Prediction of Student performance using Intuitionistic Fuzzy Mean Shift Clustering boosted with Chaotic Cheetah Chase Algorithm

Mrs.P.Menaka[1], Dr.K.Nandhini[2]

Research Scholar, PG & Research Department of Computer Science, Chikkanna Government Arts College/ Assistant Professor, Dr.N.G.P. Arts and Science College.
Research Supervisor, PG & Research Department of Computer Science, Chikkanna Government Arts College, Tirupur

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online:28April 2021

Abstract
It is an urgent desire for all the education institutions to predict the student's academic performance at the earliest. Though, there are many factors are involved in discovering their performance, the most important factor focused in this paper is learning style of the individual student. In this work a novel method is constructed to cluster students learning style and with other facilities available to the students. The inconsistencies and impreciseness in identifying students by standard clustering models as Assimilator /Divergent / Converger / Accomadator is overcome by applying the intuitionistic fuzzy Mean shift clustering (IFMSC). This algorithm will intelligently tackle the outliers and the students whom lies in the border of the clusters by defining each student in terms of intuitionistic fuzzification with the assistance of mean shift method. The standard clustering models selects the centroids in an arbitrary manner which may have the possibility of failing to select the optimal instances as centroids during the initial stage itself. This problem is overcome in this proposed work by inheriting the Chaotic Cheetah Chase Algorithm(C3A) , which in turn is optimized by the chaotic theory while choosing the parameters and centroids. Thus, by applying the proposed IFMSC-C3A, the prediction of learning style of the students which influence their academic performance is accomplished more precisely while comparing the other clustering models.

Keywords: Learning Style, Students performance, Cheetah Chase, Chaotic, Intuitionistic fuzzy Mean shift, clustering, prediction

Introduction

Providing education with high quality is very essential for the developing countries. Determining the academic performance of the students is the toughest challenge as it pivots on varied factors such as socio-economic, personal, psychological and some other environmental elements [1]. Evaluating the student's performance is a vital part for the education institutions. The high quality of the educational institutions is greatly depending on the excellent record of academic achievement. Student's academic performance can be identified by evaluation their learning assessment. There are many methods available to analyze the student's performance, one of the most powerful and useful technique which can able to handle the large volume of data is data mining [3]. While data mining is applied in the field of education then it is termed as education mining.

In the present years, most of the researchers focuses on investigating the students learning style and its allegations. They also reported that the students who have diverse personality types tend to have various learning styles, which in turn influence their performance over entire course. In Education mining, it extracts the useful patterns and infer hidden knowledge about the students from the high-volume education database. By predicting the student's performance, it will be helpful for the management as well as teachers to provide effective teaching approach. The education mining provides the advantage such as students can progress their learning activities, teachers can monitor the achievements of students in their academics and it allows the administration to improvise the education system [4].

This research work aims to predict the student's academic performance by clustering them based on the learning style as the major factor and exploring the utilization of small size student's dataset in education domains.

Related Work

This section discusses about the some of the exiting works which involves in prediction of student's academic performance is explained.

Elaf et al [5] developed ensemble approaches to predict the student's academic performance by considering behavioral features with other features and they also performed analysis after behavioral features are removed. The result shows that it greatly increases the accuracy.

et al [6] used different classification models such as KNN, ANN, Naïve Bayes and J48 classifier for predicting the performance of the students. They used ten-fold cross validation and this work categorize the student's dropouts.

Asif et al [7] focuses on investigating the importance of clustering and visualization in determining the indicators from small dataset. They performed the prediction process on undergraduate students based on their learning analytics.

Sheeba et al [8] they concentrated on developing strong decision-making rules which will increase the strength of the standard decision tree to precisely classify the learning style of the students in order to improve the accuracy of classification. The used web log files of learning management system by preprocessing the dataset and constructing decision tree using Felder Silverman learning style model.

Hashmia et al [9] designed two different prediction models namely decision tree and fuzzy genetic algorithm which uses internal marks. In this work only, the marks are considered for predicting the students, their personal details and their learning style is not considered for improved accuracy

Abu Zohair et al [10] in his work reported that modeling a small dataset and training it is more feasible to produce accurate result. The possibility of the using small dataset in developing prediction, clustering and visualization is the key factor of this work and they used support vector machine and discriminant analysis approach is used for training small size dataset with acceptable accuracy.

Ramesh et al [11] investigation about the factors which influence the student's academic performance in their final exams by determining suitable data mining model to predict it. This helps to give timely warning to the students those who are at risk. The results revealed that the type of school is not influencing the performance of the students and occupation of parents in predicting their grades.

Background study

Mean Shift Clustering Algorithm

The standard mean shift clustering like input dataset it maintains a set of data points which are copied from the input. During each iteration this set is replaced by mean of those points with the reachable distance. Mean shift algorithm first defines a spherical window with the radius and calculates mean of points which lie within the window [12]. Then it shifts the window to the mean and it continues till it converge. During each iteration window will be shifted to densely populated area of dataset till data is equally distributed.

Each iteration comprised of following steps:

1. Assign estimate dt
2. Compute kernel function

$$KF(dt_i - dt),\ KF(dt_i - dt) = e^{e\| dt_i - dt\|}$$

3. Calculate weighted mean of density in window

$$\text{mean}(dt) = \frac{\sum_{dt_i \in N(dt)} KF(dt_i - dt)dt}{\sum_{dt_i \in N(dt)} KF(dt_i - dt)dt}$$

   where N(dt) is the neighborhood of x, set of point $KF(d) \neq 0$

4. Assing $dt_i \leftarrow \text{mean}(dt)$
5. Repeat until mean(dt) converges

The mean shift algorithm procedure is illustrated in the figure. The shaded dots are data points and black dots are successive window centers. Here, mean shift operation begins at Y1 data point, b defining spherical window with the radius r this algorithm computes the data points mean which presented within the window and shifts the window to the mean and the procedure continues till the algorithm meets the convergence. During each iteration window is shifted to more densely population portion.

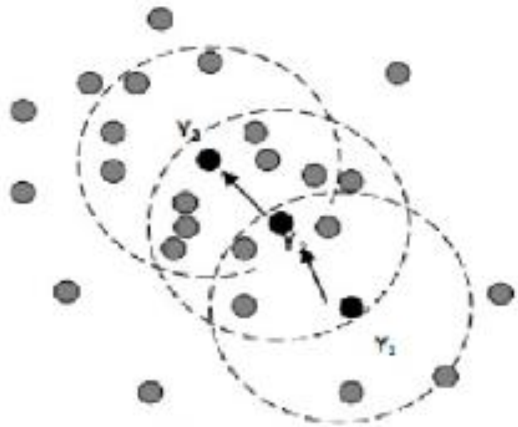

Fig.1. Process of Mean Shift Algorithm

Cheetah chase algorithm (CCA)

Cheetah is the giant, speediest land vertebrate and energetic predator which has monsters speed of more than 60 mph for a minimum make span time [15]. It has outstanding visual perception so that it chases its prey between 10 to 30 meters away. Cheetah's body is light, thin it influences hasty acceleration with high speed blast and executes excellent alters while moving at high speed. These are the unique features of cheetah to capture the prey which is moving very fast.

The cheetah hunting starts by initializing parameters in the start node by considering the parameters namely speed, velocity, acceleration, time and distance. It starts its chase moving to next node. It updates parameters and move towards prey. If it captures the prey then measure its distance, speed, total time and details of the route. Else if it does not capture the prey it moves to the next node. Once all cheetah captures prey then update node details and their final values. Now select the best shortest path. End the process. The pseudo code for cheetah chase algorithm is described as follows

Proposed Intuitionistic Fuzzy Mean Shift clustering enriched with Chaotic Cheetah Chase Algorithm for Student Performance Prediction

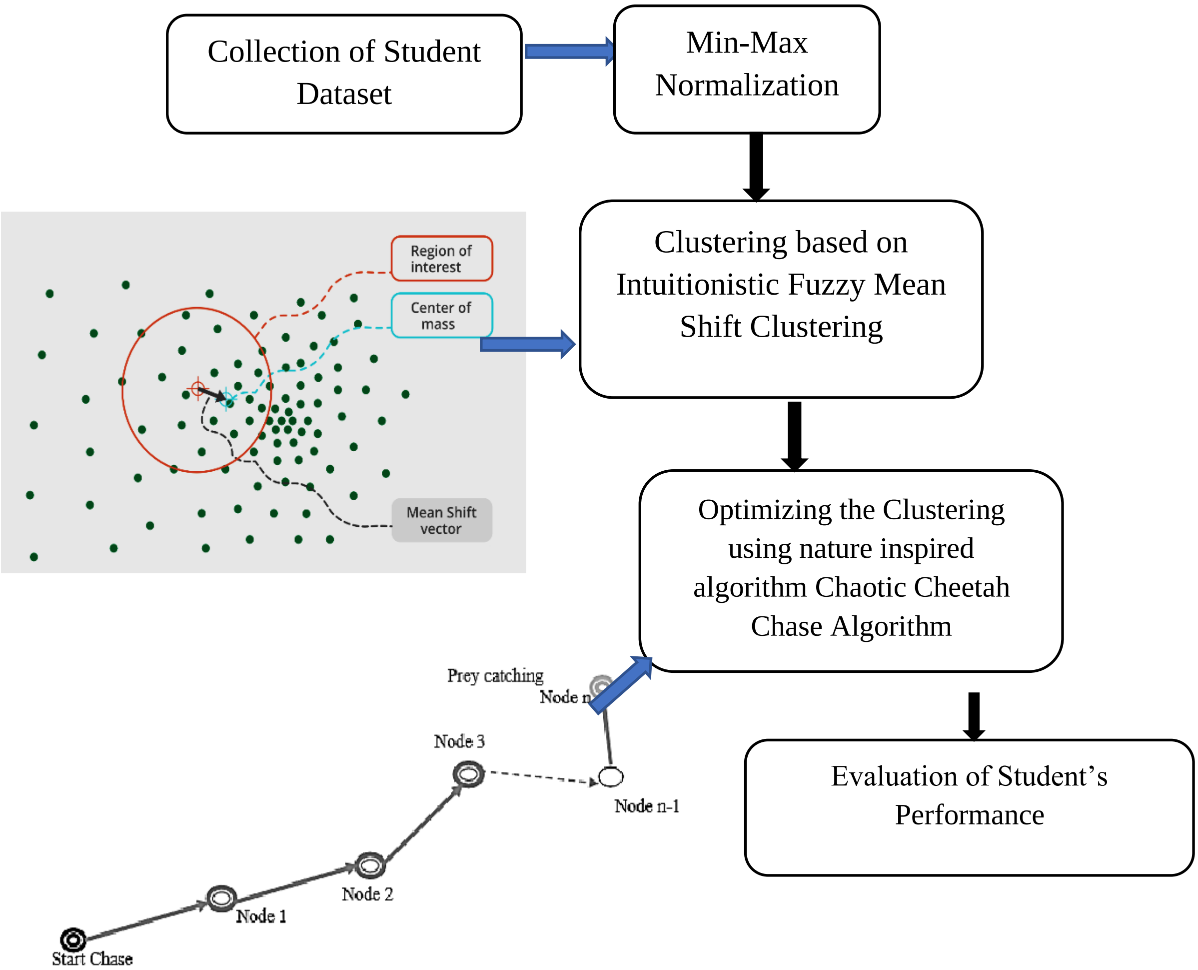

Figure Proposed Intuitionistic Fuzzy Mean Shift clustering enriched with Chaotic Cheetah Chase Algorithm for Student Performance Prediction

In this proposed method, students learning style is examined by applying the unsupervised clustering process known as Intuitionistic Fuzzy Mean Shift Clustering is constructed. This proposed model aims to handle the indeterminacy in predicting the student's academic performance based on their learning attitude. Instead of performing centroids selection in a random manner, this work uses nature inspired metaheuristic search algorithm known as chaotic cheetah chase algorithm performs the discovery of optimized centroids to cluster the similar patterns of students learning style and predict them accordingly.

Dataset used for predicting the student's academic performance using proposed clustering model based on learning style of college students. The dataset contains 61 student's information along with 17 attributes such as their personal details, education details, previous school background details, facilities available to access the resources for studying, percentage of attendance and 1st semester mark details. As in our previous work [17], only five potential attributes are selected by the Tuned Relief Feature (TuRF) subset selection is used for clustering process.

The detailed explanation of the proposed algorithm is explained in the following subsections:

Intuitionistic Fuzzy Means Shift Clustering

Intuitionistic fuzzy [13] represents each element in terms of membership μ(x) and non-membership ν(x) grades to handle them in case of hesitancy or indeterminacy in clustering students. The μ(x) and ν(x) values lies between 0 and 1. The summation of both the grades membership and non-membership values will be always less than one.

$$\mu_D(x) + \nu_D(x) \leq 1$$

With these to grades the presence of indeterminacy is handled by calculating the degree of hesitation represented as $\pi_D(x) = 1 - \mu_D(x) - \nu_D(x)$ and the $\pi_D(x)$ value lie between [0,1].

In Intuitionistic fuzzy Mean shift clustering, prior to clustering process all the instances of student dataset will be converted to intuitionistic fuzzy representation as follows

$$<\mu_{ins=1..n}(att_{j=1..m}),\nu_{ins=1,..,n}(att_{j=1..m})>$$

Where n represents number of records of the students and m represented number of attributes in each record.

Assume that the students records r_i (i= 1,2,3,..., n) all these have different importance where wt = $wt_1, wt_2, wt3, \dots wt_n$) and to determine the similarity among the instances and adding them to the nearest cluster is done by finding the weighted euclidean distance as calculated as shown

$$\mathrm{WED(F,G)} = \left(\frac{1}{2}\sum_{i=1}^{n} wt_i((\mu_F(r_i) - \mu_G(r_i) + (\nu_F(r_i) - \nu_G(r_i))^2\right)^{\frac{1}{2}}$$

IFMSC objective function is defined as

$$\begin{cases} \min OJ_m(U,V) = \sum_{j=1}^{q}\sum_{i=1}^{t} \mu_{ij}^m dt_1^2(Y_j, V_i) \\ \sum_{i=1}^{t} \mu_{ij} = 1, \quad 1 \le j \le q \\ \mu_{ij} \ge 0, \ 1 \le i \le t; \ 1 \le j \le q \\ \sum_{j=1}^{q} \mu_{ij} > 0, \quad 1 \le i \le t \end{cases}$$

Where Y = $\{Y_1, Y_2, Y_3, \dots Y_q\}$ are q instances represented in IFS and t is the number of clusters $(1 \le i \le t)$ and V = $\{V_1, V_2, \dots, V_t\}$ is the centroids of the clusters. m is the fuzzy factors whose value is greater than 1 and μ_{ij} is the j^{th} instance and i^{th} cluster's degree of membership and U =$(\mu_{ij})_{txq}$ refers to the Intuitionistic fuzzy matrix t x q [14]

The weighted measure wt of the vectors is calculated as

$$\boldsymbol{V_i = f(Y, wt^{(i)}) = \left\{< x_{l,} \sum_{j=1}^{q} wt_j^{(i)} \mu Y_j(x_l), \sum_{j=1}^{q} wt_j^{(i)} \vartheta Y_j(x_l) > | 1 \le l \le n\right\}, 1 \le i \le t}$$

Initialization of Cluster Centroids using chaos strategy

Initial selection of directions is an important factor to enhance the process of cheetah hunting the prey, in standard clustering models the initial centroids are selected in a random manner and in further iterations only the centroids are correctly selected. This process may skip the most promising instances to be selected as centroids which may produce better clustering accuracy. The randomness may not always guarantee the handling of impreciseness is grouping the instances which may leads to be an outlier. Chaos method is also known as butterfly effect, as if there is any slight change in a stage will affect the later stages very highly.

In this proposed work to overcome the random selection of centroids by the cheetah leads to earlier convergence to local optima while adapting the techniques of chaotic and mean shift among the most densely populated portions, the global optimization is achieved by applying the following condition of logistic chaotic mapping [16]

$$g_{i+1} = \begin{cases} 4\mu g_i(0.5 - g_i) & , \quad 0 \le g_i < 0.5 \\ 1 - 4\mu g_i(0.5 - g_i)(1 - g_i), & 0.5 \le g_i, \le 1 \end{cases}$$

where $3.57 \le \mu \le 4$, $\mu = 4$, $g_0 = \text{rand} \in (0,1)$.

Procedure for Intuitionistic Fuzzy Mean Shift Clustering enriched with Chaotic Chase Clustering

Input: Students Dataset (Y)

Output: Clustering Students (Assimilator or Divergent or Converger or Accomadator)

Steps

1 Set centroid vectors V(0), a=0, and $\in > 0$

2 Select the centroids by applying Chaotic Cheetah chase algorithm

2 Calculate U(a) = $(u_{ij}(a))_{txq,}$

(a) If $dt_1\left(Y_{j,} V_r(a)\right) > 0$, then

$$u_{ij}(a) = \frac{1}{\sum_{r=1}^{c}\left(\frac{dt_1\left(Y_j, V_i(a)\right)}{dt_1\left(Y_j, V_r(a)\right)}\right)^{\frac{2}{m-1}}} \qquad 1 \le i \le t; 1 \le j \le q$$

(b) Compute kernel function

$$\mathrm{KF}dt_1\left(Y_j, V_i(a)\right) = e^{t\|dt_i - dt\|}$$

6. Calculate weighted mean of density in window

$$\mathrm{mean(dt)} = \frac{\sum_{\mathrm{dt}_i \in \mathrm{N(dt)}} \mathrm{KF}(\mathrm{dt}_i - \mathrm{dt})\mathrm{dt}}{\sum_{\mathrm{dt}_i \in \mathrm{N(dt)}} \mathrm{KF}(\mathrm{dt}_i - \mathrm{dt})\mathrm{dt}}$$

where N(dt) is the neighborhood of x, set of point KF(d) ≠ 0

(c) If $dt_1\left(Y_j, V_r(a)\right) = 0$, then

$u_{rj}(a) = 1$ and $u_{rj}(a) = 0 \ \forall$ i $\neq r$.

3 Compute V(a+1) = $\{V_1(a+1), V_2(a+1), \ldots V_c(a+1)$

$$V_i(a+1) = f\left(Y, wt^{(i)}(a+1)\right), 1 \le i \le t$$

$$wt^{(i)}(\mathrm{a+1}) = \left\{\frac{u_{i1}(a)}{\sum_{j=1}^{q} u_{ij}(a)}, \frac{u_{i2}(a)}{\sum_{j=1}^{q} u_{ij}(a)}, \ldots, \frac{u_{ip}(a)}{\sum_{j=1}^{p} u_{ij}(a)}\right\}, 1 \le i \le t$$

4 If $\sum_{i=1}^{t} dt_1\left(V_i(a), V_i(a+1)\right)/t < \epsilon$, then go to step 5
else a = a+1, and return to Step 2.
5 End

Algorithm 2: Procedure for Chaotic Cheetah Chase
Begin
Steps

1. Initialize speed spd, acceleration acc, velocity vel, distance dst and time tm
2. Initialize total time tt, total distance travelled tdt
3. While i < t_max
4. In the start node compute the parameters spd, acc, vel
5. Move to next node, update its parameters by applying equation
6. If cheetah captured its prey then
    - Compute tt = tt1 + tt2
      Where tt1 is the top speed accelerated and tt2 is travel with top speed for specific distance
    - Compute distance tdt = dt1+ dt2
    - Prey's maximum distance travelled calculated as Max_dt = tdt – dtp

   Else Repeat the step 4
7. End while
8. Choose the best possible shortest path

End

The Students datasets is given as input to the Intuitionistic fuzzy mean shift clustering enriched by applying chaotic cheetah chase algorithm. In this proposed IFMSC-C3A, the centroids are selected by adopting chasing behaviour of cheetah whose goal is to select most feasible prey. Here, the candidates of student instances which are to be considered as centroids are elected by cheetah chase algorithm. Once initial set of instances are selected for clustering, mean shift window is used for finding centroids from densely populated potions. To overcome local optima during centroid selection and clustering in this work chaotic strategy is applied when the parameters of cheetah chase algorithm reaches its optima. Inserted of selecting centroids randomly, the intelligence of cheetah with chaotic theory achieve global optima in student performance prediction by clustering them based on their learning style and other factors.

Results and Discussions

This section discusses about the performance of the IFMSC-C3A of student performance prediction using matlab software. The student dataset used in this work is collected from 61 students with 17 attributes, among them the most potential features Gender, Computer at Home, Preferred Learning Style, Attendance Percentage and Learning Style are considered in this work for clustering the students with similar pattern. Four different clusters are generated such as Assimilator or Divergent or Converger or Accomadator. The performance of the proposed IFMSC-C3A is compared with three other clustering models namely DBSCAN, Fuzzy Means Clustering and Intuitionistic fuzzy C-Means Clustering

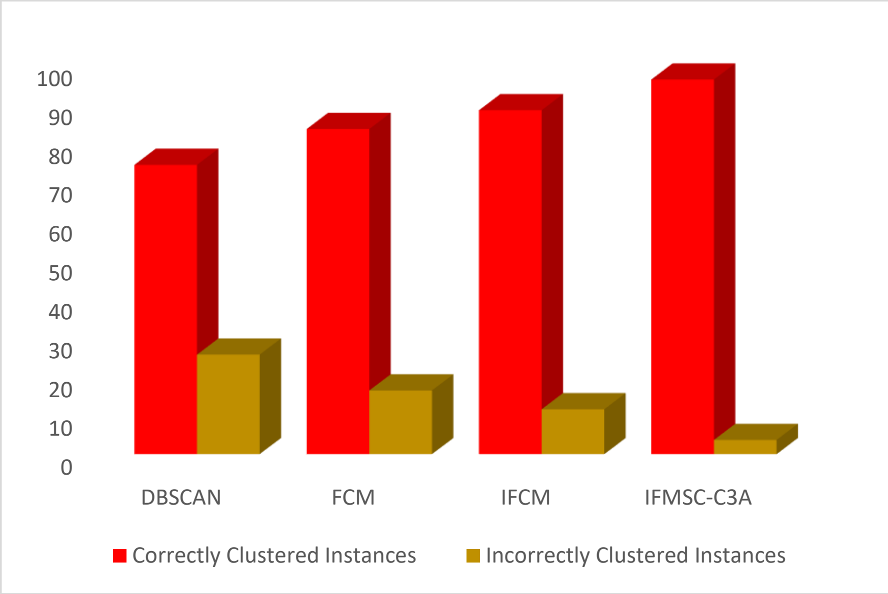

Figure Comparative analyses based on correctly and incorrectly clustered students' performance

The figure explores correctly and incorrectly clustered students' academic performance based on the similar pattern of learning style by applying four different clustering models. The performance of proposed IFMSC-C3A produce higher accuracy of clustering as it does not require any prior knowledge about number of clusters. Each instance is defined in terms of membership and non-membership grade. Thus, IFMSC-C3A has highest rate of 96.35% as correctly clustered instances while comparing with other three clustering models such as DBSCAN obtains 74.4%, FCM produce 83.62% and IFCM obtains 88.46%. while using mean shift in intuitionistic fuzzy it does not have any constrain of its shape.

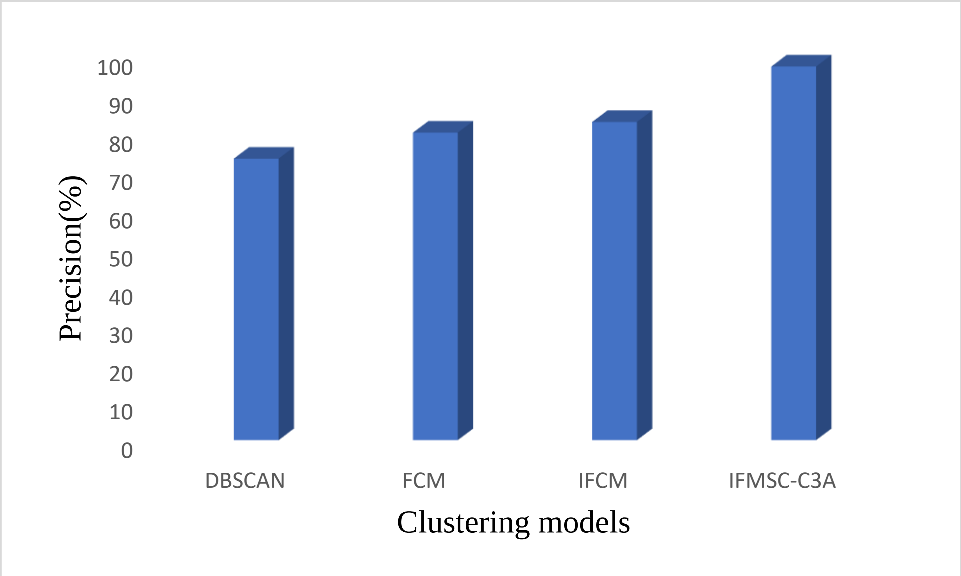

Figure Comparative analyses based on Precision

The precision obtained by four clustering models is depicted in the figure to predict students' performance. The proposed model intuitionistic fuzzy based mean shift clustering with its ability of shifting instances to the best centroids and clusters, it overcome the problem of outlier in student dataset. The proposed IFMSC-C3A produce highest recall rate of 97.7%, DBSCAN obtains 73.6%, FCM produce 80.4% and IFCM obtains 83.2%. Each centroid is selected by applying chaotic cheetah chase algorithm which selects instances with highest fitness value to produce optimized clustering. Thus, IFMSC-C3A achieves highest precision rate of student's performance prediction by examining their learning style. The conventional clustering models finds difficult to handle vagueness in handling outliers and thus they produce less precision rate.

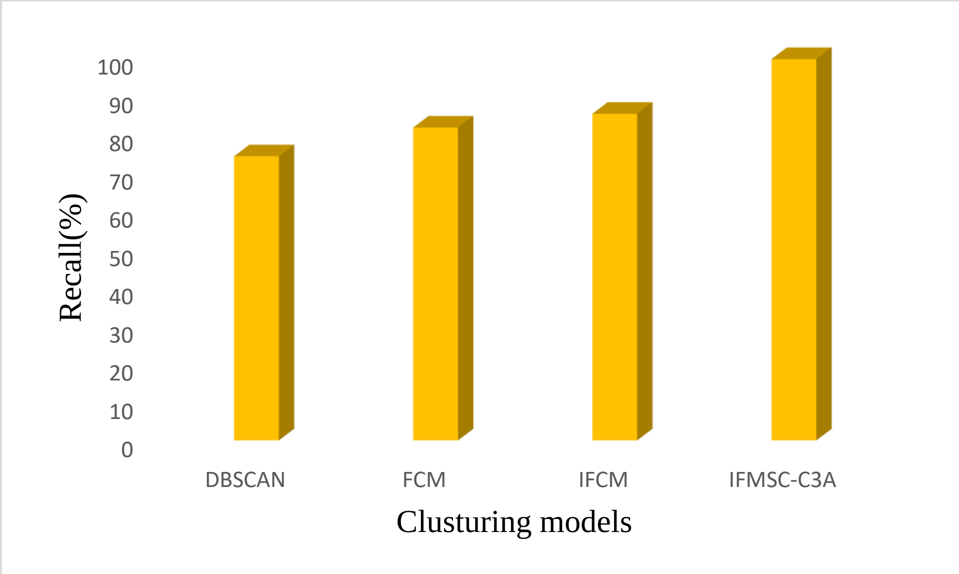

Figure Comparative analyses based on Recall

The figure shows the output of clustering models in terms of recall for predicting students' academic performance by analyzing their learning style, mode of learning, attendance percentage and their previous education. The proposed IFMSC-C3A produce highest recall rate of 99.6 %, DBSCAN obtains 74.2 %, FCM produce 81.7 % and IFCM obtains 85.3 %. The IFMSC-C3A performance better because the mean shift based intuitionistic based clustering has the ability to handle arbitrary instances while performing clustering of students with similar style of learning. The centroids of IFMSC-C3A are optimized by applying chaotic cheetah chase approach.

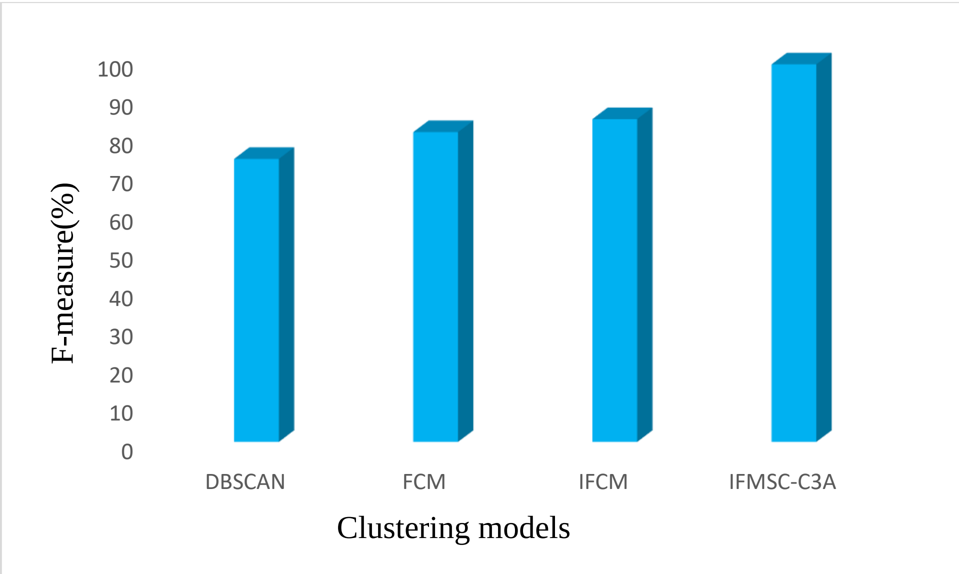

Figure Comparative analyses based on F-measure

The figure depicts performance of four clustering models based on F-measure obtained during prediction of student academic performance. By adapting chaotic cheetah chase algorithm, the process of clustering greatly increases its performance as it reaches the global optimization in clustering the similar instances more precisely. The means shift algorithm also additionally prove the efficacy of the intuitionistic fuzzy clustering by shifting the instances towards their more prominent cluster and centroids. Thus, the proposed IFMSC-C3A produce highest recall rate of 98.6% whereas DBSCAN obtains 73.9%, FCM produce 80.9% and IFCM obtains 84.3%.

Conclusion

This paper focuses on constructing an optimized unsupervised model using Intuitionistic fuzzy theory with Mean shift clustering which is boosted by chaotic cheetah chase algorithms (IFMSC-C3A). In this work instead of only considering the internal marks and semester marks scored by the students other factors like gender, learning style preferred, learning style, attendance percentage and computer at home are taken into the account. The impreciseness in determining the individuals who are considered as outliers are handled well in this work by representing them in terms of membership and non-membership grades. The mean shift clustering moves to the densely populated regions to select the most optimal centroids. The chaotic cheetah chase algorithm involved in selection of initial centroids and initial instances for producing optimized clustering results with its prey hunting strategy. The simulation results proved that the proposed IFMSC-C3A produce highest accuracy of classification while comparing with DBSCAN, Fuzzy C Means and Intuitionistic fuzzy C-means, because the proposed work has the ability to handle the outliers in an optimized way.

References

1. Juan Feldman, Ariel Monteserin, Analía Amandi, "Automatic detection of learning styles: state of the art", Artificial Intelligence Review, Vol.44, Issue.2, pp. 157--186, August 2015.
2. Bin Mat, N. Buniyamin, P. M. Arsad, R. Kassim, An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention, in: Engineering Education (ICEED), 2013 IEEE 5th Conference on, IEEE, 2013, pp. 126–130.
3. T. M. Christian, M. Ayub, Exploration of classification using NBtree for predicting student's performance, in: Data and Software Engineering (ICODSE), 2014 International Conference on, IEEE, 2014, pp. 1–6.
4. Chang YC, Wen-Yan Kao, Chih-Ping Chu, Chiung-Hui Chiu, "A learning style classification mechanism for e-learning", Computer Education, Vol.53, Issue.2, pp. 273--285, 2009
5. Elaf Abu Amrieh , Thair Hamtini, Ibrahim Aljarah, Mining Educational Data to Predict Student's academic Performance using Ensemble Methods, International Journal of Database Theory and Application, Volume 9, issue 8, pp.119-136, 2016.
6. Erman Yukselturk, Serhat Ozekes , Yalın Kılıç Turel, Predicting dropout student: an application of data mining methods in an online education program. European Journal of Open, Distance and eLearning, vol.17(1), pp 118- 133, 2014.
7. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. Computers in Education, 113, 177–194.
8. T. Sheeba and Reshmy Krishnan. 2018. Prediction of student learning style using modified decision tree algorithm in e-learning system. In Proceedings of the 2018 International Conference on Data Science and Information Technology (DSIT '18).
9. Hashmia Hamsa, Simi Indiradevi, Jubilant J Kizhakkehottam, Student academic performance prediction model using decision tree and fuzzy genetic algorithm Global colloquium in recent advancement and effectual researches in engineering, science and technology, Procedia Technology, 326-332, 2016
10. Lubna Mahmoud Abu Zohair, Prediction of Student's performance by modelling small dataset size, International Journal of Educational Technology in Higher Education (2019) 16:27, pp 1-18
11. V.Ramesh, P.Parkavi, K.Ramar, Predicting Student Performance: A Statistical and Data Mining Approach, International Journal of Computer Applications (0975 – 8887) Volume 63– No.8, February 2013
12. Aliyari Ghassabeh, Youness (2015-03-01). "A sufficient condition for the convergence of the mean shift algorithm with Gaussian kernel". Journal of Multivariate Analysis. 135: 1–10.
13. Atanassov, K.T.: Intuitionistic Fuzzy Sets: Theory and Applications. Studies in Fuzziness
14. and Soft Computing 35 (1999)
15. Zeshui Xu, Junjie Wu, Intuitionistic fuzzy C-means clustering algorithms, Journal of Systems Engineering and Electronics Vol. 21, No. 4, August 2010, pp.580–590
16. M. Goudhaman, Cheetah chase algorithm (CCA): a nature-inspired metaheuristic algorithm, International Journal of Engineering & Technology, 7(3) (2018) 1804-1811
17. Fan J. L, Zhang, X. F, Piecewise logistic chaotic map and its performance analysis. Acta Electronica Sinica, 37(4), 720-725, 2009
18. P. Menaka, K. Nandhini, Enriched model of Students Academic Performance Prediction based on their Learning Life Style using improved Fuzzy Min-Max Neural Classifier with Firefly Algorithm. International Journal of Advanced Science and Technology, 29(05), 14301 – 14316, 2020.