ISSN NO: 2279-543X

Effective Performance of Integrated K-Family Clusters and Decision Tree Structures on Quality Assessment of Institutional Data

Dr.Prakashkumar.S[1] , Mr.R.Thiagarajan[2] , Mr.E.M.Murugan[3]

[1]Assistant Professor, Department of Computer Science, St. Joseph's College (Arts & Science) , Kovur, Near Porur, Chennai-600 128.

[2]Assistant Professor Cum Head of the Department, Department of Computer Science, St. Joseph's College(Arts & Science), Kovur, Near Porur,Chennai-600 128.

[3]Assistant Professor Cum Head of the Department, Department of visual communication, St. Joseph's College(Arts & Science), Kovur, Near Porur, Chennai-600 128.

***Email(s):** [1]drsp1974@gmail.com, [2]**thiagarajan.nvr@gmail.com**, [3]**emm8016@gmail.com**

Abstract

The paper titled "Effective Performance of Integrated K-family Clusters and Decision Tree Structures on Quality Assessment of Institutional Data" focuses primarily on the demand for data mining in educational institutions, which is constantly going up due to the development of information systems in this age of globalization. The quality of data in educational institutions has not only immense influence but also plays a key role in science and technology. Analysis of the reasons of low data quality on system engineering theory results in the establishment of a new method known as data mining. Any data quality problem can be solved with the help of a Meta synthesis method, software designing management and data mining testing. In application, a fair practicality brings about data mining problems in low quality data and so any institution may take any decision at random. Data mining absolutely needs high quality data because no sufficient method is available to get quality data in the institutional zone and, therefore, data mining and fuzzy based approach can be applied to find relationships between the attributes of student and staff, and get firm conclusion. The growth of educational and institutional systems of course depends upon the quality of services under critical circumstances. The faculty profile, student performance and infrastructure requirements are institutional performance parameters. Upon the decision-making procedures such as planning, counseling, assessment evaluation and conformation, the highest level of quality can be attained. Above all, this can be achieved and utilized by means of managerial and succinct decisions based upon implicit knowledge. The knowledge that remains hidden in the educational data set can be extracted from data mining technology.

***Keywords**: Fuzzy K-Family Cluster, decision tree, quality assessment, meta synthesis*

1. Introduction

A predictive model analysis in data mining is a process by which a model is generated or selected to predict the best possibility of an outcome. In certain scenarios, the model is selected on the basis of detection theory to guess the probability of an outcome given with a group of input data. For example, given institution's data, ranging from student, faculty and infrastructure which determine how likely a student selects that institution.

1.1 Predictive Analysis In Data Mining

A predictive analysis in data mining, such as classification, starts from a given classification of the data items. From that it derives a situation based on the properties of the data objects that permit to predict the membership to a specific class. For example, the prediction could be based on a partitioning of the attribute values along with each dimension. Predictive data mining comprises of combining the predicted classifications from different models, or from similar type of models for the purpose of different learning data. At the same time, predictive data mining is also used to tackle the inherent instability of outcome when applying composite models to compare small data sets. Suppose, if the task of data mining is to construct a model for classification of predictive types of data, and the data set that is involved in mining is relatively small and then the dataset can be sub-sampled repeatedly from the data set and a tree classifier to successive samples.

1.2 Objectives

(i) To introduce an approach for quality assessment of institutional data.(ii) To use K-Family cluster validation for clustering institutional information(iii) To propose a decision tree model from the clustered institutional data for improving the quality of obtained data using student, faculty and infrastructure datasets(iv) To propose a framework to assess the quality of data obtained using K-Family Cluster and decision tree.

1.3 Contribution

(i) Design and development of K-Family Cluster Validation to identify soft clusters, where a specific point belongs to more than one cluster with probability occurrences when compared to the conventional K-Means which discovers hard clusters.(ii) Improve the cluster validation by updating the centroids in incremental manner when compared to K-Means, where the centroids are updated at the end of the assignment.(iii) Construct decision tree model for predictive mining evaluate the quality of the institution with the hidden data attributes of the student profile, grade, faculty profile, curriculum, etc., extracted from the educational data bases.(iv) To develop a cluster based decisive data mining to measure the quality of educational institution.

2. Related work

Related works on clustering techniques and its model for institutional quality assessment using data mining. In addition articles related to decision tree structure in predictive mining for quality assessment with clusters and decision trees were studied to identify the issues. Clusters for assessment of institutional quality using institutional data were also interpreted from the recent journals to evaluate the problem statement of the proposal.

3. Proposed work

The proposed work of efficient cluster validation scheme, evaluates the quality of clusters formed in the process of mining hidden data of education data sets. The validity is verified with cluster object cohesiveness and its precision value. The cluster validation is helpful in evaluating the quality index of the data set along with a decision tree algorithm to present the usability of K-Family cluster validation efficiently. The decision tree obtained from clusters show how close are the results to the real partitions of the dataset (assuming that the data set presents clustering tendency). The proposed work defined the validity index, for assessing the results of clustering using K-Family of clusters being generated A predictive based decision tree is presented from cluster data of educational system metrics such as student performance, faculty skill set and institutional infrastructure requirements. Student assessment is a sub-process on the real educational system. By gaining a deep understanding of student enrollment patterns and tendencies in this course, it is enabled to predict which students are less likely to perform well in that specific course, or those who are less likely to be successful in it. The K-Family cluster

validation provides an effective analytical predictive model framework using data mining clusters in evaluating the quality of the educational institution in terms of student performance, curriculum, faculty skill sets, facilitation provided and life a of ray of success. The main idea is organized into a model proposed to represent how this technology is used in the educational system to improve the efficiency and effectiveness of the traditional processes. The model is also presented as a guideline for educational system to improve their decision-making processes. The educational data mining clustering technologies allow educators to study how students learn (descriptive studies) and which learning strategies are the most effective (causal/predictive studies). Since educational systems are capable of collecting vast amounts of student profile data, data mining and knowledge discovery techniques can be applied to find interesting relationships between attributes of students, assessments, and the solution strategies adopted by students. The major focus of this work is to introduce an approach for predicting performance quality of education, use clustering ensembles to build an optimal framework for clustering assessment resources and propose a framework for the discovery of interesting decision rules within an educational system. The next work introducing predictive based decision tree using different institutional metrics with benchmark dataset like University dataset, Teaching Assistant Evaluation dataset and non-benchmark dataset like student, faculty and infrastructure using the algorithmic representation.

3.1 K- Family Cluster Validations On Institutional Data

The proposed work has efficiently presented a K-family Clustering algorithm in the formation of student, faculty and infrastructural clusters based on the performance, skill set and facilitation availability correspondingly. The clustering is completed using fuzzy K-means clustering algorithm. The stability of the system also improves since the data is partitioned based on their characteristics. The validation metrics is also better using the proposed approach. The proposed approach shows better performance over the existing approaches. The next work direction for using a decision tree approach for classification of institutional dataset.

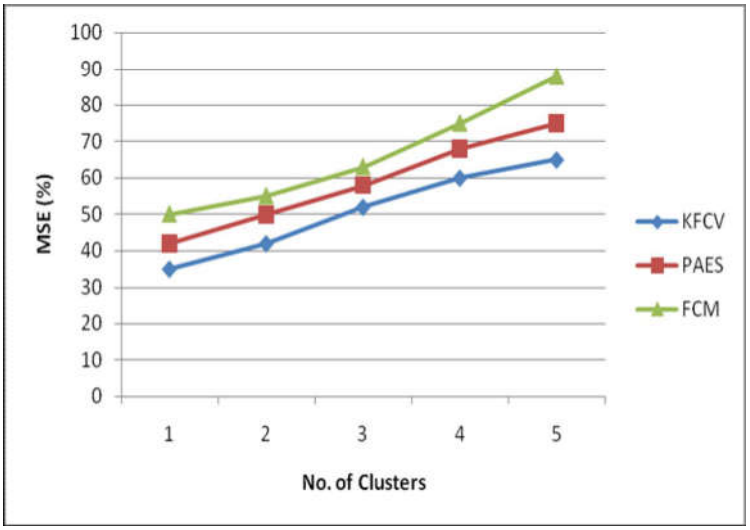
3.2 Decision Tree Model for Predictive Mining Using Institution Data

Clustering and analysis are completed with three initial parameters, student based values, faculty based values and infrastructure based values that affect the quality assessment. The quality evaluation representation also handles the infrastructural acquirements. The system has instructed course coordinators to fully access the students' educational records based on their characteristics and performance in their studies. The work is strong in terms of data recorded as it is handled along with the hash key. With this access, they are capable to evaluate the problems presented in the course after the students have used the educational materials, through some statistical reports. It has also provided a rapid evaluation of students' submissions for every problem in a course. The trees constructed for each attribute contain information that clearly describes the outcome of that particular attribute. It can be observed that the dataset used forces a lot of rules that can be matched using decision trees. The results are better as it uses the entropy integrated decision tree algorithm. The next work quality assessment of the educational institution via sub sampling, and adaptive techniques using K-Family Cluster Validation (KFCV)
.

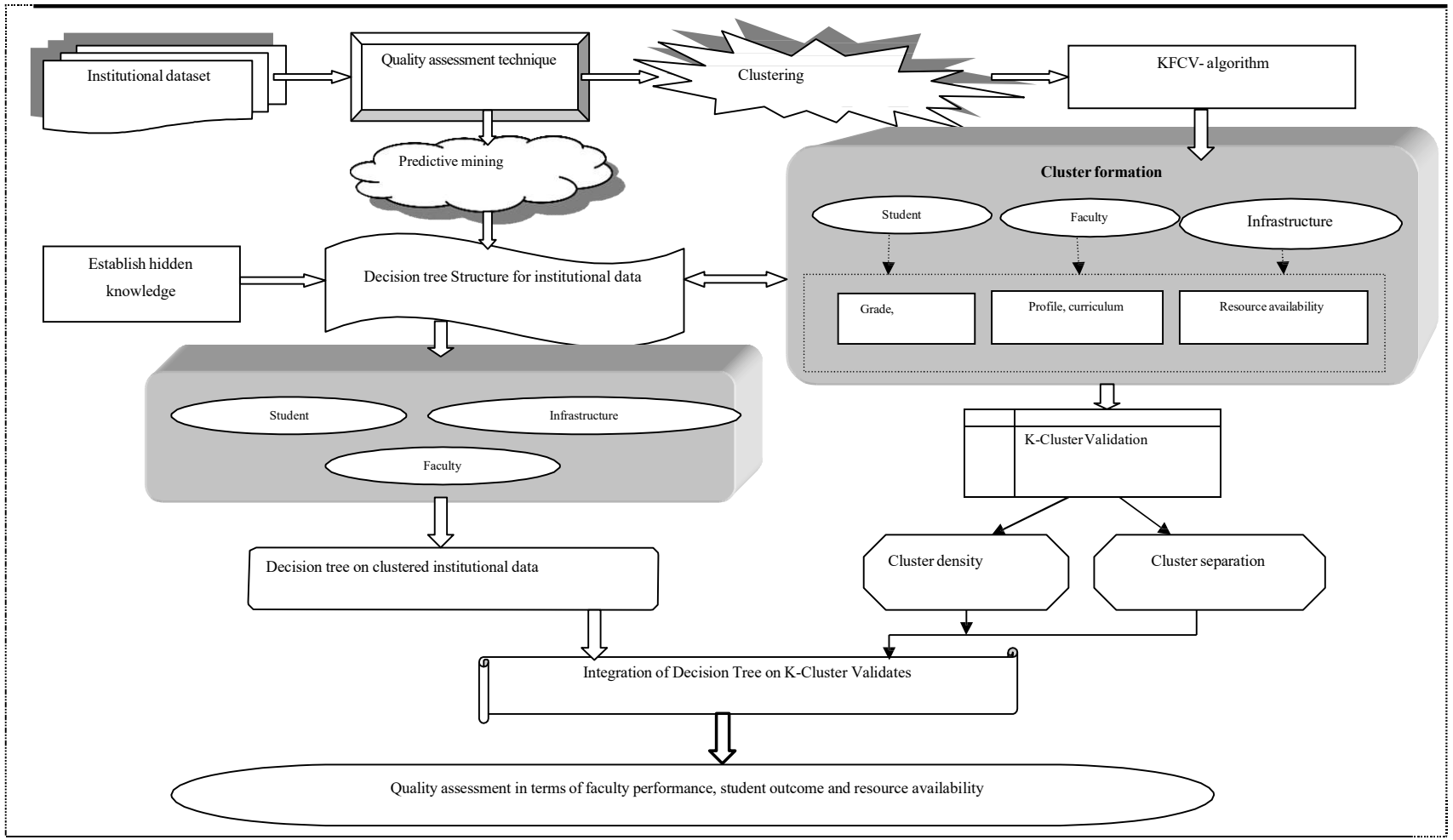
ISSN NO: 2279-543X

3.3 Quality Assessment Using K-Family clustering and Decision Tree In Institutional data (Qa-Kfdt)

The quality assessment technique presented integrates the Fuzzy K-Means Clustering algorithm with Entropy Integrated Decision Tree Algorithm to provide an efficient quality analysis of Institutional data sets. A capacity matrix is used for assessment. The cluster validation scheme evaluates the quality of clusters formed in the process of mining hidden data of education data sets. The validity is verified with cluster object cohesiveness and its precision value. The cluster validation is evaluated with quality index of the institutional data set obtained from decision tree algorithm. The decision tree is derived to assess institutional quality by utilizing intrinsic attributes of the institutional data. The validity index is used for assessing the results. The index is optimized for benchmark and non-benchmark data sets that include compact and well-separated clusters. The compactness of the data set is measured by the intra-cluster density. The next work presents an overview on the results and discussion of K-family cluster for institutional data and measures the reliability of data using K-Family Cluster and Decision Tree model for Predictive Mining.



ISSN NO: 2279-543X



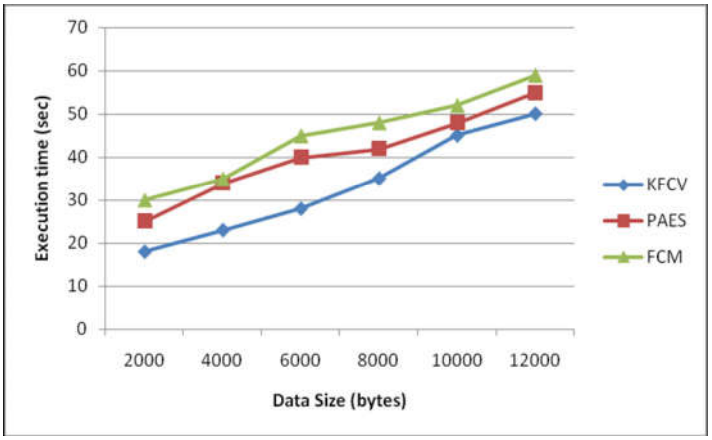
The entire contribution for overall framework architectural diagram for integrated k-family clusters and decision tree structures on quality assessment of institutional data

ISSN NO: 2279-543X

4.Results and discussion

The quality assessment for two benchmark (university evaluation assistant) and three non-benchmark (student, faculty and infrastructure) dataset were validated based on the data collected from UCI repository and Master of Computer Applications Department, Erode Sengunthar Engineering College, Thudupathi, Erode, Tamil Nadu, South India using K-Family Clustering and Decision Tree (QA-KFDT). The results obtained using the benchmark and non-benchmark datasets can be used as generalization by other institutions to improve the quality of student, staff andin a way finally improving the institutional performance by providing quality students for the uplift of the society. On the basis of the results obtained, the performance of the students was improved by providing updated course materials, improving lab facilities for the students. On the other hand, by measuring the number of journals published by the faculty, the managem entprovided training and organized workshops to both the faculty and student to improve the success ratio in the Master of Computer Applications Department, Erode Sengunthar Engineering College, Thudupathi, Erode, Tamil Nadu, South India. Several attributes including university, number-of students, and student-faculty ratio and so on are analyzed with the bench marked dataset. Again with the benchmark teaching assistant evaluation dataset the overall score is evaluated over three regular semesters and two summer semesters of teaching assistant assignments are analyzed. With the help of inferences provided in this work, the results obtained can be used by other institutions to improve the performance of student/faculty which results in the benefits for the institution and to the society as a whole.

DataSize (bytes)	Execution time (sec)		
	KFCV	PAES	FCM
200	18	25	30
400	23	34	35
600	28	40	45
800	35	42	48
1000	45	48	52



ISSN NO: 2279-543X

5. Conclusion

The problem of quality assessment to provide an efficient quality analysis for benchmark and non-benchmark dataset is discussed by integrating the K-Family Cluster with decision tree model. Especially, this paper work has addressed the issue by clustering and decision tree based on university, teaching evaluation assistant dataset. In particular, we have proposed three algorithms have been proposed, Fuzzy K-Means cluster algorithm, Entropy Integrated Decision Tree Algorithm, Integrated Decision Tree algorithm, for clustering, predictive mining and quality assessment, respectively. Fuzzy K-Means cluster algorithm extends the traditional k-means algorithm which clusters the K-Family cluster comprising of three institutional datasets namely, student, staff and infrastructure. Next, Entropy Integrated Decision Tree algorithm uses entropy for decision tree approach that preserves the information. Finally, Integrated Decision Tree algorithm is applied over the clustered data obtained for predictive mining using decision support tree mechanism to measure the quality of the educational institution with the hidden data attributes presented and analyzed using the bench mark and non-benchmark dataset.

The quality evaluation is made in terms of student performance, curriculum, faculty skill sets, facilitation provided and success. In our context of educational system quality, contrast rules help to identify attributes characterizing patterns of performance disparity between various groups of students, faculty and infrastructural facilities. Examining these contrasts can improve the educational systems for teachers, students and the management allowing for more accurate assessment and more effective evaluation of the educational quality. The educational data mining clustering technologies allow educators to derive how students learn and which learning strategies are the most effective. Since educational systems are capable of collecting vast amounts of student profile data, data mining and knowledge discovery techniques can be applied to find interesting relationships between attributes of students, assessments, and the solution strategies adopted by students. The major focus of this work is that it introduces an approach for predicting performance quality of education, use clustering ensembles and also proposes a framework for the discovery of interesting decision rules within an educational system. The proposed work of K-Family Cluster Validation scheme, evaluates the quality of clusters formed in the process of mining hidden data of educational data sets. The validity is verified with cluster object cohesiveness and its precision value that is helpful in evaluating the quality index of the data set along with Entropy Integrated Decision Tree algorithm. The decision tree obtained from clusters measures the effectiveness of the validity index, for assessing the results of clustering fuzzy k-means. The compactness of the data set is measured by the intra-cluster density whereas the separation is performed by the density between clusters. The decision tree is derived to assess institutional quality by utilizing intrinsic attributes of the institutional data.

The discovery of decisive rule by mining interesting contrast rules, are sets of conjunctive rules describing interesting characteristics of different institutional parameters. In the context of educational system quality, contrast rules help to identify attributes characterizing patterns of performance is parity between various groups of students, faculty and infrastructural facilitations. Examining these contrasts improve the educational system for teachers, students and management allowing for more accurate assessment and more effective evaluation of the educational quality. The proposed framework model is used to analyze the existing works, and identify existing gaps and recommend the scope for further works. The paperless may use the model to identify the existing area of paper in the field of data mining in other institutions and university to improve the quality of student and staff. The experimental results are evaluated using the Weka. The experimental result shows that integrated fuzzy k-means with decision tree shows better quality assessment compared to traditional k-family clustering techniques. From the experimental results, the execution time calculated for cluster object are almost

linear which is presented graphically for fuzzy k means on institutional quality metrics such as student and faculty performance. When compared to the existing processes for clustering, the execution time of the proposed fuzzy k-means algorithm for enhancing the educational institutions is linear. The existing work does not clearly describe each faculty and each student grade and performance. But, the proposed work clearly describes each student's performance and the faculty quality efficiently. The proposed work has provided an easy way to evaluate the educational institutions quality. When compared to existing model, the proposed work for enhancing the educational institutions quality is 80%effective in terms of analyzing the students' performance, faculty quality and so on.

References

1. Amershi, S. and Conati, C. "Automatic Recognition of Learner Groupsin Exploratory Learning Environments" in Proc.of the 8th InternationalConference on Intelligent Tutoring Systems, Vol.4053,pp.463-472,2006.

2. Amirzadeh, V., Mashinchi, M. and Yaghoobi, M.A. "Construction of Control Charts Using Fuzzy Multinomial Quality", Journal of Mathematics and Statistics, Vol.4, No.1, pp.26-31, 2008.

3. Amjed, S. "Quality Assessment of ECG Compression TechniquesUsing a Wavelet-Based Diagnostic Measure", IEEE Transactions onInformation Technology in Biomedicine, Vol.10, No.1, pp.182-191,2006.

4. Anandhavalli, M., Ghose, M.K. and Gauthaman, K. "Association Rule Mining in Genomics", International Journal of Computer Theory andEngineering, Vol.2, No.2, pp.1793-8201, 2010.

5. Andreas Geyer-Schuluz and Micheal Hahsler "Comparing tworecommender Algorithms with the help of recommendation by peers", Springer, Lecture Vol.2703, pp. 137-158,2003.2003.

6. Anthony Tung, K.H., Hongjun Lu, Jiawei Han and Ling Feng"Breaking the Barrier of Transactions: Mining Inter-TransactionAssociation Rules", Proc.of the fifth ACM SIGKDD internationalconference on Knowledge discovery and data mining, pp.297-301,1999.

7. Anthony, J., Lee, T., Chun-Sheng Wang, Wan-Yu Weng and Yi-AnChen, Huei-Wen Wu "An efficient Algorithm for mining closed intertransactionitemsets", Data And Knowledge Engineering, Vol. 66,No. 1, pp.68-91, 2008.

8. Archana Tomar, Vineet Richhariya and Pandey, R.K. "AComprehensive Survey of privacy preserving algorithm of AssociationRule Mining in Centralized Databases" , International Journal ofComputer Applications (0975-8887), Vol.16, No.5, pp. 23-47, 2011.