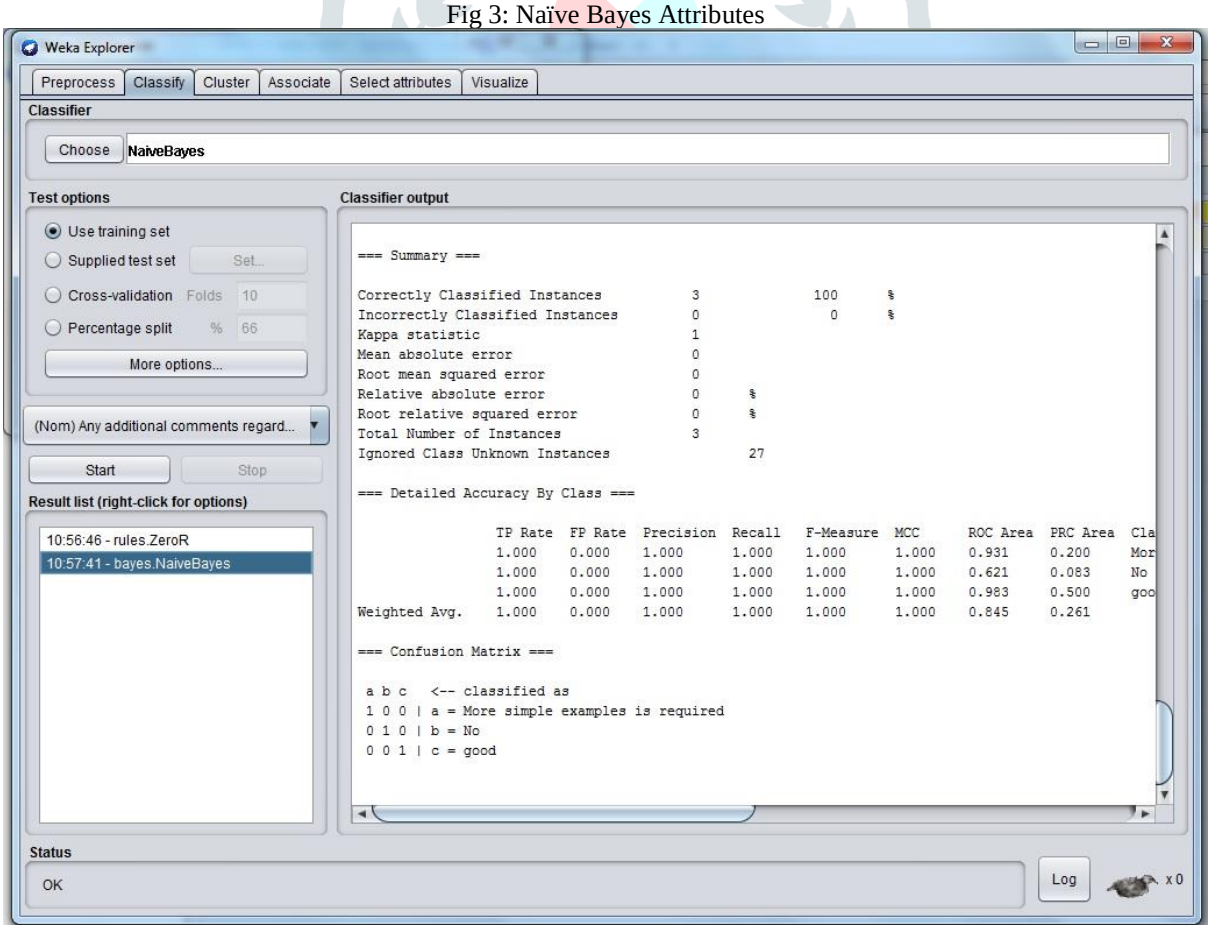


Fig 4: Summary and Confusion Matrix

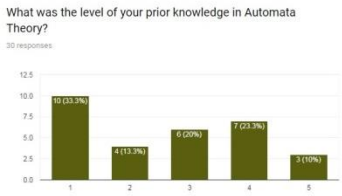


Fig 5: Prior Knowledge of subject

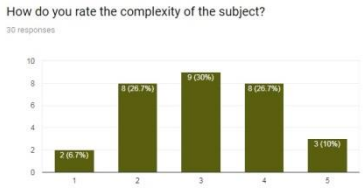


Fig 6: Complexity of Subject

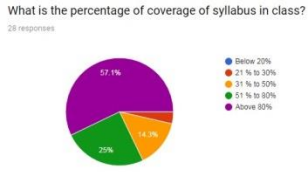


Fig 7: Coverage of syllabus

There were about 20 attributes in the questionnaire. 33% of the students did not have any prior knowledge about the subject. Around 26.7% of students have said the subject is more complex and around 30%have told it is complex. 57.1% of students have said coverage of syllabus was above 80%. The feedback about different topic covered is also taken. The graph is plotted with the results.

Fig. 1 shows the input file generated by Google Forms. Fig. 2 is the snapshot of applying the classifier Naïve Bayes on the data set. The Run information on WEKA is shown. Fig. 3 shows the different attributes used. Fig.4 shows the detailed summary along with the confusion matrix. Figures 5,6 and 7 are the snapshots of the graphs created using inputs of Google Forms.

Conclusion

The present work used WEKA tool. Preprocessing was automatically taken care of. Of the other classifiers available, Naïve Bayes classifier is suitable for analysis of the present data set. Thisclassifier gives 100% correctness for the input data.Further, classification and clustering methods have to be applied on the same training data set. The inferences have to be drawn. A comparative study has to be made. The present study will be continued with more number of attributes and on a large dataset, with different Machine Learning Techniques.

Bibliography

[1] ManolisChalaris*, StefanosGritzalis, ManolisMaragoudakis, Cleo Sgouropoulou and AnastasiosTsolakidis, "Improving Quality of Educational Processes Providing NewKnowledge using Data Mining Techniques", **ScienceDirect,** Procedia - Social and Behavioral Sciences 147 (2014) 390 – 397.

[2] HemaMalini B. H, Dr. L Suresh, "Data Mining in Higher Education System and the Quality of Faculty Affecting Students Academic Performance: A Systematic Review ", International Journal of Innovations & Advancement in Computer Science, IJIACS, ISSN 2347 – 8616, Volume 7, Issue 3, March 2018, page 66-70.

[3] PoojaThakar, Anil Mehta, Manisha, "Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue", International Journal of Computer Applications (0975 – 8887) Volume 110 – No. 15, January 2015.

[4] Kamal Bunkar, Rajesh Bunkar, Umesh Kumar Singh, BhupendraPandya, "Data Mining: Prediction for Performance Improvement of Graduate Students using Classification", 978-1-4673-1989-8/12©2012 IEEE

[5] Tjioe Marvin Christian, MewatiAyub, "Exploration of Classification Using NBTree forPredicting Students' Performance", 978-1-4799-7996-7/14 ©2014 IEEE

[6] AshishDutt, MaizatulAkmar Ismail, And TututHerawan, "A Systematic Review on Educational Data Mining", 2169-3536 2017 IEEE., VOLUME 5, 2017, page 15991- 16005

[7]https://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html

[8]https://towardsdatascience.com/naive-bayes-in-machine-learning-f49cc8f831b4

[9] https://machinelearningmastery.com/naive-bayes-for-machine-learning/