Algorithms for Computer Aided Diagnosis – An Overview

Dr.A.Padmapriya,M.C.A.,M.Phil.,Ph.D[#1], K.Silamboli Chella Maragatham[#2]
[#]*Department of Computer Science and Engineering, Alagappa University*
Karaikudi

***Abstract*— In medicine, two types of resources are becoming widely used: the Content-based Image Retrieval (CBIR) and the Computer- Aided Diagnosis (CAD) systems. The purpose of CAD is to increase the accuracy of diagnosis, as well as to improve the consistency of image interpretation by using the computer results as a second opinion. Similar to CAD systems, CBIR uses information extracted from images to represent them. However, the main purpose of a CBIR system is to retrieve ‘‘cases” or images similar to a given one. Analyzing past similar cases and their reports can improve the radiologist’s confidence on elaborating a new image report, besides making the training and the diagnosing process faster. Moreover, CAD and CBIR systems are very useful in medicine teaching. Currently, image mining has been focused by many researchers in data mining and information retrieval fields and has achieved prominent results. A major challenge in the image mining field is to effectively relate low-level features (automatically extracted from image pixels) to high-level semantics based on the human perception. Association rules has been successfully applied to other research areas, e.g. business, and can reveal interesting patterns relating low-level and high-level image data as well. In this work, association rules are employed to support both CAD and CBIR systems.**

Keywords- **Image mining, Association rules, Classification, Prediction.**

I .INTRODUCTION

Data Mining is the automated extraction of patterns representing knowledge implicitly stored in large databases, data warehouses, and other massive information repositories. Standard data mining methods may be integrated with information retrieval techniques and the construction or use of hierarchies specifically for text data as well as discipline oriented term categorization systems such as in chemistry, medicine, law, or economics. Image databases are databases that contain descriptions for objects.

This paper is focusing to develop a method based on association rule-mining to enhance the diagnosis of brain images. This is to implement a computer-aided decision support system for an automated diagnosis and classification of images. The method uses association rule mining to analyze the medical images and automatically generates suggestions of diagnosis. It combines automatically extracted low-level features from images with high-level knowledge given by a specialist in order to suggest the diagnosing. Then feature extraction process is applied to extract the features and from that extracted features only the relevant features are selected. So feature selection and discretization process is done on the extracted features that reduce the mining complexity. Association rule mining is a well-known data mining technique that aims to find interesting patterns in very large databases.

The widely used and well-known data mining functionalities are Characterization and Discrimination, content based analysis, Association Analysis, Categorization and Prediction, Outlier Analysis, Evolution Analysis [21].

The General Prediction solution plan to create mining flows that complete most tasks that require numerical or categorical prediction. Prediction models predict values for unknown properties of our data based on historical data or data from similar cases. To create mining flows that performs *regression* or *classification*. Regression predicts numerical values, such as the yearly revenue of a particular store branch, and classification predicts categorical values, such as whether a customer purchased a five-year warranty, a one-year warranty, or no warranty with a particular purchase.

Predictive analysis is performed in three steps:

1. The training step, which analyzes training data that contains known values to generate a prediction model
2. The test step, which applies the model to testing data that also contains known values to ensure that the model can be generalized to unknown data
3. The scoring step, which applies the model to our data to predict the unknown values of your target field

The General Prediction solution plan creates a mining flow that analyzes our data and can produce the following output:

- A table with detailed results of the analysis, such as a description of the nodes of a decision tree

- A table with the predicted values for records with unknown values
- A classification or regression model that describes patterns in the training data This model can be used to:
  - Predict values of the target property for new records
  - Visualize the prediction model to understand why certain predictions are made. This includes analyzing which input fields have the greatest influence on the outcome of the prediction.
  - Generate quality information such as the percentage of correct and incorrect predictions, the average deviation of the predictions from the true values or gains chart information.
- A test model that describes the outcome of a quality test of the prediction model. This model can be used to:
  - Compare the quality of several models that predict the same property.

.

The Prediction operator requires as input tables that contain aggregated data. For example, if your database has a record for every purchase made by every customer, and you want to predict which customers will purchase an item, you need to construct a table that contains a record for each customer that aggregates all purchase data. You can also merge several ta1bles and select rows and columns from these tables or add new columns by performing calculations using the values of other columns.

II.BACKGROUND

A.Feature Extraction:

When dealing with medical images, the earliest phase of a CAD system demands to extract the main image features regarding a specific criterion. Essentially, the most representative features vary according to the image type (e.g. mammogram, brain or lung) and according to the focus of the analysis (e.g. to distinguish nodules or to identify brain white matter). M.X. Ribeiro [12] defined as The Image Diagnosis Enhancement through Association rules (IDEA) method can work with various types of medical images, and with different focus of analysis. However, for each type of image and goal, an appropriate feature extractor should be employed.

B.Feature Selection:

Feature Selection is a process that attempts to select a subset of features, satisfying a combination of application and methodology-dependent criteria: minimizing the cardinality of the feature subset; ensuring classification accuracy does not significantly decrease; and approximating the original class distribution with the class distribution given the selected features.

R. Agrawal et al. [1] proposed a method based on association rule-mining to enhance the diagnosis of medical images. It combines low-level features automatically extracted from images and high-level knowledge from specialists to search for patterns. The proposed method analyzes medical images and automatically generates suggestions of diagnosis employing of association rules. The suggestions of diagnosis are used to accelerate the image analysis performed by specialists as well as to provide them an alternative to work on.

In machine learning and statistics, feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features for use in model construction. The central assumption when using a feature selection technique is that the data contains many *redundant* or *irrelevant* features. Irrelevant features provide no useful information in any context. Feature selection techniques are a subset of the more general field of feature extraction.

Feature extraction creates new features from functions of the original features, whereas feature selection returns a subset of the features. Feature selection techniques are often used in domains where there are many features and comparatively few samples (or data points). Feature selection techniques provide three main benefits while constructing predictive models:

- Improved model interpretability,
- Shorter training times,
- Enhanced generalization by reducing over fitting.

Feature selection is also useful as part of the data analysis process in which features are important for prediction, and how these features are related. Feature Selection reduces the number of variables in a problem by removing those variables that are inconsequential to the classification problem. There are two main arguments for removing useless variables from problem domains Rich Caruana and Dayne Freitag [4]. There are four large groups of feature selection methodologies.

1. Idea: Attempt to locate a feature subset of minimal cardinality that is necessary and sufficient for the classification task at hand [5].

2 .Data Cardinality: Select a feature subset R of the set of features S (R _ S) such that a quality metric is optimized over all other feature subsets of the same cardinality as R [8].

3. Improving prediction accuracy: Choose a subset of features such that a learner training on the subset improves its prediction accuracy or its prediction accuracy does not drop significantly [6].

4. Approximate original class distribution: Chose a subset of features such that the class distribution, given only the values for features in this subset, approximates the class distribution for the entire, original feature set [7].

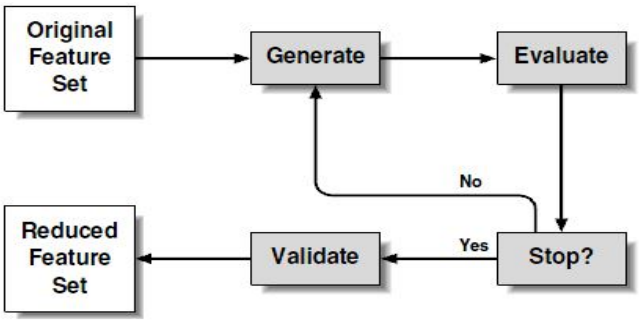

Fig 1: A block diagram of a typical Feature Selection algorithm

It shows how feature subsets are generated and evaluated iteratively until a stopping criterion is met, at which point a final validation step is carried out .FS architecture may be abstracted into a generate-and-test iterative process: feature subsets are somehow generated and evaluated until a stopping criterion is met, at which point the subset if (optionally) validated. This is illustrated in figure1.

In terms of the generation step, there exists extreme diversity in the field. Any methodology that can be applied to searching can be applied to generating candidate feature subsets. Approaches range from brute force generation and heuristically guided generation to evolutionary and stochastic generation of subsets. As regards the evaluation step, FS algorithms may be segregated into two large families.

'Filters' are pure preprocessors. They rely on evaluating the information content of features, and thus draw heavily from Information Theory. Filters are very generic, but employ no knowledge of the classifying properties of the data [3]. 'Wrappers' gauge the quality of subsets of features based on how well those features classify training samples. This is performed in a symbiotic relation with a classifier system. Due to this co-operation, wrappers yield better results than filters, but sacrifice much efficiency and generality for their increased accuracy.

C.Feature Selection Metrics:

The following are the metrics used for Feature Selection.

1. Distance metrics: To improve separability, the distance metric is maximized during feature selection so that the distance between samples is kept maximal.

2. Information measures: These determine the information content gained by adding features and attempt to maximize that.

3.Dependence/Correlation metrics: Such metrics identify redundant features by calculating the correlation between them and other features. Such redundant features can then be removed.

4. Consistency metrics: This relatively new class of metrics [2] employs the training data to assess their consistency, given the subset of features currently evaluated.

Hussein Almuallim [23], is a wrapper-type FS algorithm that performs forward selection to locate a variable subset of minimum cardinality that maintains the consistency of the original data.

Rich Caruana and Freitag's generalization [22], however, reduces the algorithm's computational costs by running two different variants of FOCUS concurrently. One works by forward selection, the other utilizes depth-first search to perform backward elimination. Executing both variants of the algorithm concurrently allows the super-algorithm to terminate early if a small variable subset yields consistency, as the original sub-algorithm will find it efficiently. If, on the other hand, a large variable subset yields consistency, the backward elimination sub-algorithm will find it efficiently.

Subset selection algorithms can be broken into Wrappers, Filters and Embedded. Wrappers use a search algorithm to search through the space of possible features and evaluate each subset by running a model on the subset. Filters are similar to Wrappers in the search approach, but instead of evaluating against a model, a simpler filter is evaluated. Embedded techniques are embedded in and specific to a model.

Other available filter metrics include:

- Class separability
  - Error probability
  - Inter-class distance
  - Probabilistic distance
  - Entropy
- Consistency-based feature selection
- Correlation-based feature selection

Kira and L.A Rendell [11] describe an algorithm called RELIEF that uses instance based learning to assign a relevance weight to each feature. Each feature's weight reflects its ability to distinguish among the class values. Features are ranked by weight and those that exceed a user-specified threshold are selected to form the final subset. The algorithm works by randomly sampling instances from the training data. For each instance sampled the nearest instance of the same class (nearest hit) and opposite class (nearest miss) is found. An attribute's weight is updated according to

how well its values distinguish the sampled instance from its nearest hit and nearest miss. An attribute will receive a high weight if it differentiates between instances from different classes and has the same value for instances of the same class.

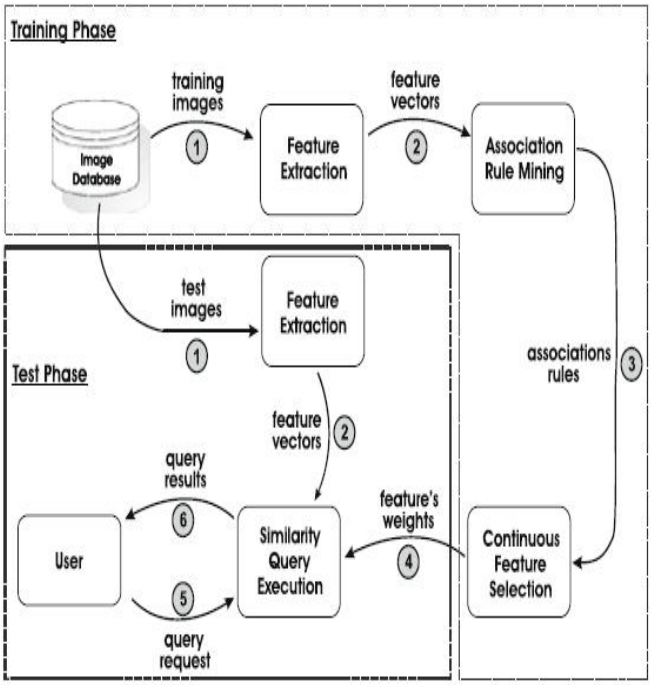

Fig 2: Pipeline of feature selection method

In the training phase: features are extracted from the images, and the feature vectors are used to represent the images (line 1). The feature vectors and the class of each training image are submitted to Omega, which removes irrelevant features from the feature vector and gives a discretization of the remaining features (line 2). The class is the most important keyword chosen by the specialist to describe the image. In the training phase, a processed feature vector is merged.

With the diagnosis keywords about the training images, producing the transaction representation of each image. The transaction representations of all training images are submitted to the Apriori algorithm [13] for association rule mining (line 3), limiting the minimum confidence to high values. In the test phase (lines 4–6), the feature vector of the test image is extracted (line 4) and submitted to the Association classifier engine algorithm (ACE) (line 5).

III.EXISTING METHODOLOGIES

A. Association Rules:

Association rule mining, one of the most important and well researched techniques of data mining. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories. Association rule mining is to find out association rules that satisfy the predefined minimum support and confidence from a given database.

Generally, an association rules mining algorithm contains the following steps:

• The set of candidate k-itemsets is generated by 1-extensions of the large (k -1)-itemsets generated in the previous iteration.
• Supports for the candidate k-itemsets are generated by a pass over the database.
• Itemsets that do not have the minimum support are discarded and the remaining itemsets are called large k-itemsets.This process is repeated until no more large itemsets are found.

Increasing the Efficiency of Association Rules Algorithms:

The computational cost of association rules mining can be reduced in four ways:
* by reducing the number of passes over the database
* by sampling the database
* by adding extra constraints on the structure of patterns through parallelization.

FP-Tree [18] proposed, frequent pattern mining, is another milestone in the development of association rule mining, which breaks the main bottlenecks of the Apriori Tree Projection is another efficient algorithm recently proposed in R.Agarwal [10]. The general idea of Tree Projection is that it constructs a lexicographical tree and projects a large database into a set of reduced, item-based sub-databases based on the frequent patterns mined.

Tien Dung Do et al [19] proposed a specific type of constraints called category-based as well as the associated algorithm for constrained rule mining based on Apriori. The Category-based Apriori algorithm reduces the computational complexity of the mining process by bypassing most of the subsets of the final itemsets.

Wang and Tjortjis [9] proposed an efficient algorithm for mining association rules. Their approach reduces large itemset generation time, known to be the most time-consuming step, by scanning the database only once and using logical operations in the process.

The IDEA method employs the Apriori algorithm [13] to mine association rules. The output of the Omega algorithm and the keywords of the report of the training images are submitted to the Apriori algorithm. A constraint, which restricts the diagnosis keywords to the head of the rules, is added to the mining process. The body of the rules is composed of indexes of the features and their intervals. The values of minimum confidence are set to be high (greater than 97%). The mined rules are used as input to the ACE algorithm.

B.Apriori Algorithm:

This method combines the low level features automatically extracted from images with high level knowledge given by the specialist in order to suggest the diagnosis of a new image. This is to implement a computer-aided decision support system for an automated diagnosis and classification of images. The proposed system is divided into two phases namely

i) The training phase and
ii) The test phase.

Algorithm:

Input - training images and a test image
Output –brain tumors category classification

1. Input image is preprocessed.
2. Extract the required features.
3. Relevant features are extracted through feature selection process.
4. Execute apriori algorithm.
5. Generate association rules.
6. Classify the image based on generated association rules.

C. Association Classifier Engine (ACE) Algorithm:

Associative Classifier Engine (ACE) algorithm, it is necessary to clarify some terms. An image matches a rule, if the image features satisfy the whole body of the rule. An image partially matches a rule, if the image features only satisfy part of the rule's body. An image does not match a rule, if the image features do not satisfy any part of the rule's body. ACE is a special classifier able to return multiple classes (keywords) when processing a test image.

IV.SUGGESTIONS OF IMAGE MINING

This paper suggests the following three methods such as Support Vector Machine, Nearest Neighbor Classification and Naïve Bayes Classification for image mining.

A. Support Vector Machine:

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis.

The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier.

- Given a set of training examples, each marked as belonging to one of two categories, a SVM training algorithm builds a model that assigns new examples into one category or the other.
- A SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

J. Cervantes *et al.*, [15**]** recommended SVM classification for large data sets by considering models of classes' distribution. N. Cristianini and J. Shawe-Taylor[14] presented a Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. A schematic example is shown in the illustration below. In this example, the objects belong either to class C1or C2. The separating line defines a boundary on the right side of which all objects are C1 and to the left of which all objects are C2.

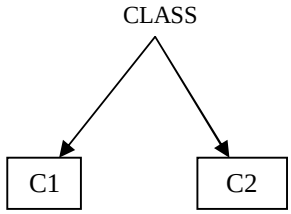

Figure 3 Classifications of two classes

SVM models can be classified into four distinct groups:

- Classification SVM Type 1 (also known as C-SVM classification)
- Classification SVM Type 2 (also known as nu-SVM classification)
- Regression SVM Type 1 (also known as epsilon-SVM regression)
- Regression SVM Type 2 (also known as nu-SVM regression)

B. Nearest Neighbor Classification:

In instance-based classification [17], each new instance is compared with existing ones using a distance metric, and the closest existing instance is used to assign the class to the new one. This is called the ***nearest-neighbor*** **classification method.**

Sometimes more than one nearest neighbors used, and the majority class of the closest *k* neighbors (or the distance weighted average if the class is numeric) is assigned to the new instance. This is termed the *k-nearest-neighbor* method [16].

Example

Just one numeric attribute: It is just the difference between the two attribute values. It is almost as straightforward when there are several numeric attributes: Generally, the standard Euclidean distance is used. However, this assumes that the attributes are normalized and are of equal importance, and one of the main problems in learning is to determine which are the important features.

EX: values *red, green,* and *blue*

Usually, a distance of zero is assigned if the values are identical; otherwise, the distance is one. Thus, the distance between *red* and *red* is zero but the distance between *red* and *green* is one. However, it may be desirable to use a more sophisticated representation of the attributes.

For example,
Given a single instance of each of two classes, the nearest-neighbor rule effectively splits the instance space along the perpendicular bisector of the line joining the instances. Given several instances of each class, the space is divided by a set of lines that represent the perpendicular bisectors of selected lines joining an instance of one class to one of another class.

C. Naïve Bayes for Document Classification

Naïve Bayes is a popular technique for this application because it is very fast and quite accurate. However, this does not take into account sthe number of occurrences of each word, which is potentially useful information when determining the category of a document. Instead, a document can be viewed as a *bag of words*—a set that contains all the words in the document, with multiple occurrences of a word appearing multiple times. Word frequencies can be accommodated by applying a modified form of Naïve Bayes called multinominal Naïve Bayes.

D. Prediction

Prediction and classification also differ in the methods that are used to build their respective models. As with classification, the training set used to build a predictor should not be used to assess its accuracy. An independent test set should be used instead. the attribute can be referred to simply as the predicted attribute predictor is estimated by computing an error based on the difference between the predicted value and the actual known value of y for each of the test tuples, X.

Data prediction is a two step process, similar to that of data classification. However, for prediction, instead of using the term "class label attribute" the attribute can be referred to simply as the predicted attribute.

Preparing the Data for Classification and Prediction

Data cleaning: This refers to the preprocessing of data in order to remove or reduce noise and the treatment of missing values. This step can help reduce confusion during learning [21].

Data transformation and reduction: Normalization involves scaling all values for a given attribute so that they fall within a small specified range, such as 0:0 to 1:0.

Relevance analysis: Many of the attributes in the data may be redundant. A database may also contain irrelevant attributes [20].

E.Comparing Classification and Prediction Methods

Accuracy: Accuracy of a classifier refers to the ability of a given classifier to correctly predict the class label of new

Speed: This refers to the computational costs involved in generating and using the given classifier or predictor.

Robustness: This is the ability of the classifier or predictor to make correct predictions given noisy data or data with missing values.

Scalability: This refers to the ability to construct the classifier or predictor efficiently given large amounts of data.

Interpretability: This refers to the level of understanding and insight that is provided by the classifier or predictor.

V.CONCLUSION

This paper studied various methods for image mining. The Existing methodologies employed are apriori and association classifier engine algorithm. This paper is also suggest that classifications and predictions may be applied for image mining. This paper identified the metrics for comparing of classification and prediction methods. This work can be further extended to implement a fuzzy based roughest theory for feature selection for efficient image mining.

REFERENCES

[1] R. Agrawal, T. Imielinski, and A. N. Swami, ―Mining association rules between sets of items in large databases, in

Proc. 1993 ACMSIGMOD Int. Conf. Manage. Data - SIGMOD 93 SIGMOD 93, Washington, DC, 1993, pp. 207–216.

[2] M. Dash and H. Liu. Feature selection for classification. Intelligent Data Analysis — An International Journal, 1(3), 1997. http://www.public.asu.edu/_huanliu/papers/ida97.ps. 10, 12, 13, 14

[3] Juzhen Z. Dong, Ning Chong, and Setsuo Ohsuga. Using rough sets with heuristicsto feature selection. In Ning Zhong, Andrzej Skowron, and Setsuo Ohsuga, editors, Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing (RSFDGrC-99), volume 1711 of Lecture Notes inArtificial Intelligence, pages 178–187, Berlin, November 9–11 1999. Springer. 11, 21, 42.

[4] Rich Caruana and Dayne Freitag. Greedy attribute selection. In Proceedings of the 11thInternational Conference on Machine Learning, pages 28–36. Morgan Kaufmann, 1994. 10, 12, 14, 42, 43, 44.

[5] K. Kira and L. A. Rendell. The feature selection problem: Traditional methods and a new algorithm. In Proceedings of the Ninth National Conference on Artificial Intelligence, pages 129–134. AAAI Press, 1992. 10

[6] George H. John, Ron Kohavi, and Karl Pfleger. Irrelevant features and the subset selection problem. In Proceedings of ICML-94, the Eleventh International Conference on Machine Learning, pages 121–129, New Brunswick, USA, 1994. 10 48

[7] Daphne Koller and Mehran Sahami. Toward optimal feature selection. In Proceedings of ICML-96, the Thirteenth International Conference on Machine Learning, pages 284–292, Bari, Italy, 1996. 10

[8] P. M. Narendra and K. Fukunaga. A branch and bound algorithm for feature subset selection. IEEE Transactions on Computers, 26:917–922, 1977. 10Z. Pawlak: Rough sets, International Journal of Computer and Information Sciences, 11, 341-356, 1982.

[9] Wang, C., Tjortjis, C., PRICES: An Efficient Algorithm for Mining Association Rules.

[10] Agarwal, R. Aggarwal, C. and Prasad V., A tree projection algorithm for generation of frequent itemsets. In J. Parallel and Distributed Computing, 2000.

[11] K. Kira and L. A. Rendell. A practical approach to feature selection. In *Machine*Learning: Proceedings of the Ninth International Conference, 1992.

[12] M.X. Ribeiro, A.J.M. Traina, C.T. Jr., N.A. Rosa, P.M.A. Marques, How to improve medical image diagnosis through association rules: The idea method, in:The 21th IEEE International Symposium on Computer-Based Medical Systems, Jyvaskyla, Finland, 2008, pp. 266–271.

[13] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Intl. Conf. on VLDB, Santiago de Chile, Chile, 1994, pp. 487–499.

[14] N. Cristianini and J. Shawe-Taylor. Support Vector machines. Cambridge Press, 2000.

[15] J. Cervantes, Xiaoou Li, and Wen Yu, "SVM Classification for Large Data Sets by Considering Models of Classes Distribution," Sixth Mexican International Conference on Artificial Intelligence - Special Session, Pp. 51 – 60, MICAI 2007.

[16] Aha, D. (1992). Tolerating noisy, irrelevant, and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies*, 36(2), 267–287.

[17] Bay, S. D. (1999). Nearest neighbor classification from multiple feature subsets. *IntelligenData Analysis*, 3(3), 191–209.

[18] Han, J. and Pei, J. 2000. Mining frequent patterns by pattern-growth: methodology andimplications. ACM SIGKDD Explorations Newsletter 2, 2, 14-20.

[19] Tien Dung Do, Siu Cheung Hui, Alvis Fong, Mining Frequent Itemsets with Category-Based Constraints, Lecture Notes in Computer Science, Volume 2843, 2003, pp. 76 – 86

[20] Content based analysis (Hayes, 1990), Association Analysis, Categorization and Prediction (Han,2001), Outlier Analysis, Evolution Analysis (Lewis, 1990).

[21] Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol. Modell. 157: 101_ 118. Austin, M. P. and Cunningham, R. B. 1981. Obse.

[22] Rich Caruana and Dayne Freitag. Greedy attribute selection. In Proceedings of the 11th International Conference on Machine Learning, pages 28–36. Morgan Kaufmann, 1994. 10, 12, 14, 42, 43, 44

[23] Hussein Almuallim and Thomas G. Dietterich. Learning Boolean concepts in the presence of many irrelevant features. Artificial Intelligence, 69(1–2):279–305, 1994. 12, 14, 43