




Ensemble learning-based prediction of contentment score using social multimedia in education

Maninder Kaur¹ · Himika Mehta¹ · Sukhchandan Randhawa¹ · Pradip Kumar Sharma² · Jong Hyuk Park³ 

Received: 18 September 2019 / Revised: 25 November 2020 / Accepted: 10 March 2021 /
Published online: 24 August 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Over the years, social multimedia has gained credibility as a source of information and a reliable platform on which organizations, students and employees can interact with expert audiences. In the areas of education and teaching, the technique, multimedia and network of use computerized methods to build a creative environment for learners and broaden our perspective on a variety of topics. Getting new information and sharing it with others has become much easier with social multimedia. To boost the productivity and growth of any university, student contentment is a critical factor. Student contentment level is the need of the hour that is necessitated to be analyzed every year for the progress of the university. In this paper, we use a social multimedia technique to collect data from the students of the university based on a designed questionnaire circulated. The collected information embraces different aspects like academics, research, recreational, and technology that portray the image of the university. The current work relies on developing a stacking ensemble machine learning model for prediction of student's overall contentment score, an indicator to perceive overall, how much the university gets the thumbs up from its current students. The work employs the cuckoo search meta-heuristic based wrapper method for feature selection from the original dataset with 78 features. The proposed ensemble model portrayed a lowest RMSE value of 0.373 by the combination of Self Organizing Map, Multilayer Perceptron, Boosted Generalized Linear Model and Gaussian Process with Polynomial Kernel along with Partial Least Squares as meta-learner, showcasing its ability to accurately predict student contentment levels of a University. The proposed machine learning framework acts as a great developmental tool for foreseeing and analyzing student contentment for its university.

Keywords Student contentment · Regression · Machine learning · Ensemble approach

✉ Jong Hyuk Park
jhpark1@seoultech.ac.kr

Extended author information available on the last page of the article

1 Introduction

With a progressively increasing competition in the education sector, student contentment has emerged as an essential factor in magnetizing and retaining the best students, who, in turn, perk up the rank and status of the university. The term ‘Student Contentment’ is a measure that quantifies the student’s appraisal of the services offered by the colleges and universities [6]. Student satisfaction is of convincing enthusiasm to schools and colleges as they try to consistently improve the learning condition for undergraduates, meet the desires for their constituent gatherings and authoritative bodies, and exhibit their institutional viability. It is an important indicator of the quality of learning experiences. Higher education establishments consider student satisfaction as one of the significant components in deciding the nature of different projects in the present markets [14, 38, 39]. Resources such as effort, money and time are being spent by the students at the university, and thus it matters to students as they aim to achieve quality education for their higher studies. Contentment level, at a particular Institute/University, alters the motivational level of the students and ultimately changes the rate of retention of the students at a specific University. It allows universities to build a system for continuously keeping an eye on how effectively University caters student needs. From a university point of view, satisfied students are more likely to do well academically, which further augments the financial status and reputation of the University. Student contentment reflects not only the relish time of a student at University, but it also focuses on his performance within and outside University. A new study has found that student contentment has a remarkable influence on the apparent uniqueness of an institution, which has a direct impact on the achievement of student’s recruitment efforts [20]. Student contentment is vital for promoting institution life as well as it is catering to an essential influence for standing the university in global rankings.

In recent times, differing research has featured the significance of examining studying the life satisfaction of students, for the scholastic and individual repercussions [26, 32]. Most of the research target a broad range of topics, addressing student contentment. For example, some practitioners attempt to explore which aspects influence student contentment [28, 35]. Others investigate the correlation involving student contentment and other factors such as learning environment, service quality or instructional design and management style [20]. The authors [17] focused on finding the viewpoint of the students on sustainability. The authors [33] presented research work to determine whether Facebook likes are sufficient to determine student contentment in Open Distance Learning (ODL) or not. Some research work analyzed student contentment for courses like Computer Architecture and Organization by the use of multiple e-Learning tools [3]. The researchers [9] discussed how student contentment is related to teaching, along with other university experiences.

Further, the research work [18] investigated the contentment level at the School of Technology and Applied Sciences, Mahatma Gandhi University; Kerala. The authors [29] concluded that there is a rise in the contentment level of the students from the year 2007 to 2013, with the improvement in laboratory conditions that included laboratory notes, the computers, and the other engineering equipment. The researchers [40] presented that how meaningful is the participation and contentment of the students towards the approach of online learning. The research also focused on finding the methods the faculty followed to promote online learning among the students [1, 23]. Some researchers [13] worked on a project that aims to enhance the level of student contentment using a Virtual Learning Environment. The

work [25] aimed to trace out the aspects affect the happiness of the pupils in universities in Pakistan and to evaluate these factors' relationship, either positive or negative with the satisfaction. The research [31] predicted the levels of student contentment in face-to-face learning. The work [19] examined the extent to which dealings and other predictors put in to student contentment in online learning. The researchers [21] determined the predictors of student contentment focusing on recreational sports and cultural facilities within the campus. The research [12] discovered that log files from student course activities play a significant role in predicting student contentment with modules from a virtual learning environment. The authors [11] investigated that data mining techniques can choose a lesser number of constructs that need consideration to handle student contentment. The work [15] concluded that the two most significant aspects to student course contentment are both the count of enrolled students to a course and their extraordinary merit rate in their final grading.

In recent times, machine learning has been associated with the field of educational studies. In the report of 2018 New Media Consortium Horizon (NMCH), the sector of machine learning is appealed to be embraced as a part of adaptive skills integrating artificial intelligence in the succeeding 2–3 years, and as a part of analytics skills less than a year [5, 8, 16, 22, 24].

Utmost studies on an eminence in the higher education sector pay more attention to academic features rather than administrative factors [10, 27, 30, 34, 41]. These studies mostly concentrate on the quality of courses, their teaching methodologies, and effective course delivery tools. The author [7] for example, suggests several determinants influencing the image of a higher education institution, such as average class strength, diversity of courses, reputation from academic perspective, students' qualification, as well as their personal qualities, staff students interaction, etc. On the other hand, the author [4] highlights the significance of student consultation, well-being of the student, library facilities, teaching environment, technical facilities, etc.

Until now, research usually focuses on determining correlations among student contentment and academic success of students or the final scores and individual psychological elements. This work is the first step towards the prediction of student contentment score from different aspects of Academics, Faculty zone, General, Research zone, Extra-Curricular, Hostel, Technology, Teaching Practices and Recreational zones. The purpose of this work is to exploit machine learning practices in the field of educational research. The paper presents an ensemble machine-learning model for predicting student contentment score based on the data procured using a designed questionnaire circulated through online means [37].

1.1 Major research contribution

The following are the main contributions:

1. The proposed predictive analytic ensemble model integrates the capabilities of each predictive model thereby generating improved and consistent predictions. The proposed framework performed notably better than the best performance models viz. 'Gaussian Process with Polynomial Kernel'
2. From a computational point of view, the ensemble approach is more in demand and Wrapper feature selection using heuristic inspired by cuckoo to optimize is the generalization performance of a predictive model

3. The superiority of the novel ensemble prediction model was validated based on the data procured using a designed questionnaire circulated through online means
4. In contrast to the baseline models, the proposed ensemble prediction established its supremacy and the results revealed superior student satisfaction predictive strength in comparison to the base models.
5. The comparison of the prediction efficiency of the proposed model along with the explorations of the significance of the proposed prediction model was carried out.

1.2 Organization

The remaining part of paper involves the following sections: Section 2 gives the detailed background of materials and methods for the proposed analytical model; Section 3 demonstrates the experimental results of the proposed model along with comparative glance with base models based on various performance metrics, and Section 4 concludes the work with highlights of future research directions.

2 Materials and methods

This section presents a detailed description of the proposed methodology along with the features of procured data. The proposed methodology encompasses the prediction module that predicts the contentment rate of the students by employing multiple ensemble machine learning models on the ratings received from the students through Google forms.

2.1 Data collection

In the current work, the data related to the evaluation of student contentment level is originally collected by circulating the questionnaire to the students of the university through online means. The survey named The Overall Contentment Predictor Questionnaire (TOCPQ) is designed to know the perceptions of the students of the university that can further aid in forecasting the student's contentment level associated with that university as depicted as Table 2. The survey focused on distinct groups like Academics, Campus environment (Hostel facility, library facility, within the class environment), General, and Research as shown in Table 1. The data collected through online means include 78 parameters and 4386 distinct instances. The participants involved pursuing graduation, post-graduation, and doctorate degrees from the university (Table 2).

2.2 Data preprocessing

Data collection step is followed by data pre-processing wherein the data is pre-processed to exclude irrelevant data from the dataset by cleaning the noisy data along with imputing missing values through interpolation technique. Various categorical attributes are changed to numerical values (Fig. 1).

This step is followed by the application of cuckoo search meta-heuristics wherein the original dataset, including 78 features is reduced to an optimal subset of features.

Table 1 Broad classification of the dataset

Class No.	Parameter	Description
Class- 1	Gender	It includes the gender information of the participant.
Class- 2	Academics	It contains all the parameters related to academics such as syllabus of the course.
Class- 3	Faculty	It contains all the parameters which are related to faculty.
Class- 4	General	All the general facilities are grouped together and placed in this class.
Class- 5	Research	The research parameters are put together in this class.
Class- 6	Extra-Curricular	It contains the parameters which tell about the extracurricular activities at the campus.
Class- 7	Hostel	All the hostel facilities are included in this class such as hostel living and food etc.
Class- 8	Technology	It includes all those parameters which are related to technology such as the use of latest version of the software in labs.
Class- 9	Teaching Practices	It gives us the information about the teaching practices within the university
Class- 10	Recreational	All the recreational parameters are included in this class such as meditation classes.

2.3 Proposed methodology

The prediction module broadly consists of three broad sub-modules, namely Data Preparation, Feature selection using cuckoo search, Stacking ensemble, and Cross-Validation. The pipeline of the methodology is illustrated in Fig. 3.

2.3.1 Cuckoo search meta-heuristic

The cuckoo search, the best suited meta-heuristic for solving the combinatorial optimization problem relies on the lazy nature of cuckoos that lay their eggs in the host nest. The eggs look same as that of the eggs of host bird. If the host bird differentiates foreign eggs, the bird either discards that nest or kicks out those eggs. On the basis of this lazy characteristic of cuckoos’ species, the authors [36] proposed a cuckoo search algorithm that is a novel meta-heuristic optimization algorithm. The algorithm considers every nest as single solution. The algorithm iteratively improves the new solutions/cuckoo eggs by discarding/replacing the older ones based on fitness function values. The detailed pseudo code of the proposed model for finding the student’s contentment score is given in Fig. 2.

The main sections of the pseudo code are as follows.

Encoding of Solution The solution is encoded as a combination of zeros and ones with ‘m’ bits where m-representing the number of features in original data (here m = 78) with 1 signifying selection and 0 as discard of feature in the feature space.

The solution $X_{i,j}^k$ is characterized as in Fig. 3.

The new solutions are bounded to binary values using Eq. (1) .

$$Z(X_{i,j}^k) = \frac{1}{1 + e^{X_{i,j}^k}} \text{ and } X_{i,j}^{k+1} = \begin{cases} 1, & \text{if } Z(X_{i,j}^k) < \hat{u} \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $\hat{u} \sim U(0, 1)$ & $X_{i,j}^k$ represents the solution at generation k.

Table 2 The proposed overall contentment predictor questionnaire

<p>Section 1. General information</p> <p>What is your gender? •Male •Female</p> <p>What are you? •Day Scholar •Hosteller</p> <p>What is the mode of teaching that you prefer? •White Boards • Presentations •Combined</p> <p>If you are joining college yet again, would you select this campus again? •Yes •No</p> <p>Are any malpractices observed within the college campus? •Yes •No</p> <p>Section 2. Overall contentment levels: Contentment Rating 1 – Very Dissatisfied 2 - Somewhat Dissatisfied 3 - Neutral 4 - Somewhat Satisfied 5 - Very Satisfied. How contented are you by:</p> <p>The orientation program/ awareness program offered when you were a fresh student at this college (e.g. New student orientation activities, different block locations, timetable, and schedule of classes)?</p> <p>Fees structure of your college?</p> <p>The security and safety at campus?</p> <p>how much was understood in class the quality of learning</p> <p>The syllabus that is structured for your course?</p> <p>The opportunities to meet the faculty outside the classroom?</p> <p>The availability of courses to make progress towards your degree?</p> <p>The use of technology in your class (like projectors, different android applications)?</p> <p>Accessibility of your teachers (by emails, Phone calls)?</p> <p>Overall quality of examination procedure?</p> <p>The placement drives in campus?</p> <p>Section 3. Numerical questions</p>	<p>What is your course? •Graduation • Post-Graduation • Doctorate</p> <p>Which hostel do you live in?</p> <p>Are your teachers audible in your class? •Yes •No</p> <p>Does the strength of your class affects the quality of teaching? •Yes •No</p> <p>Have you been a part of learner club/ organization? •Yes •No</p> <p>Exposure to outside class experiences (e.g., participating in sports, Members of student clubs participating in organized social or cultural activities)?</p> <p>The academic procedure that is being followed?</p> <p>The quality of teaching?</p> <p>what was the quality of coverage how much was remembered (content)</p> <p>The classroom/ lab facility?</p> <p>How well are you able to correlate while the teacher is teaching?</p> <p>The schedule of your classes?</p> <p>With the campus bookstore?</p> <p>The student financial aid services (like funding of projects)?</p> <p>The campus library services?</p> <p>The food services provided in campus?</p>	<p>Which year? •1 •2 •3 •4</p> <p>What is your current CGPA?</p> <p>Is white board visible? •Yes •No</p> <p>Is discipline well maintained in your class? •Yes •No</p> <p>Is guide available to you? •Yes •No</p> <p>Your casual participation outside class (e.g. having informal student discussions, get together, watching plays, listening to music)?</p> <p>The extent to which faculty follows the methods like class presentations, assignments or discussions to increase the productivity of students?</p> <p>Your chance to take part in an independent research project with a tutor?</p> <p>Personal attention in the class how much was covered (syllabus/Topic)</p> <p>Your prospect to take part in a abroad study program?</p> <p>The approach with which the college authorities deal the grievances of the students?</p> <p>The campus resources for students (such as scholarships, reductions in fees with low yearly income or good grades scoring students)?</p> <p>The internet facilities at your campus?</p> <p>The medical services?</p> <p>The recreational facilities (gym, swimming pool) in campus?</p>
--	---	---

Table 2 (continued)

<p>What is the strength of your class? < 30 • 30–70 • 70–100 • 100–150 • > 150</p> <p>Section 4. Hostel facilities 1 – Very Dissatisfied 2- Somewhat Dissatisfied 3- Neutral 4- Somewhat Satisfied 5- Very Satisfied. Your level of satisfaction with:</p> <p>The overall hostel facilities?</p> <p>The basic requirements in hostel (like water purifiers, washing machines)?</p> <p>The internet facilities in your hostel?</p> <p>Section 5. Estimate the count of times. Zero times . 1–2 times . 3–4 times . 5–6 times . 7 or more</p> <p>Collaborative research work with the faculty extra to curriculum? (1 project = 1 time)</p> <p>Attended a cultural event (art exhibition, plays, and dance or theatre performances)?</p> <p>Participation in organizations (like IEEE, ACM), Community events?</p> <p>Section 6. How many hours do you pay? Zero 1–5 h 6–10 h 11–14 h 15 or more</p> <p>Working (for pay) on/ off campus?</p> <p>Exercising or practicing in clubs or sports?</p> <p>Section 7. Feedback /comments</p> <p>Section 8. Overall contentment score _____ BETWEEN [0 TO 5]</p>	<p>When are most of your classes scheduled? 8 am–4 pm • 10 am–5 pm • 12 noon– 7 pm</p> <p>The cleanliness of hostel rooms?</p> <p>The hostel mess food?</p> <p>The other basic facilities in your hostel (gym, lift)?</p> <p>Participated with faculty members on activities other than course work (e.g. Student organization, outside class activities)?</p> <p>Participated in spiritual/ religious activities off/on campus (meditation, worship, prayer)?</p> <p>How many class sessions do you skip in an average week?</p> <p>Studying / doing homework/ team projects outside a class?</p>	<p>How many staff members do you acquainted with well whom you can ask for a personal or a professional advice? • Zero • 1–3 • 4–6 • 7–9 • 10 or more</p> <p>The bedding provided in hostels?</p> <p>The timings of food availability in the mess of your hostel?</p> <p>Attended out of class lectures (university funded presentations/speakers)?</p> <p>Attended an athletic event?</p> <p>Voluntarily practicing in community service?</p>
---	--	--

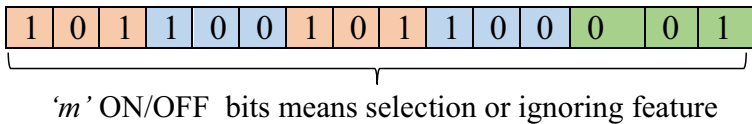


Fig. 1 Solution encoding

New Solutions generation using Lévy flight The new solution is obtained from the randomly selected solution using Levy’s flight as given in (2).

$$X_{i,j}^k = X_{i,j}^{k-1} + \alpha \times L \times (X_{i,j}^{k-1} - \text{Optimal_}X_{i,j}) \quad (2)$$

where $X_{i,j}^k$ the new solution which is generated using Lévy flight; $X_{i,j}^{k-1}$ represents arbitrarily chosen solution from the population space; α represents step size; $\text{Optimal_}X_{i,j}$ signifies the best solution found so far; and L is the step length (Lévy flight vector). The newly obtained solutions are compared with the old ones on the basis of their fitness function values and the highly fitted solution is forwarded to the population for subsequent processing.

Discovery of alien eggs For each cuckoo within the population, the probability matrix, is used for identifying the alien eggs as presented in Eq. (3)

$$\rho_{i,j} = \begin{cases} 1, & \text{if } \text{random}(0, 1) < \rho \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

Input: Read the dataset (*DS*), Max iterations (*G*), population size(*N*), Loss Parameters (α & ρ), Dimensions of Solution(*D*),

Output: Cuckoo’s Solution with best fitness (Least RMSE)

Begin: For each Host Nest in the population, $i = 1: N$

For every egg’s dimension in Nest, $j = 1: D$

Random initialization of cuckoo egg i.e. Solution, $X_{i,j}^0$

Fitness $_i^{\text{th}}$ Solution = ∞

End of For

End of For

While (iterations < *G*)

For every cuckoo egg nest, $i = 1: N$

New Solution NX_i Fitness $FT \leftarrow$ Calculate_fitness ($DS, X_{i,j}^{g-1}$);

End of For

For every cuckoo egg nest, $i = 1: N$

For every egg’s dimension in Nest, $j = 1: D$

$X_{i,j}^g \leftarrow X_{i,j}^{g-1} + \text{Lévy Flight (as in (2))};$

End of For

$New_X_i \leftarrow$ Cuckoo_Fitness ($DS, X_{i,j}^g, FT_i$);

End of For

For every Nest, $i = 1: N$

For egg’s dimension in Nest, $j = 1: D$

Make confined step size, (LS) using eq.(5) and $\rho_{i,j}$ from eq (3)

$X_{i,j}^g = X_{i,j}^{g-1} + LS * \rho_{i,j}$, (as in (4));

End of For

$New_X_i \leftarrow$ Cuckoo_Fitness ($DS, X_{i,j}^g, FT_i$);

End of For

End of While

End

Fig. 2 The pseudocode of the proposed approach

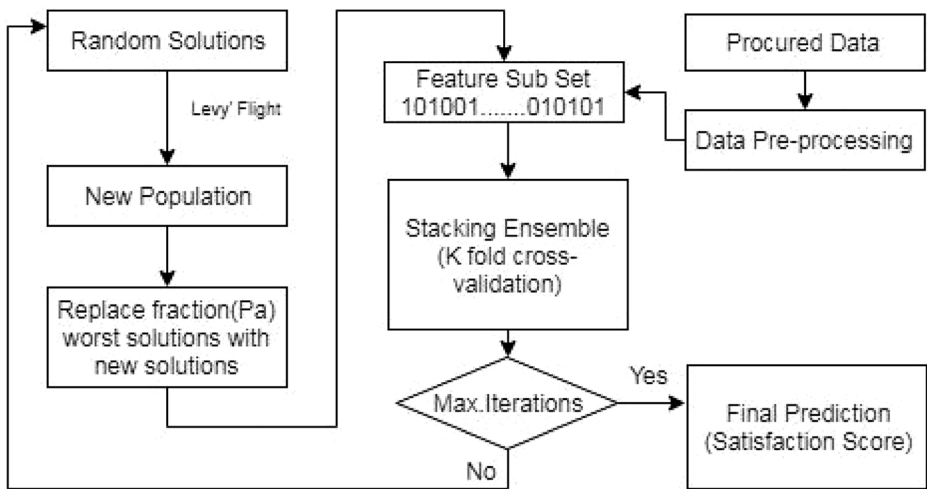


Fig. 3 Depicts the flow of the proposed scheme

where $\rho_{i,j}$ represents the probability of noticing foreign eggs in the solution i for the j th cuckoo’s dimension.

The comparative values of ρ and uniform random number generator $\text{random}(0, 1)$ determines if local random walk is required or not. After calculating the probability of discovery, new solutions are obtained using Eq.(4).

$$X_{i,j}^k = X_{i,j}^{k-1} + LSx\rho_{i,j} \tag{4}$$

Where $\rho_{i,j}$ represents the probability matrix and LS matrix signifies local step size found by using Eq. (5).

$$LS = \text{rand}(0, 1) \times (\text{rand_perm}(\text{Solution}(i))\text{rand_perm}(\text{Solution}(j))) \tag{5}$$

where $\text{rand_perm}()$ shuffles the solution arbitrarily.

The new solution substitutes the old one in succeeding iteration, in case the objective function/ fitness value of the new solution is better than the present one. The entire procedure starting from the generation of solution till searching for foreign eggs is repeated until the fixed set of iterations.

2.3.2 Flow of proposed scheme

The first step deals with data collection. It includes the formation and circulation of the questionnaire. The next step involves data preprocessing. On the original dataset with 78 features, cuckoo search based wrapper feature selection is applied to select the specific set of features. In the next level, the reduced dataset with selected features is used to train the regression models with their optimum tuning parameters employing 5 fold cross-validation [2]. These machine learning models are analyzed on various evaluation parameters viz. Correlation, R-Squared, Root Mean Square Error and Accuracy. Out of the 10 machine learning models employed, the five models combinations are chosen out of possible

combinations on the basis of RMSE value that yields better RMSE using stacking generalization. Figure 1 depicts the flow of the proposed methodology.

In stacking is ensemble the predictions from different base learners, are employed as inputs for meta-learning algorithm [5]. This meta-learner is trained to optimally integrate the predictions from base models to generate a new set of predictions.

In the stacked generalization framework, the predictions from four base models are exploited by the meta-model to predict the student satisfaction score.

The following are the steps of the generation the proposed model from stacking ensemble technique:

- Specify a set of 10 base learners (with specific model parameters).
- State a meta-learner.
- Train each base learner on the training data for student score prediction.
- Out of different combinations, the five models combinations are found with better RMSE (in a group of four-base model and one as meta-learner for making different configurations) that further refines the performance of the prediction model. The final model is chosen using the stacking ensemble. As the data traveled through the five models, the models trained the data to offer reliable and precise outcomes. The best ensemble model was based on RMSE contributes to the final prediction.

3 Experimental results

The experimental results to compute the performance of prediction models have been assessed by applying all approaches stated in Table 3. R programming [36] is employed to implement various machine learning models using the processed data.

Table 3 The models along with their tuning parameters

Underlying model	Functions	Library	Model parameters(Tuning)
Conditional inference random forest	Cforest()	party	t(Surv(time, cens) ~., data=GBSG2, control=cforest_unbiased(ntree=50)
Self organizing map	Som()	kohonen	Grid=somgrid(6,4,"rectangular")
Multilayer perceptron	Mlp()	RSNNS	size=5, learnFuncParams=c(0.1), maxit=50,
Gaussian process with polynomial kernel	gausspr()	kernlab	x, y, scaled=TRUE, type=NULL, kernel="rbfdot", kpar="automatic", var.=1, variance.model=FALSE, tol=0.0005, cross=0, fit=TRUE,
Stochastic gradient boosting	gbmt()	gbm	formula=formula(data), distribution="bernoulli", data=list(), weights, var.monotone=NULL, n.trees=100, interaction.depth=1, n.minobsinnode=10, shrinkage=0.1, bag.fraction=0.5, train.fraction=1, cv.folds=0, keep.data=TRUE, verbose=FALSE, class.stratify.cv=NULL, n.cores=NULL)
Parallel random forest	Foreach()	foreach	(a=1:1000, b=rep(10, 2))
Generalized linear model	Glmnet()	glmnet	family="binomial"
Partial least squares	Pls()	mixOmics	ncomp=2,scale=TRUE,mode=c("regression", "canonical", "invariant", "classic"),tol=1e-06,max.iter=100,near.zero.var.=FALSE, logratio="none",multilevel=NULL, all.outputs=TRUE

The work is implemented in R language on standalone machine with an Intel Core®i7 processor running at 2.80 GHz, 8 GB RAM, using 64-bit windows OS. The main stages of the proposed work are as below:

- Data procurement and Pre-processing
- Feature selection using Cuckoo Search Meta-heuristic
- Prediction using 10 machine learning approaches
- Selecting best combination of five learners (on the basis of RMSE)
- Application of stacking ensemble using four at one time as base learner and the 5th one as meta-learner
- Choosing the best ensemble for final testing.

The dataset procured is free from missing values. The dataset contained values on mixed scale, as percentage, ratios, and average scores. The ratings offered by the students at the university are dealt with the prediction module. The various parameter values settings of the underlying CS algorithm are:

Initial solution count, N - 20
 Generations count, Gen - 500
 Step size, α - 0.15
 Discovery probability, ρ - 0.25
 Levy step length, L - 1.5

The *normalize()* function in R is utilized for normalization purpose. As data is procured from real data, the work is simulated on original data without considering any outliers effect. Table 3 depicts the tuning parameters of the models used in the present study along with their required packages.

Various parameters are being analyzed to compute the overall contentment rate within the university. To evaluate the performance of the proposed methodology various evaluation parameters are calculated such as Root Mean Square Error (RMSE), Correlation, R-Squared and Accuracy.

Root Mean Square Error (RMSE) Root mean square error depicts the standard deviation of the model's prediction by finding the difference between the model predicted values and the actual values. The lower the value of RMSE, the better is the performance of the model. The RMSE is given as in Eq. 6

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

where n represents the total number of observations, y_i denotes the actual observed values and \hat{y}_i denotes the predicted values.

Correlation(r) The correlation coefficient illustrates how well are two variables linearly related. For variables X and Y, correlation(r) is computed as in Eq. 7.

Table 4 Parameters results for Dataset with features, with different training data

Kvalue	1st Fold			2nd Fold			3rd Fold			4th Fold			5th Fold							
	r	R	RMSE	Acc	r	R	RMSE	Acc	r	R	RMSE	Acc	r	R	RMSE	Acc				
Multilayer perceptron	0.71	0.50	1.67	78.02	0.71	0.50	1.96	77.93	0.78	0.61	1.55	79.51	0.80	0.64	1.45	80.12	0.81	0.65	1.48	79.52
K nearest neighbour	0.71	0.50	3.96	70.45	0.78	0.60	3.99	74.31	0.78	0.61	3.97	74.29	0.79	0.62	3.56	75.43	0.80	0.64	3.29	73.85
Parallel random forest	0.67	0.44	2.01	73.58	0.60	0.35	1.99	74.79	0.70	0.48	2.42	76.45	0.76	0.58	2.15	74.66	0.74	0.55	2.31	75.64
Boosted generalized linear model	0.76	0.57	4.96	69.22	0.76	0.57	4.88	72.02	0.77	0.59	4.73	70.15	0.76	0.59	4.09	77.19	0.78	0.61	4.97	72.01
Stochastic gradient boosting	0.66	0.44	1.91	76.04	0.65	0.43	1.79	76.07	0.63	0.40	1.94	76.32	0.69	0.47	2.30	77.41	0.69	0.47	1.83	75.58
Partial least squares	0.37	0.14	1.94	78.11	0.37	0.14	1.73	77.15	0.41	0.17	1.75	76.43	0.37	0.14	1.84	78.45	0.59	0.34	1.77	78.10
Gaussian process with polynomial kernel	0.73	0.54	0.99	81.66	0.73	0.53	1.45	81.71	0.74	0.55	1.22	81.36	0.77	0.59	1.01	81.72	0.80	0.64	1.25	81.89
Glmnet	0.72	0.52	2.33	72.90	0.73	0.53	2.78	74.36	0.77	0.59	2.33	77.38	0.79	0.62	2.23	74.01	0.73	0.54	2.26	73.11
Conditional inference random forest	0.33	0.11	2.69	73.37	0.14	0.02	3.02	73.08	0.40	0.16	3.17	75.99	0.33	0.11	2.88	73.56	0.52	0.27	2.97	74.34
Self organizing map	0.73	0.53	1.24	79.39	0.72	0.52	1.56	78.31	0.78	0.61	1.33	80.09	0.76	0.58	1.41	80.01	0.79	0.62	1.23	79.64

Table 5 Average values of evaluation parameters for 10 models (5-Fold Cross validation)

Model No	r	R	RMSE	Accuracy
Multilayer perceptron	0.76	0.58	1.62	79.02
K nearest neighbour	0.77	0.59	3.75	73.67
Parallel random forest	0.69	0.48	2.18	75.02
Boosted generalized linear model	0.76	0.59	4.73	72.12
Stochastic gradient boosting	0.66	0.44	1.95	76.28
Partial least squares	0.42	0.18	1.81	77.65
Gaussian process with polynomial kernel	0.75	0.57	1.18	81.67
Glmnet	0.75	0.56	2.39	74.35
Conditional inference random forest	0.34	0.13	2.95	74.07
Self organizing map	0.76	0.57	1.35	79.49

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}} \tag{7}$$

N-total number of samples

$\sum xy$ - summation of products of paired sample values

$\sum x$ - summation of x values

$\sum y$ - summation of y values

$\sum x^2$ - summation of squared x values

$\sum y^2$ - summation of squared y values

R-Squared (R2) R-squared determines how well the data fits to the regression line. It is square of coefficient of correlation and is given as in Eq. (8)

$$R^2 = \frac{\sum (y^{\wedge}i - \bar{Y})^2}{\sum (yi - \bar{Y})^2} \tag{8}$$

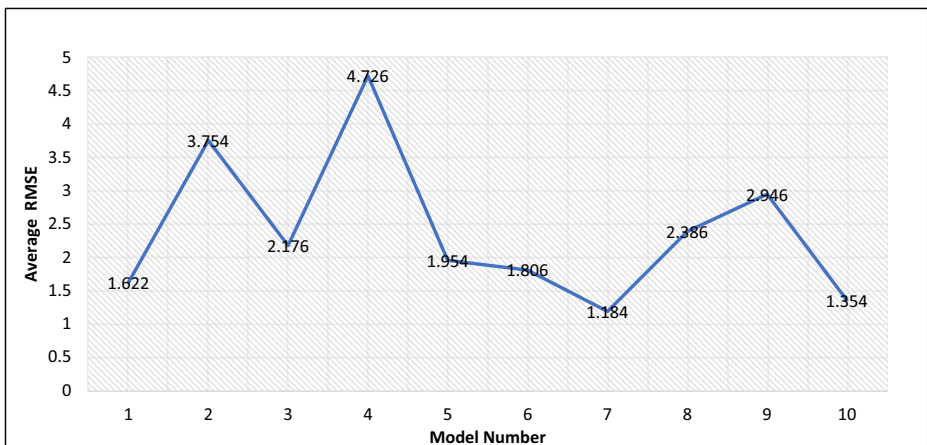


Fig. 4 Comparison of average RMSE values for different machine learning models

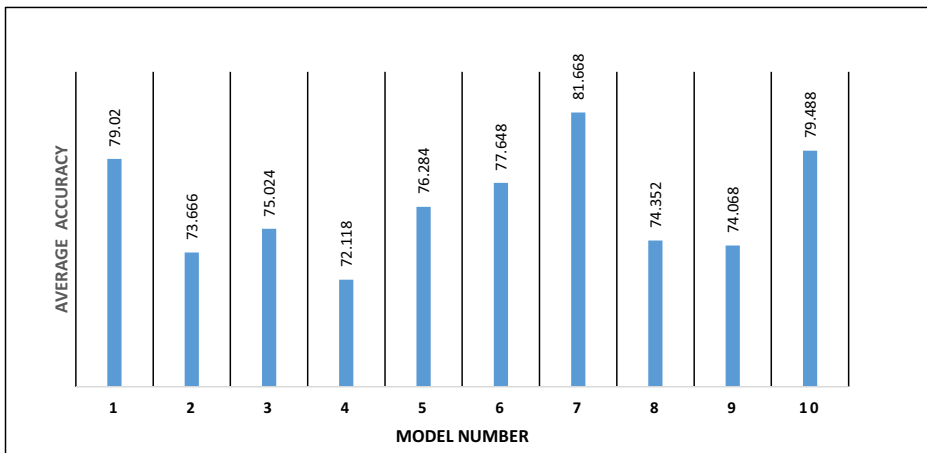


Fig. 5 Comparison of accuracy values for different machine learning models

Here, y_i denotes the observed values of the dependent variable, Y depicts the mean, and \hat{y}_i indicates the fitted value.

Accuracy It illustrates how well the model been able to predict the values of the defined target class. It is computed by comparing the actual values with the predicted values and is calculated as in Eq. (9).

$$Accuracy = (mean(abs(Actual == Predicted), 1, 0)) * 100 \tag{9}$$

The original cleansed dataset is is fed to cuckoo search meta-heuristic for feature selection. Table 4 provides the r, RMSE, R2 and accuracy results for dataset with selected features at different values of folds in 5 fold cross-validation, while Table 5 concludes the average result of 5-fold cross-validation for different evaluation parameters. The main evaluation parameter used in this work is predictive RMSE.

Figure 4 depicts the average RMSE values of 10 base machine learning models on a reduced dataset after feature selection. It is clear from the figure that the lowest RMSE of 1.18 is depicted by the Gaussian Process with Polynomial Kernel model and the highest RMSE of 4.726% is shown by Boosted Linear Model.. The K nearest Neighbour model has better performance in comparison to the Boosted Linear Model with less RMSE value of 3.75.

Figure 5 depicts the comparison between machine learning models over the accuracy evaluation parameter on the reduced dataset. The Gaussian Process with Polynomial Kernel is the best performer in terms of accuracy with an accuracy value of 81.67% while Boosted

Table 6 The 5 models combination (based on RMSE)

S.no	Model name
1	Gaussian Process with Polynomial Kernel
2	Self Organizing Map
3	Partial Least Squares
4	Multilayer Perceptron
5	Boosted Generalized Linear Model

Table 7 Different ensemble models with the 5 models

Sr. No	Possible variations	RMSE
1	Self Organizing Map + Multilayer Perceptron+ Partial Least Squares + Gaussian Process with Polynomial Kernel	1.072
2	Self Organizing Map + Multilayer Perceptron + Boosted Generalized Linear Model + Gaussian Process with Polynomial Kernel	0.373
3	Multilayer Perceptron + Partial Least Squares + Boosted Generalized Linear Model + Gaussian Process with Polynomial Kernel	0.958
4	Self Organizing Map + Partial Least Squares + Boosted Generalized Linear Model + Gaussian Process with Polynomial Kernel	0.886
5	Self Organizing Map + Multilayer Perceptron + Partial Least Squares + Boosted Generalized Linear Model	1.068

Generalized Linear Model shows the least accuracy value of 72.12% in comparison to rest of the base models.

These models are further exploited for stacking ensemble with a group of four as base learners and one as meta- learner, leading to different possible combinations of the heterogeneous learners.

All heterogeneous learners are exploited (in combination of 5) for stacking ensemble, that further aid in avoiding the ensemble to overfit the model.

Out of the ensemble of these diverse predictors combinations, 5 models (in combination) are chosen from the set of 10 base models based on the RMSE values as given in Table 6. Table 7 presents the average RMSE values of stacking ensemble of the 5 models combinations.

Figure 6 concludes that the models Self Organizing Map, Multilayer Perceptron, Boosted Generalized Linear Model and Gaussian Process with Polynomial Kernel along with Partial Least Squares as meta-learner portrays lowest RMSE of 0.373. The combination of Multilayer Perceptron, Partial Least Squares and Boosted Generalized Linear Model is not proven to better combination for predicting the contentment score of a student in university. The proposed stacking ensemble with cuckoo search meta-heuristic is proven to be beneficial for computing the overall contentment rate of the university. The proposed methodology can help to find and conclude the parameters that impact the student contentment levels the most.

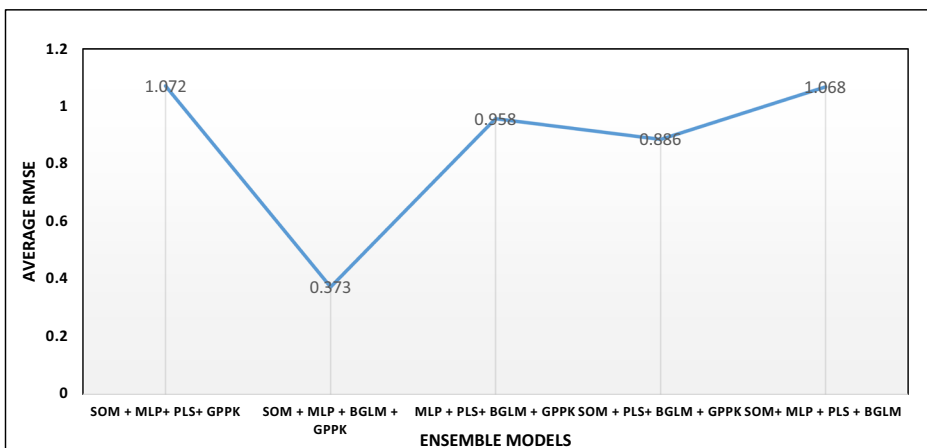


Fig. 6 Comparison of average RMSE values for different stacking ensembles

4 Conclusion and future scope

The current work focuses on the development of an ensemble machine learning model to foresee the overall student contentment score according to their perceived level of satisfaction. The proposed approach utilizes cuckoo search meta-heuristic for selecting the best combinations of features out of 78 original feature set. The work exploits the stacking ensemble to improve the efficiency of the proposed approach. The results indicate that an ensemble model of Self Organizing Map, Multilayer Perceptron, Boosted Generalized Linear Model and Gaussian Process with Polynomial Kernel along with Partial Least Squares as meta-learner gives lowest RMSE of 0.373. The current work can be extended by enhancing the dataset by capturing feedback data and reviews from social media leading.

Acknowledgments We really want to appreciate the efforts of the volunteers of Technical Society who took keen interest in encouraging UG and PG students to give their valuable time and contribution to make the data available to us.

References

1. Anastasiades PS, Elena V, Nikos G (2008) Collaborative learning activities at a distance via interactive videoconferencing in elementary schools: parents' attitudes. *Comput Educ* 50:1527–1539
2. Arlot S, Alain C (2010) A survey of cross-validation procedures for model selection. *Statistics surveys* 4: 40–79
3. Armenski G, Kostoska M, Ristov S and Gusev M 2014 April. Student satisfaction of e-learning tools for computer architecture and organization course. In *IEEE Global Engineering Education Conference (EDUCON)*, 630-637.
4. Athiyaman A (1997) Linking student satisfaction and service quality perceptions: the case of university education. *Eur J Mark* 31:528–540
5. Becker SA, Brown M, Dahlstrom E, Davis A, DePaul K, Diaz V, Pomerantz J (2018) NMC horizon report: 2018 higher, education edn. EDUCAUSE, Louisville, CO
6. Bolliger Doris U (2004) Key factors for determining student satisfaction in online courses. *International Journal on E-learning* 3:61–67
7. Carney R (1994) Building an image. In: *In Proceedings Symposium for the Marketing of Higher Education*. American Marketing Association, New Orleans, Louisiana
8. Chen Y, Xiong J, Xu W, Zuo J (2019) A novel online incremental and decremental learning algorithm based on variable support vector machine. *Clust Comput* 22(3):7435–7445
9. Choudhary M A. 2012 Factors influencing engineering students' performance and their relationship with the student satisfaction with the teaching, learning as well as overall university experiences. In *International Conference on Information Technology Based Higher Education and Training*, 1-5.
10. De Wit H (2020) Internationalization of higher education. *J Int Stud* 10(1):i–iv
11. Dejaeger K, Frank G, Antonio G, Lapo M, Bart B (2012) Gaining insight into student satisfaction using comprehensible data mining techniques. *Eur J Oper Res* 218:548–562
12. Đurđević I (2015) Predicting student satisfaction with courses based on log data from a virtual learning environment—a neural network and classification tree model. *Croatian Operational Research Review* 6:105–120
13. Faour H, Mohammad H and Ahmed Al G. 2012 Enhancing student learning experience and satisfaction using virtual learning environments. In *International Conference on Education and e-Learning Innovation* 1-2.
14. Gu K, Wang L, Yin B (2019) Social community detection and message propagation scheme based on personal willingness in social network. *Soft Comput* 23(15):6267–6285
15. Guo W (2010) Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction. *Expert Syst Appl* 37:3358–3365
16. Huang C, Liu B (2019) New studies on dynamic analysis of inertial neural networks involving non-reduced order method. *Neurocomputing* 325:283–287

17. Kamis T, Zainah O, Nurul S (2015) Satisfaction level among diploma in medical electronic students on the sustainability of physical facilities in Politeknik sultan Salahuddin Abdul Aziz Shah. In *Innovation & Commercialization of Medical Electronic Technology Conference (ICMET)*:114–120
18. Kumar KK, Sagi TM (2012) Assessment of overall satisfaction level of students in a technical institution. In *IEEE International Conference on Engineering Education, Innovative Practices and Future Trends (AICERA)*, pp 1–6
19. Kuo YC, Andrew EW, Brian B, Kerstin S (2013) A predictive study of student satisfaction in online education programs. *Int Rev Res Open Distribut Learn* 14:16–39
20. Lee HJ, Ilju R (2009) Influence of structure and interaction on student achievement and satisfaction in web-based distance learning. *J Educ Technol Soc* 4:372
21. Lee HJ, Ilju R (2009) Influence of structure and interaction on student achievement and satisfaction in web-based distance learning. *J Educ Technol Soc* 12:372
22. Li F, Zhou SR, Zhang JM, Zhang DY, Xiang LY (2013) Attribute-based knowledge transfer learning for human pose estimation. *Neurocomputing* 116:301–310
23. Littlejohn A, Isobel F, Lou M (2008) Characterizing effective eLearning resources. *Comput Educ* 50:757–771
24. Long M, Zeng Y (2019) Detecting iris liveness with batch normalized convolutional neural network. *Comput Mater Contin* 58:493–504
25. Manzoor H. 2013 Measuring student satisfaction in public and private universities in Pakistan. *Global Journal of Management and Business Research*.
26. Méndez-Giménez A, Cecchini-Estrada JA, Fernández-Río J, Mendez-Alonso D, Prieto-Saborit JA (2017) Achievement goals 3×2 , self-determined motivation and satisfaction with life in secondary education. *Journal of Psychodidactics* 22(2):150–156
27. Mo H, Wang F (2009) Linguistic dynamic systems based on computing with words and their stabilities. *Science in China Series F: Information Sciences* 52(5):780–796
28. Negricea Costel I, Tudor E, Emanuela Maria A (2014) Establishing influence of specific academic quality on student satisfaction. *Procedia Soc Behav Sci* 116:4430–4435
29. Nikolic S, Christian R, Peter James V, Montserrat R, Stirling D (2015) Decoding student satisfaction: how to manage and improve the laboratory experience. *IEEE Trans Educ* 58:151–158
30. Onditi EO, Wechuli TW (2017) Service quality and student satisfaction in higher education institutions: a review of literature. *Int J Sci Res Publ* 7(7):328–335
31. Rjaibi N, Latifa Ben A R, and Mohamed L. 2012 Modeling the prediction of student's satisfaction in face to face learning: an empirical investigation. In *International Conference on Education and e-Learning Innovations* 1-6.
32. Salvador-Ferrer C (2017) The relationship between gratitude and life satisfaction in a sample of Spanish university students. *Anales De Psicología/Annals of Psychology* 33(1):114–119
33. Sohoraye M, Poomalay P, Meera G. (2014) Are facebook likes enough to assess student satisfaction in open distance learning (ODL)? An incursion into students' experience of ODL through online social networks (OSNs). In *IEEE IST-Africa Conference Proceedings*, 1-8.
34. Song H, Lee K, Moon N (2019) User modeling using user preference and user life pattern based on personal bio data and SNS data. *J Inform Process Syst* 15(3)
35. Tessema T, Kathryn R, Yu W (2012) Factors affecting college students' satisfaction with major curriculum: evidence from nine years of data. *Int J Humanit Soc Sci* 2:34–44
36. Warnes G R 2010 Gtools: various R programming tools. R package version 2.6. 2.
37. Witten I H., Eibe F, Mark A. H, and Christopher J. P 2016. Data mining: practical machine learning tools and techniques. Morgan Kaufmann
38. Yin C, Ding S, Wang J (2019) Mobile marketing recommendation method based on user location feedback. *Human-centric Computing and Information Sciences* 9(1):14
39. Yukselturk, E., & Yildirim, Z .2008. Investigation of interaction, online support, course structure and flexibility as the contributing factors to students' satisfaction in an online certificate program. *J Educ Technol Soc*, 11.4., 51–65.
40. Zarirah Z and Ridzuan A. 2015 An investigation on student's participation and satisfaction towards online learning. In *IEEE Conference on e-Learning, e-Management and e-Services (IC3e)* 143-147.
41. Zhang X., Xiao WX (2012, May) Active semi-supervised framework with data editing. In *2012 international conference on systems and informatics (ICSAI2012)* (pp. 46-50). IEEE

Affiliations

Maninder Kaur¹ · Himika Mehta¹ · Sukhchandan Randhawa¹ · Pradip Kumar Sharma² · Jong Hyuk Park³

Maninder Kaur
manindersohal@thapar.edu

Himika Mehta
himika19@gmail.com

Sukhchandan Randhawa
sukhchandan95@gmail.com

Pradip Kumar Sharma
pradip.sharma@abdn.ac.uk

¹ Department of Computer Science and Engineering, Thapar University, Patiala 147004, India

² Department of Computing Science, University of Aberdeen, Aberdeen AB24 3FX, UK

³ Department of Computer Science and Engineering, Seoul National University of Science and Technology, (SeoulTech), Seoul 01811, South Korea