International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August 2019

An Effectual Ga Based Association Rule Generation and Fuzzy Svm Classification Algorithm for Predicting Students Performance

E.Chandra Blessie, K R Vineetha,

Abstract— This investigation provides outcome of utilizing educational data mining [EDM] to design academic performance of students from real time and online dataset collected from colleges. Data mining is determined to examine non-academic and academic data; this model utilizes a classification approach termed as Fuzzy SVM classification with Genetic algorithm to attain effectual understanding of association rule in enrolment and to evaluate data quality for classification, which is identified as prediction task of performance and academic status based on low academic performance. This model attempts to predict student's performance in grading system. Academic and student records attained from process were considered to train models estimated using cross-validation and formerly records from complete academic performance. Simulation was performed in MATLAB environment and show that academic status prediction is enhanced while hybrid dataset are added. The accuracy was compared with the existing models and shows better trade off than those methods.

Index Terms— Educational Data Mining, Fuzzy SVM, Genetic, academic performance, academic records

I. INTRODUCTION

Education is extremely a significant issue that is related to the development of country, specifically in growing countries like India, where education is a factor powerfully related to social mobility; henceforth, to recognize students at risk while analyzing academic performance effectually, and to recognize the factors has higher influence over this [1]. For this, data mining approach is more appropriate tool to deal with these tasks.

Analytics application in education has raised context in few decades. In [2], et al. has three drivers for evaluating it: initially, volume of data that has been collected from any educational institutions has significantly augmented, whether from learning management systems or course management or student information systems. Next, e-learning utilization: although it helps in aggregating data and brought certain learning issued like possible motivation lack and complexity for educators to attain feedback considering level of interest, mood or students understanding; at last, countries are attaining superior understanding of significance of education in development and have an focus to enhance better learning opportunities that causes to superior outcomes.

Revised Manuscript Received on August 30, 2019.
* Correspondence Author
E.Chandra Blessie*, Professor, Department Of Computer Application, Nehru College of Management, Coimbatore,
K R Vineetha, PhD Research Scholar, Nehru College of Management, Coimbatore,

Based on these circumstances, DM approaches have been used in administrative crisis and Learning crisis, however with superior concentration on automated discovery. In latter process, consider educator- or learner oriented.

In initial process, concentration is predominantly on students to study successfully by recommending novel contents. At last, the aim is to offer educators with tools to authorize, therefore direct learner effectually. Subsequently, in application towards administrative issues in data warehouses or business intelligence tools helps to assist DM procedure.

DM application in Education field is specified as Education Data Mining (EDM) and it is described as International Education DM Society as, "emerging field, related to approaches for exploring various data that emerges from educational settings, and utilizing these techniques to superior recognize students and learn in". Based on these approaches of EDM, the author in [3] examines classification as trails: Clustering, recognition, data distillation for judgment, prediction and model discovery. In [4] et al, recommends diverse taxonomy sourced on educational tasks: Offering feedback, examining and visualization, predicting performance, suggestion, student modelling, detecting behaviour, grouping students, developing concept map, social network analysis, constructing courseware, planning and scheduling [5]. Moreover, in EDM, data mining application and its task are similar: Clustering, Classification and Association rules analysis are included in KDD.

Here, two data mining classification are anticipated to analyse students' academic performance at time with socio-economic data, however based on academic records. Numerous circumstances are analyzed based on data utilized. It is an extended version of investigation titled DM to analyze performance of students.

The work is structured as trails: Section II offers certain EDM background and works; section III offers datasets, pre-processing, classification models for calculating academic success is explained. Section IV problem statement and section V demonstrates experimental and evaluation outcomes of proposed model; section VI shows conclusion a future direct of research work.

II. RELATED WORKS

An essential feature of classification approach is that is can be constructed based on the part of data, and as well termed as training set, which is utilized to learn model. Based on attributes subset, these are considered as class. After construction of model, it is utilized to allocate label to records, class attributes are unidentified So as to construct this model, two extensively utilized approaches are used: Bayesian classifier and decision trees. Black box models like SVMs and ANNs are also found in literature however it was not utilized in this work.

In general, DT is representation comprises of arcs and nodes in which internal node offers decision sourced on attribute and arcs specifies node. It terminates on leaf, which specifies label to be allocated [6]. To categorize record with DT, it commences by root and moves level at time based on conditions test outcomes on every node; when it terminates on leaf, record is categorized to label of leaf node. Here, C4.5 algorithm is sourced on Hunts procedure [7]. It is an essential feature that has ability to deal continuous and discrete attributes.

Subsequently, Bayesian classifier utilizes a probabilistic association amongst attributes and classes, indeed of deterministic where attributes set provided are not possess an identical label outcome [8]. Classification is to categorize based on attributes values and expressed as record probability from class 'Y', provided set of attributes

Subsequent methodology to investigate success to analyze performance in course; it utilizes same approach for diverse outcome and analyzing failure at completing course, it evaluates failures at course [9]. Both utilize information regarding students' present and past to identify academic success in class, or complete program.

Applications of DM approaches are extensively utilized to analyze problem from certain decade. In [10], author utilized diverse approaches for identifying dropouts from class sourced on demographic and students data with NB as best option. In [11], the author illustrated the occurrence of academic failure of students. It offer variables with extremely correlated to success on model utilized by [12], which demonstrates academic results are influenced by three factors: students' perceptions, history and involvement in studies [13]. As well, this work comprises of an application of DM to categorize first year students into three categories: high, medium and low risk students. In [14], the author utilized school records to identify GPA in engineering studies at Wellington University [15]. Here, diverse model is learned for undergraduate engineering programs and it is significantly superior.

III. DATA SET AND PREPROCESSING

The students academic data set measured here can be merged in three categorized as described below:

Students information: high school (for instance: private, public), access (for instance: admission program, regular), option for selecting program (1, 1 to 3) and program exists.

Socio-economic & Demographic: Age, admission, city, 'estrato', ethnicity ((Socio-economic classification).

Academic efficiency: Test score in modules (i.e. math, text, sciences, Image & social studies) classification for technical studies.

All the academic information system dataset comprises field to recognize period, student and program where student is enrolled. Enrolment report is of students reports at academic period. It comprises of diverse fields based on student data, number of enrolments and certain academic performance like weighted GPA and GPA.

Grade report comprises of course data in every period and final outcomes. Some related fields regarding courses are: number of credits, course section, numeric grade (0 to 5), alphabetic grade (not approved or approved) and typology of course, that is, foundation, professional, optional electives and levelling courses. Academic status loss of students registers while academic report is blocked. Certain corresponding fields are blocking code, description, academic period and date, if it is still in functionality or not. If available, information of unlocking students' academic history is performed which comprises description, code and academic period. Based on the lower performance of academic status, student may have some academic history blocked due to numerous reasons which are codified in Information systems. Types of blocking were classified as academic, owing to low academic performance, that is, inappropriate GPA, .fails in two subjects or insufficient credits/non-academic credits, when academic performance needs were still satisfied however student has not enrolled in academic period, for instance, transfer to other program, withdrawal or suspension for not renewing enrolment. Academic category is only one consideration in this work as an attempt to isolate components of academic performance. To carry out classification task, two datasets were merged into one file. This process comprises of numerous steps. Initially, grades report was summed up for each student in certain academic period. This comprises of percentage of approved credits, number of credits, average grade both in general and as well specific typology of course, that is, foundation, professional and optional electives. There is some information regarding performance in two levelling of course: best performance and worst performance.

IV. PROBLEM STATEMENT

The ultimate objective of this research is to in co-operate prior knowledge of Fuzzy Support Vector Machine to assist in learning process by integrating certain constraints derived from experts' knowledge about target function. This section offers forma definition of problem statement. Initially, the concept of Fuzzy based SVM is defined and explains how it assists in learning process of FSVM. Subsequently, certain constraints of FSVM are overcome by genetic algorithm for recognizing academic performance. Thirdly, in classification model, dual optimization problem of FSVM-GA is formulated and theoretically analyze its classification property. At last, the constraints of existing models are overcome and assist in improving learning performance using dataset instances.

V. PROPOSED METHOD

This section explains in detail about genetic algorithm and fuzzy based SVM classification for predicting student's performance.

a. Genetic based Association

In accordance to basic concept of genetic algorithm, some specific operations like species grouping like encode, select, interest and variation. At last, some useful association rules can also be drawn. The fundamental motivation lies in considering support and confidence as computational standard of fitness function in meantime and determining the influence of two individuals.

In primary species grouping, it will construct more effectual individual copy, which increases the quality of species group. In accordance to mutation operator on variation, and regulate individuals in two small species groups based on fitness size, moreover it guarantees quality of population.

b. Implementation of algorithm model

Step 1: In initialization phase, it randomly produce an initial species group which comprises of 'N' individuals after coding and compute average fitness and fitness of every individual in species group in accordance to computation of fitness which is defined as above;
Step 2: sorting every individual in species group in accordance to its fitness from large to small;
Step 3: validating individuals in accordance to fitness size. 20% individuals, and does not copy rest of 80% and computing reserved individuals as 'M';
Step 4: if 'M' then haphazardly computes N-M individuals, and maintains total amount 'N' of species group else move to step 4;
Step 5: Evaluate individuals in accordance to average fitness and fitness, and validate mutation rate;
Step 6: computing least fitness value → 20% of individuals and highest fitness value → 80% of individuals, if least fitness value is lesser than highest fitness value, then substitute individuals, else move to step 7; repeat step till the value is not exchanged, then go to step 7;
Step 7: choosing two individuals from 20% individual group haphazardly; then perform numerous single points over two individuals, at last, select most finest individual from attained individuals and accumulating novel species group;
Step 8: terminate till fulfills every conditions, else move to step 2;
Step 9: construct association rules.

c. Association rule generation

Based on the above algorithm, student academic performance is constructed based on association rules as follows: (100% confidence, 35% support)

Name <Rank: junior; domestic development level: developed, gender: male>

(100% confidence, 25% support)

Name <Rank: senior, domestic development level: general, gender: female>

(95% confidence, 15% support)

Name <Rank: senior, domestic development level: general, gender: male>

(94% confidence, 8% support)

Name <Rank: senior, gender: male, domestic development level: highly developed>

(95% confidence, 10% support)

Based on the above extracted rules, the result extracted shows that higher level performance of students depicts effectual performance of individuals in grade exam; Students performance also depends on monthly students report; individuals who secures very lesser grade will have lesser confidence and therefore efforts to deal with confidence enhancement has to be done. In the group of contemporary individuals computation of grade level, majority of student performance is evaluated, therefore it can be seen that higher grade scoring individuals are more attractive for grade computation.

d. FSVM

This section explains in detail about the virtual computation of instances acquired from the dataset. This approach is modelled for some non-linear cases, wherein linear models are derived from elimination of kernel with identity function.

Number FSVM model were constructed for numerous classification crisis, certain training points are significant. Some of the training points are extremely important for classification purposes, whereas less significant training points are not considered or misclassified. Followed by this, fuzzy based membership function is defined for every training points, x_i is measured as significance and classified as one class, while $1 - S_i$ is a degree where meaningless.

Some observed student's dataset (academic performance) $\{(x_k, y_k, s_k)|k = 1,2, \ldots N\}$, with respect to these data, FSVM is presented as in Eq. (1):

$$\min J(W, E) = \frac{1}{2} W^T W + C \sum_{k=1}^{N} s_k e_k \qquad (1)$$

The above derived function transforms input to high dimensional feature space, while classification crisis are permissible, or chances are classification is superior. $\varphi(X)$ is identity function, which is specifically used in case of linearly separable crisis.

From Eq. [1], target function has form $y\,(x) = sign\,[W^T \varphi(X) + \mathrm{b}]$, which is monotonic in case if it without loss of generality. Monotonic constraint is expressed as in Eq. (2):

$$w^T \varphi(X) \le w^T \varphi(X) \qquad for\ all\ \ x \le \tilde{x} \quad (2)$$

Consider a set of virtual academic performance of students or observed pair of data, with these observed virtual instances, classification can be done with monotonic constraints based inequality.

From the dataset used, the academic performance of students is considered as attributes, along with the grade reports of students periodically. Grade report increases the monotonic constraint nature whereas the student's personal information reduces the monotonic constraint nature. Therefore, the association between the attribute relationships is increased or decreased based on the monotonic constraints. Let an academic performance of student example $x_1$and x_2 be two students correspondingly. Suppose the comparison of these two students performance in grade examination has shorter duration for grade 1, grade 2, grade 3 and so on, these students possess different GPA, then based on monotonic nature provided by experts, student x_1 is assigned to class which is less than or equal to class of student x_2. Linear inequality based monotonic property is provided as in Eq. (3):

$$w^T\ \varphi(X_1) \le w^T \varphi(X_2) \qquad (3)$$

By adding this monotonic constraint to FSVM, the model can be constructed as in Eq. (4):

$$\min J(w, e) = \frac{1}{2} w^T w + C \sum_{k=1}^{N} s_k e_k \qquad (4)$$

Subject to $y_k(w^T \varphi_k) + b) \ge 1 - e_k, k = 1,2, \ldots N$

Algorithm 1:
Input: Students dataset with fuzzy membership
Output: Classification based on students performance
Step 1: Determine the monotonic constraint using Eq. (2)
Step 2: Compute the target function using SVM
Step 3: Solve the loss of generality using quadratic program solve such as MATLAB
Step 4: Apply grid search and find optimal parameter
Step 5: Output optimal grade level using linear inequality
Step 6: Determine FSVM classifier
Step 7: Construct FSVM using monotonic constraint using Eq. (4)

VI. SIMULATION SETUP AND DISCUSSIONS

In order to carry out a fair comparison, all SVM classifiers incorporated with Fuzzy, and same MATLAB is adopted for GA approaches. Codes are executed in MATLAB R2011a with Intel Core i7-3770 CPU of 16 GB RAM Server 2008. Datasets are provided as training and test sets to analyze out-of-sample prediction performance. Cross-validation is used in training to find optimal factors for proposed techniques.

a. Performance Measures

In this investigation, we compare the proposed GA-FSVM algorithm with other prevailing classifiers, examining their performances in terms of accuracy, recall, precision, F-measure.

Table Ia: Tabular representation for P-value outcomes of GA-FSVM and existing k-NN, SVM, DT in online dataset

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC	FSVM
1	0.5558	0.5431	0.5249	0.4637	0.355	0.360
2	0.5556	0.5514	0.5060	0.4931	0.341	0.345

Table I b: Tabular representation for P-value outcomes of GA-FSVM and existing k-NN, SVM, DT in real time dataset

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC	FSVM
1	0.5500	0.5425	0.5239	0.4627	0.345	0.350
2	0.551	0.541	0.5050	0.4921	0.331	0.335

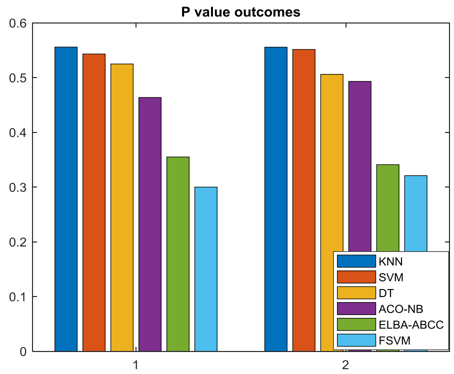

Fig 1: Graphical representation of 'P' value in online and real time dataset

Table II a: Tabular representation for T-value outcomes of GA-FSVM and existing k-NN, SVM, DT, ACO-NB in online dataset

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC	FSVM
1	25.5546	25.1147	25.0510	23.9635	21.96	22.3
2	27.4008	24.6391	21.2804	20.1384	19.85	19.90

Table II b: Tabular representation for T-value outcomes of GA-FSVM and existing k-NN, SVM, DT, ACO-NB in real time dataset

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC	FSVM
1	25.5536	25.1137	25.0520	23.9625	21.86	21.90
2	27.4001	24.6381	21.2794	20.1374	19.75	19.80

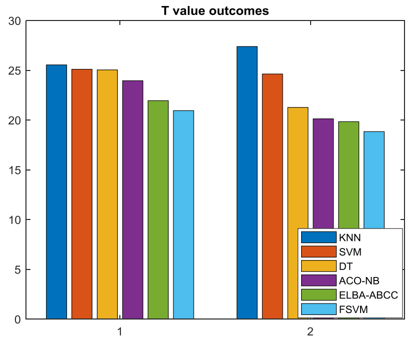

Fig 2: Graphical representation of 'T' value in online and real time dataset

Table III a: Tabular representation for F-measure outcomes of GA-FSVM and existing k-NN, SVM, DT, ACO-NB in online dataset

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC	FSVM
1	84.7702	87.3614	92.0183	92.6802	94.56	95.66
2	88.6957	92.9263	93.2766	93.3519	94.35	95.68

Table III b: Tabular representation for F-measure outcomes of GA-FSVM and existing k-NN, SVM, DT, ACO-NB in real time dataset

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC	FSVM
1	84.7602	87.3514	92.0162	92.6780	94.40	95.45
2	88.6950	92.9326	93.2750	93.3500	94.20	95.02

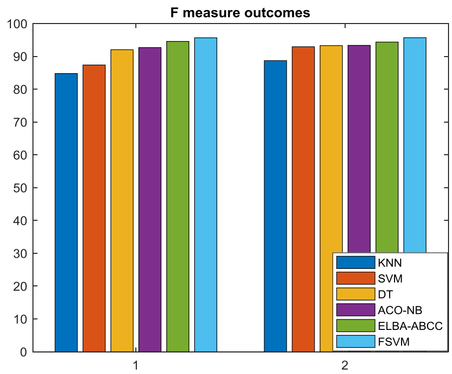

Fig 3: Graphical representation of F-measure value in online and real time dataset

Table IV a: Tabular representation for precision outcomes of GA-FSVM and existing k-NN, SVM, DT, ACO-NB in online dataset

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC	FSVM
1	91.1721	94.2630	94.7021	97.8204	98.2	98.9
2	93.8838	96.2255	96.4893	97.1901	98.5	98.9

Table IV b: Tabular representation for precision outcomes of GA-FSVM and existing k-NN, SVM, DT, ACO-NB in online dataset

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC	FSVM
1	91.1700	94.2600	94.7000	97.8100	98.20	98.9
2	93.8700	96.2200	96.4800	97.1800	98.50	98.89

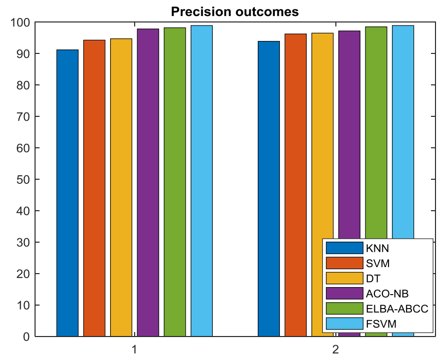

Fig 4: Graphical representation of Precision in online and real time dataset

Table V a: Tabular representation for Recall outcomes of GA-FSVM and existing k-NN, SVM, DT, ACO-NB in online dataset

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC	FSVM
1	78.8785	82.9769	83.3104	83.8203	85.6	86
2	78.8910	83.5338	85.2581	86.7424	84.5	85

Table V b: Tabular representation for Recall outcomes of GA-FSVM and existing k-NN, SVM, DT, ACO-NB in online dataset

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC	FSVM
1	78.8700	82.9665	83.3000	83.8081	85.0011	86.11
2	78.8801	83.5238	85.2500	86.7300	84.4001	85.5

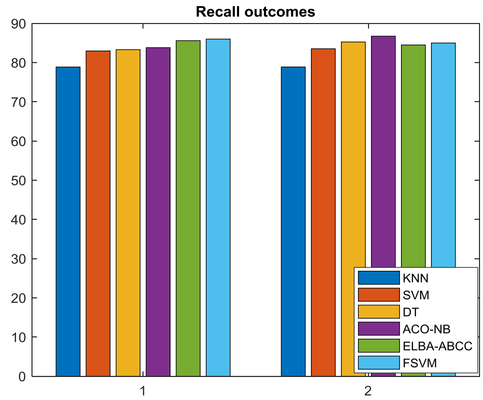

Fig 5: Graphical representation of recall value in online and real time dataset

Table VI a: Tabular representation for accuracy outcomes of GA-FSVM and existing k-NN, SVM, DT, ACO-NB in online dataset

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC	FSVM
1	78.7342	80.5064	82.3165	85.6190	88.72	89.27
2	79.4602	81.1438	85.3756	85.3888	88.9	89.5

Table VI b: Tabular representation for accuracy outcomes of GA-FSVM and existing k-NN, SVM, DT, ACO-NB in real dataset

S.NO	K-NN	SVM	DT	ACO-NB	ELBA-ABCC	FSVM
1	78.7242	80.4098	82.3262	85.6001	88.60	89.65
2	79.4502	81.1004	85.3680	85.3700	88.80	89.90

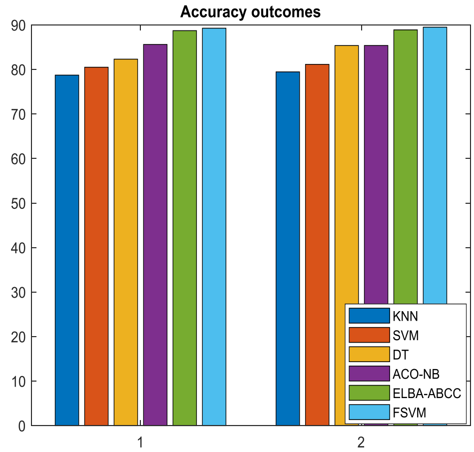

Fig 6: Graphical representation of Accuracy value in online and real time dataset

Accuracy is intuitive measurement strategy, which openly describes predictive ability on tested data which are classified appropriately, and defined as in Eq. (5):

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

where TP, FP, TN and FN are values utilized to describe predictive power of DM, and defined as below:

TP = instances with positive outcomes which are classified appropriately.

FP = instances with positive outcomes which are misclassified.

TN = instances with negative outcomes which are appropriately classified.

FN = instances with negative outcomes which are misclassified.

Recall depicted as ratio of positives that are appropriately identified and defined as in Eq. (6):

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

Precision is ratio of instances with positive predictive outcomes with appropriate prediction. It is defined as in Eq. (7):

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

F-measure is harmonic mean of precision and recall (sensitivity), and depicted as in Eq. (8):

$$F-measure = 2 * \frac{PPV.sensitivity}{PPV+sensitivity} \quad (8)$$

With F-measure, precision and recall are considered to avoid situation with high precision and low recall or vice versa.

Figure 1 shows graphical representation of P-value of the proposed method with the existing methods such as -NN, SVM, and DT. The outcome attained using GA-FSVM is 0.300, 0.321 respectively.

Table I1 a and I b depicts the iterative outcome attained for the proposed method with the existing work such as k-NN, SVM, DT in online and real time dataset. The P-Value attained for prevailing methods are 0.5558, 0.5556, 0.5431, 0.5514, 0.5249, and 0.5060 correspondingly.

Figure 2 depicts graphical representation of T-value of proposed method with the existing methods such as -NN, SVM, and DT. The outcome attained using GA-FSVM is 20.96, 18.85 respectively.

Table II a and b depicts the iterative outcome attained for the proposed method with the existing work such as k-NN, SVM, DT over online and real time dataset. The T-Value attained for prevailing methods are 25.5546, 27.4008, 25.1147, 24.6391, 25.0510, and 21.2804 correspondingly.

Figure 3 shows the graphical representation of F-measure of anticipated method with the existing approaches such as -NN, SVM, and DT. The outcome attained using GA-FSVM is 95.66 and 95.68respectively.

Table III a and b depicts the iterative outcome attained for the proposed method with the existing work such as k-NN, SVM, DT over online and real time dataset. The F-measure attained for prevailing methods are 84.7702, 88.6957, 87.3614, 92.9263, 92.0183, 93.2766correspondingly.

Figure 4 depicts graphical representation of precision of anticipated approach with the existing methods such as -NN, SVM, and DT. The outcome attained using GA-FSVM is 98.9 and 98.9 respectively.

Table IV a and b depicts the iterative outcome attained for the proposed method with the existing work such as k-NN, SVM, DT over online and real time dataset. The precision attained for prevailing methods are 91.1721, 93.8838, 94.2630, 96.2255, 94.7021, and 96.4893 correspondingly.

Figure 5 shows the graphical representation of Recall % of the proposed method with the existing methods such as -NN, SVM, and DT. The outcome attained using GA-FSVM is 86 and 85 respectively.

Table V a and b depicts the iterative outcome attained for the proposed method with the existing work such as k-NN, SVM, DT over online and real time dataset. The Recall attained for prevailing methods are 78.8785, 78.8910, 82.9769, 83.5338, 83.3104, and 85.2581correspondingly.

Figure 6 depicts graphical representation of Accuracy % of anticipated method with existing methods such as -NN, SVM, and DT. The outcome attained using GA-FSVM is 89.27 and 89.5 respectively.

Table VI a and b depicts the iterative outcome attained for the proposed method with the existing work such as k-NN, SVM, DT over online and real time dataset. The accuracy attained for prevailing methods are 78.7342, 79.4602, 80.5064, 81.1438, 82.3165, and 85.3756 correspondingly. As per the investigation, accuracy of 85.6190% is attained in our anticipated replica through validation phase. The outcome of validation phase shows reliability of anticipated replica.

VII. CONCLUSION

In this investigation, the proposed model is examined using the hybrid online and the collected dataset. The proposed model impact over the performance of using proposed Fuzzy membership is generated using FSVM and monotonic constraints of training students' data. Analytical outcomes illustrates that the anticipated methods have benefits over FSVM and GA algorithm, which shows the effectiveness and feasibility of the model over the prevailing techniques. As well, formulation of these constraints attains kernel estimation which can be extended and can be utilized as framework for developing classifiers. In future, loss of generality can be investigated in detail to eliminate redundant features and to reduce complexity.

REFERENCES

1. S. Kotsiantis, "Educational data mining: A case study for predicting dropout—prone students," Int. J. Know. Eng. Soft Data Paradigms, vol. 1, no. 2, pp. 101–111, 2009.
2. S. Kotsiantis, "A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education," Knowl. Based Syst., vol. 23, no. 6, Aug. 2010.
3. C. Romero, "Educational data mining: A review of the state of the art," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40, Nov. 2010.
4. M. N. Quadril, "Drop out feature of student data for academic performance using decision tree techniques," Global J. Comput. Sci. Technol., Feb. 2010.
5. F. Araque, C. Roldán, "Factors influencing university drop out rates," Comput. Educ., 2009.
6. R. Alkhasawneh, "Modeling student retention in science and engineering disciplines using neural networks," in EDUCON, 2011.
7. R. D. Reason... "Student Variables that Predict Retention: Recent Research and New Developments," Journal of Student Affairs Research and Practice, 2009.
8. D. A. Carnegie, "Prediction of success in engineering study," in 2012 IEEE Global Engineering Education Conference 2012.
9. G. W. Dekker,. "Predicting Students Drop Out: A Case Study," in Proceedings of 2nd International Conference on Educational Data Mining, 2009.
10. S. Kotsiantis "Educational data mining: a case study for predicting dropout-prone students," International Journal of Knowledge Engineering and Soft Data Paradigms, 2009.
11. G. Siemens, "Learning analytics and educational data mining: towards communication and collaboration," in Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, 2012.
12. R. Ferguson "The State of Learning Analytics in 2012: A Review and Future Challenges," Knowledge Media Institute, 2012.
13. Y. Kurniawan, "Use data warehouse and data mining to predict student academic performance in schools: A case study (perspective application and benefits)," IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), 2013.
14. S. Ferreira, "Academic Analytics: Mapeando o Genoma da Universidade," VAEP, RITA, IEEE, 2013.
15. R. Hosseinn, "Student Learning Outcomes Assessment and Curriculum Continuous Improvement Processes," in Education, Research and Development: Proceedings of Latin American and Caribbean Conference for Engineering and Technology, Tegucigalpa, Honduras, 2008.