International Journal of Computer Science Engineering & Technology
APPROVED BY
NATIONAL SCIENCE LIBRARY (NSL)
NATIONAL INSTITUTE OF SCIENCE COMMUNICATION AND INFORMATION RESOURCES (NISCAIR)
COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH (CSIR)– NEW DELHI INDIA .
ISSN :2455-9091

PREDICTION OF STUDENTS PERFORMANCE USING EDUCATIONAL DATA MINING AND LEARNING ANALYTICS (EDMLA)

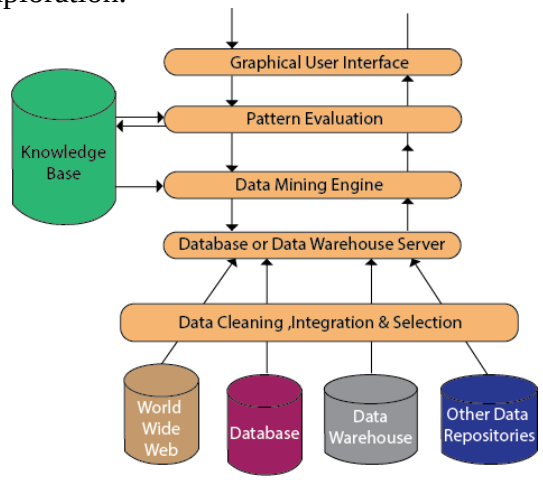
[1] Dr. S. P. Priyadharshini
[1] Assistant professor,
[1] Department of computer science
[1] Kamadhenu Arts and Science College,
[1] TamilNadu, India.

ABSTRACT-As we know that net is one of the most appropriate technology to share and collect information and the utilization of the net is rising every day. Computers are a repository of knowledge that contains a huge amount of information; web mining i.e. one of the operations of the data mining process that is utilized for finding information and knowledge present in the Web data. Taking the multifunctional Hardtop big data mining platform as an example, this article analyses the internal workflow of big data mining and its advantages as well as challenges, which can provide a reference for users to cognize and apply the big data mining. Its computing mode and type of storage is single, which makes it focus more on the data quality and accuracy. However, there are uncertain factors in the large and fast-moving data.

Keywords-[data mining, big data, KDD, clustering, Fuzz-Logic, EDMLA.]

1. INTRODUCTION

The development of Information Technology has produced an enormous number of databases and tremendous data in different regions. The examination in databases and information innovation has brought about a way to deal with store and control this valuable data for additional dynamics. Data mining is a cycle of extraction of valuable information and patterns from gigantic data. It is additionally called a knowledge revelation measure, knowledge mining from data, knowledge extraction, data or example analysis. Data mining is a sensible interaction that is utilized to look through a lot of data to discover helpful data. The objective of this procedure is to discover patterns that were beforehand obscure. When these patterns are discovered they can additionally be utilized to settle on specific choices for the development of their organizations or exploration.



Three steps involved are

- Exploration
- Pattern identification
- Deployment

Pattern Identification

Whenever data is investigated, refined and characterized for the specific variables the subsequent advance is to frame design ID. Distinguish and pick the patterns which make the best prediction.

Data Mining Algorithms and Techniques

Different algorithms and procedures like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithms, the Nearest Neighbor technique and so forth, are utilized for knowledge discovery from data bases.

Classification

Classification is the most generally applied information mining strategy, which utilizes a bunch of pre-ordered guides to foster a model that can classify the population of records on the loose. Misrepresentation detection and credit hazard applications are especially appropriate to this kind of examination. This methodology regularly utilizes decision trees or neural organization based classification algorithms. The information classification measure includes learning and classification. In Learning the preparation information are dissected by the classification algorithm. In classification test information are utilized to assess the precision of the classification rules. On the off chance that the precision is adequate, the standards can be applied to the new information tuple. For an extortion detection application, this would incorporate total records of both deceitful and legitimate exercises decided on a record-by-record premise. The classifier-preparing algorithm utilizes these pre-characterized guides to decide the arrangement of parameters needed for legitimate segregation. The algorithm then, at that point encodes these parameters into a model called a classifier.

Types of classification models

- Classification by decision tree induction
- Bayesian Classification Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

Clustering

Clustering can be said as the identification of comparative classes of items. By utilizing clustering methods it can additionally recognize thick and scanty areas in object space and can discover generally speaking distribution examples and relationships among information attributes. The classification approach can likewise be utilized for powerful methods for recognizing gatherings or classes of items however it turns out to be expensive so clustering can be utilized as preprocessing approach for attribute subset choice and classification. For instance, to frame a gathering of clients dependent on buying designs, to order qualities with comparative functionality.

Types of clustering methods

- Partitioning Methods
- Hierarchical Agglomerative (divisive) methods
- Density based methods
- Grid-based methods
- Model-based method

Predication

The regression procedure can be adjusted for prediction. Regression examination can be utilized to demonstrate the connection between at least one autonomous factor and ward factor. In information mining, free factors are ascribed definitely known and response factors are what we need to anticipate. Shockingly, some true issues are

not just predictions. For example, deals volumes, stock costs, and product failure rates are generally exceptionally hard to foresee on the grounds that they may rely upon complex connections of different indicator factors. Hence, more intricate methods (e.g., logistic regression, choice trees, or neural nets) might be important to conjecture future qualities. Similar model sorts can regularly be utilized for both regression and classification. For instance, the CART (Classification and Regression Trees) choice tree calculation can be utilized to assemble both classification trees (to order downright response factors) and regression trees (to figure consistent response factors). Neural networks also can make both classification and regression models.

Types of regression methods

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

Neural networks

A neural organization is a bunch of associated input/yield units and every connection has a weight present with it. During the learning phase, the organization learns by changing loads in order to have the option to foresee the right class marks of the info tuples. Neural networks have the remarkable capacity to get importance from confounded or loose information and can be utilized to extricate designs and identify patterns that are too intricate to possibly be seen by one or the other people or other computer techniques. These are appropriate for consistent esteemed data sources and yields. For instance, handwritten character revamping, for preparing a computer to articulate English content and some true business problems and have effectively been effectively applied in numerous enterprises. Neural networks are best at distinguishing examples or patterns in information and are appropriate for expectation or forecasting needs.

Types of neural networks

- Back Propagation

Tasks	Supervised	Unsupervised
Classification	Memory based reasoning, genetic algorithm, C&RT, link analysis, C5.0,ANN	Kohonen nets
Estimation	ANN, C&RT	
Segmentation	Market basket analysis, memory based reasoning, link analysis, rule induction	Cluster detection, K-means, generalized rule induction, APRIORI
Description	Rule induction, market basket analysis	Spatial visualization

Current Trends in Data Mining

Developing out of traditional statistics, data mining began as a free arrangement of apparatuses. To an ever increasing extent, visualization and database data mining are received. Conventional visualization methods are focused on the leaders who are information purchasers. Spatial visualization gives visual plots portraying individuals from the populace in their element space

Fuzzy Logic

Another data mining calculation being created is fuzzy logic, which can be applied to both guideline acceptance techniques and neural networks.

Genetic Algorithm

A genetic algorithm (GA) is an optimization algorithm developed by John Holland at the University of Michigan. It depends on the two essential rules that oversee the tremendous natural world, choice and variety. Genetic algorithm utilizes determination, hybrid, and change boundaries in developmental computation in arriving at the solution.

Applications of Data Mining

Data mining has been as of late found by the scholarly world yet was first put to full use by the Fortune 500, who have since profited hugely. Data mining was behind various effective market campaigns and quality assurance. Data mining was first executed for marketing outside advanced education. It unquestionably has equal ramifications and worth in advanced education. Marketing concerns the service region, enlistment, yearly mission, graduated class, and school picture. Joined with institutional examination, it ventures into understudy criticism and fulfillment, course accessibility, and personnel and staff employing. A college service region presently incorporates online course contributions, in this way bringing the idea of mining course data to another measurement. Data mining is rapidly turning into a mission-critical segment for the dynamic and information the executive's processes.

2. EXISTING METHODOLOGIES

1. ngda Li, Hongzhi Wang, and Jianzhong Li(2020),et.al proposed Mining Conditional Functional Dependency Rules on Big Data. Current Conditional Functional Dependency (CFD) discovery algorithms consistently need a decidedly ready preparing dataset. This condition makes them hard to apply on enormous and inferior quality datasets. To deal with the volume issue of large data, testing algorithms to acquire a little representative preparing set is created. Flaw open minded standard discovery is planned and compromise algorithms to resolve the bad quality issue of large data. Boundary determination procedure to guarantee the viability of CFD discovery algorithms is introduced. The technique that can find successful CFD rules on billion-tuple data inside a sensible period has given the result productively. For large data, rule discovery in data cleaning brings new difficulties. To take care of this issue, a novel CFD discovery technique is proposed for enormous data. For the volume highlight of enormous data, an examining calculation is intended to get run of the mill tests by filtering data just a single time. Then, at that point, on the example set, existing CFD discovery algorithms are adjusted to tolerate the shortcoming. By coordinating these adjusted strategies, a primer CFD set has been found. To build the nature of the found guideline set, a diagram based principle determination calculation is planned. Taking into account that a client may have various necessities for CFD discovery, a procedure to choose boundaries as per the prerequisites of clients is introduced. The exploratory outcomes showed that the proposed calculation is appropriate for huge data and outperforms existing algorithms.

Merits

1. Utilizing physical experiments and tests to get fundamental designing data for design can be costly.
2. CFD reenactments are generally reasonable, and costs are probably going to diminish as computers become all the more remarkable.

Demerits

1. Errors may happen because of basic stream models or worked on boundary conditions.
2. Possible uncertainties brought about by too little computing esteems per cell and consequently, accordingly, bringing about interpolation blunders.

2. Fan Jiang, Carson K. Leung, and Adam G. M. Pazdor(2016),et.al proposed Big Data Mining of Social Networks for Friend Recommendation.In the current time of big data, high volumes of important data can be handily gathered and created. Social networks are instances of creating wellsprings of this big data. Users in these social networks are regularly connected by some interdependency like friendship. As these big social networks continue to develop, there are situations in which an individual client needs to discover famous gatherings of friends so he can prescribe similar gatherings to different users. A big data analytic arrangement is introduced that utilizes the Map-Reduce model in mining these big social networks for discovering gatherings of regularly associated users for friend recommendations. The quantity of common friends may differ starting with one Facebook client then onto the next. Besides, a Facebook client can likewise buy in or follow the public postings of some other Facebook users without the need of adding them as friends. Moreover, the "like" button permits users to communicate their appreciation of content, for example, announcements, remarks, photographs, and advertisements. Assessment results show the productivity and common sense of our data analytic arrangement in mining big social networks, discovering famous users, and suggesting friends. BigPFM assists social with networking users to find gatherings of regularly associated users from big social networks by utilizing the Map Reduce model. Assessment results show the effectiveness and reasonableness of BigPFM in mining big social networks, discovering famous users, and suggesting friends. As continuous work, more comprehensive assessment and correlations are finished up with cutting edge algorithms.

Merits

1. **Cost Savings -** The implementation of Real-Time Analytics instruments might be costly, it will ultimately set aside a great deal of cash. A few instruments of it like Hadoop and Cloud-Based Analytics can carry cost benefits to business when large amounts of information are to be put away and these apparatuses additionally help in recognizing more efficient methods of working together.

Demerits

1. **Incompatible Tools** Hadoop is the most regularly utilized device for examination. Be that as it may, the standard version of Hadoop isn't at present ready to deal with ongoing analysis.

3.Ms.TismyDevasia ,Ms.Vinushree T P , Mr.Vinayak Hegde (2016),et.al proposed Prediction of Students Performance using Educational Data Mining. Data mining assumes a significant part in the business world and it assists the instructive foundation with anticipating and settles on decisions identified with the understudies' academic status. With advanced education, these days exiting understudies' has been expanding; it influences the understudies' career as well as the standing of the establishment. The current framework is a framework that keeps up with the understudy data as mathematical qualities and it simply stores and recover the data that it contains. So the framework has no intelligence to examine the data. An electronic application that utilizes the Naive Bayesian mining procedure is proposed for the extraction of helpful data. Result demonstrates that the Naive Bayesian algorithm gives more exactness over different techniques like Regression, Decision Tree, Neural networks and so forth, for comparison and prediction. The framework targets expanding the achievement graph of understudies utilizing Naive Bayesian and the framework which keeps up with all understudy affirmation subtleties, course subtleties, subject subtleties, understudy marks subtleties, participation subtleties, and so on it accepts understudy's scholarly history as info and gives understudies' forthcoming exhibitions based on the semester. The order is utilized in

understudy data to foresee the understudies' division is introduced on the reason of past data. As there are a few methodologies that region units utilized for information grouping, the Naive hypothesis is utilized here. Data like gathering activity, class test, workshop and task marks were gathered from the understudies' past data, to anticipate the presentation at the highest point of the semester. This examination can work with the understudies and the teachers to help the understudies, all things considered, to perform well. This investigation assists with spotting out those understudies who require unique consideration; limit the disappointment proportion and making a satisfactory move for the forthcoming semester assessment.

Merits

1. Choice Tree is utilized to take care of both classification and regression issues.
2. Less data cleaning is needed for the processing.

Demerits

They are shaky, meaning that a little change in the information can prompt an enormous change in the structure of the optimal decision tree.

4.Matsumoto, T., Sunayama, W., Hatanaka, Y., &Ogohara, K. (2017), et.al proposed Data Analysis Support by Combining Data Mining and Text Mining.Data mining and text mining techniques have been every now and again utilized for investigating polls and survey data. Data mining techniques like association analysis and bunch analysis are utilized for promoting analysis since they can find connections and rules covering up in huge numerical data. Then again, text mining techniques like catchphrases extraction and assessment extraction are utilized for poll or survey text analysis, in light of the fact that those can uphold us to examine purchasers' opinions in text data. Be that as it may, data mining tools and text mining tools can't be utilized in a solitary climate. Subsequently, data that has both numerical and text data isn't very much broke down in light of the fact that the numerical part and text part can't be associated for understanding. In this paper, a mining system that can treat both numerical and text data is proposed. It can repeat data therapist and data analysis with both numerical and text analysis tools in the remarkable system. In light of experimental outcomes, the proposed framework was adequately utilized for data analysis for audit texts. A framework that can treat both numerical and text data for data analysis is proposed. In view of the experimental outcomes, clients of the proposed framework might have made substantial thoughts. The things and those qualities are imported. Since TETDM can't treat such info data presently, just the text some portion of the data is given to TETDM as standard information. Then, at that point, the numerical piece of the data is ready in CSV organization and given to the data mining apparatus straightforwardly.

Merits

1.Data analytics helps an organization make better decisions.

Demerits

1. Low quality data and concerns.

5. Wang, Z., WU, B., BAI, D., & QIN, J. (2018),et.al proposed Distributed Big Data Mining Platform for Smart Grid.With the fast development of information technology and the web, a wide range of industry data detonated causing it difficult to investigate and mine helpful information from huge data. Traditional analysis framework has bottlenecks of execution and scalability in the huge data processing. The innovative work of novel and productive huge data analysis and mining platforms has become the focal point, all things considered. Alongside the development of the brilliant grid, power data with qualities of the force business needs more designated and effective data mining analysis. A circulated large data mining platform

dependent on dispersed framework foundation, for example, Hadoop and Spark is introduced. The platform creates and executes an assortment of quick exceptionally parallel mining algorithms by Spark and Tensor stream, including AI, statistics and analysis, profound learning, etc. Utilizing the OSGI technology to construct a low coupling segment model, the platform works on the reusability of the segment calculation, presents the work process motor and easy to understand GUI, decreases the intricacy of the client activities, support client characterized data mining undertakings. For the attributes of savvy grid enormous data, the platform creates and works on the many calculation parts about data processing and analysis. Also, planning an adaptable algorithms library and the part library enormously works on the scalability of large data mining platforms and processing savvy grid data. Our platform has effectively been dispatched in a state grid Company, fulfilling the interest of different brilliant grid data analysis organizations. For the interest of force large data business analysis and mining alongside the development of keen grid, this paper plans and fosters a conveyed parallel data digging platform for power huge data. The parallel platform embraces a profoundly reusable circulated system and plans a versatile parallel calculation library. It incorporates almost 100 well-running exhibitions and exceptionally parallelized algorithms, which are mostly better than existing open-source calculation libraries MLib and Hive devices. An assortment of profound neural organizations, AI, factual analysis and different classifications of parallel general mining analysis algorithms and exceptional algorithms for power data analysis business needs are associated with the library. What's more, the graphical interactive interface Studio given by the parallel platform shows the data stream bearing and reliance relationship with the work process chart. It additionally outwardly shows the framework analysis measure, the current execution status and the analysis result.

Merits

1. Easy to build architecture.

Demerits

1. Local data sources have restricted availability due to privacy.

3. PROPOSED METHODOLOGY

Prediction of Students Performance using Educational Data mining and learning analytics in higher education

Analyzing the huge amount of data to shape summed up valuable information is a dreary errand for humanity. Data Mining is the region that dissects huge storehouses of data to extract important or helpful information. PCs can handle any sort of data like numbers, writings, pictures and realities. This undertaking plays out the examination dependent on the examples, affiliation, relations among every one of these data to get the information. The prediction with high exactness in understudies' presentation is useful as it helps in distinguishing the understudies with low scholarly accomplishments at the beginning phase of scholastics. In colleges, understudy maintenance is identified with scholarly execution and the enlistment framework. The means to help the low scholastic entertainers with better education are:

a) Generation of the data wellspring of predictive variables.
b) Identification of different features or factors which influences the performance of understudy's picking up during the scholarly vocation.
c) Construction of an expectation model with the assistance of classification data mining procedures based on predictive variables which are promptly distinguished.
d) Approval of the model which is created for Universities with understudies' performance.

The general issue in institutions of advanced education across the world is scholarly achievement and the maintenance of understudies. This is an especially pressing

matter with regards to enlarging support for under-represented understudy gatherings, increasing understudy variety and educational quality affirmation and responsibility processes. As institutions of advanced education collect increasingly more data about their understudies and as understudies' record databases have grown more unpredictable and available. This new period of utilizing data to further develop understudy achievement, smooth out processes, and all the more effectively use assets. When the data is broke down it guarantees better understudy placement processes; more precise enrolment conjectures, and early admonition frameworks that distinguish and help understudies in danger of failing or exiting.

With all being said, however frameworks and structures of data analytics have been utilized and applied in numerous data frameworks (IS) research considers, the basic causal system stays indefinable. Using plan science research procedure, the examination coordinates different analytical frameworks to foster a more exhaustive large data engineering for learning analytics. A review was directed to take a gander at comparable and differentiating assessments from various individuals with various perspectives and interests. This was done to solidify the exactness, legitimacy and reliability of the findings.

Data mining is the knowledge discovery process from a gigantic data volume. The mechanism works in a huge dataset where the understudy execution in the end semester assessment is assessed. Data preparation Student related data were gathered from the school Amrita School of Arts and Sciences, Mysore on the inspecting technique for the software engineering division from the meeting 2013 to 2016. In this progression, data put away in various tables were joined into a solitary set. B. Data selection and change in this progression just those fields were chosen that were needed for the data mining process. The understudy register number, tenth, twelfth, degree marks in every semester astute, task, sex, parent's education, pay was taken as the qualities for predictions. Naïve Bayes Algorithm.

Step 1: Scan the student Data set
Step 2: Calculate the probability of each attribute value. [n, n_c,m, p]
Step 3: Apply the formulae
P (attribute value (ai) subject value (vj))=(n_c+mp)(n+m)
Where:
N= the number of training data item for which v=vj
Nc= number of examples for which v=vj and a=ai
P= a priori estimate for P (ai,vj)
M= the parallel size of the sample
Step 4: Multiply the probabilities by p
Step 5: Compare the values and classify the attribute values to one of the predefined set of class

EDUCATIONAL DATA MINING

The Educational Data Mining people group site characterizes educational data mining as "an arising discipline, worried about developing methods for investigating the interesting kinds of data that come from educational settings, and utilizing those methods to more readily get understudies, and the settings which they learn in". As indicated by Romero educational data mining is worried about developing, exploring, and applying automated methods to distinguish designs in huge assortments of educational data – designs that would somehow be hard or difficult to break down because of the enormous volume of data they exist inside.

Learning analytics and educational data mining are very comparative as far as data type, data sources, examination space, cycles, and destinations and objectives. Nonetheless, learning analytics and EDM appears to contrast in the procedures they apply for data investigation. EDM fundamentally centers on the utilization of common data mining methods including bunching, grouping, and affiliation rule

mining. These data mining procedures support educators and understudies in examining the learning interaction.

Learning analytics

Learning analytics has gotten expanded attention, for the most part since it offers numerous advantages to institutions of advanced education, including expanding understudy achievement, further developing understudy maintenance and furthermore offers responsibility. Learning analytics has been characterized in a few endeavors in literature and the web. The accompanying segment will examine the various definitions of learning analytics and furthermore the learning analytics apparatuses accessible in the literature. As indicated by Ackerson (2006), learning analytics tries to profit by the displaying capacity of analytics: to predict conduct, follow up on predictions, and afterward feed those outcomes back into the process to further develop prediction over the long haul.

Siemens claims that learning analytics is the utilization of shrewd data, student delivered data, and investigation models to find information and social connections, and to predict and exhort on learning, while, Agudo-Peregrina learning analytics is the examination of electronic data which permits educators, course architects and chairmen of virtual learning conditions to look for unseen examples and basic information in learning processes. The primary point of learning analytics is to further develop learning results and the general learning process in electronic learning virtual study halls and PC upheld instruction. The above definitions vary in subtleties however they all underline the investigation of instructive data to find examples and afterward change the examples into noteworthy insight. This current investigation looks to find the elements of analytics in advanced education to foster a learning analytics framework and engineering. With the end goal of this exploration study, the term learning analytics will be utilized all through the examination.

Data Storage and Management System: The data storage framework comprises of a major database management framework with all capabilities like buffering and continuous inquiry streamlining. This stage is likewise answerable for data pre-handling and data cleaning. At the end of the day, the principle usefulness of the data storage and management framework is to measure and change over raw data into a structure that can be productively prepared by the analytics engine.

Data analytics system: This is the core of the whole system. The data analytics system incorporates keen processing algorithms for effectively extracting significant and important data from the raw streaming or static data.

Data Visualization: The objective of the data visualization system is to make a visual depiction of the analysis results so that appropriate decision can be taken instantly.

Action: The objective of the learning analytics framework is to give learners, managers and speakers with alarms and alerts, directing and systematic improvement, educational program plan and instructing intervention.

Data gathering devices: This framework incorporates all articles and gadgets that are answerable for gathering crude data at each phase of advanced education measures. Different data assortment gadgets like students' cards, social networks, learning management systems, sensors, and understudy data systems will fill in as the data source. The organized and unstructured data created from individual students are IEEE African 2017 Proceedings 954 disregarded the data management systems to examination.

4. EXPERIMENT RESULT

Accuracy

Values	CFD	Big Data Mining	Proposed EDMLA
1	79.4	82.4	84.3
2	87.1	89.3	91.9
3	94.7	95.6	97.6
4	97.3	98.5	99.2

Table 1. Comparison table of Accuracy

Comparison table 1 of Accuracy of Value explains the different values of existing algorithms C4 CFD, Big Data Mining and proposed improved EDMLA Algorithm. While comparing the Existing algorithm, the proposed improved EWQRF provides better results. The existing algorithm (CFD, Big Data Mining) values start from 79.4 to 97.3, 82.4 to 98.5 and the proposed EDMLA Algorithm starts from 84.3 to 99.2, provides great results.

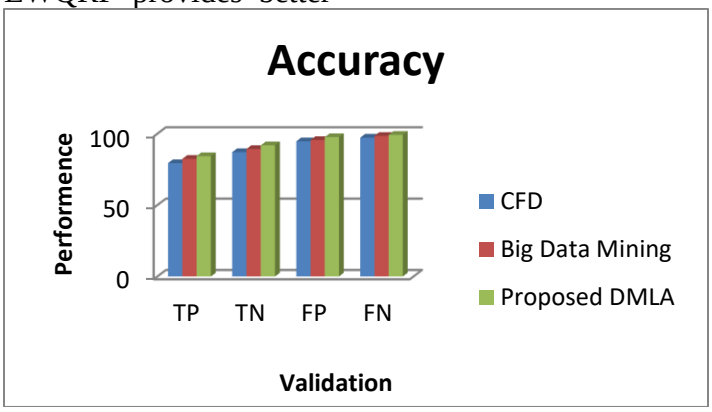


Figure 1 Comparison chart of Accuracy

Figure 1 shows the comparison chart of Specificity values demonstrates the existing systems CFD, Big Data Mining and proposed EDMLA algorithm. The proposed algorithm is better than the existing algorithm. The X-axis denotes the validation values and Y-axis denotes the performance in %. The existing algorithm (CFD, Big Data Mining) values start from 79.4 to 97.3, 82.4 to 98.5 and the proposed EDMLA Algorithm starts from 84.3 to 99.2, provides great results.

F1-Score

Values	CFD	Big Data Mining	Proposed EDMLA
1	0.78	0.81	0.85
2	0.76	0.79	0.83
3	0.81	0.83	0.86
4	0.85	0.88	0.91

Table 2. Comparison table ofF1-Score

The Comparison table 2 of F1-score of Value explains the different values of existing algorithms C4 CFD, Big Data Mining and proposed improved EDMLA Algorithm. While comparing the Existing algorithm, the proposed improved EWQRF provides better results. The

existing algorithm (CFD, Big Data Mining) values start from 0.78 to 0.85, 0.79 to 0.88 and the proposed EDMLA Algorithm starts from 0.83 to 0.91, provides great results.

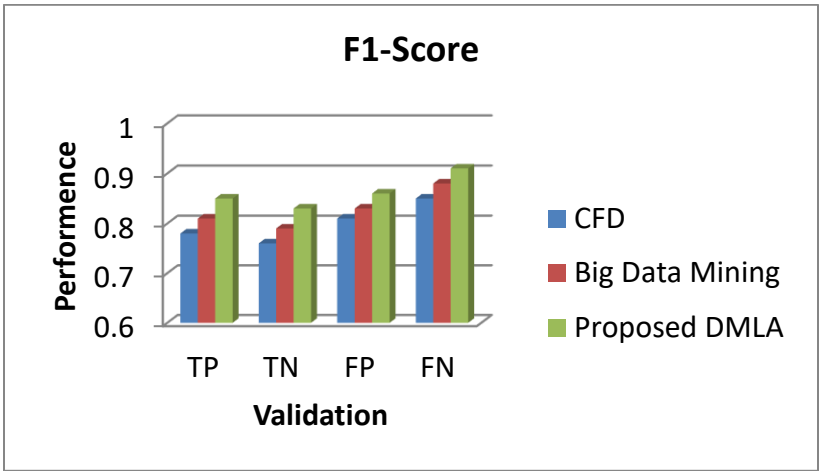


Figure 2 Comparison chart of F1-Score

Figure 2 shows the comparison chart of Specificity values demonstrates the existingsystems CFD, Big Data Mining and proposed EDMLAalgorithm. The proposed algorithm is better than the existing algorithm. X axis denotes the validation values and Y axis denotes the performance in %. The existing algorithm (CFD, Big Data Mining) values start from 0.78 to 0.85, 0.79 to 0.88 and proposed EDMLAAlgorithm starts from 0.83 to 0.91, provides the great results.

CONCLUSION

For big data, rule discovery in data cleaning brings new challenges. To solve this problem, In existing CFD discovery algorithms to tolerate the fault. By integrating these modified methods, we discovered a preliminary CFD set. New knowledge and practical tools have been come into existence to help users and webmasters; web mining is one among them. EDMLA algorithm provides the best result.

REFERENCE

[1]. T. Devasia, Vinushree T P and V. Hegde, "Prediction of students performance using Educational Data Mining," 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), 2016, pp. 91-95, doi: 10.1109/SAPIENCE.2016.7684167.

[2]. F. Matsebula and E. Mnkandla, big data architecture for learning analytics in higher education," 2017 IEEE AFRICON, 2017, pp. 951-956, doi: 10.1109/AFRCON.2017.8095610.

[3]. Z. Wang, B. WU, D. BAI and J. QIN, "Distributed Big Data Mining Platform for Smart Grid," 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 2345-2354, doi: 10.1109/BigData.2018.8622163.

[4]. T. Matsumoto, W. Sunayama, Y. Hatanaka and K. Ogohara, "Data Analysis Support by Combining Data Mining and Text Mining," 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), 2017, pp. 313-318, doi: 10.1109/IIAI-AAI.2017.165.

[5]. T. Devasia, Vinushree T P and V. Hegde, "Prediction of students performance using Educational Data Mining," 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), 2016, pp. 91-95, doi: 10.1109/SAPIENCE.2016.7684167.

[6]. M. Li, H. Wang and J. Li, "Mining conditional functional dependency rules on big data," in Big Data Mining and Analytics, vol. 3, no. 1, pp. 68-84, March 2020, doi: 10.26599/BDMA.2019.9020019.

[7]. Y. Cong, "Research on Data Association Rules Mining Method Based on Improved

Apriori Algorithm," 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), 2020, pp. 373-376, doi: 10.1109/ICBASE51474.2020.00085.
[8]. R. K. Shukla, P. Sharma, N. Samaiya and M. Kherajani, "WEB USAGE MINING-A Study of Web data pattern detecting methodologies and its applications in Data Mining," 2nd International Conference on Data, Engineering and Applications (IDEA), 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170690.