

AN ANALYSIS OF MODEL FOR HIGHER EDUCATION ORGANIZATION USING DATA MINING TECHNIQUES

Poonam Singh, MSIT, MATS University, Raipur, India
(Poonammatsuniversity@gmail.com)

Bhavana Narain, MSIT, MATS University, Raipur, India (narainbhawna@gmail.com)

ABSTRACT

A student satisfaction ratio is a feedback mechanism used to measure how satisfied students are with a college and university's overall operations and teaching practices. The survey is usually made up of open-ended and tick questions that cover different aspects of the college and university curriculum and management. It is important for colleges and universities to administrative, academic, and placement this survey regularly to refocus and reshape management actions and strategies. The student survey should be specific and relevant to college accurate feedback. The survey helps colleges and universities to improve and adjust accordingly in the landscape of higher education criteria. Surveys can also be helpful for management measuring the pace and workload of the organization which side is weak for the student side and for getting feedback on specific changes for the organization. It is important to keep the survey brief, clean, and specific to give students free space to weigh in on things that might not have been considered.

Keywords: Machine Learning, Education, Student Satisfaction Ratio, Higher Education.

1. Introduction

Satisfaction is a feeling of happiness that is obtained when a person is fulfilled. His or her needs and desires. Students' satisfaction is a short-term attitude, Resulting from an evaluation of a student's educational experiences. It is a Positive antecedent of student loyalty and is the result and outcome of an educational system.

Student satisfaction can be defined as a function of the relative level of experience and perceived performance of educational service during the study period. satisfaction refers to the evaluation of a service and comprises Cognitive, affective, and attitudinal components. Higher satisfaction

is associated with consumer loyalty and a positive effect, in research that has mostly tested cross-sectional associations. Therefore, the institutional outcome such as word-of-mouth recommendation, enrolment, and retention relate to student satisfaction but reciprocal relations are possible. Student satisfaction can also have important implications for teaching due to its bidirectional associations with student engagement and achievement.

It is a positive antecedent of student loyalty and is the result and outcome of an educational system (Amran Rasli et al.,2012). Student satisfaction can be defined as a function of the relative level of experiences and perceived performance in educational service during the study period. Satisfaction refers to the evaluation of service and comprises cognitive, affective, and attitudinal components. Higher satisfaction is associated with consumer loyalty and a positive effect, in research that has mostly tested cross-sectional associations. Therefore, institutional outcomes such as word-of-mouth recommendations, enrolment, and retention relate to student satisfaction; these associations are generally interpreted as indicating consequences of satisfaction but reciprocal relations are possible (Farida Haji et.al.,2020) Student satisfaction can also have important implications for teaching due to its bidirectional associations with student engagement and achievement. The national Higher Education Regulatory Authority (Nhera), will be set up to regulate in a 'light but tight' and facilitative manner, meaning that a few important matters – particularly financial probity, good governance, and fully online and offline public disclosure of all finance, procedures, faculty/staff, course, and educational outcomes will be very effectively regulated while leaving the rest to the judgment of the HELs, which is essential to institutional autonomy, innovation, and pursuit of excellence.

1.1. HIGHER EDUCATION

Quality Universities and Colleges: A New and Forward-looking Vision for India's Higher Education System divided into 5 reforms.

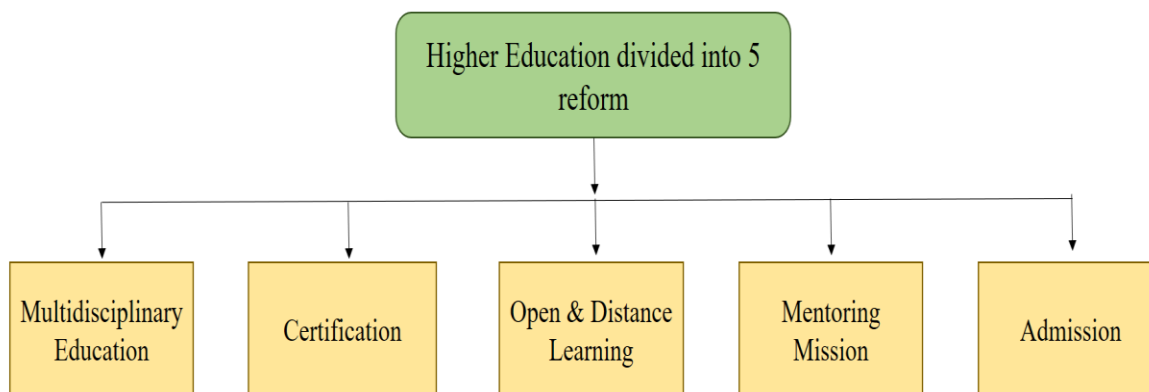


Fig. 1. Higher Education System

This reform is used in multidisciplinary education in Multidisciplinary used in a flexible curriculum, creative combination of subjects, and Integration of vocational education. Reform 2 used in certification in UG education can be for 3 or 4 years with multiple exit options. Appropriate certification within this period (Sadia Ijaz et al., 2022). Emphasis on online course digital repositories improved student service credit-based recognition of MOOC, etc. A large pool of outstanding senior retired faculty. Including those with the ability to teach in Indian Languages. Be willing to provide short and long-term mentoring/ professional. To University/ College teachers (Sadiq Hussain et al., 2018). 3.5 crore new seats will be added to higher education institutions—single entrance exams for other than medical and legal students.

1.2. Optimal Learning Environments and Support for Students

- Optimal learning requires a comprehensive approach that involves appropriate curriculum-engaging pedagogy, continuous formative assessment, and adequate student support.
- The curriculum must be interesting and relevant and updated regularly to align with the latest knowledge requirements and to meet specified learning outcomes.
- High-quality pedagogy is necessary to successfully impart the curricular material to students (Saurabh Bhagvatula et al., 2015).
- The assessment methods must be scientific, designed to continuously improve learning and test the application of knowledge.

- The development of capacities that promote student wellness such as fitness, good health, psycho-social well-being, and sound ethical grounding are also critical for high-quality learning.

1.3. Weka in Research Work

- Weka is a collection of machine-learning algorithms for data mining tasks. The algorithm can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization (**Mahesh Singh et al., 2014**).
- Easy Simple yet powerful tool for data mining weka is easy to learn. I can directly fit the data to any machine learning algorithm available in Weka which makes my job simple without worrying about coding.
- Found only on the islands of New Zealand, the weka is a flightless bird with an inquisitive nature. The name is pronounced like this, and the bird sounds like this.
- Weka is open-source software issued under the GNU General Public License.
- We have put together several free online courses that teach machine learning and data mining using Weka. The videos for the courses are available on YouTube.

1.4. Random Forest Algorithm

Random Forest Algorithm (RFA) is widely used to be a powerful new approach to data exploration. RFA is a collection of CART-like trees for growing, combination, testing, and post-processing (**Dr. N. Venkatesan et al., 2015**).

1.5. J-48 Algorithm

J48 algorithm is one of the most widely used machine learning algorithms to examine the data categorically and continuously. The C4.5 algorithm (J48) is mostly used in many fields for classifying data for example interpreting the clinical data for the diagnosis of coronary heart disease, classifying E-governance data, and many more (**Saurab K. Ghosh et al., 2021**).

2. Literature Review

This paper reviews the part of pupil satisfaction and factors that may impact pupil satisfaction and presents the empirical exploration with the case of the University. assessing the satisfaction of manly and womanish scholars (**Hoang Thai Son et al., 2018**). In the 21st century, those countries will lead the world who have quality advanced education systems because advanced education not only generates larger particular and social wealth, but all the aspects of development similar to education, The Government of India has started Rashtriya Uchcharat Shiksha Abhiyan (RUSA), a Centrally Sponsored Scheme in 2013 and as a result, at present the registration rate is 25 (**Jwel Hoque., 2018**).

In this competitive world, the educational setting also uses Data Mining tools to explore and assay pupil performance, prognosticate their results to help drops out and concentrate on both good and academically poor players, feedback for the faculties and preceptors, visualization data, and have a better assessment of literacy process (**Aaditya Desai et al.,2012**). The data cover three groups of characteristics- particular characteristics, academic terrain, and social factors. Grounded on the review of the literature, four MLAs have been named, which give good results in analogous exploration NB, MLP, SMO, and J48. Attribute selection ways are applied and 14 attributes are named and ranked. The processing is done with Weka open-source software (**Veselina Nedeva et al., 2020**).

In this exploration, trouble is made to find the impact of our proposed features on pupil performance vaticinator with the help of bracket models. This model uses data mining ways similar to Decision Tree, Random Forest, and Naïve- Bayes classifier to prognosticate the delicacy of our model. originally by considering many attributes the model prognosticated the delicacy (**P.Ajay et al., 2020**). In conclusion, we point out that the scholars' attributes enforced in our work aren't bounded rather more new attributes can be introduced in our database to ameliorate the quality of our model. New attributes and more experts can be added to get further sapience into factor conditions (**Matthias Schonlau et al., 2020**). Performance analysis for scholars is a major problem. It's important that they're combated. The work reported in this thesis indicates the machine literacy ways with supervised literacy algorithms to understand the performance of the algorithm with respect to pupil records where we analyze the performance of scholars and

distributed it into three classes as high, average, and low with a delicacy of 79% (**Gorinkala Hemasri et al., 2022**).

3. Learning Environment

Conceptually, the learning environment refers to diverse components and activities within which teaching and learning occur. Hence, it takes into consideration several variables that directly and indirectly affect students. Productive teaching and learning require the blending of many factors which include the classroom seats and sitting arrangement, painting and lighting, classroom climate, air quality, and ventilation. had evaluated the mathematics achievement of students with regard to the learning environment. The study proved that the learning environment is a vital key factor for their academic accomplishment. pointed out that high achievers are likely to have undergone curriculum content under a conducive and ideal environment. found that the essential factors for a healthy and positive environment are school climate conditions such as the structure of the school building and the relationship between students and teachers. Apart from that, said that a good learning environment must have good features of facilities itself. Research has proven that schools with a friendly and conducive environment will have students achieve higher in comparison to schools with dull and poor learning environments.

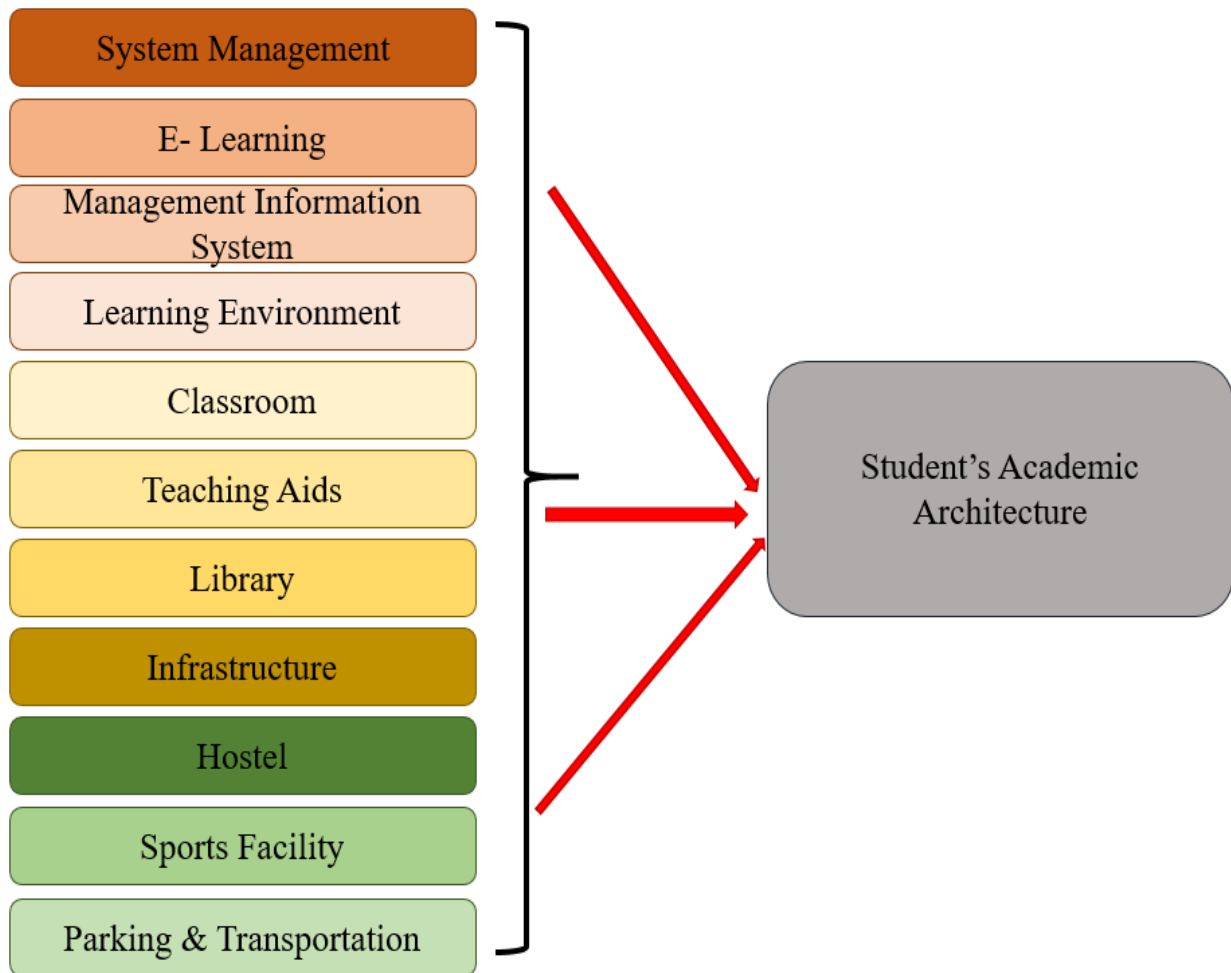


Fig. 2. Concept Framework of the Study

3.1. Classifier Evaluations and Error Measurement Techniques

The performance measures are derived from confusion matrix. A confusion matrix is formed based on the four outcomes of binary classification. In binary classification, the dataset usually has two labels positive (P) and negative (N). The outcomes are true positive (TP) i.e. correct positive prediction, true negative (TN) i.e. correct negative prediction, false positive (FP) i.e. incorrect positive prediction and false negative (FN) i.e. incorrect negative prediction.

a. Sensitivity (Recall or True positive rate)

Recall is the number of correct classifications divided by the total number of positives. So, $R = TP / (TP + FN) = TP / P$ -----(1)

b. Precision

Precision is the number of correct positive classifications divided by total number of positive classifications.

So, $P = TP / (TP + FP)$ -----(2)

c. F-score

F-score is harmonic mean of precision and recall. So, $F = 2PR / (P+R)$ -----(3)

Accuracy is the number of all correct classifications divided by the total numbers of cases. So,

Accuracy = $(TP+TN) / (TP+TN+FN+FP) = (TP+TN) / (P+N)$ -----(4)

The following section explains different error measures used for classification methods.

e. Mean Absolute Error (MAE)

MAE estimates how far the predictions or forecasts differ from the actual values.

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x|$$
 -----(5)

where n = the number of errors, $|x_i - x|$ = the absolute errors.

f. Root Mean Square Error (RMSE)

RMSE is an evaluator of the differences between the predictor values and the actual observed values.

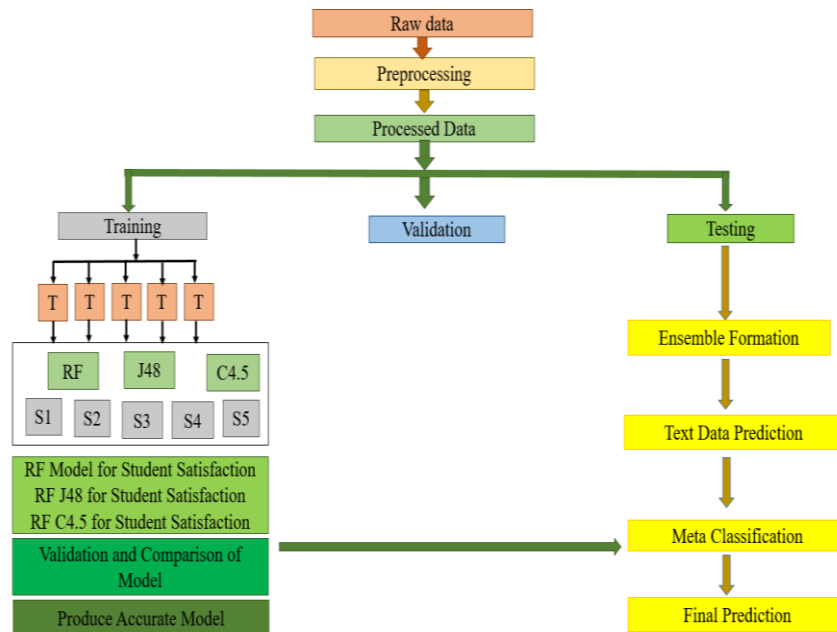


Fig. 3. Flowchart of our Work

4. Preparation of Dataset

For our work we have designed a google form and circulated in various colleges and universities of Chhattisgarh. Raw data was collected and preprocessing was applied to remove noisy data. Sample of google form is given below. To excel our research, we have developed a questionnaire based on the results of the literature review measuring students' satisfaction. In total 30 items were created in the categories of teachers' roles (such as fairness, and mentoring) course curriculums (such as relevant assignments and project works given to help us learn the subject), college facilities administrative staff (such as the administrative staff responds to my needs promptly), future prospects (such as the college organization guest speakers to help us be informed about job market), college location (e.g. the college is accessible to transportation) and overall satisfaction (such as I have gained positive environment from the institute). Later on, demographic characteristics such as gender, age, course, and year of students were measured.

Datamining is a technique used for the analysis of a large amount of data. This technique is helpful in our domain as our domain details millions of data ringlets. As we know classification is one of

the methods to analyze a big amount of dataset. With labels. In our work quality ratio is the mail label and so we have selected a classification method for the analysis of our dataset. We have developed a program in Java. This program is for Random Forest, J48, and C4.5 classifiers. The flow of our work starts from the collection of raw data, Preprocessing, and development of the prepared dataset. In the next step, we go for the training, testing, and validation part. Training of data is done in five rounds. This training is for the given three algorithms. Trained data is then tested with the given dataset.

This figure is based on an academic Google form on student satisfaction. This form was used in a survey to find out the academic performance of the institute. In the above-designed form, we have taken four questions. All questions are related to student academic performance. These questions will help us to find students' academic performance. We have also gone through National Education Policy 2020.

Academics

SURVEY TO FIND OUT ACADEMIC PERFORMANCE OF THE INSTITUTE .

FAIRNESS

YES

No

ASSIGNMENT DISCUSSION

Yes

No

Fig. 4. Academics Google Form

MENTORING

YES

NO

PROGRAM OUTPUT *

YES

No

Fig. 5. Infrastructure Google Form 1

This figure based on an Infrastructure Google form in student satisfaction. This form used in survey to find out Infrastructure performance of the institute. In the above designed form, we have taken

four questions. All questions are related to Infrastructure. These questions will help us to find students' view about institute infrastructure. We have also gone through National Education Policy 2020.

The image shows a Google Form titled "INFRASTRUCTURE" with the subtitle "INFRASTRUCTURE OF THE INSTITUTE". It contains two sections: "BASIC FACILITY" and "LIBRARY". Each section has two radio button options: "YES" and "No".

Fig. 6. Infrastructure Google Form 2

In the above designed form, we have taken twelve questions. All questions are related to students of higher education. These questions will help us to find students view about higher education.

We have also gone through National Education Policy 2020.

Id	Stream	College	Gender	Water	Washroom	Campus	Teaching	Pedegory	Holistic	National	International	Package	Percenta...
84101648	PG	Pragti College	Female	1	0	1	1	0	0	0	1	1	55
84101680	PG	Mats College	Male	0	1	0	1	0	0	1	1	1	55
84101695	PG	Mats College	Female	0	0	0	0	0	0	0	0	0	0
84101727	UG	Pragti College	Female	1	1	1	1	1	1	1	0	0	77
84101742	PG	Mats College	Male	1	1	0	0	0	0	1	0	0	33
84101773	UG	Pragti College	Female	0	1	1	0	0	0	1	1	0	44
84101789	UG	Pragti College	Female	1	0	1	1	1	1	1	1	0	77
84101820	UG	Pragti College	Female	0	1	0	1	0	1	1	1	0	55
84101836	Phd Scholar	Pragti College	Female	0	1	0	1	1	1	0	0	0	44
84101867	UG	Mats College	Male	1	0	0	1	1	0	1	0	1	55
84101883	UG	Pragti College	Male	1	0	0	0	0	0	1	1	0	33
84101914	UG	Mats College	Male	0	1	1	1	1	0	1	0	1	66
84101930	PG	Mats College	Male	1	1	1	1	1	1	1	0	1	88

Fig 7. Student Satisfaction Data 2018-2019

This model has been installed WEKA tab there is a package manager in classification. This model is appearing we have also install the GitHub and package URL is given in WEKA. Given figures is the screenshot of our model in WEKA.

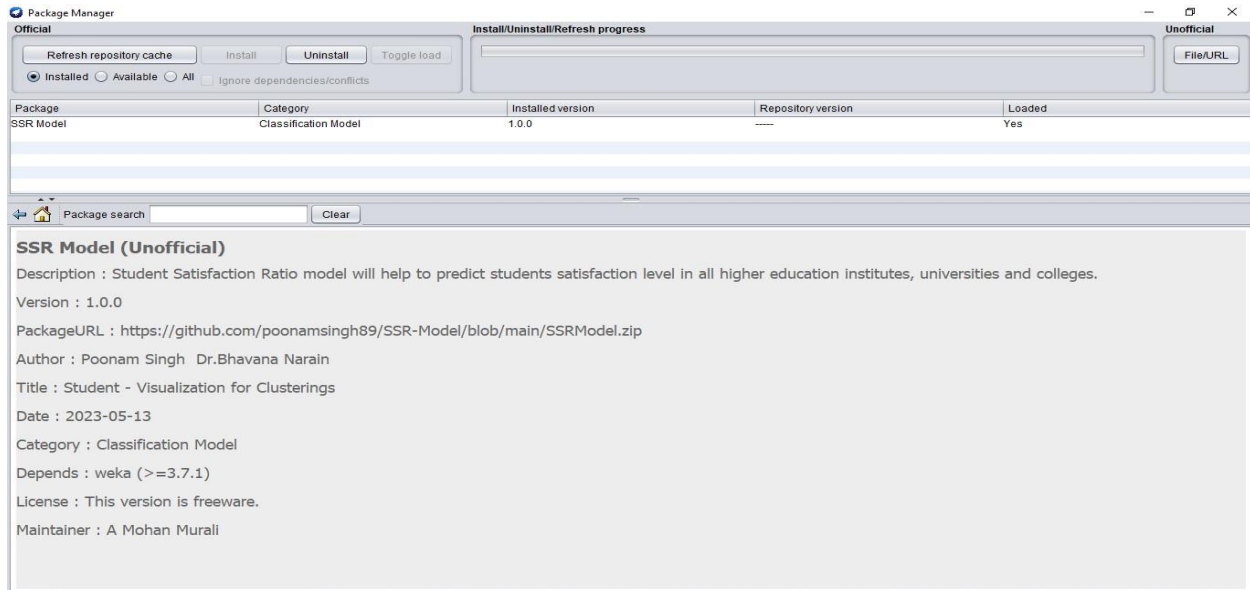


Fig. 8. Shown this screenshot of URL Package in uploaded GitHub

Given below fig. 4.11 is the screen short of error list of our SSR model. We have taken error list all instance of our model and separated instance of our model.

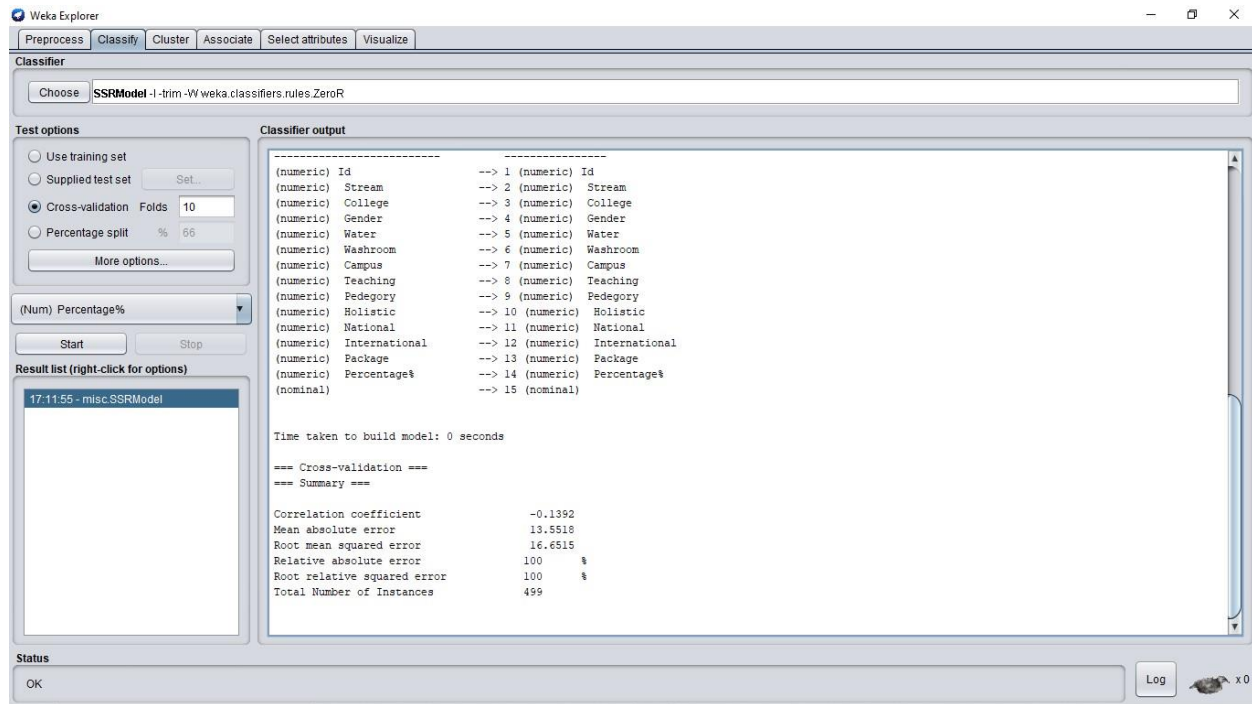


Fig 9. This Classifier Show the Error List

5. Result and Analysis

This is an output and data collection sheet from our model. This sheet stores all panel questions student response details.

SSRM (Student Satisfaction Ratio Model) is chosen for implementation of model. The objective of selecting this model is to understand the basic concepts of student satisfaction in an organization and also to apply these models to define the percentage of student satisfaction ratio in real time. SSRM is helpful in learning the basic concepts of college and university basic facility for students with different options and analyzes the output that is being produced.

Implementation Procedure Used in SSRM (Student Satisfaction Ratio Model):

The dataset file is fed into the Java model and the classification for SSRM is implemented as defined in the following steps:

- The first step enroll students in the SSRM model for student detail.
- The second step is to feed the organization or college details, session, and password.
- Third step is tick the different penal for satisfaction ratio in student.
- All panel show percentage for student satisfaction ratio.

e) This model main purpose improvement quality for college and university for student satisfaction.

Our proposed model is student satisfaction Ratio Model. We have created two panel

1. Student Panel
2. Organization Panel



Fig. 10. Level of Administrative Satisfaction in Administrative panel

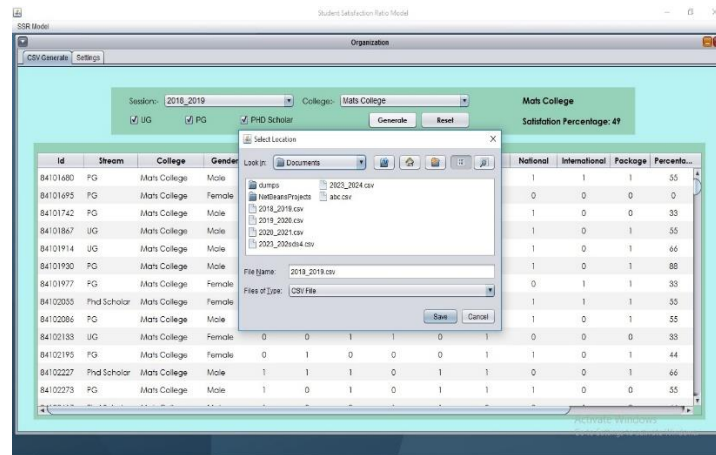


Fig 11. Student Satisfaction Data Set Selected Location

There is the first option Administrative which has three options first water second washroom third campus in this if the student clicks on one option then the percentage shows 35% which means that the student is not satisfied this student satisfaction ratio is less than 50% if the student clicks on two or three option then his percentage ratio more than 50%.

This sheet show dataset stores in laptop drive location in session wise. Store dataset in use file name and type of file name.

6. Comparative Study

Adaptive Environment and Users in SSRM. Data mining is integrating multiple areas, which provide coverage to users and Educational Systems in number of recent Case studies.

Table.1. provides a comparison of all objectives in Recent Case studies.

Action	Student Satisfaction	Educational System	Case Study
Cristobal et al.,	Learning, Educators and Course Developers, Organization, System Administrator	Traditional education LCM, ITS, AIWBE System, Web-based Course Ventura	Cordoba University
Fliguni et al.,	Students, Educational Researchers, Learning Providers	Traditional Education	Educational Classroom
George et al.,	Research Communities, Tool Developers, Analytics Practitioners	Learning Analytic and Knowledge	Society Learning Analytic and International Data Mining Society Knowledge (SoLAR & IEDMS)
Algami	Students and Course Developers	Traditional Education, LMS	KKU (King Khalid University)
Nikhil et al.,	Student, Teachers	Learning Content management System	(Shree Rayeshwar Institute of

			Engineering & IT) SRIEIT
Sadiq et al.,	Administrations, Students, Teachers	Learning Content management System	Dibrugarh University
Jiechao Cheng	Learning, Teachers, Researchers, Institutions	Learning Content management System, Web-Based System	International School of Software Wahan Universidad National de Colombia
Kalaivani et al.,	Administrators, Students	Web-Based System	Universided National de Colombia

7. Conclusion and Future Work

The proposal to design the status is to find out how certified the student is from the different facilities of the college and university and we can improve the report of that college by knowing what is lacking in any college and university through this student satisfaction ratio model. In future work, we will convert this model using deep learning techniques. So that the model can be trend-accurate result can be Gant's new model will be using Python language. We will pattern our model and make it available for maximum higher education organizations like (universities, and colleges) so that they can find out the loop point of their organization higher education organizations remove all the luckuna so that a maximum number of students can be satisfied from their organization our model will be available for of app.