Journal of Ambient Intelligence and Humanized Computing (2021) 12:7095–7105
https://doi.org/10.1007/s12652-020-02375-3

ORIGINAL RESEARCH

Towards developing an ensemble based two-level student classification model (ESCM) using advanced learning patterns and analytics

R. Vidhya[1] · G. Vadivu[2]

Received: 23 May 2020 / Accepted: 21 July 2020 / Published online: 28 July 2020

Abstract
In recent decade, learning analytics has gained more attention and several advanced data mining models are developed for deriving the hidden sources from educational databases. The extracted data helps the Educational Institutions or Universities to enhance the teaching methodologies of faculties and student's learning process in efficient manner. For improving the student performance and better educational results, the student data evaluations based on their learning modes are significant. With that note, the proposed model develops a new model called ensemble based two-level student classification model (ESCM) for effectively analysing and classifying the student data. With the student data pursuing technical higher education, the ESCM is performed with three traditional classification model with ensemble classifier techniques for enhancing the classification accuracy. The model utilizes support vector machine, Naive Bayesian and J48 classifier that are combined with Ensemble classification methods as modified meta classifier such as bagging and Stacking. Here, the technical higher education student data collected from SRM student database based on the feature set contains the student learning factors that support performance enhancement. The results are evaluated with the SRM student datasets and compared based on the classification accuracy and model reliability. Furthermore, the obtained results outperform the existing models. Based on the accurate predictions, special attentions and measures are taken to improve the student results and institutional reputation.

Keywords Learning analytics · Ensemble based student classification model (ESCM) · Support vector machine (SVM) · Naive Bayesian (NB) · J48 classifier · Bagging · Stacking · Student behaviour

1 Introduction

In recent days of education management models, learning analytics (LA) provides major contributions, specifically with Internet Development. Moreover, LA combines the educational domains that are collected and evaluated from digitalized student records. While discussing about student data, it may include the student academic performances, extra-curricular enrolments, personal and also financial content that are obtained from single individual or the educational institutions (van Barneveld et al. 2012). LA has become the fast growing domain of educational research in present scenario, which has been applied in universities for enhancing the learning patterns of students and teaching methodologies in effective concerns (Ferguson 2012; Siemens 2013). The Fig. 1 presents the contribution of LA in handling things combined with educational research and others like E-learning evaluations, big data processing and stack management (Long and Siemens 2011; Yadav et al. 2011).

For effectively evaluating the student performance with academic results, LA is used with mining methodologies. With that concern, the proposed model develops a novel method called ensemble based two-level student classification model (ESCM) for evaluating the technical higher education student data, who are playing significant role in society. The Fig. 2 portrays the data mining process in educational systems including learning analytics.

✉ R. Vidhya
vidhya.cse63@gmail.com

G. Vadivu
vadivukar@gmail.com

[1] Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

[2] Department of Information Technology, SRM Institute of Science and Technology, Chennai, India

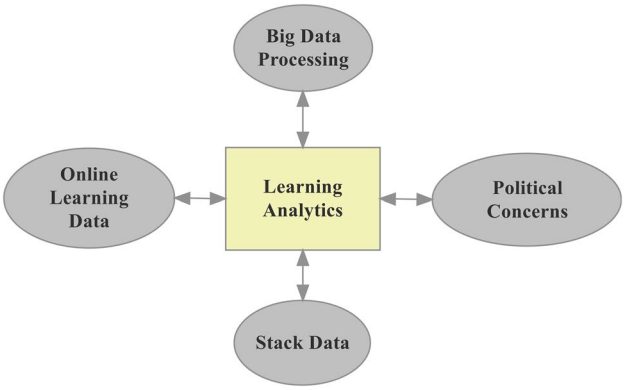


Fig. 1 Involvement of LA in data processing

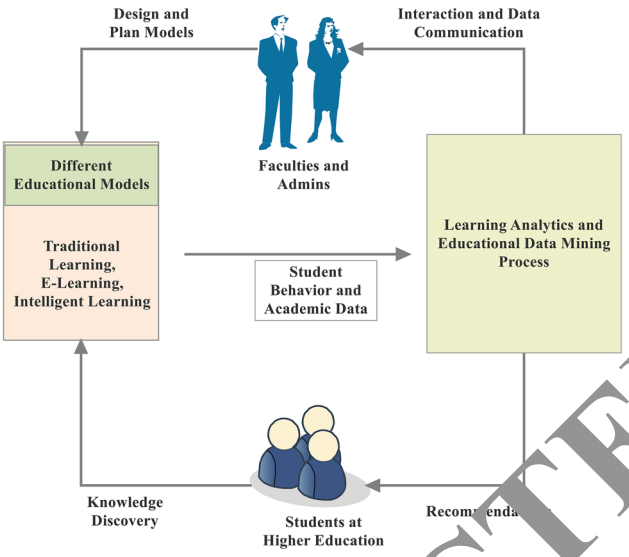


Fig. 2 Process of data mining in education systems along with learning analytics

In previous works, the student performances in higher education levels are predicted with different attributes such as academic results, background of family, personal data, income, etc. (Osmanbegovic and Suljic 2012; Alapati and Sindhu 2016; Yair al El-Halees 2012). In the proposed model, learning analytics and patterns of student data is used for evaluating their performances in better manner and classify them accordingly. Mainly, the novelty of the derived work is mainly presented at the Ensemble based classification and the attributes that are obtained from the effectively framed Questionnaires in a wider manner. This increases the accuracy rate of classification results considerably and aids in framing better solutions for enhancing the performance level of students (Shah et al. 2020; Lekshmy and Rahiman 2020). Hence, the contributions of the proposed ESCM is presented below.

1. Framing the LA based effective attribute set that covers all dimensions of student learning factors includes their personal data, learning patter, behaviour analysis, emotional factors, multiple intelligence and cognitive abilities.
2. Enhancing the classification accuracy by using Ensemble based classification, instead using single classification model.
3. Effective utilization of integration of classification techniques such as support vector machine (SVM), Naive Bayesian (NB) and J48 classifier with the Effective Attribute Set.
4. Bagging and stacking are the ensemble classifiers used here for classifying the higher education students under EXCELLENT, GOOD, AVERAGE, GOOD, BETTER and POOR.
5. For the sake of providing performance evaluations and comparative analysis, benchmark datasets are used, and operations are carried out based on the analysis factors such as Precision, Recall and Accuracy rates.

The rest of this paper is organized as follows: Sect. 2 contains some description about the related works in Educational Data Mining and Learning Analytics based student performance evaluations. Section 3 describes about the work process of the proposed work that includes LA based Effective Feature Set construction. Section 4 comprises the evaluation results and the work is concluded at Sect. 5 with some routes for future enhancements.

2 Related works

There are many research work have been done for evaluating the student performances in different dimensions. An informative literature survey work has been presented in Ramaswami and Bhaskaran (2010) comprised descriptions about traditional education systems and web based data management. Moreover, a different model has been proposed for identifying weak students using association rule based mining algorithm. Genetic algorithm based student classification was used for categorizing the student into three levels based on their grade. It was given that the utilization of combined classifiers can produce accurate classification results. A regression model has been employed for detecting student performances based on their test reports. Moreover, the instigators have used rule induction classifier and NB classifier for classifying student grades using clustering data.

Probabilistic graphic model has been used for determining the performances of students and tutors abilities thereby enhancing the future outcomes. The model also considered about the democratic data and personal features of students for providing appropriate results. In the same manner, data

mining models have been used for evaluating the student performances in Engineering Colleges. The process of artificial neural networks (ANN) has been used in Oyedotun et al. (2015) for student performance analysis and course repetition with a case study explanation. In a combined manner, the model used decision tree and clustering model for classifying the data samples. Moreover, the authors for global model for classification (GMC) (Anwar et al. 2014) used Supervised Learning approach for enhancing the result precision rate. Bound model for clustering and classification (BMCC) has been developed in Anoopkumar and Zubair Rahman (2018) with the integration of J48 decision tree classification technique and k-means clustering.

3 Procedure of ensemble based student classification model

In the developing experimentations with Learning Analytics, extracting student data for improving the learning abilities and personalities of students and also the teaching pattern of tutors is being the major concern. Moreover, in recent times, the learning pattern of students is depending on several factors and the analysis is a more complicated process. For that, this paper proposes an ensemble based student classification model (ESCM) that incorporates the efficiencies of both the base classifiers and Meta classifiers. The process depends on the functions such as, data acquisition from student databases, data pre-processing, first level classification, second level classification, result evaluations and knowledge representation, the pictorial representation about the proposed work is given in Fig. 3.

3.1 Data acquisition from students

In this work, the main focus is specified on this part called data acquisition from students. The performance of students in technical higher education depends on several factors such as personal, financial, environmental conditions of students and so on. Here, concerning those factors and a questionnaire set is framed in a pattern that covers all aspects of learning factors of students, who are pursuing higher education. Moreover, the questionnaire set is framed with the following six major factors,

1. Student's personal data
2. Learning pattern
3. Behaviour
4. Emotional factors
5. Multiple intelligence
6. Cognitive abilities

Based on the above mentioned factors, the data are obtained from the students and the sample questionnaire set is presented in the Table 1.

According to the answers obtained for the questionnaire set, the dataset is prepared and processed for training and testing. The incorporation of personalized features in the data acquisition process is one of the enhancement parts in Learning Analytics to improve the prediction results of student performance. Hence, the students are accurately classified under categories such as EXCELLENT, GOOD, AVERAGE, and POOR, thereby, helping the tutors to concentrate more for result enhancement and student betterment.

3.2 Data pre-processing

Data pre-processing includes two functions such as data cleaning and feature set construction. Data cleaning is the process to reduce irrelevant, duplicate and repeated contents from the obtained data from student database. For Example, the factors such as financial status of student's family or blood group are not required for evaluating the academic excellence of them. Though they are unavoidable in the dataset, they are not having greater impact on evaluating student performances. In similar manner, the dataset may have some missing values that are to be eliminated for reducing computational complexities. Following that, feature set construction for training is processed.

In order to perform dimensionality reduction, the irrelevant instances are removed from the obtained data and appropriate features are selected. Here, for constructing the feature set, Chi square attribute evaluation is used. For that, the Chi square rate (CRR) is estimated between each attribute from student sample and the target and the required features with better Chi square values are selected for feature set. The computation is given as follows,

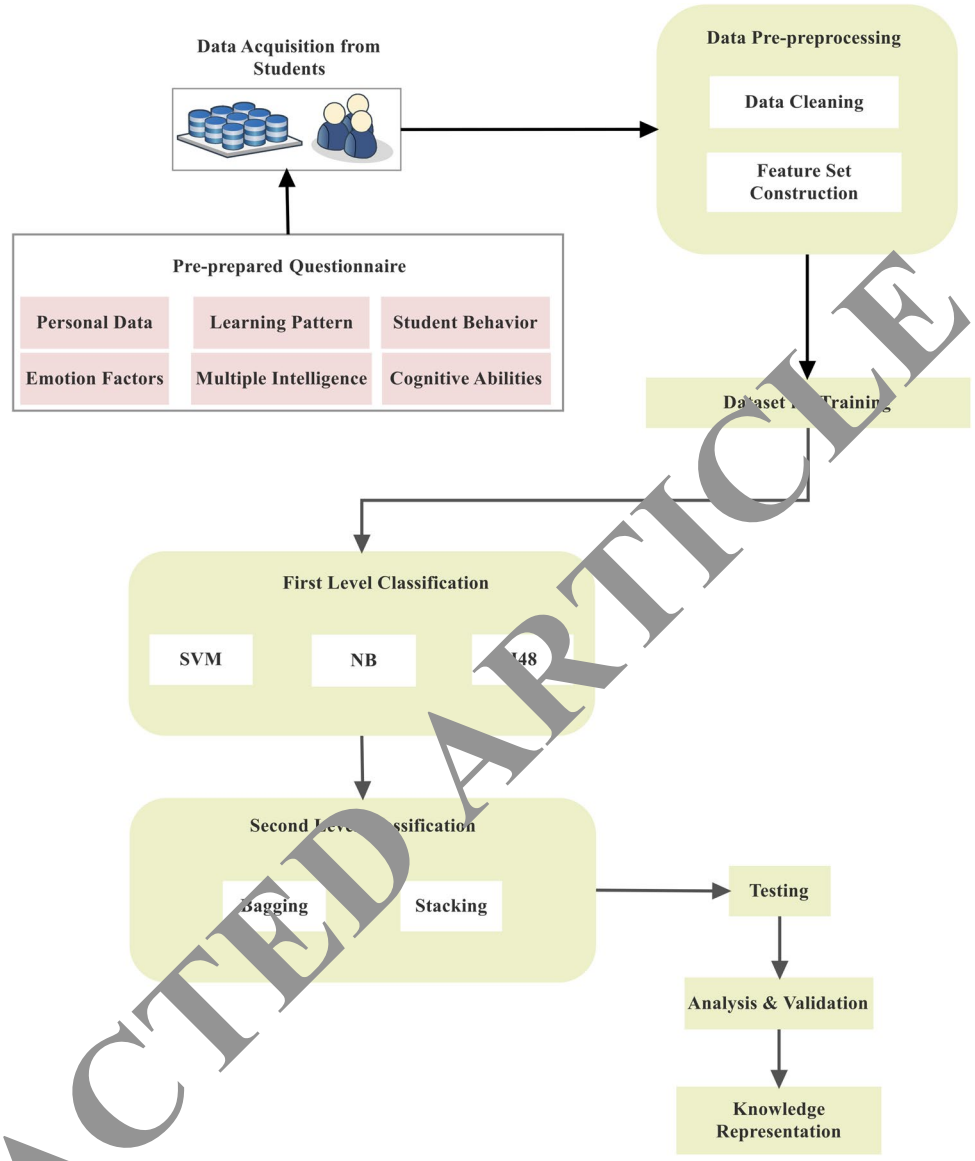
$$CRR(X^2) = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

where 'n' is the total number of instances, 'O_i' number of observations of samples and 'E_i' denotes the number of expected observations based on the target and feature relationships. Based on the results the feature set is constructed and given for the training process.

3.3 First level classification process with student dataset

In this first level classification with the obtained student dataset, the following base classifiers are used in the proposed model, support vector machine (SVM), Naive Bayesian (NB) and J48 classifier. And, the work process of each classifier is described below.

Fig. 3 Operations in ESCM



3.3.1 SVM based student dataset classification

In support vector machine, the nearest data vectors are determined using the hyperplane separation for appropriate decision making. The steps in SVM in the proposed ESCM are given as follows:

Step 1: When it is considered that there are two classes Student Class (SC_1 and SC_2), the indefinite feature vertex (V) may belong to SC_1 or SC_2.

Step 2: Perform linear discriminant function as follow,

$$g(V) = w^T(V) + b \tag{2}$$

where '$w^T(V)$' is the transpose value of weight vector in which 'V' is the input feature vector term. 'b' denotes the bias rate for the defined two-dimensionality space.

Step 3: In a two dimensional vector, when the input feature vector is the 2D vector, the linear discriminate equation results with a straight line, represents, $w^T(V) + b = 0$.

Step 4: In a case that input vector is a three dimensional one, the linear equation results in forming a plane.

Step 5: When the dimension of the input feature vector is greater than 3, then hyperplane is framed, in which the weigh vector is perpendicular to the hyperplane.

Step 6: The SVM classification rules for student classification is described as,

For each feature vector (V), linear function is computed,

(i) When the vector lies on the hyperplane-positive side, then,

Table 1 Sample questionnaire set from student data acquisition

Factors	Description	Possible values
1. Student personal data		
Gender	Student's sex	{Male, female}
Age group	Under 18, 19–20, 21–22, 23 and above	{A, B, C, D}
Type of education	Engineering/management/arts	{1, 2, 3}
Family income	Financial status students for scholarship purposes	{< 10,000, (10,000–30,000), (30,000–50,000), > 50,000}
Location	Urban/semi-urban, sub-urban, rural	{1, 2, 3, 4}
2. Learning pattern		
Visual Based	Student's interest on learning though visual patterns	Rating (1–10)
Kinaesthetic based	Activities can do along with studies	Rating (1–10)
Physical based	Activities in learning	Rating (1–10)
Aural based	Listening oriented observation	Rating (1–10)
3. Behaviour		
Attitude of students	Behaviour with strangers	Rating (1–10)
Dependability	Dependent on internet/faculties	Rating (1–10)
Integrity	Mindset when certain things happen	Rating (1–10)
4. Emotional factors		
Anger	Frequency	Rating (1–10)
Anxiety	Tension rate on certain things	Rating (1–10)
Stress	Stress level during exams	Rating (1–10)
5. Multiple intelligence		
Inter-personal	How interactive and friendly	Rating (1–10)
Intra-Personal	Willingness	Rating (1–10)
Logical	Mathematical skills	Rating (1–10)
6. Cognitive ability		
Thought process	Capability to do innovations	Rating (1–10)
Attention	Metal ability	Rating (1–10)
Memory	Ability to handle things with protocols	Rating (1–10)

$$g(V_1) = w^T(V_1) + b, \quad \text{where} \quad w^T(V_1) + b > 0$$

(ii) When the vector, 'V_1' lies on the hyperplane-negative side, then, $w^T(V_1) + b < 0$.

(iii) In the remaining case, when the feature vector lies on the hyperplane, then it can be stated that, $w^T(V_1) + b = 0$.

Step 7: The student classification in SVM is done by determining the hyperplane that divides all data points from one to another.

3.3.2 Contribution of NB in ESCM

Naive Bayesian classification is a kind of supervised learning model that performs classification function using statistical knowledge. For producing better classification results, Bayes theorem is used for computing the probability of classes based on the feature vector, and given as,

$$p(SC_i|V) = \frac{p(SC_i)p(V|SC_i)}{p(V)} \tag{3}$$

And, the steps involved are described below,

Step 1: Let 'S' be the training set of samples and their corresponding student classes and each class is given by an n-Dimensional attribute vector, for example, for feature vector V = {sem 1, sem2,…, semN} and there are 'm' number of subjects, which is given as, {sj1, sj2,…, sjM}.

Step 2: When the classification purpose is to acquire the highest posterior, that is,

MAX p(*sji*|*sem*), can be obtained from the Eq. (3).

3.3.3 J48 classifier description in the proposed model

The main work process of J48 classifier is to develop a classification model from the dataset having appropriate student class labels here. Decision tree pruning is the major advantage of using J48 classifier. Moreover, in WEKA tool, there

are some effective options for tree pruning the produces précised results. The contribution is established progressive generalization of tree till it reaches high accuracy in classification. Furthermore, the operations in J48 classifier includes,

(i) When there is a case that the sample are belonging to similar class the tree denotes a leaf, and the leaf is also returned with same class label.
(ii) The potential value for each feature is computed.
(iii) Gain is also computed for each attribute and the best attribute is further chosen for branching.

3.4 Second level with modified meta classifier (MMM) based student data classification in ESCM

For producing precise student classification, Ensemble classifier techniques are incorporated in the proposed model, which combines multiple classification models as modified meta classifier and producing united results. The operation of combining multiple classification models in ensemble classifier is based on the following objectives,

(i) Enhancing the complete classification accuracy compared to single classification model.
(ii) Obtaining better generalization based on the combined classifiers.

The major goal of the proposed work is that to select a set of hypotheses based on the available results and combines their identifications into one. Moreover, this second level ensemble classifier uses bagging and stacking techniques, which are explained in detail below.

3.4.1 Bagging in modified meta classifier

The function of bagging is performed with the bootstrap aggregation. The base classifier models in the ensemble model are taken for consideration and provided equal weights for all. Further, voting concept is utilized for selecting the final classification result in bagging model, which is explained pictorially in Fig. 4.

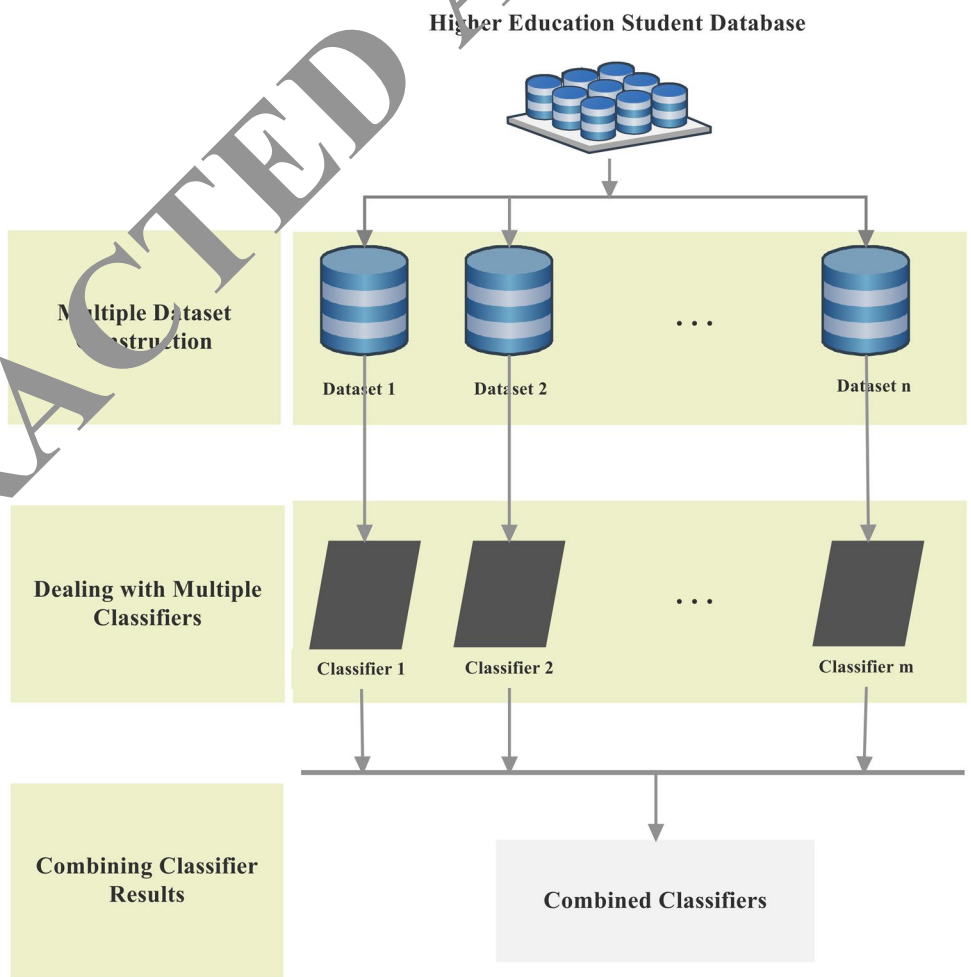


Fig. 4 Work process in bagging

1. When there are 'N' number of student samples and 'M' number of features.
2. The feature sets are used to develop the model with classification samples and sub sets.
3. The feature that produces the best split is selected during the training data.
4. This is repeated in each process that is trained in parallel.
5. Results are provided based on the combination of predictions all classifiers.

Moreover, for producing accurate results, weights are assigned for all the samples in the model. After the evaluations, the wrongly classified data is provided with larger weights; hence, it can be focussed more in further process. The steps are given as follows,

1. Weight are assigned to the training samples based on the incorrect classifications
2. Define the hypothesis
3. Rate the hypothesis with weights

The final results are derived based on the weight based voting. The calculation for determining final classification (fc) is given as,

$$fc = \left(\sum c_i wt_i / * \sum wt_i \right) / n \tag{4}$$

where $\{c_1, c_2, \ldots, c_m\}$ classifiers used, 'wt_i' denotes the weights for each and 'n' represents the number of classifiers in this model.

3.4.2 Process of stacking in modified meta classifier

Stacking is the process in which single dataset is given to several models to train. Here, the obtained training dataset divided into multiple subsets and the resultant model is derived. From the base classifiers that are used for first level classification, the stacking ensemble is fit to be combined using modified meta classifier. The steps are presented below,

(i) The training data set is divided into twofold
(ii) The base classifiers are used to fit them to the samples at the first fold
(iii) For each base classifier, the predictions are made based in derivations in the second-fold
(iv) Fit the MMM at the second fold with respect to the results obtained by the base classifiers as inputs

In the aforementioned steps, the dataset are divided into two-folds using the observations on student data that have been used for training the base classifiers. By performing that, the model produces accurate results with minimal time and error for the obtained real-time student dataset.

3.5 Factors for performance evaluation

In the proposed model, the results obtained from the ensemble classifier are evaluated based on the rates of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Moreover, the performance evaluations are performed by the following factors based on the aforementioned results.

1. Sensitivity rate is defined as the prospect of results to be positive, when there is appropriate classification occurs. The computation is given as,

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{5}$$

2. Rate of Specificity can also be termed as True Positive Rate and Recall, which is given as the classification results are obtained to be negative in a specific SC, computed as,

$$Specificity = \frac{True\ Negative}{False\ Positive + True\ Negative} \tag{6}$$

3. Precision is the important factor to be determined for performance evaluation of the proposed model, which can be defined as the acquisition of positive predictions. And, the formula for precision is given as,

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{7}$$

4. Accuracy Rate is defined as the rate of total number of exactly classified instances among total number of obtained samples and the formula is denoted as,

$$Accuracy\ Rate(AR) = \frac{TP + TN}{FN((TP + 1) + (1 + TN))} \times 100 \tag{8}$$

5. F-measure is computed as,

$$F{-}Measure = \frac{2 * precision * recall}{precision + recall} \tag{9}$$

4 Results and discussions

For the performance evaluations of the proposed ensemble based two level student classification model, the student data are collected from SRM student database based on the Questionnaire set presented in Table 1. The dataset contains

233 samples with 45 features, that is, the total number of instances is about 10,485. In order to make the dataset feasible for using in WEKA tool, it is saved as comma separated value (CSV) format and converted to ARFF. Furthermore, the model evaluations are carried out based on the factors described in Sect. 3.5. And, obtained results are compared with the base classifiers such as SVM, NB and ANN.

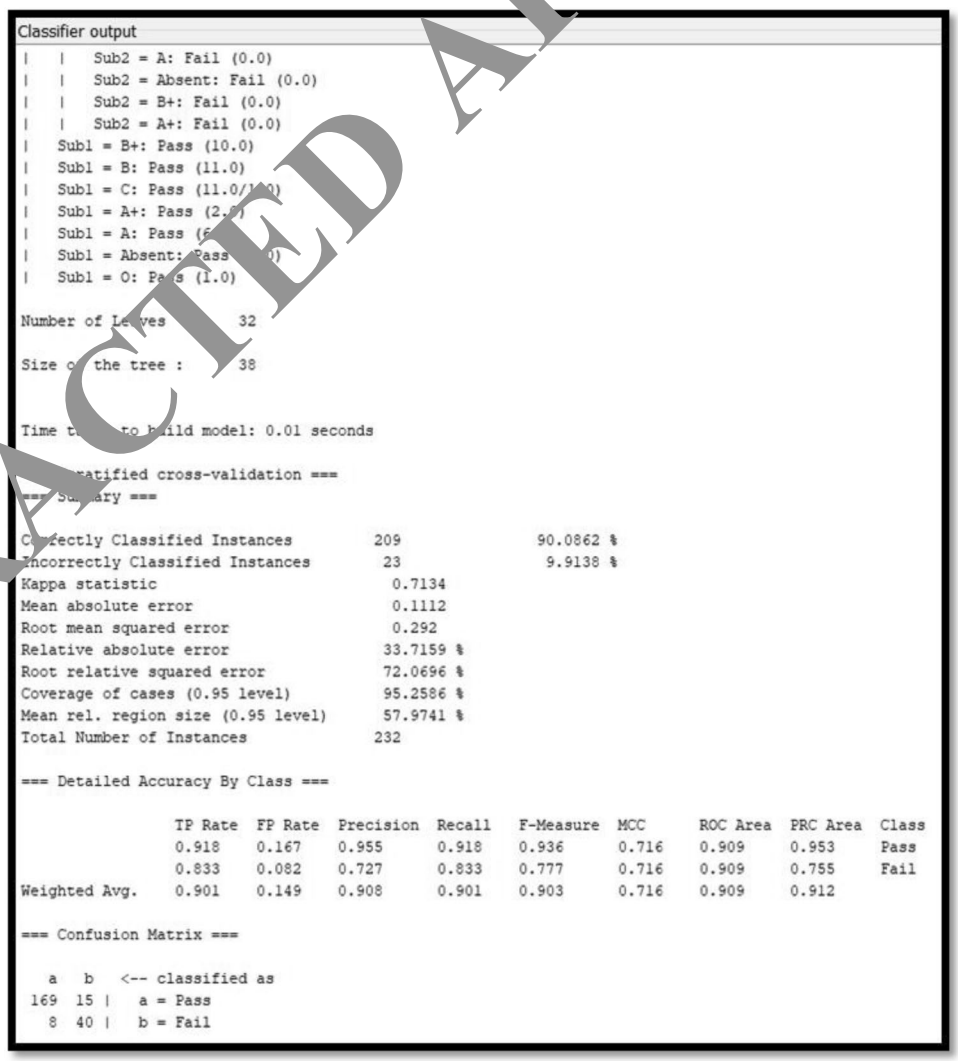
However, the proposed ESCM is developed in such a manner to improve the result accuracy by integrating heterogeneous classifiers. In the proposed model, two levels of classifications are used, where, the first level classification process contains the base classifiers, and the results are becoming more accurate with the second level classification that develops modified meta classifier with bagging and stacking. The obtained features from the student dataset are divided into two and processed parallel with modified Bagging and Stacking techniques.

The experimental results depict that the proposed model give appropriate classification results of student data under classes such as EXCELLENT, GOOD, AVERAGE, and POOR. The Fig. 5 shows the results obtained, when the dataset crosses over the base classifiers for classifying the students based on their academic performance.

The Table 2 contains the results of the proposed ESCM with the second level classification called Ensemble classifiers such as Stacking and Bagging. For RF implementation, the results show 94.3% of accuracy, and for bagging technique, it shows 97.4% of accuracy. In average, the proposed model produces 95.85% of accuracy in classifying the data based on academics. The Figs. 6 and 7 shows the results obtained at WEKA tool for ensemble classifier bagging and stacking.

From the below screen shot of the execution of the proposed model for ensemble classifier it is observed that there are 35 leaves and the tree size is 44 (Fig. 6) and tree size is 3 with 2 leaves in (Fig. 7). By the efficient combination of the heterogeneous classifiers in the proposed model, the model produces more appropriate results and the novelty

Fig. 5 Academic performance based student classification with base classifiers



```
Classifier output
|   |   Sub2 = A: Fail (0.0)
|   |   Sub2 = Absent: Fail (0.0)
|   |   Sub2 = B+: Fail (0.0)
|   |   Sub2 = A+: Fail (0.0)
|   Sub1 = B+: Pass (10.0)
|   Sub1 = B: Pass (11.0)
|   Sub1 = C: Pass (11.0/1.0)
|   Sub1 = A+: Pass (2.0)
|   Sub1 = A: Pass (6.0)
|   Sub1 = Absent: Pass (0.0)
|   Sub1 = O: Pass (1.0)

Number of Leaves          32

Size of the tree :        38


Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         209               90.0862 %
Incorrectly Classified Instances        23                9.9138 %
Kappa statistic                          0.7134
Mean absolute error                      0.1112
Root mean squared error                  0.292
Relative absolute error                 33.7159 %
Root relative squared error             72.0696 %
Coverage of cases (0.95 level)          95.2586 %
Mean rel. region size (0.95 level)      57.9741 %
Total Number of Instances              232

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.918     0.167     0.955      0.918     0.936      0.716    0.909     0.953     Pass
               0.833     0.082     0.727      0.833     0.777      0.716    0.909     0.755     Fail
Weighted Avg.  0.901     0.149     0.908      0.901     0.903      0.716    0.909     0.912

=== Confusion Matrix ===

   a   b   <-- classified as
 169  15 |   a = Pass
   8  40 |   b = Fail
```

Table 2 Results obtained for ensemble classifiers bagging and stacking

Total number of students = 233					
Ensemble classifier-RF					
Correctly classified samples = 214			92.24%		
Incorrectly CLASSIFIED samples = 19			7.85%		
TP	FP	Precision	Recall	F-measure	SC
0.932	0.225	0.946	0.925	0.941	Pass
0.885	0.035	0.788	0.855	0.833	Fail
Ensemble classifier-bagging					
Correctly classified samples = 227			97.4%		
Incorrectly classified instances = 6			2.57%		
TP	FP	Precision	Recall	measure	SC
0.978	0.112	0.949	0.979	0.	Pass
0.867	0.023	0.945	0.875	0.91	Fail

Fig. 6 Academic performance based student classification with ensemble classifier-bagging

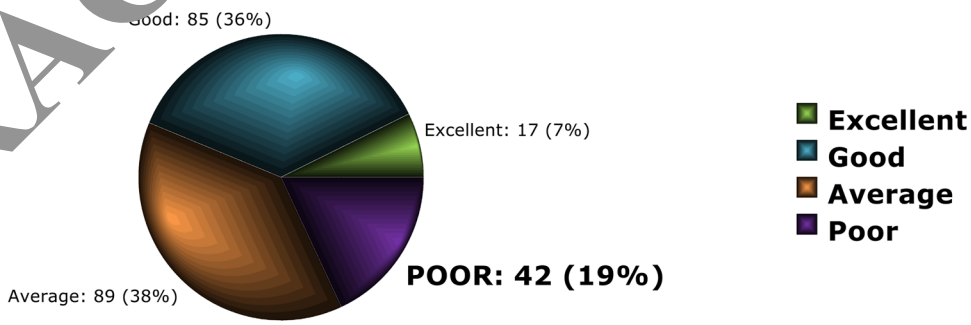
Classifier output

```
|   Sub9 = B: Pass (1.99/0.12)
|   Sub9 = A+: Fail (0.0)
|   Sub9 = A: Fail (0.0)
|   Sub9 = C: Pass (5.17)
|   Sub9 = P: Fail (15.13)
Physics Mark = FE: Fail (1.65)

Number of Leaves  :     35

Size of the tree :      44


Weight: 4.34

Number of performed Iterations: 10


Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         217               93.5345 %
Incorrectly Classified Instances        15                6.4655 %
Kappa statistic                          0.8074
Mean absolute error                      0.0652
Root mean squared error                  0.2477
Relative absolute error                 19.7589 %
Root relative squared error             61.142  %
Coverage of cases (0.95 level)          94.8276 %
Mean rel. region size (0.95 level)      51.0776 %
Total Number of Instances              232

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.951     0.125     0.967      0.951     0.959      0.808    0.981     0.995     Pass
               0.875     0.049     0.824      0.875     0.848      0.808    0.981     0.952     Fail
Weighted Avg.  0.935     0.109     0.937      0.935     0.936      0.808    0.981     0.986

=== Confusion Matrix ===

   a   b   <-- classified as
 175   9 |   a = Pass
   6  42 |   b = Fail
```


of the proposed model is to be highlighted that the model focuses on all aspects of student education factors for effective classification. Based on the defined method, the students of SRM Student Database is classified under major 4 classes Excellent, Good, Average and Poor.

Fig. 7 Academic performance based student classification with ensemble classifier-stacking

Fig. 8 Classification results achieved with sample dataset from SRM student database



The pie chart depicted in Fig. 8 contains the overall classification results obtained for the dataset with 233 student samples. As given earlier, the complete classification process is made with the student data prepared according to the questionnaire set comprises, student's personal data, learning pattern, behaviour, emotional factors, multiple intelligence and cognitive abilities. From the presented outcomes, the tutors and the management perform appropriate decision making for improving the students under POOR class, thereby, enhancing the results of the institution and also reputation. Moreover, the graph presented in Fig. 9 portrays the comparison of the proposed work in student classification with other existing models. It is explicit from the comparison graph, that the proposed model produces better rate of accuracy than other compared models, which evidences the efficiency of the proposed model.

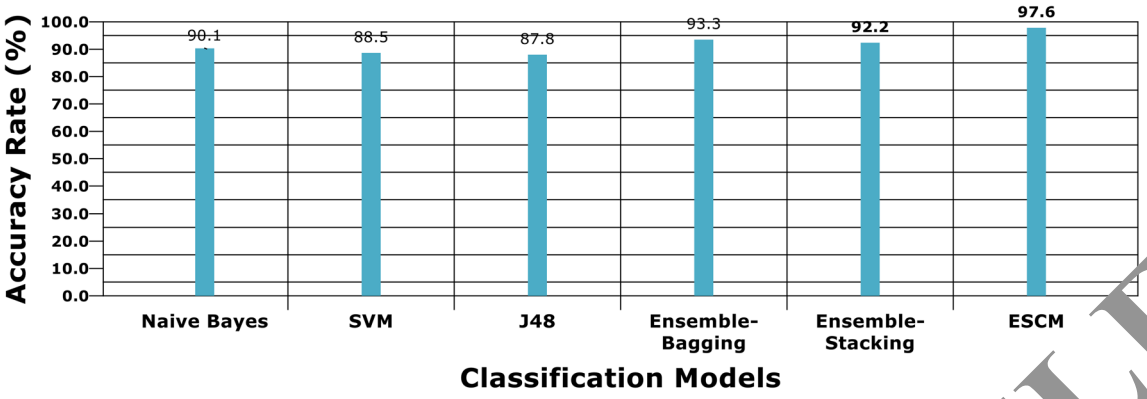


Fig. 9 Accuracy rate comparison between classification models

5 Conclusion and future work

Learning Analytics is providing more contribution in developing the student quality and the overall results of the Educational Institutions combined with educational data mining. this paper presents a novel ensemble based two level student classification model (ESCM) for classifying students under technical higher education in four major classes such as Excellent, Good, Average and Poor. For that, the model integrates base and ensemble classifiers and performs two level classifications. Moreover, higher rate of accuracy is obtained with the modified meta classifier. The UCI student database is used for the experimentation and the dataset is prepared based on the effective questionnaire set that comprises all the factors that impacts student education. Student's personal data, learning pattern, behaviour, emotional factors, multiple intelligence and cognitive abilities are the major factors considered for developing the student dataset. By the effective integration of heterogeneous classifiers such as NB, SVM and J48, along with ensemble classifier bagging and RF, the proposed model produces accurate classification results. Performance evaluations are carried out with respect to the rate of accuracy and precision. The proposed model produces 97% of accuracy (in average), which is greater than other compared models.

In future, the model can be focussed in developing a new approach for handling dynamic student profiling from deep Web in Learning Analytics and effective model for evaluating critical learners.

References

Alapati YK, Sindhu K (2016) Combining clustering with classification: a technique to improve classification accuracy. Int J Comput Sci Eng 5(6):336–338

Anoopkumar M, Zubair Rahman AM, J (2018) Bound model of clustering and classification (BMCC) for proficient performance prediction of educational outcomes of students. Int J Adv Comput Sci Appl 9(10):243–246

Anwar H, Qamar O, M (2014) Global optimization ensemble model for classification methods, Hindawi Publishing Corporation Sci World J 2014:1–9

Ferguson R (2012) Learning analytics: drivers, developments and challenges. Int J Technol Enhanc Learn 4(5/6):304–317

Lekshmy PL, Rahiman MA (2020) A sanitization approach for privacy preserving data mining on social distributed environment. J Ambient Intell Hum Comput 11:2761–2777. https://doi.org/10.1007/s12652-019-01335-w

Long P, Siemens G (2011) Penetrating the fog: analytics in learning and education. Educ Rev 46(5):31–40

Osmanbegović E, Suljic M (2012) Data mining approach for predicting student performance. Econ Rev J Econ Bus 10(1):3–12

Oyedotun OK, Tackie SN, Olaniyi EO (2015) Data Mining of students' performance: Turkish students as a case study. Int J Intell Syst Appl 7(9):20–27

Ramaswami M, Bhaskaran R (2010) A CHAID based performance prediction model in educational data mining. Int J Comput Sci 7(1):10–18

Shah AM, Yan X, Shah SAA et al (2020) Mining patient opinion to evaluate the service quality in healthcare: a deep-learning approach. J Ambient Intell Hum Comput 11:2925–2942. https://doi.org/10.1007/s12652-019-01434-8

Siemens G (2013) Learning analytics: the emergence of a discipline. Am Behav Sci 50(10):1380–1400

Tair MMA, El-Halees AM (2012) Mining educational data to improve students' performance: a case study. Int J Inf Commun Technol Res 2(2):140–146

van Barneveld A, Arnold KE, Campbell JP (2012) Analytics in higher education: establishing a common language. Educ Learn Initiat 1:1–11

Yadav SK, Bharadwaj B, Pal S (2011) Data mining applications: a comparative study for predicting student's performance. Int J Innov Technol Creat Eng 1(12):13–19