IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 9, Issue.9, September 2021 E-ISSN: 2347-2693

Emotion Analysis and Performance Perdiction Using Cluster Based LDA

Kajal Devi[1*], Harjinder Kaur[2]

[1,2]Department of Computer Science, Swami Sarvanand College of Engineering and Technology, Punjab Technical University, Jalandhar, India

[*]*Corresponding Author: dkajal077@gmail.com, Tel.: +91-9501497689*

DOI: https://doi.org/10.26438/ijcse/v9i9.1624 | Available online at: www.ijcseonline.org

Received: 21/Sept/2021, Accepted: 23/Sept/2021, Published: 30/Sept/2021

***Abstract*—** Kajal Devi, Harjinder Kaur, for a productive life, Marketing plays a critical role to fill individual life with value and excellence. Marketing is compulsory to provide things that individuals partake in to compete in the modern world. Predicting the academic performance of the Business is the most successive research in this era. A different set of approaches and methods are incorporated to increase Business performance. However, this is a challenging task due to the wrong course selection. In this paper, we have used the Cluster-based Linear Discriminant Analysis (CLDA) and Artificial Neural Network (ANN) based approaches for the prediction and classification of Business performance. The proposed study will provide the prospective business with the motivational comments and the video recommendations by which Business can choose the right subject and the comments will facilitate the Business with the insight reasons of dropout opted by other Business for this course. The outcomes of this study will help in the reduction of the number of dropouts. The Business will be able to choose an appropriate course for performance enhancement and carrier excel.

***Keywords*—** Cluster-based Linear Discriminant Analysis (CLDA), Business performance, Dropouts, Classification, Prediction, and machine learning.

I. INTRODUCTION

Data mining is one of the most cardinal areas in recent technologies for retrieving valid information from the huge amount of unstructured and distributed data using parallel processing of data [7]. Data mining techniques are applied in various fields to find novel information from the huge data set. The researchers are seeking interest in the Marketing filed to investigate new research [19]. Recently, data mining technology has been used in the Marketing filed to extract the hidden information from Marketing data sets. It also supports the classical Marketing system facilitating teachers to analyse what Business know and what learning techniques are most effective for Business. Marketing data mining and learning analytics employ technologies from statistics, computer science, and machine learning to extract useful information from collected Marketing data, gain valuable insight into learning and find out solutions to improve learning performance and teaching effectiveness.

Figure 1: machine learning steps

Nowadays, the digitization is used by the universities in teaching-learning and other academic processes causing the generation of a huge amount of digital data. This data is helpful for teachers, policymakers, and administrators for decision making if it is effectively transformed. By providing timely information to different stakeholders, it advances the quality of Marketing processes. It is found that in many countries (including developing and developed countries) the number of dropouts is on the rise every year [11]. All counties have taken important initiatives in this direction to tackle and overcome this Marketing flaw. Past years have appeared developing interest and worry in a few nations about the issue of school failure and the assurance of its principle contributing elements [14].

In modern Marketing systems, the use of Information and Communications Technology (ICT) and other teaching-learning techniques are widespread. They enhance the teaching and learning capabilities and improve the overall process of the study. At the same time, the use of these modern tools and techniques helps in capturing the day to day Business-related data. The Business performance analysis and suggesting a future course of action is the need of the hour. To enhance the quality of Marketing system Business performance analysis plays an important role in decision support. It is an important factor for Business, if they choose the wrong course it will create problems for them.

In this research, our focus is to predict the performance of Business and to provide an automatic recommendation to the Business which will ultimately help them to avoid (overcome) the problem of dropouts. In the period of exploring the existing literature on classifying the characteristics of dropout Business, it is found that most studies are employing text mining techniques. But, these methods itself showing the major disadvantage that the clusters formed by these heuristics are overlapped and boundary values are not identified. To rectify this issue, we proposed the CLDA approach is applied. It is observed that the outcome of the proposed system will assist the Business to choose the right subject. The comments will also facilitate the Business with the insight reasons for the dropout opted by other Business for the specific course. More specifically, this will allow the Business to know about the key issues of dropout Business.

The paper is arranged in such a manner that Section 1 introduction about the problem. In section 2 the related work. Section 3 describes the data analysis and methodology. Section 4 delineates the performance and results. Section 6 presents the discussion in Section 7 conclusion and future work are given.

II. RELATED WORK

To prove the worth of the proposed work, we study some research work described in this section. In most of the literature, the different data mining techniques are used to extract the data from the database. The common observation that we perceive while conducting this literature survey that the dropout ratio prediction is inaccurate and can be further improved. Few of the studies that we found relevant with our existing work are:

Decision trees are used to classify the give data into a set of classes. The nodes within the decision tree represent attributes to be classified. The attribute classification is based on the highest correlation values. The decision tree mechanism used to determine dropout Business with good classification accuracy (Santos *et al.*, 2019). Naïve Bayes's approach is based on an inductive learning mechanism. The mechanism operates in two distinct phases. In the first phase, training is performed and in the second phase, testing is performed (Manhães and Zimbrão, 2014). However, the mechanism predicts the dropout ratio but it cannot predict dropout prone Business. An artificial neural network is another model used for Business dropout prediction that is based upon inductive learning mechanism. This is based upon the working of biological neurons. Training and testing mechanisms are incorporated to predict Business dropouts (Lesinski et al., 2016; Divyabharathi and Someswari, 2018; Mason et al., 2018). Hierarchical clustering is the mechanism for finding the relationship between the attributes and passes out Business. The training is performed according to the pass percentage attribute within the dataset. The test results in the dropout as prediction (Thammasiri et al., 2014). The problem of dropout prone Business detection is not handled and the clusters are overlapped.

Suljic [5] proposed the data mining approach to analyse Business performance. Three distinct data mining approaches are followed for the same. Data collection which is a preliminary approach is accomplished by the use of surveys and questionnaires. After the preliminary approach obtained data is fed into a classifier to obtain predictions. naive Bayes classifier outperforms all other classifiers. Liu et al. [2] described a technique that analyses the behavioural pattern of Business that join the course. It found a relationship by analysing the login attempt of Business and the duration of watching videos. Then it will predict the drop out ratio and number of a Business interested in a particular field. It utilizes social learning network and regression analysis for prediction. Aleman and Garza [3] proposed a quantitative method that uses statistics and probabilistic models to analyse the MOOC dataset. It gives levels of retention, desertion, and completion of courses in which Business are enrolled. The result shows low terminal efficiency and also gives the percentage of Business that positively responded. Wang and Graduate [14] described a deep neural network that combines convolution neural networks and recurrent neural networks. This model automatically extracts the features of raw MOOC data and it does not need manual feature extraction. It gives a better prediction rate that is given by the use of a feature that is extracted from raw data but it does not consider classification problems so the accuracy is affected. Chen et al. [4] proposed a hybrid algorithm that is based on a decision tree and extreme learning mechanism that used for predicting dropout ratio. It uses a decision tree for selecting features and then weights are assigned to the selected features for enhancing their classification ability. It can analytically determine the prediction results that are updated after every iteration automatically. The overall result is better and prediction accuracy is high.

Feng et al. [16] suggested a context-aware feature interaction network that used to predict the dropout rate. It utilizes two dataset KDDCUP and XuetangX that are analyzed for predicting the results. It utilizes a context-smoothing technique to smooth the values of activity features using the convolution neural network. It also provides an attention mechanism that combines user and course information for predicting the values. Peng and Member [18] proposed an LDA based mechanism to improve the classification accuracy from the extracted data in the dataset. The size of the data could be large. Forming clusters from the dataset reduces execution time and hence abnormal patterns could be discovered effectively. Zheng et al. [20] introduced a trust-based mechanism in the trust-based recommendations. The recommendation mechanism uses content-based filtering. The content-based mechanism uses KNN clustering and recommendations are based on these mechanisms are quick with high classification accuracy

.

Márquez-Vera et al. [19] proposed a business failure ratio within the school using the application of a genetic algorithm. The genetic algorithm can operate on a large dataset since it operates in phases. The genetic approach provides better accuracy but the convergence rate is poor. Kuo et al. [22] proposed a stacked auto-encoder mechanism to predict Business dropout from any course at the undergraduate or postgraduate level. The mechanism that is considered can be used the person at UG or PG level to choose the course according to their capabilities. Fortin et al. [12] proposed a dropout prediction mechanism by the use of gender categorization mechanisms. This mechanism gives the dropout by forming clusters based on gender attributes. This means two different clusters are formed using the nearest neighbour based approach. The edge detection is missing in this approach thus classification accuracy is a problem.

Ktona, Xhaja and Ninka in [2] stated that data mining is an area in computer science that can be used in Marketing and at the same time can provide findings that will help to increase the Marketing quality. It's a combination of tools of statistics with database management and artificial intelligence. Carter [27], and Dojrulwkp [22] used a mixed-effects analysis of variance (ANOVA) models to evaluate the impact of using Twitter on college Business engagement and learning outcomes. The engagement was measured with a dedicated instrument called the National Survey of Business Engagement. Results showed a significant increase in both engagement and grades for the experimental group, in which Business used Twitter for various types of academic discussions.

Pinjuh [25] and Bekele et al. [20] focused on the utilization of Business' commitment in a discussion gathering as a displayer of Business execution. Information was taken from 114 Business that are tried out an early on a software engineering course. They utilized the gathering that is incorporated into Moodle LMS for talking about the substance, posing inquiries, or giving assistance to peers, and took a test at the remainder of the semester. The creators proposed to gauge whether Business passed or bombed the course dependent on their gathering utilization, in states of quantitative, subjective, and informal organization pointers. A contrast between traditional classification and clustering algorithms implemented in Weka was performed, together with various approaches for instance, and attribute selection.

Heroic-Bozic et al. [23] and Feng [18] explored Business' usage data in Moodle LMS as a forecaster for their grade of exam. 438 Business from seven engineering courses were included in the study. Eight attributes related to learner action on quizzes, assignments, and forum messages were calculated for every Business. For classifying Business with similar grades, authors applied various data mining techniques. Performance comparisons were carried out, with various pre-processing techniques. On the whole, the accuracy obtained is not very high (around 65%), representing the complexity of the prediction task. Varga et al. [35] use the technique of hybrid cloud. Where clusters are formed together and clusters are overlapped in the hybrid mechanism
.
The existing literature focuses on the dropout rate but the mechanism by which it can decrease is not highlighted. Also, there is no provision to handle noisy data as it can affect accuracy. Comment recommendation is the basic need by which the Business dropout rate can be controlled is also not present. We precisely observed that in [35] the clustering mechanism is used but the boundary value analysis is missing by which overlapping of clusters is there which leads to a major issue of inaccurate prediction of the Business performance. We propose a solution in which the dropout rate is decreased with an increase in the accuracy of the solution. We also aimed to provide comment recommendations to the Business that will ultimately help them to choose the right course.

To prove the worth of the proposed work, we study some research work described in this section. In most of the literature, the different data mining techniques are used to extract the data from the database. The common observation that we perceive while conducting this literature survey that the dropout ratio prediction is inaccurate and can be further improved. Few of the studies that we found relevant with our existing work are:

Decision trees are used to classify the give data into a set of classes. The nodes within the decision tree represent attributes to be classified. The attribute classification is based on the highest correlation values. The decision tree mechanism used to determine dropout Business with good classification accuracy (Santos *et al.*, 2019). Naïve Bayes's approach is based on an inductive learning mechanism. The mechanism operates in two distinct phases. In the first phase, training is performed and in the second phase, testing is performed (Manhães and Zimbrão, 2014). However, the mechanism predicts the dropout ratio but it cannot predict dropout prone Business. An artificial neural network is another model used for Business dropout prediction that is based upon inductive learning mechanism. This is based upon the working of biological neurons. Training and testing mechanisms are incorporated to predict Business dropouts (Lesinski et al., 2016; Divyabharathi and Someswari, 2018; Mason et al., 2018). Hierarchical clustering is the mechanism for finding the relationship between the attributes and passes out Business. The training is performed according to the pass percentage attribute within the dataset. The test results in the dropout as prediction (Thammasiri et al., 2014). The problem of dropout prone Business detection is not handled and the clusters are overlapped.

Suljic [5] proposed the data mining approach to analyse Business performance. Three distinct data mining approaches are followed for the same. Data collection which is a preliminary approach is accomplished by the use of surveys and questionnaires. After the preliminary approach obtained data is fed into a classifier to obtain

predictions. Naive Bayes classifier outperforms all other classifiers. Liu et al. [2] described a technique that analyses the behavioural pattern of Business that join the course. It found a relationship by analysing the login attempt of Business and the duration of watching videos. Then it will predict the drop out ratio and number of a Business interested in a particular field. It utilizes social learning network and regression analysis for prediction. Aleman and Garza [3] proposed a quantitative method that uses statistics and probabilistic models to analyse the MOOC dataset. It gives levels of retention, desertion, and completion of courses in which Business are enrolled. The result shows low terminal efficiency and also gives the percentage of Business that positively responded. Wang and Graduate [14] described a deep neural network that combines convolution neural networks and recurrent neural networks. This model automatically extracts the features of raw MOOC data and it does not need manual feature extraction. It gives a better prediction rate that is given by the use of a feature that is extracted from raw data but it does not consider classification problems so the accuracy is affected. Chen et al. [4] proposed a hybrid algorithm that is based on a decision tree and extreme learning mechanism that used for predicting dropout ratio. It uses a decision tree for selecting features and then weights are assigned to the selected features for enhancing their classification ability. It can analytically determine the prediction results that are updated after every iteration automatically. The overall result is better and prediction accuracy is high.

Feng et al. [16] suggested a context-aware feature interaction network that used to predict the dropout rate. It utilizes two dataset KDDCUP and Xuedong that are analysed for predicting the results. It utilizes a context-smoothing technique to smooth the values of activity features using the convolution neural network. It also provides an attention mechanism that combines user and course information for predicting the values. Peng and Member [18] proposed an LDA based mechanism to improve the classification accuracy from the extracted data in the dataset. The size of the data could be large. Forming clusters from the dataset reduces execution time and hence abnormal patterns could be discovered effectively. Zheng et al. [20] introduced a trust-based mechanism in the trust-based recommendations. The recommendation mechanism uses content-based filtering. The content-based mechanism uses KNN clustering and recommendations are based on these mechanisms are quick with high classification accuracy

.

Márquez-Vera et al. [19] proposed a business failure ratio within the school using the application of a genetic algorithm. The genetic algorithm can operate on a large dataset since it operates in phases. The genetic approach provides better accuracy but the convergence rate is poor. Kuo et al. [22] proposed a stacked auto-encoder mechanism to predict Business dropout from any course at the undergraduate or postgraduate level. The mechanism that is considered can be used the person at UG or PG level to choose the course according to their capabilities. Fortin et al. [12] proposed a dropout prediction mechanism by the use of gender categorization mechanisms. This mechanism gives the dropout by forming clusters based on gender attributes. This means two different clusters are formed using the nearest neighbour-based approach. The edge detection is missing in this approach thus classification accuracy is a problem.

Ktona, Xhaja and Ninka in [2] stated that data mining is an area in computer science that can be used in Marketing and at the same time can provide findings that will help to increase the Marketing quality. It's a combination of tools of statistics with database management and artificial intelligence. Carter [27], and Dojrulwkp [22] used a mixed-effects analysis of variance (ANOVA) models to evaluate the impact of using Twitter on college Business engagement and learning outcomes. The engagement was measured with a dedicated instrument called the National Survey of Business Engagement. Results showed a significant increase in both engagement and grades for the experimental group, in which Business used Twitter for various types of academic discussions.

Pinjuh [25] and Bekele et al. [20] focused on the utilization of Business' commitment in a discussion gathering as a displayer of Business execution. Information was taken from 114 Business that are tried out an early on a software engineering course. They utilized the gathering that is incorporated into Moodle LMS for talking about the substance, posing inquiries, or giving assistance to peers, and took a test at the remainder of the semester. The creators proposed to gauge whether Business passed or bombed the course dependent on their gathering utilization, in states of quantitative, subjective, and informal organization pointers. A contrast between traditional classification and clustering algorithms implemented in Weka was performed, together with various approaches for instance, and attribute selection.

Heroic-Bozic et al. [23] and Feng [18] explored Business' usage data in Moodle LMS as a forecaster for their grade of exam. 438 Business from seven engineering courses were included in the study. Eight attributes related to learner action on quizzes, assignments, and forum messages were calculated for every Business. For classifying Business with similar grades, authors applied various data mining techniques. Performance comparisons were carried out, with various pre-processing techniques. On the whole, the accuracy obtained is not very high (around 65%), representing the complexity of the prediction task. Varga et al. [35] use the technique of hybrid cloud. Where clusters are formed together and clusters are overlapped in the hybrid mechanism

.

The existing literature focuses on the dropout rate but the mechanism by which it can decrease is not highlighted. Also, there is no provision to handle noisy data as it can affect accuracy. Comment recommendation is the basic need by which the Business dropout rate can be controlled

is also not present. We precisely observed that in [35] the clustering mechanism is used but the boundary value analysis is missing by which overlapping of clusters is there which leads to a major issue of inaccurate prediction of the Business performance. We propose a solution in which the dropout rate is decreased with an increase in the accuracy of the solution. We also aimed to provide comment recommendations to the Business that will ultimately help them to choose the right course.

III. METHODOLOGY

The ability to predict the Business dropout and the failure rate is very important in an advanced -Marketing scenario. The prediction of academic performance of Business could provide an early warning to them who are at risk of dropout [5]. Moreover, the prediction result can also be useful in investigating instructors' performance [17]. Marketing Data Mining (EDM) can be used to develop a prediction model by exploring Marketing data and extracting hidden patterns for predicting Business' academic performance using machine learning techniques.

A. The workflow of the proposed system

The detailed workflow of the conducted study is shown in Fig 1. At the primary step, the two datasets including benchmark and synthetic dataset are taken and stored within the buffer. These datasets are used one at a time. From the dataset, data is fetched and passed through the pre-processing phase. This phase eliminates missing values if any from the dataset. After pre-processing clustering is performed using the CLDA mechanism. Clustering, group the Business according to their performance. After the clustering is over-classification using ANN is accomplished. This classification labels the Business if they are likely to be drop out or not. This prediction is conveyed to tutor for performance enhancement of weak Business.

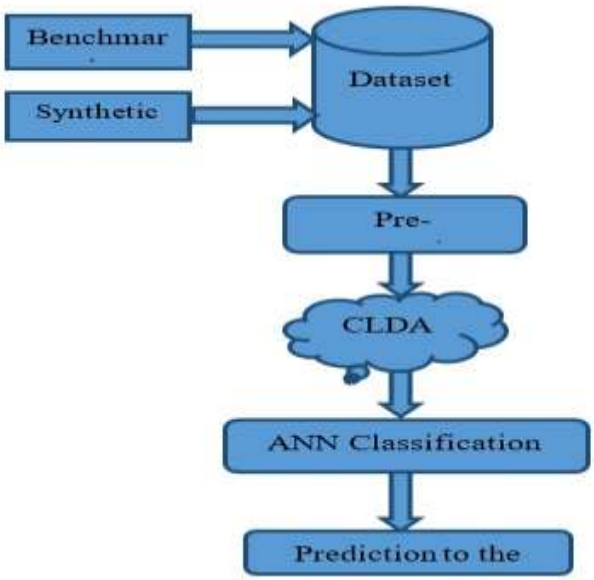

Figure 2. Proposed methodology

B. Data Collection

Data both synthetic and the benchmark is used for testing the proposed work. The informatics course is selected for evaluation. The collected data is checked for missing values under the pre-processing module.

C. Data Preprocessing

Pre-processing includes handling the missing values. Datasets may contain null values. These values cause a reduction in classification accuracy. To handle such a situation, a model is employed and the highest frequency values corresponding to attributes are replaced with the missing values. The highest frequency values belong to a similar ID. This means that attribute values corresponding to the same business are evaluated. The set of attributes are divided into two groups. The first group corresponds to the curriculum-based group and the second group is known as the Business's performance group. The curriculum-based group consists of attributes mainly Business age, gender, computer knowledge (see Table 1), and the performance group consists of attributes marks, attendance shown in Table 2.

Table 1. Curriculum-based group

Attribute	Description
Age	20-26
Gender	Male, female
Occupation	No, Part-time, Full time
Computer Knowledge	Yes, No

Table 2. Performance-based group

Attribute	Description
Subject Marks	0-100
Attendance	10-100%
Assignment	Submitted, Not submitted
Comments	Text

D. Applying Cluster-based linear discriminant analysis (CLDA**)**

The methodology of the proposed mechanism consists of three distinct phases. The first phase identifies the significance of the machine learning approach to dropout Business detection. The labelling information is generated using the CLDA approach.

Figure 2 gives the flow of the proposed work. The algorithm proceeds by beginning with the data collection phase. The data collection is vital for the success of this system. Any discrepancy may cause classification accuracy to decay significantly. To achieve high classification accuracy, a pre-processing mechanism using mode based replacement strategy is used. This way null values are eliminated from the dataset. The next step is clustering, to increase the speed of prediction. In the case of a large dataset, clustering becomes critical. The CLDA approach is used within the clustering. In this approach, frequent items or patterns are grouped using the hierarchical approach. The items sorting and then the path-building approach is used to build a cluster.

Linear discriminant analysis mechanism employed helps in the boundary analysis. In case the boundary is identified clearly than the problem of fetching the data from the wrong cluster is rectified. The mechanism of clustering followed in the proposed approach identifies the boundary and assign distinct centroid value to the cluster. The boundary value analysis performed in CLDA able it to distinctly identifies every cluster with no overlapping. Once clustering is done, ANN is applied to predict the result to the teacher.

E. Artificial neural network (ANN)

The next phase consists of ANN that is used to classify the Business from test data. The last phase applies to video recommendations. ANN is the layered approach in which input, processing, and output layers are used. The input layer takes the pre-processed data after those weights are decided to make the data fall within a particular class. Processing layer checks for the weight adjustment and processing layer finally predict the class. ANN-based classification employed in this work presents effective classification accuracy.

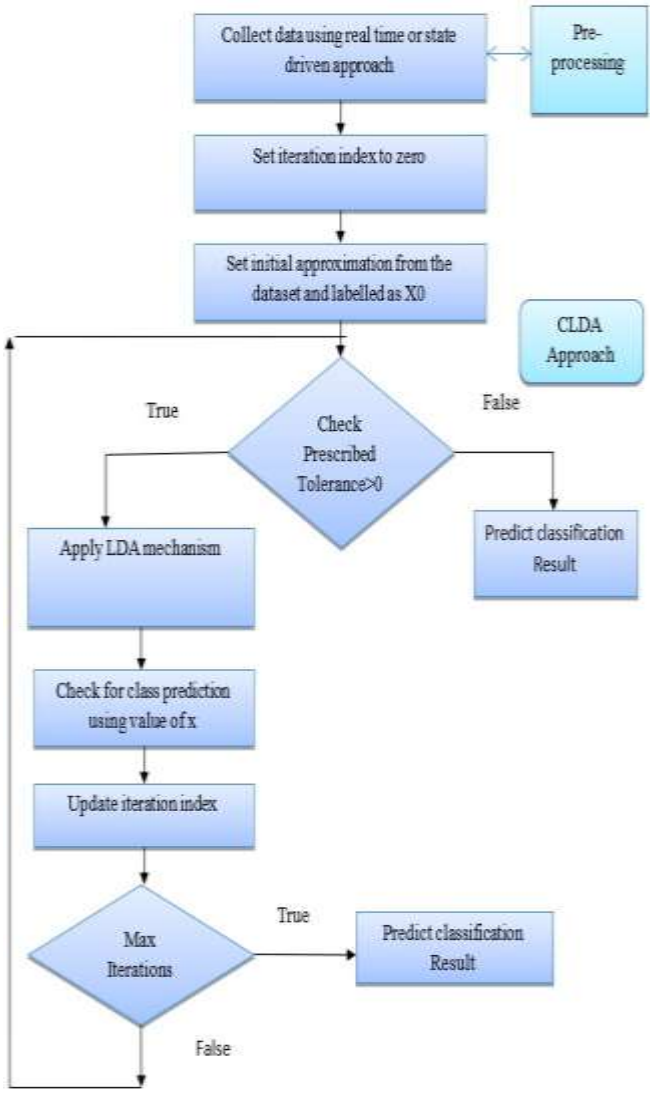

Figure 3. Workflow of CLDA

The algorithm of the proposed mechanism is given as under

Algorithm

- Input Dataset and store it within
  For i=1:n by 1 do
  D_i=Dataset$_i$ // Store dataset Rows into D variable
  End of for
- Pre-processing Mechanism
  For i=1:n
  K=Mode($D_{i(id)}$) finding frequent value on the basis of ID
  If(D_i==NULL)
  D_i=K Replacing it with black values
  End of if

  End of for
- Applying clustering mechanism and LDA
  Perform clustering based on user score
  For i=1:n by 1 do// n is the total number of items in the dataset
  If(Score$_i$>=Score$_{i+1}$-K) //K is threshold value for clustering
  CLUSTER$_i$=i
  End of if
  End of For

  Finding patterns using LDA----------
  Find unique items from each cluster
- <Transaction_ID, Item>=unique(Cluster$_i$)
- Calculate frequency of each item
  For i=1: N// N is the items within cluster
  If(Item$_i$==item$_{i+1}$)
  Count$_i$=Count$_i$+1
  End of if
- Perform sorting of each item, present within the current transaction id using frequency. Placing the highest frequency item in the first place.

  ----------ANN Based Classification----------
- Apply ANN for classification of results
- Build a classification Tree having NULL at root and patterns from the frequency table
- Highest frequency pattern is promoted to the weak business

IV. PERFORMANCE ANALYSIS AND RESULTS

The study is conducted in two distinct phases. During the first phase, the training algorithm extracts the information from the dataset and performs pre-processing. The individual Business are analysed and labelling information is generated and associated with individual clusters generated with the LDA mechanism. The items sorting and then the path-building approach is used to build a cluster. Linear discriminant analysis mechanism employed helps in the boundary analysis. The boundary identification rectifies the problem of fetching the data from the

misclassified data. Using the dataset, three distinct clusters identified and labelled as dropout, pass, and failed Business. These BS Business are prone to dropout.

In the next phase, the test dataset prepared by the teacher is uploaded and checked with the trained system. The information generated helps tutor in focusing more on Business at risk. The recommendations and videos are added synthetically within the dataset. Cluster labelled as passed is used to fetch the comments and video recommendations for the dropout prone business. Video and comment recommendation fetching uses the mechanism of ANN word cloud. Word cloud distinguishes highest-ranked comments in the first place.

We validate the approach on 10 different test datasets. The first five datasets are prepared from the training benchmark dataset and the next five using university Business information. The result of Naïve Bayes, Hierarchical clustering, Neural network, and CLDA is compared. The results in terms of classification accuracy of the first five datasets are given in Table 3 and Fig 3.

Table 3. Comparison of classification accuracy (in %) of Naïve Bayes, Neural network, Hierarchical clustering with CLDA (proposed method)

Dataset (Benchmark)	Naïve Bayes	Neural network	Hierarchical	CLDA
Dataset 1	73	85	85	92
Dataset 2	75	86	87.82	93
Dataset 3	77.56	84	80	91
Dataset 4	76	85	81	95
Dataset 5	75	86	85.2	91.23

Benchmark dataset contains certain fields that are undesirable such as Name, Age, and social_security_number. All of these are eliminated in the proposed work. This is critical as this will increase the speed of execution. The classification can be performed at greater speed using this approach.

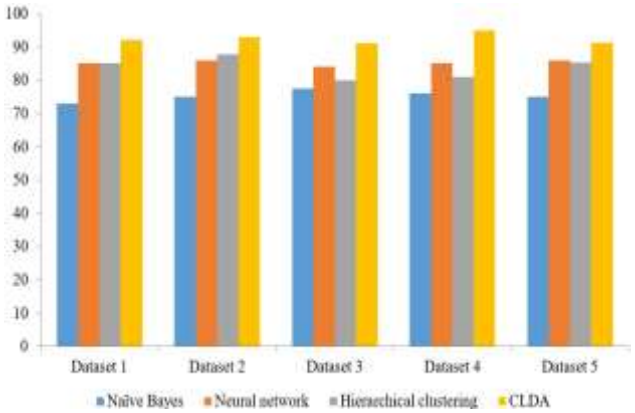

Fig 4. Classification accuracy of distinct machine learning approaches and CLDA

Table 3, indicates that the classification accuracy of CLDA is better and averaged to 93%. The primary cause of this improvement is the clustering approach with LDA. Clustering used in this work clearly distinguish all category of Business and hence accuracy improves. Using this misclassification is greatly handled.

The dataset formed from information gathering is the next phase of evaluation. Five datasets are prepared for evaluation. Table 4 gives the classification accuracy result corresponding to the university Business dataset.

Table 4. Classification accuracy using machine learning and CLDA approach

Dataset (University)	Naïve Bayes	Neural network	Hierarchical clustering	CLDA
Dataset 1	65	75	85	91
Dataset 2	64	75	86	92
Dataset 3	63	74	82	91
Dataset 4	65	75	81	94
Dataset 5	66	76	85	91

The proposed approach gives better results as compared to existing machine learning approaches. The result is decayed by 1 to 2% when the dataset built from university Business is used. The plot for table 4 is given in figure 4.

The dataset accuracy is at stakes since it is prepared synthetically but classification accuracy is improved with the proposed approach. The recommendation generation is the last phase. The patterns discovered if normal is rejected meaning pass Business and only abnormal patterns requiring recommendations are shown through the proposed system. The prediction accuracy is much higher in the proposed system as compared to the existing system.

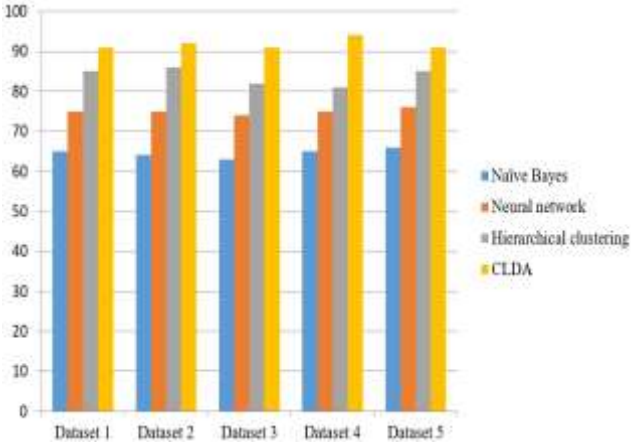

Figure 5: Classification accuracy using university dataset

V. DISCUSSION OF RESULTS

The result obtained from the proposed mechanism with the CLDA includes classification accuracy. This metric is base for checking against the existing mechanism. The improvement of nearly 10% is achieved which is significant and shows the worth of study. The future prediction depends greatly upon the true and false positive categorizations. The true positive rate for the proposed work is much greater as compared to the cluster without CLDA. Patterns detected abnormal, are removed and other

patterns based on true positive rate are maintained causing classification accuracy is improved. The implementation of this system is in MATLAB 2018b that provides tools and methods for quick code segments formation.

Business performance prediction in advance helps them choose the right course for future endeavours. Predicting performance by identifying Business into distinct categories is proposed using clustering but overlapping is an issue. Due to this problem cluster identification causes low classification accuracy and high error rate. To tackle this issue, the proposed system uses the CLDA mechanism. as the clustering is clear and concise, identifying failed and dropout Business are hassle-free. To provide Business with the motivation, the CLDA mechanism uses word cloud to extract comments from passed out Business.

A two-phase approach is used for validation. First of all benchmark dataset is used with CLDA and after that synthetic dataset is used for evaluation purposes. Classification accuracy decays when the synthetic dataset is used but still shows great improvement over other machine learning approaches.

The detection is faster as edges are detected clearly and no cluster overlaps with each other. Also Comments from the past, Business are used to motivating weak and drop out Business. These comments allow weak Business to choose activities that they required to perform to get better results. Also, video recommendations are used in the proposed framework which is missing in the existing system.

VI. CONCLUSION AND FUTURE SCOPE

A prototype-based model is constructed that is capable of insisting the tutor in detecting the Business at risk of dropping out and allowing the tutor to focus more on those Business. This aspect can greatly reduce dropout that is an issue associated with distance Marketing programs. However, it can be concluded that this methodology can be used to help Business and teachers to improve Business's performance; reduce failing ratio by taking appropriate steps at the right time to improve Business performance. In the future, we will use IoT with wearable devices to collect information regarding Business online. Thus, we will try to implement our prototype model on a real-time large dataset and check the validity of it in detecting potential dropout prone Business.

REFERENCES

[1]. Aleman, L., & Garza, D. E. L. A.. RESEARCH ANALYSIS ON MOOC COURSE DROPOUT AND. IEEE Access, **April, 3–14, 2016.**

[2]. Battin-Pearson, S., Newcomb, M. D., Abbott, R. D., Hill, K. G., Catalano, R. F., & Hawkins, J. D. (2000). Predictors of early high school dropout: A test of five theories. Journal of Marketingal Psychology, **92(3), 568–582, 2000.** https://doi.org/10.1037/0022-0663.92.3.568

[3]. Chen, J., Feng, J., Sun, X., Wu, N., Yang, Z., & Chen, S. (2019). MOOC Dropout Prediction Using a Hybrid Algorithm Based on Decision Tree and Extreme Learning Machine. IEEE Access, **2019.**

[4]. De Santos, K. J. O., Menezes, A. G., De Carvalho, A. B., & Montesco, C. A. E. (2019). Supervised learning in the context of Marketingal data mining to avoid university Businesss dropout. Proceedings - IEEE 19th International Conference on Advanced Learning Technologies, ICALT 2019, 207–208. https://doi.org/10.1109/ICALT.2019.00068

[5]. Feng, W., Tang, J., & Liu, T. X. (2015). Understanding Dropouts in MOOCs. IEEE Access.

[6]. Fortin, L., Lessard, A., & Marcotte, D. (2010). Comparison by gender of Businesss with behavior problems who dropped out of school. Procedia - Social and Behavioral Sciences, 2(2), 5530–5538. https://doi.org/10.1016/j.sbspro.2010.03.902

[7]. Fox, C. K., Barr-Anderson, D., Neumark-Sztainer, D., & Wall, M. (2010). Physical activity and sports team participation: Associations with academic outcomes in middle school and high school Businesss. Journal of School Health, 80(1), 31–37. https://doi.org/10.1111/j.1746-1561.2009.00454.x

[8]. Haraty, R. A., Dimishkieh, M., & Masud, M. (2014). An Enhanced k -means Clustering Algorithm for Pattern Discovery in Healthcare Data. International Journal of Distributed Sensor Networks, **1–18, 2015.**

[9]. Hasbun, T., Araya, A., & Villalon, J. (2016). Extracurricular activities as dropout prediction factors in higher Marketing using decision trees. Proceedings - IEEE 16th International Conference on Advanced Learning Technologies, ICALT 2016, 242–244. https://doi.org/10.1109/ICALT.2016.66

[10]. Kingdom, U. (2015). DROPOUT RATES OF MASSIVE OPEN ONLINE COURSES : BEHAVIOURAL PATTERNS MOOC Dropout and Completion : Existing Evaluations. IEEE Access.

[11]. Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. ACM International Conference Proceeding Series, 170–179. https://doi.org/10.1145/2460296.2460330

[12]. Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC Dropout over Weeks Using Machine Learning Methods. IEEE Access, 60–65.

[13]. Kuo, J. Y., Pan, C. W., & Lei, B. (2017). Using Stacked Denoising Autoencoder for the Business Dropout Prediction. Proceedings - 2017 IEEE International Symposium on Multimedia, ISM 2017, 2017-January, 483–488. https://doi.org/10.1109/ISM.2017.96

[14]. Liang, J., Yang, J., Wu, Y., Li, C., & Zheng, L. (2016). Big data application in Marketing: Dropout prediction in Marketing MOOCs. Proceedings - 2016 IEEE 2nd International Conference on Multimedia Big Data, BigMM 2016, 440–443. https://doi.org/10.1109/BigMM.2016.70

[15]. Limsathitwong, K., Tiwatthanont, K., & Yatsungnoen, T. (2018). Dropout prediction system to reduce discontinue study rate of information technology Businesss. Proceedings of 2018 5th International Conference on Business and Industrial Research: Smart Technology for Next Generation of Information, Engineering, Business, and Social Science, ICBIR 2018, 110–114. https://doi.org/10.1109/ICBIR.2018.8391176

[16]. Liu, F., Wang, L., Qian, Y., & Wu, Y.. Analysis of MOOCs Courses Dropout Rate Based on Businesss ' Studying Behaviors. IEEE Access, 83(Hss), **139–144, 2017.**

[17]. Manhães, L. M. B., Da Cruz, S. M. S., & Zimbrão, G. (2014). WAVE: An architecture for predicting dropout in undergraduate courses using EDM. Proceedings of the ACM Symposium on Applied Computing, 243–245. https://doi.org/10.1145/2554850.2555135

[18]. Márquez-Vera, C., Cano, A., Romero, C., & Ventura, S. (2013). Predicting Business failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. Applied Intelligence, 38(3), 315–330. https://doi.org/10.1007/s10489-012-0374-8

[19]. Mohan, A., Sun, H., Lederman, O., Full, K., & Pentland, A. (2018). Measurement and Feedback of Group Activity Using Wearables for Face-to-Face Collaborative Learning. 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT), 163–167. https://doi.org/10.1109/ICALT.2018.00046
[20]. Onah, D., & Sinclair, J. (2014). Dropout Rates of Massive Open Online Courses : Behavioural Patterns DROPOUT RATES OF MASSIVE OPEN ONLINE COURSES : BEHAVIOURAL PATTERNS. IEEE Access, July. https://doi.org/10.13140/RG.2.1.2402.0009
[21]. Osmanbegovic, E., Suljic, M. (2012). Data mining approach for predicting Business performance. Journal of Economics and Business, X(1), 3–12.
[22]. Pal, S. (2012). Mining Marketingal Data to Reduce Dropout Rates of Engineering Businesss. International Journal of Information Engineering and Electronic Business, 4(2), 1–7. https://doi.org/10.5815/ijieeb.2012.02.01
[23]. Romero, C., & Ventura, S. (2007). Marketingal data mining: A survey from 1995 to 2005. Expert Systems with Applications, 33(1), 135–146. https://doi.org/10.1016/j.eswa.2006.04.005
[24]. Tinto, V. (1975). Dropout from Higher Marketing: A Theoretical Synthesis of Recent Research. Review of Marketingal Research, 45(1), 89–125. https://doi.org/10.3102/00346543045001089
[25]. Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2019). Implementing AutoML in Marketingal Data Mining for Prediction Tasks. Applied Sciences, 10(1), 90. https://doi.org/10.3390/app10010090
[26]. Wang, W., & Graduate, I. (2015). Deep Model for Dropout Prediction in MOOCs. IEEE Access.
[27]. Wibisono, A., Jatmiko, W., Wisesa, H. A., Hardjono, B., & Mursanto, P. (2016). Knowledge-Based Systems Traffic big data prediction and visualization using Fast Incremental Model Trees-Drift Detection (FIMT-DD). Knowledge-Based Systems, 93, 33–46. https://doi.org/10.1016/j.knosys.2015.10.028
[28]. Zepke, N., Leach, L., & Prebble, T. (2006). Being learner-centered: One way to improve Business retention? Studies in Higher Marketing, 31(5), 587–600. https://doi.org/10.1080/03075070600923418
[29]. Zheng, X., Member, S., Chen, C., Hung, J., He, W., Hong, F., & Lin, Z. (2015). A Hybrid Trust-based Recommender System for Online Communities of Practice. IEEE Transactions on Learning Technologies, 1382(c), 1–13. https://doi.org/10.1109/TLT.2015.2419262
[30]. Hector Varga, Ruben Heradio, Jesus Chacon, Luis LA Torre, Gonzalo Faria (2019)"Automated Assessment and Monitoring Support for Competency-Based Courses" IEEE.
[31]. Cios, K.J., Pedrycz W., Swiniarski, R.W. & Kurgan, L.A. (2007), Data Mining: A Knowledge Discovery Approach, Springer, New York.
[32]. Klosgen, W. & Zytkow, J. (2002), Handbook of data mining and knowledge discovery oxford university press, New York.
[33]. Edin Osmanbegović , Mirza Suljić DATA MINING APPROACH FOR PREDICTING BUSINESS PERFORMANCE
[34]. Kalpesh P. Chaudhari1, Riya A. Sharma2, Shreya S. Jha3, Rajeshwari J. Bari4 Business Performance Prediction System using Data Mining Approach.
[35]. Hector Varga, Ruben Heradio, Jesus chacon, Luis LA Torre, Gonzalo Faria,"Automated Assessment and Monitoring Support for Competency-Based Courses" IEEE,2019.

AUTHORS PROFILE

Miss kajal Devi pursed Bachelor of Science from Swami Sarvanand college of Engineering and Technology,Dinanagr since 2018 and currently pursuing Master of Science from Swami Sarvanand Group of Institutes, DinaNagar Punjab since 2018. This is my first paper that I am publishing in international journals computer science and Engg**..**

Mrs Harjinder Kaur pursed Bachelor of Science from PTU campus, Kapurthala and Master of Science in CSE from BCET, Gurdaspur. She is currently working as Assistant Professor in Dept. of CSE, Swami Sarvanand Institute of Engineering and Technology, Dinanagar Punjab. She has 12 years of teaching Experience in the Field of Computer Science and Department

