

Comparative Study on Machine Learning Algorithms using Weka

G . Naga Rama Devi
Associate Professor in PEWT, Hyd
Research Scholar
Department of Computer Science
S.P.M.V.V., Tirupati..

1. INTRODUCTION

Abstract:- Data mining is the analysis step of Knowledge Discovery in Databases. It focuses on the discovery of unknown properties in the data. Machine learning is focuses on prediction. Machine learning is based on known properties learned from the training data. The performance is evaluated in machine learning with respect to the ability to reproduce known knowledge. The key task in Knowledge Discovery and Data Mining (KDD) is the discovery of previously unknown knowledge. Dimensionality reduction is a process of reducing the number of random variables under consideration,^[1] which finds a lower dimensional representation of a dataset and can be divided into feature selection and feature extraction.^[2] Feature selection is to find a subset of the original variables. Feature extraction is used to transform the data in the high-dimensional space to a space of fewer dimensions. The data transformation can be done by using linear methods like principal component analysis (PCA). There are two categories of machine learning, 1) Supervised learning is the machine learning task of inferring a function from labeled training data.^[1] The given training data consist of a set of training examples. 2) Unsupervised learning is trying to find hidden structure in unlabeled data. In unsupervised learning, the examples given to the learner are unlabeled and there is no error or reward signal to evaluate a potential solution. In this paper are analyzing the machine learning algorithms using WEKA tool. WEKA is a data mining and machine learning algorithms tool.

Keywords: WEKA Machine learning, Supervised learning, Unsupervised learning, Dimensionality reduction.

WEKA means Waikato Environment for Knowledge Analysis. Which is a popular machine learning software, which was written in Java and developed at the University of Waikato, New Zealand. The Weka is free open source java package available under the GNU General Public License. The Weka has a workbench ^[1] contains a collection of visualization tools and algorithms for data analysis and predictive modeling with graphical user interfaces for easy access.

- **Explorer:** It provides an environment for exploring data with WEKA

- **Experimenter:** It provides environment for performing experiments and conducting statistical tests between learning schemes.

- **KnowledgeFlow:** It has a drag-and-drop interface. It supports incremental learning.

- **SimpleCLI:** It provides a simple command-line interface that allows direct execution of Weka commands for operating systems.

Explorer: It is necessary to open (and potentially preprocess) a data set before starting to explore the data. It has 6 tabs are as follows

1. **Preprocess:** Choose and modify the data being acted on.
2. **Classify:** Train and test learning schemes that classify or perform regression.
3. **Cluster:** Learn clusters for the data.
4. **Associate:** Learn association rules for the data.
5. **Select attributes:** Select the most relevant attributes in the data.
6. **Visualize:** View an interactive 2D plot of the data.

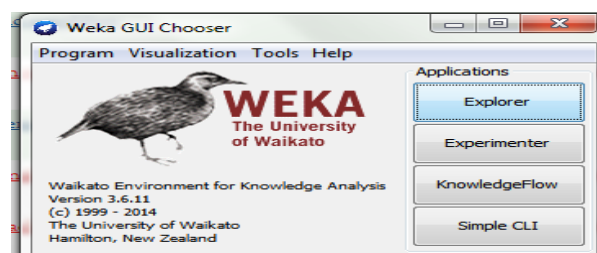


Figure1: Weka GUI chooser

II. DIMENSIONALITY REDUCTION

Dimensionality reduction is the process of reducing high dimensional dataset into low dimensional dataset. The principal component analysis is the main linear technique for dimensionality reduction. PCA performs a linear mapping of the high dimensional data to lower dimensional space in such a way that the variance of the data on the low-dimensional representation is maximized. It can be divided into feature selection and feature extraction.^[2] The correlation matrix of the data is constructed first, then eigenvectors on this matrix are computed next. The eigenvectors that correspond to the largest eigenvalues i.e., the principal components can now be used as a large fraction of the variance of the original data. Dimensionality reduction techniques are Factor Analysis, canonical-correlation analysis (CCA), independent component analysis (ICA), Linear discriminant analysis (LDA), Non-negative matrix factorization (NMF), Principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE).

Factor analysis (FA): Observed random variables are a linear combination of correlated gaussian factors.

Canonical-correlation analysis (CCA): In statistics, CCA is a way of making sense of cross-covariance matrices.

Independent component analysis (ICA): The observed random variables are a linear combination of independent components or factors that are non-gaussian.

Linear discriminant analysis (LDA): LDA and the related Fisher's linear discriminant are methods used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects.

Non-negative matrix factorization (NMF): NMF, also non-negative matrix approximation^{[1][2]} is a group of algorithms in multivariate analysis and linear algebra where a matrix V is factorized into two matrices W and H , with the property that all three matrices have no negative elements.

Principal component analysis (PCA): PCA is a linear dimensionality and statistical technique that uses an

V. MACHINE LEARNING

Machine learning is a branch of artificial intelligence, based on mathematical algorithms and automation. It concerns the construction and study of systems that can learn from data. A machine learning system could be trained on email messages to learn to distinguish between spam and non-spam messages. After learning, it can then be used to classify new email messages into spam and non-spam folders. Machine learning algorithms can also be grouped into generative models and discriminative models. Machine learning is used to predict which medical patients will respond to which treatments, by analyzing experience captured in databases of online medical records. We also study mobile robots that learn how to successfully navigate based on experience they gather from sensors as they roam their environment,

Supervised learning algorithms: These are trained on labeled examples, to the given input where the desired output is

orthogonal transformation to convert a set of observations of correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables.

t-Distributed stochastic neighbor embedding (t-SNE): t-SNE is a machine learning algorithm for dimensionality reduction. It is a nonlinear dimensionality reduction technique. It is particularly well suited for embedding high-dimensional data into a space of two or three dimensions and visualized in a scatter plot.

III. FEATURE SELECTION

Feature selection is used to find a subset of the original variables and also called as features or attributes. Two strategies are filter (e.g. information gain) and wrapper (e.g. search guided by the accuracy) approaches. In some cases, data analysis such as regression or classification can be done in the reduced space more accurately than in the original space.

IV. FEATURE EXTRACTION

It transforms the data in the high-dimensional space to a space of fewer dimensions. The data transformation can be done by using linear methods like principal component analysis (PCA), Independent component analysis (ICA), Singular value decomposition (SVD) and Factor analysis. but many nonlinear dimensionality reduction techniques like SOP also exist.^{[3][4]} and Principal curves and manifolds gives the natural geometric framework for nonlinear dimensionality reduction. In dimensionality reduction, tensor representation can be used on for multidimensional data through multilinear subspace learning^[5].

known. It attempts to generalize a function or mapping from inputs to outputs which can then be used to generate an output for previously unseen inputs.

Unsupervised learning algorithms: It operates on unlabelled examples to the given input where the desired output is unknown. The objective is to discover structure in the data not to generalize a mapping from inputs to outputs.

Semi-supervised learning algorithms: It combines both labeled and unlabelled examples to generate an appropriate function or classifier.

Transduction inference algorithms: It tries to predict new outputs on specific and fixed (test) cases from observed, specific (training) cases.

VI. MACHINE LEARNING TYPES

1. Supervised Learning
2. Statistical Classification
3. Unsupervised Learning
4. Association rule learning.
5. Hierarchical Clustering
6. Partitioned Clustering
7. Outlier Detection
8. Reinforcement Learning
9. Deep Learning
10. Others

1.SUPERVISED LEARNING

Supervised learning[5] is the machine learning task of inferring a function from labeled training data. The set of training examples included in the training data. In order to solve a given problem of supervised learning one has to perform the following steps.

- Determine the type of training examples.
- Gather a training set.
- Determine the input feature representation of the learned function.
- Determine the structure of the learned function and corresponding learning algorithm.
- Complete the design and evaluate the accuracy of the learned function.

supervised learning major issues:

Decision tree learning: Decision tree learning uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. It is one of the predictive modeling approaches used in statistics, data mining and machine learning. More descriptive names for such tree models are classification trees or regression trees. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making.

1. Ensembles (Bagging, Boosting, Random forest).
2. k-NN.
3. Linear regression.
4. Naive Bayes.
5. Neural networks.
6. Logistic regression.
7. Perceptron.
8. Support vector machine (SVM)

Artificial neural networks[9]: An artificial neural network (ANN) learning algorithm, usually called "neural network" (NN), is a learning algorithm that is inspired by the structure and functional aspects of biological neural networks.

Types of Neural Networks:

1. Feedforward neural network

- Bias-variance tradeoff.
- Complexity of functions and amount of training data.
- Dimensionality of the input space.
- Noise in the output values.

Supervised Learning Data types: A learning algorithm include the following data types

- Heterogeneity of the data
- Redundancy in the data.
- Presence of interactions and non-linearity.

Supervised Learning Approaches and Algorithms:

1. Analytical learning&Artificial neural network
2. Back propagation&Boosting i.e., meta-algorithm.
3. Bayesian statistics&Case-based reasoning
4. Decision tree learning&Inductive logic programming
5. Gaussian process regression&Group method of data handling
6. Kernel estimators&Learning Automata
7. Minimum message length.
8. Multilinear subspace learning&Support vector machines
9. Naive bayes classifier&Nearest Neighbor Algorithm
10. Probably approximately correct learning (PAC) learning
11. Ripple down rules, a knowledge acquisition methodology
12. Symbolic machine learning algorithms
13. Subsymbolic machine learning algorithms.

2. Radial basis function (RBF) network
3. Kohonen self-organizing network

Learning paradigms :

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

Real-life applications of neural networks

1. Function approximation, or regression analysis, including time series prediction, fitness approximation and modeling.
2. Classification, including pattern and sequence recognition, novelty detection and sequential decision making.
3. Data processing, including filtering, clustering, blind source separation and compression.
4. Robotics, including directing manipulators, prosthesis.
5. Control, including Computer numerical control

Types of Neural Network Models

1. Inductive logic programming:
2. Support vector machines:
3. Bayesian networks:

Supervised Learning Applications

1. Bioinformatics
2. Cheminformatics
- 2.1. Quantitative structure–activity relationship

3. Database marketing
4. Handwriting recognition
5. Information retrieval
- 5.1. Learning to rank
6. Object recognition in computer vision.
7. Optical character recognition
8. Spam detection
9. Pattern recognition
10. Speech recognition.

1. Time-consuming. Given fixed resources, it is often better to spend more time collecting additional training data and more informative features than it is to spend extra time tuning the learning algorithms.
2. Computational learning theory.
3. Inductive bias&Overfitting.
4. Class membership probabilities&Version spaces.

Disadvantages of Neural Network:

WEKA SUPPORT S CLASSIFICATION ALGORITHMS(Supervised learning):

WEKA supports the widely used machine learning classification algorithms like i.e., Support Vector Machines, Linear regression, Logistic regression, Naive Bayes, Linear discriminant analysis, Decision trees, k-nearest neighbor algorithm, and Neural Networks (Multilayer perceptron). I have applied the most popular classification algorithms on cancer dataset with 99 instances and 16 attributes, I found the following result. Depending upon the nature of the dataset , Naïve Bayes ,J48, ,simple logistic and Zero_r algorithms are giving efficient results.

Classification Algorithms	Correctly classified Instances
Naïve Bayes	44%
ZERO-R	44%
Simple logistic	72%
Bagging	66%
J48	68%

Table1: classification algorithms results using weka

J48:Class for generating a pruned or unpruned C4.5 decision tree. Test mode:10-fold cross-validation

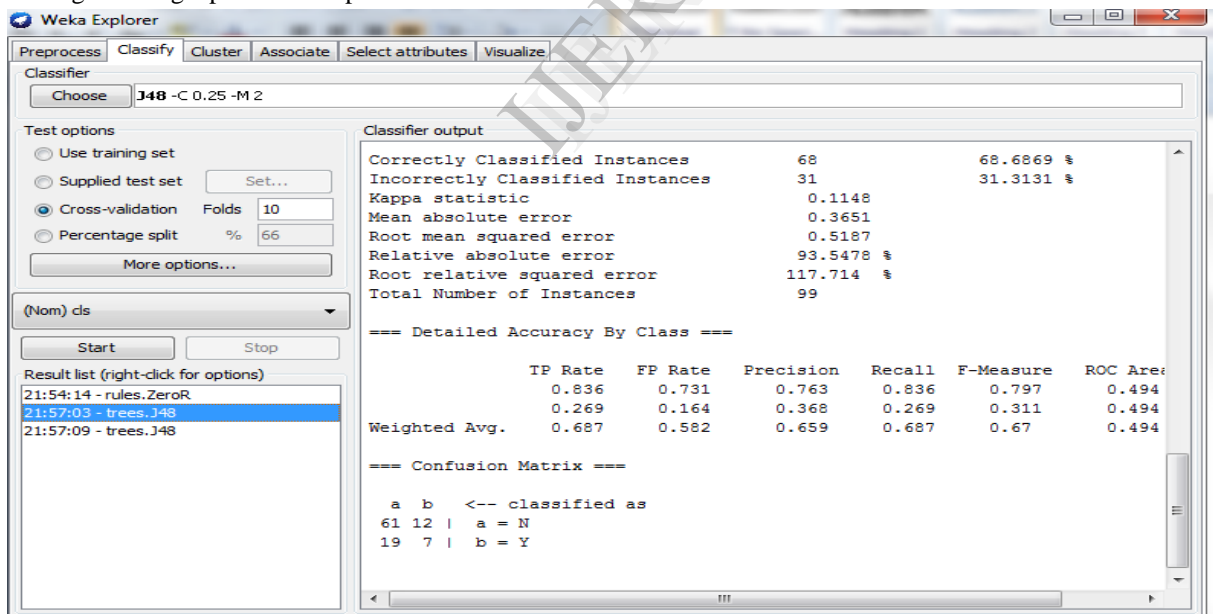


Figure1: J48 Classification Algorithm

ZERO-R:Class for building and using a 0-R classifier. Predicts the mean (for a numeric class) or the mode (for a nominal class). Test mode:10-fold cross-validation

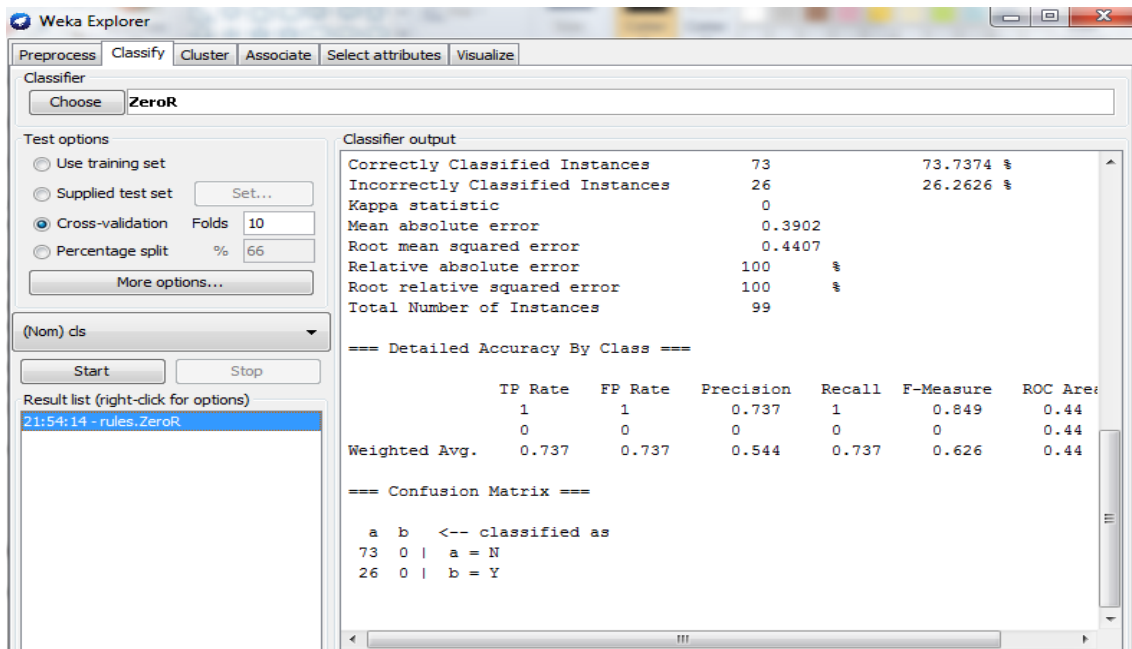


Figure2:Zero-R algorithm

3. UNSUPERVISED LEARNING

Unsupervised learning techniques are widely used to reduce the dimensionality of high dimensional genomic data sets that may involve hundreds of thousands of variables. In machine learning, the problem of unsupervised learning is that of trying to find hidden structure in unlabeled data. Since the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution.

Approaches to unsupervised learning include:

- clustering (e.g., k-means, mixture models, hierarchical clustering),^[2]
- hidden Markov models,
- blind signal separation using feature extraction techniques for dimensionality reduction (e.g., principal component analysis, independent component analysis, non-negative matrix factorization, singular value decomposition).^[3]

Applications:

- Genomics: Unsupervised learning techniques are widely used to reduce the dimensionality of high dimensional genomic data sets that may involve hundreds of thousands of variables.

Unsupervised learning Methods:

- Artificial neural network
- Data clustering
- Expectation-maximization algorithm

- Self-organizing map
- Radial basis function network
- Vector Quantization
- Generative topographic map
- Information bottleneck method

Unsupervised learning types:

- Association rule learning
- Hierarchical clustering
- Partitioned clustering
- Outlier Detection.

4. ASSOCIATION RULE LEARNING

Association rule learning is a method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness.^[9]

- Apriori algorithm
- Eclat algorithm
- FP-growth algorithm.

Apriori Algorithm: Class implementing an Apriori-type algorithm. Iteratively reduces the minimum support until it finds the required number of rules with the given minimum confidence. The algorithm has an option to mine class association rules. It is adapted as explained in the second reference.

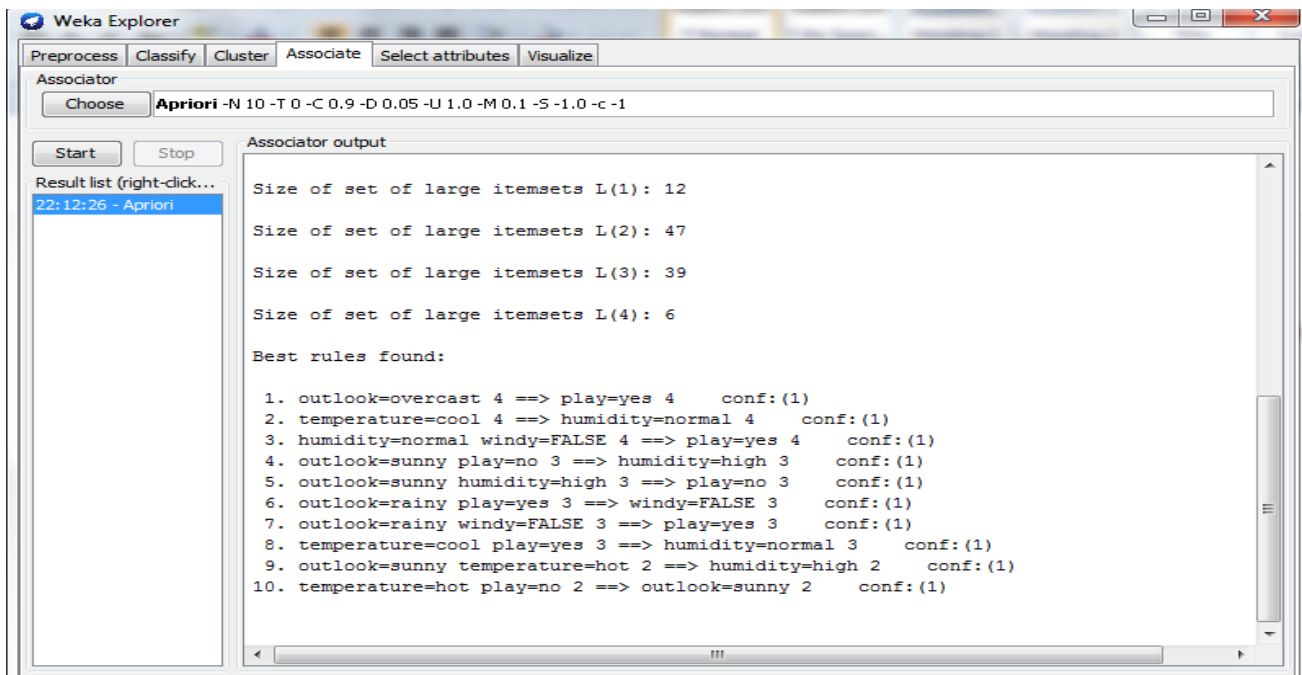


Figure 3: Apriori Algorithm result on weather..nominal data set using WEKA

5. HIERARCHICAL CLUSTERING:

1. Single-linkage clustering.
2. Conceptual clustering(COBWEB).

5.1.Clustering:

Cluster analysis is the assignment of a set of observations into subsets (called clusters) so that observations within the same cluster are similar according to some predesignated criterion or criteria, while observations drawn from different clusters are dissimilar. It is a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics. Other methods are based on estimated density and graph connectivity. It is an unsupervised learning, and a common technique for statistical data analysis.

Specialized types of cluster analysis:

- Clustering high-dimensional data
- Conceptual clustering
- Consensus clustering
- Constrained clustering
- Data stream clustering
- Sequence clustering
- Spectral clustering.

Techniques used in cluster analysis

- Artificial neural network (ANN)
- Nearest neighbor search
- Neighborhood components analysis
- Latent class analysis.

Data projection and preprocessing

- Dimension reduction
- Principal component analysis
- Multidimensional scaling.

Other

- Cluster-weighted modeling
- Curse of dimensionality
- Determining the number of clusters in a data set
- Parallel coordinates
- Structured data analysis.

Clustering Applications

- Biology, computational biology and bioinformatics
- Medicine&Business and marketing
- World wide web
- Computer science
- Social science and Field robotics
- Mathematical chemistry
- Climatology
- Physical geography.

5.2. Conceptual Clustering [9]:

Conceptual clustering is machine learning paradigm for unsupervised classification developed mainly during the 1980s. Conceptual clustering methods are capable of generating hierarchical category structures. It is closely related to formal concept analysis, decision tree learning, and mixture model learning. COBWEB is an incremental system for hierarchical conceptual clustering. COBWEB was invented by Professor Douglas H. Fisher, at Vanderbilt University. COBWEB incrementally organizes observations into a classification tree. Each node in a classification tree represents a class and is labeled by a probabilistic concept that summarizes the attribute-value distributions of objects classified under the node. This classification tree can be used to predict missing attributes or the class of a new object..

5.3.Cobweb Operations: There are 4 basic cobweb operations.

- Merging two nodes: Merging two nodes means replacing them by a node whose children is the union of the original nodes' sets of children and which summarizes the attribute-value distributions of all objects classified under them.
- Splitting a node: replacing a node with its children when a node split.
- Inserting a new node: while a node being inserted into the tree, a new node is created. Passing an object down the hierarchy: Effectively calling the COBWEB algorithm on the object and the sub tree rooted in the node.

6. PARTITIONAL CLUSTERING

- K-means algorithm
- Fuzzy clustering
- DBSCAN
- OPTICS algorithm.

7. OUTLIER DETECTION

- Local Outlier Factor.

8. CLUSTERING ALGORITHMS ANALYSIS USING WEKA

1. Cobweb
2. EM
3. Farthest First
4. Filtered Clusterer
5. Hierarchical Clusterer
6. Make Density Based Clusterer
7. Simple K Means.

1. COBWEB

The COBWEB algorithm was developed by machine learning researchers in the 1980s for clustering objects in a object-attribute data set. The COBWEB algorithm yields a clustering dendrogram called classification tree that characterizes each cluster with a Probabilistic description. Cobweb generates hierarchical clustering [2], where clusters are described probabilistically.

Advantages Of Cobweb

1. COBWEB uses a heuristic evaluation measure called category utility to guide construction of the tree.
2. It incrementally incorporates objects into a classification tree in order to get the highest category utility.
3. A new class can be created based on attribute is one of big difference between COBWEB and K-means methods.
4. COBWEB provides merging and splitting of classes based on category utility, this allows COBWEB to be able to do bidirectional Search.

Disadvantages Of Cobweb

1. It might be very sensitive to the outliers in the data.
2. probability distributions on separate attributes are statistically independent of one another.
3. This is especially so when the attributes have a large number of values because the time and space complexities depend on both the number of attribute and number of values for each attribute.

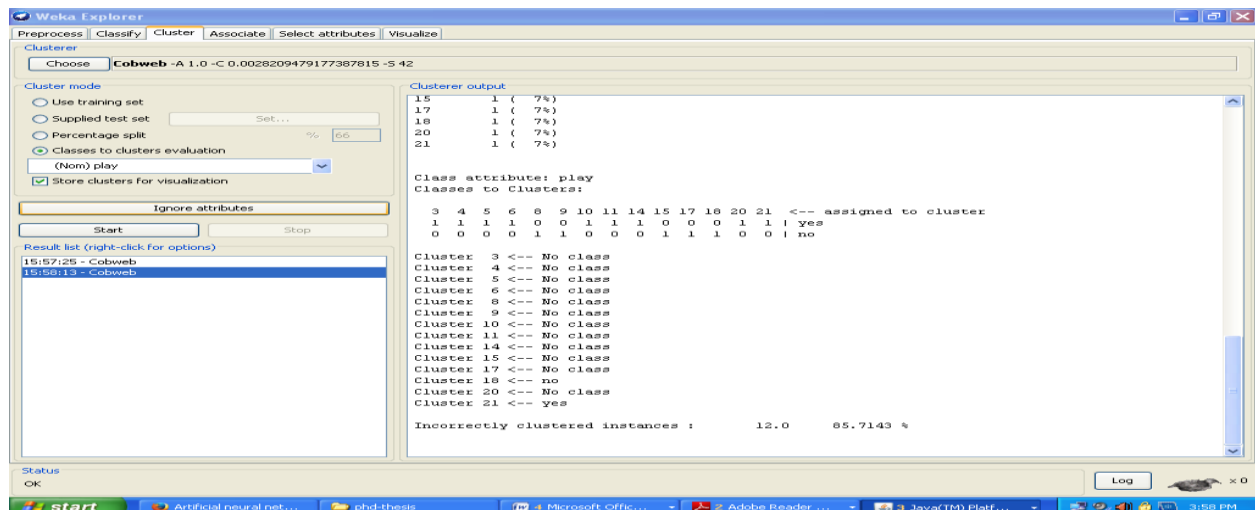


Figure4: Cobweb Clustering Algorithm

2. EM (Expectation–Maximization)

EM algorithm is a clustering algorithm of data mining. When we are not satisfied the result of k-means methods, then we use this algorithm. An expectation– maximization (EM)

algorithm is an iterative method for finding maximum likelihood or maximum a posteriori (MAP) in statistical models to estimate the parameters, where it depends on unobserved latent variables. The EM [11] iteration alternates between performing an expectation (E) step, which computes

the expectation of the loglikelihood evaluated using the current estimate for the parameters, and maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step.

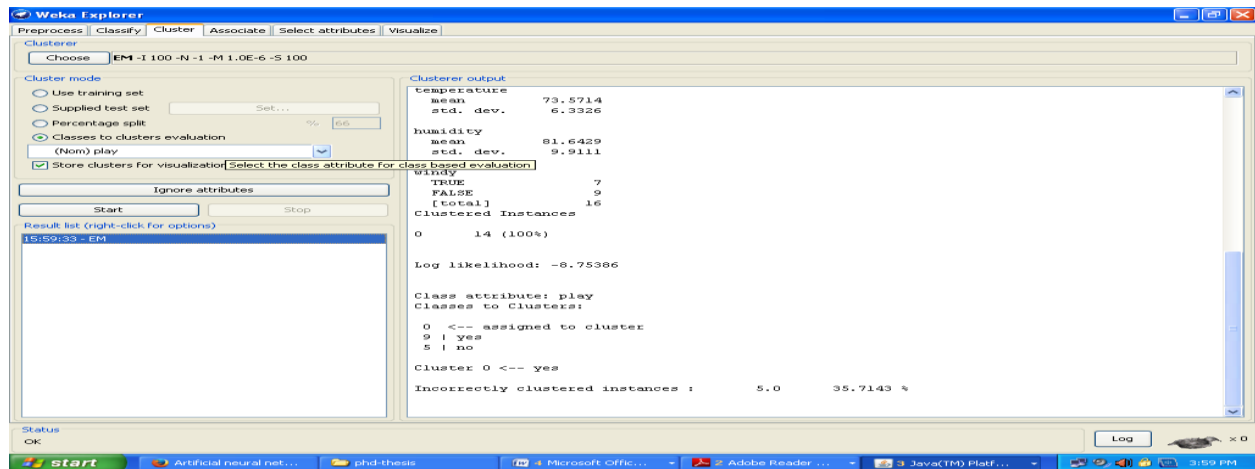


Figure5: EM Algorithm

Advantages of EM

1. Gives extremely useful result for the real world data set.
2. Use this algorithm when you want to perform a cluster analysis of a small scene or region-of interest and are not satisfied with the results obtained from the k-means algorithm.

Disadvantages of EM

1. Algorithm is highly complex in nature.

3. FARTHEST FIRST ALGORITHM

Farthest first is a Variant of K means. It starts with any point, and then iteratively adds in the point furthest from the ones chosen so far. This point must lie within the data area. Farthest-first traversal takes time $O(k|S|)$, which is fairly efficient. Its solution might not be perfect, but is always close to optimal this greatly sped up the clustering in most cases since less reassignment and adjustment is needed. Hochbaum and Shmoys 1985, Implements the "Farthest First Traversal Algorithm".

N-Specify the number of clusters to generate.

S-Specify random number seed.

Result of farthest first algorithm is shown in the figure. It is divide the whole data set in two clusters. Each cluster had shown the lowest and higher value of the data sets.

Advantages of Farthest First

Farthest-point heuristic based method has the time complexity $O(nk)$, where n is number of objects in the dataset and k is number of desired clusters. Farthest-point heuristic based method is fast and suitable for large-scale data mining applications

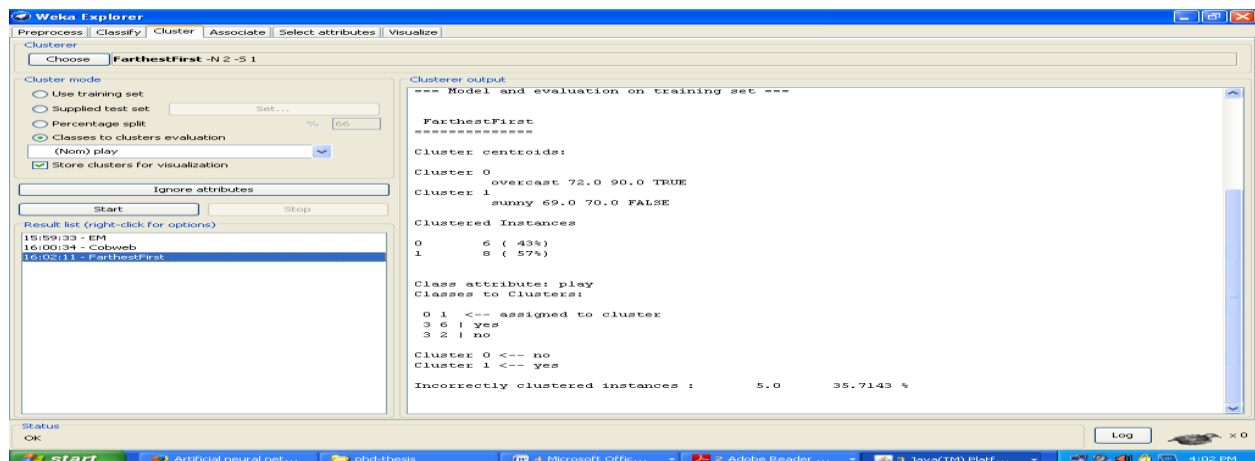


Figure5: Farthest First Algorithm

4. HIERARCHICAL CLUSTERING

Hierarchical cluster analysis or hierarchical clustering is a general approach to cluster analysis to build a hierarchy of clusters. We group similar objects that are close to one another. The main task of this analysis is repeated calculation of distance measures between objects and between clusters once objects begin to be grouped into

clusters. The result is represented graphically as a dendrogram. The dendrogram is a graphical representation of the outcome of hierarchical cluster analysis. The two main categories of methods for hierarchical cluster analysis are divisive methods and agglomerative methods. The agglomerative methods are of wider use in practice. On each step, the pair of clusters with smallest cluster-to-cluster distance is fused into a single cluster..

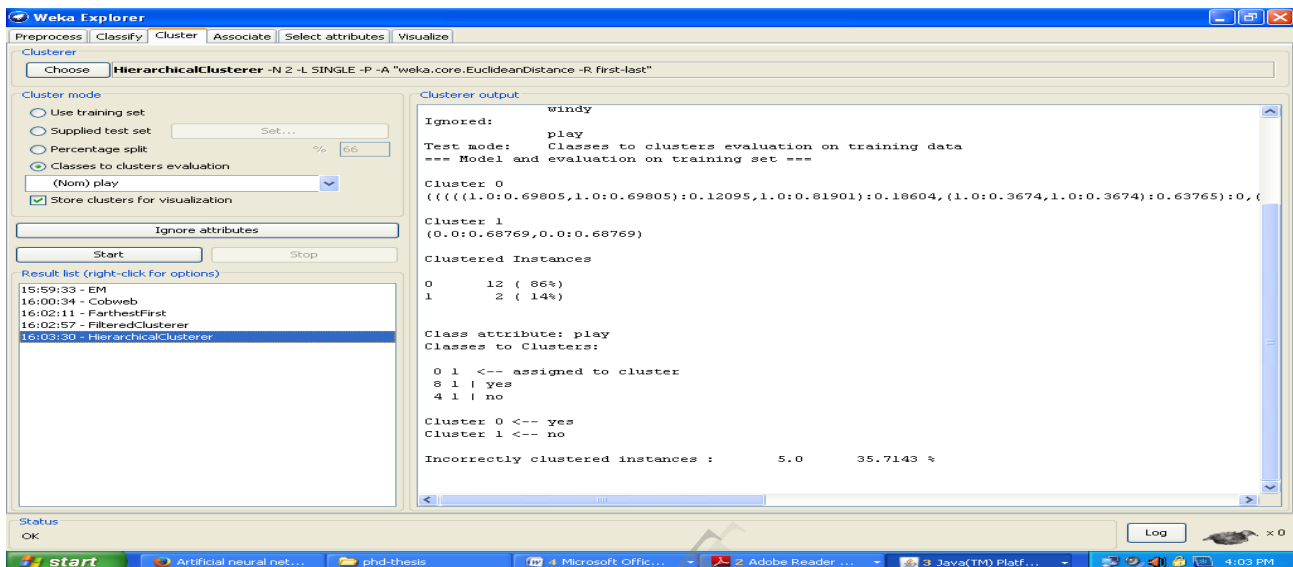


Figure6:HierarchicalClustering

Applications Of Hierarchical Clustering:

1. where hierarchical clustering would be useful is a study to predict the cost impact of deregulation.
2. where hierarchical clustering would be useful is for automatic control of urban road traffic with both adaptive traffic lights and variable message signs. Using hierarchical cluster analysis we can specify the needed number of stationary road traffic sensors and their preferable locations within a given road network.
3. where hierarchical clustering is to take a file that contains nutritional information for a set of breakfast cereals.

5.DBSCAN CLUSTERING ALGORITHM

DBSCAN(for density-based spatial clustering of applications with noise) is a data clustering algorithm proposed by Martin Ester, Hans-Peter Kriegel, Jorge Sander and Xiaowei Xu in 1996 It is a density-based clustering algorithm because it finds a number of clusters starting from the estimated density

distribution of corresponding nodes. DBSCAN is one of the most common clustering algorithms and also most cited in scientific literature.

Advantages:

1. A priori, DBSCAN does not require knowing the number of clusters in the data. (In k-means, a priori know clusters required).
2. DBSCAN can find arbitrarily shaped clusters.
3. DBSCAN has a notion of noise.
4. DBSCAN requires just two parameters and is mostly insensitive to the ordering of the points in the database.

Disadvantages:

1. DBSCAN can only result in a good clustering [8] as good as its distance measure is in the function region Query (P,ε).
2. DBSCAN cannot cluster data sets well with large differences in densities, since the MinPts combination cannot be chosen appropriately for all clusters.

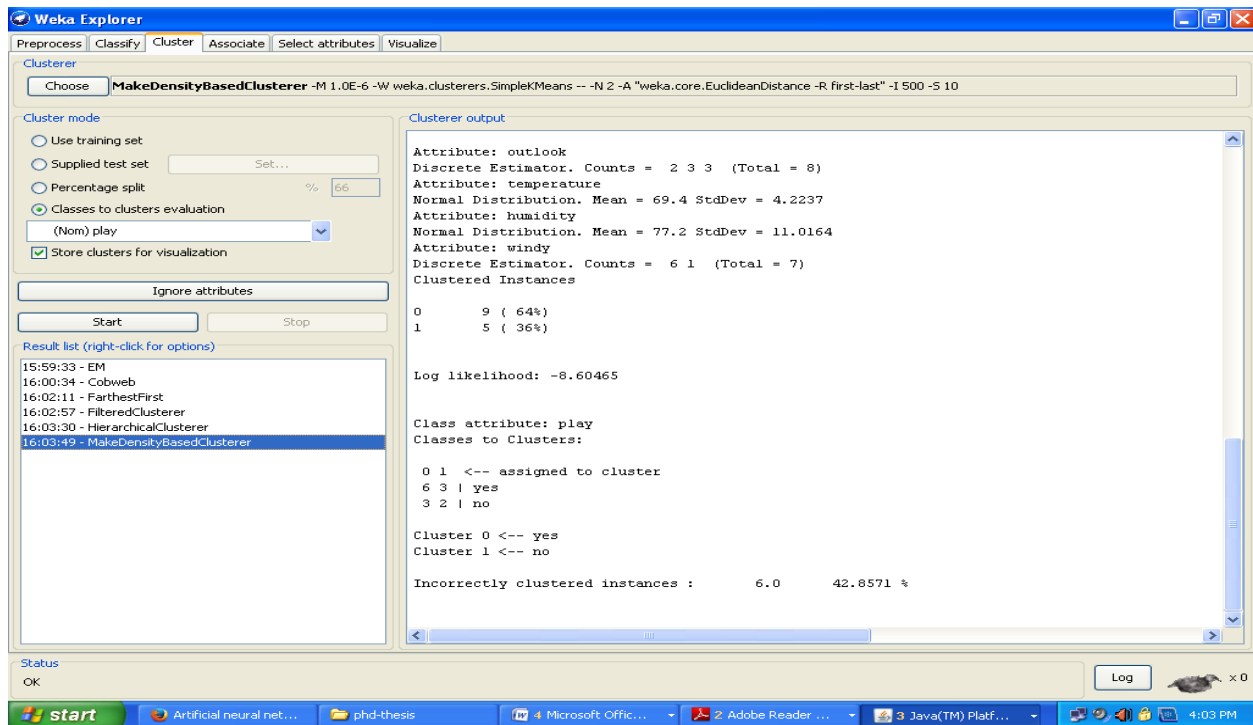


Figure7:MakeDensiyBasedCluster

6. K-MEANS CLUSTERING ALGORITHM

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. In data mining, k-means [15] clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the nearest mean cluster. This results into a partitioning of the data space into Verona cells. K-means is simple and easy way to classify a given data set through a certain number of known clusters (assume k clusters). First we define k centroids, one for each cluster. Different location of k -centroids will cause different result. Place randomly selected centroid far away from each other. The next step is associate each point belonging to a given dataset to the nearest centroid. When there is no more point exists, the first step is completed. At

this point we need to recalculate k new centroids again, which are bar centers of the clusters resulting from the previous step. A new binding has to be done between the same data set points and the nearest new centroid, With these k -new centroids,. Repeat the loop. As a result of this loop, the k centroids change their location step by step until no more changes are done.

Properties of k-means:

- Finds a local optimum
- Converges often quickly (but not always)
- The selection of initial points can have large influence in the result

The k-means algorithm follows sequence of steps:

1. Select randomly the K -objects that are being clustered, Placed these k points into the space. These points are considered as initial group centroids.
2. Assign each point associate to the given dataset to this closest centroid.
3. The positions of the K centroids are recalculated, When all objects have been assigned,
4. Repeat Steps 2 and 3 until the centroids no longer move. Calculate the minimized metric.

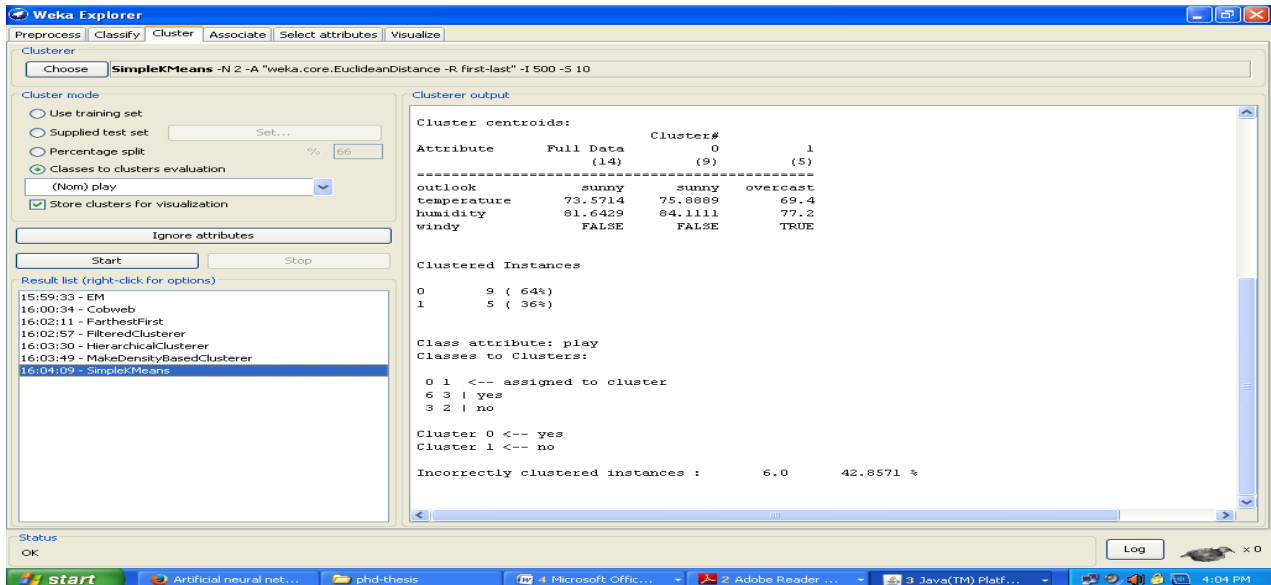


Figure8: K-Means

Advantages to K-means Technique

- K-Means computationally faster than hierarchical clustering with large no on variables (if K is small).
- If the clusters are globular , K-Means may produce tighter clusters than hierarchical clustering,

Disadvantages to K-means Technique

- The quality of the clusters produced comparison is difficult
- Different initial values of k ,will affect the outcome i.e., Clusters of different densities and Clusters of different sizes.
- It has NP-hard problem with k-centre
- It is difficult to predict what k should be, to the fixed no of clusters
- It does not work well for non-globular clusters.
- Outliers can also cause a problem

8.REINFORCEMENTLEARNING

Reinforcement learning is an area of machine learning inspired by behaviorist psychology, concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward.

- Temporal difference learning
- Q-learning
- Learning Automata
- Monte Carlo Method
- SARSA.

9.DEEP LEARNING

12.Comparative Studyon Machine Learning Clustering Algorithms

Using Weka Tool Version 3.7.3 we have worked on cancer dataset Notterman Carcinoma Data.The dataset we have taken is a non linear .It contains 2 nominal attributes and 36

It is a set of algorithms in machine learning that attempt to model high-level abstractions in data by using architectures composed of multiple non-linear transformations.^[1]Deep learning is a part of machine learning methods based on learning representations.

- Deep belief networks
- Deep Boltzmann machines
- Deep Convolutional neural networks
- Deep Recurrent neural networks

10.Others:Data preprocessing.

11.Applications For Machine Learning

- Machine perception
- Computer vision, including object recognition
- Natural language processing^[11]
- Syntactic pattern recognition
- Search engines&Medical diagnosis
- Bioinformatics&Brain-machine interfaces
- Detecting credit card fraud&Stock market analysis
- Classifying DNA sequences
- Sequence mining
- Speech and handwriting recognition
- Game playing&Software engineering
- Adaptive websites&Robot locomotion
- Computational advertising&finance
- Structural health monitoring
- Sentiment analysis
- Affective computing&Information retrieval and Recommender systems.

numeric attributes. After comparison study on clustering algorithms, we have concluded that COBWEB algorithm is the best algorithm than compare to other algorithms.

Algorithm	Error Rate
EM	46.4
Farthest First	41.4
K-Means	38.3
Filtered Clusterer	38.3
Hierarchical	31.3
Make Density Based	41.1
LCOBWEB	30.7

13. RESULT AND CONCLUSION:

We are using data mining techniques in mainly in the medical, banking, insurances, education etc. before start working in the with the data mining models, it is very necessary to knowledge of available algorithms. The main aim of this paper to provide a introduction of Weka and Machine learning algorithms. Weka is the data mining tools. It is the first model for provide the graphical user interface of the user. With the help of figures we are showing the working of various algorithms used in weka. we are showing advantages and disadvantages of some of the machine learning algorithms supported by Weka. Every algorithm has their own importance and we use them on the behavior of the data, but on the basis of this research we found that k-means clustering algorithm is simplest algorithm as compared to other algorithms. I have applied the dimensionality reduction techniques like PCA and linear regression on my non linear cancer dataset. After that I have applied the COBWEB algorithm on that reduced set. Finally I have concluded that COBWEB algorithm works well on the reduced dataset.

14. REFERENCES

1. Fodor, I. (2002) "A survey of dimension reduction techniques". Center for Applied Scientific Computing, Lawrence Livermore National, Technical Report UCRL-ID-148494
2. Cunningham, P. (2007) "Dimension Reduction" University College Dublin, Technical Report UCD-CSI-2007-7
3. Zahorian, Stephen A.; Hu, Hongbing (2011). "Nonlinear Dimensionality Reduction Methods for Use with Automatic Speech Recognition". SpeechTechnologies. doi:10.5772/16863. ISBN 978-953-307-996-7.
4. Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar (2012) Foundations of Machine Learning, The MIT Press ISBN 9780262018258.
5. Duda, Richard O.; Hart, Peter E.; and Stork, David G. (2001); Unsupervised Learning and Clustering, Chapter 10 in Pattern classification (2nd edition), p. 571, New York, NY: Wiley, ISBN 0-471-05669-3.
6. Zoubin Ghahramani (September 16, 2004). "Unsupervised Learning". Hastie, T., Robert, T., Jerome, F. (2009). The Elements of Statistical Learning: Data mining, Inference, and Prediction. New York: Springer. pp. 485–586. ISBN 978-0-387-84857-0.
7. <http://www.cse.unsw.edu.au/~billw/mldict.html#activnfn>
8. Mitchell, T. (1997). Machine Learning, McGraw Hill. ISBN 0-07-042807-7, p.2.
9. Christopher M. Bishop (2006) Pattern Recognition and Machine Learning, Springer ISBN 0-387-31073-8.
10. Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar (2012) Foundations of Machine Learning, The MIT Press ISBN 9780262018258.
11. <http://ai.stanford.edu/~ang/papers/nips01-discriminativegenerative.pdf>
12. Lu, Haiping; Plataniotis, K.N.; Venetsanopoulos, A.N. (2011). "A Survey of Multilinear Subspace Learning for Tensor Data". Pattern Recognition 44 (7): 1540–1551. doi:10.1016/j.patcog.2011.01.004.
13. Fisher, Douglas (1987). "Knowledge acquisition via incremental conceptual clustering". Machine Learning 2 (2): 139–172. doi:10.1007/BF00114265.
14. William Iba and Pat Langley. "Cobweb models of categorization and probabilistic concept formation". In Emmanuel M. Pothos and Andy J. Wills., Formal approaches in categorization.
15. Biswas, G.; Weinberg, J. B.; Fisher, Douglas H. (1998). "Iterate: A conceptual clustering algorithm for data mining".