International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 25 Issue 6 – MAY 2018.

Prediction and Frequent Pattern in Data Mining with Brain Dominance

Dr.K.Venkatasalam[#1]

[#] *Associate Professor, Department of Computer Science & Engineering, Mahendra Engineering College(Autonomous), Namakkal, India*

***Abstract*— This research work focuses the brain dominance of the students who can achieve the performance of cognitive skill of knowledge by the category of logical Reasoning, Numerical Ability and Perceptual Speed of ability. It can be analyzed by Naïve Bayes Classification technique. Apart from the academic performance of the students, the cognitive skill analysis will stimulate their attitude and skill to motivate them for their higher studies and career. An outcome of the result can analyze the students at present the status of their ability skill which is either high or very high or medium or low or very low. This research is to offer inclusive model hypothetically by conducting offline test based on the cognitive model through Human Computer Interface. In this research analysis, the brain dominance helps to analyze the student's ability level with respect to left brain dominant, or right brain dominant or whole brain dominant based on frequent pattern mining using FP tree algorithm.**

***Index Terms*— cognitive skill; Naïve Bayes Classification; FPGrowth Algorithm; Association Rule; Accuracy of classification; Brain Dominance hemisphere**

I. INTRODUCTION

Educational data mining (EDM) is the use analytical techniques to better understand relationships, structure, patterns, and causal pathways in datasets. In Educational systems are increasingly engineered to capture and store data on users' interactions with a system. These data can be analyzed using statistical, machine learning [15], and data mining techniques [20]. In this research to predicting students' future learning by creating models that incorporate information of students' knowledge, thinking, behavior, analyse the performance apart from academic environment analysed by the techniques of data mining. Data mining techniques such as K-nearest neighbor, decision tree, Naïve Bayes, Neural network, Fuzzy, Genetic and other techniques [16] are applied in various environments [1] [10].

II. CLASSIFICATION TECHNIQUES

A classification technique is an approach to building classification models from an input of data. It includes Naïve Bayes classifier [], decision tree classifiers, support vector machines, neural networks and rule-based classifiers [2]. Each techniques employs a learning algorithm identify a model that best fits the relationship between the attribute set and class label of the input data. Evaluation of the performance of a classification model is based on the counts of test records correctly and incorrectly predicted by the model. These counts are tabulated in a table is known as a confusion matrix. The confusion matrix provides the information needed to determine how well a classification model performs, summarizing this information with a single number[3] [14] .

III. SCOPE OF THE RESEARCH

In this paper, with the use of Naïve Bayes classification algorithm, assigned the Class to the records of a training data will require research on data mining techniques to predict the brain Dominar of students would involves an analysis of skills in Perceptual speed, Numerical Ability and Logical reasoning.

The classifier will predict the brain Dominar , perceptual speed, Numerical ability and logical reasoning belong to which class that should have highest posterior probability, which is used to identify students skill level and provide decision support for their future career and motivation.

An association technique will analyse the high score of ability based on frequent pattern mining with respect to support and confidence.

IV. DATA FOR RESEARCH

Cognitive processes is the process that involve knowledge, attention, memory, producing and understanding the language, problem solving and decision making. All these are very important for human behavior. Based on the GOMS model, KLM model for end-user through testing for knowledge task analysis [12], the collection of data can be stored in the database based on usability criteria which can be targeted in the system design at the stage of effectiveness, learning ability, and flexibility, attitude where the student skill can analysis effectively.

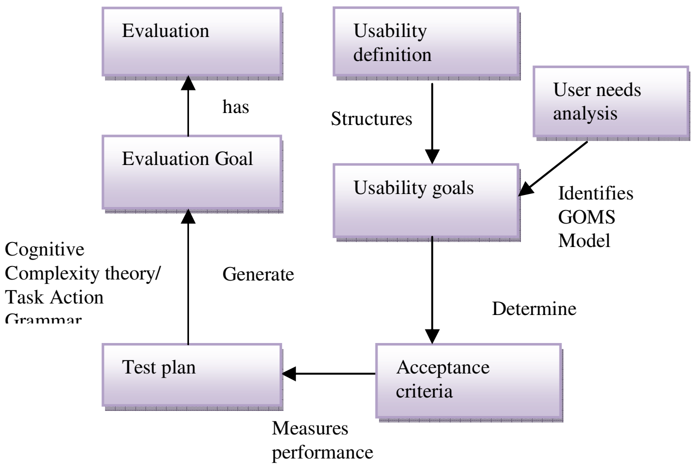

Fig.1 Task Analysis by cognitive model

Fig.1 refers an appropriate timing to each action based on usability criteria gather for knowledge measuring by the performance of human knowledge factor [11]. The end result will be a prediction of the time to perform a task in the optimal way using the cognitive complexity theory, by an experienced user to attain the goal using that particular interface design specification.

The variety of domain values are collected from the students through conduct offline test which relates with brain dominance, logical reasoning,Numerical ability and Perceptual speed accuracy based on cognitive model. From fig.2, represents the part of the research in methodology for analysing the performance cognitive skill for students based on Naïve Bayes classification method and also analyse an ability level with respect to left brain dominant, right brain dominant and whole brain dominant by creating association rule.

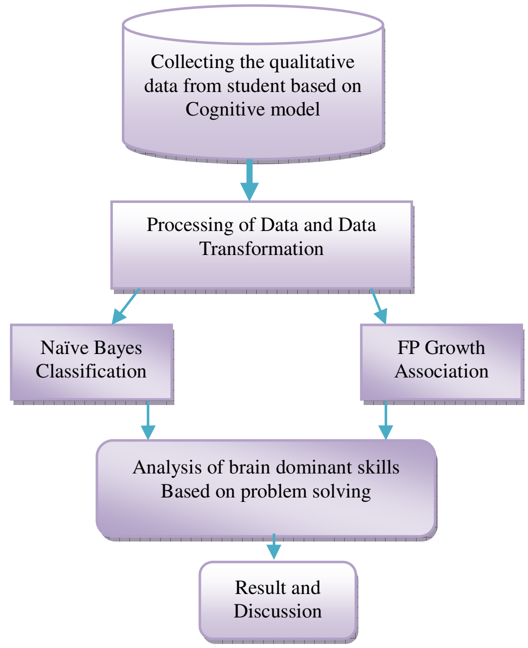

Fig.2 Methodology for proposed research

From Table.1, illustrates the collection of domain values such as Student Id, Age, Gender, Speciality, Logical Reasoning (LR), Numerical Ability (NA), Personality (P), Left brain dominant (LBD), Moderate preference for the left (ML), Slight preference toward the left (SL),

Table.1 Collection of Qualitative Training data

Attributes	Possible Values
Student Id	Id of the student
Age	15, 16, 17, 20 and 21
Gender	Male (m) and Female (F)
Speciality	**Higher Secondary School** **Arts and Science** Computer Science, Computer Technology, Information technology, Commerce, BCA **Engineering** Computer Science, Electonic and Communication and Information Technology
Logical Reasoning (L)	Above 95 = Very High {>76 and <=95} = High {>56 and <=76} = medium {>36 and <=56} =Low <=36 = Very Low
Numerical Ability (NA)	Above 95 = Very High {>76 and <=95} = High {>56 and <=76} = medium {>36 and <=56} =Low <=36 = Very Low
Personality (P)	Above 95 = Very High {>76 and <=95} = High {>56 and <=76} = medium {>36 and <=56} =Low <=36 = Very Low
Dominance Option	1 to 15 questions (a, b, and c)
Left brain dominant (LD)	-15 to -9
Moderate preference for the left (ML)	-8 to -5
Slight preference toward the left (SL)	-4 to -1
Whole-brain dominance (WB)	0
Slight preference toward the right dominance (SR)	+1 to +4
Moderate preference for the right (MR)	+5 to +8
Right brain dominant (RBD)	+9 to +15

Table.2 Training data set

Id	Age	Gender F/M	Dominance Option	LR	NA	P
1	15	f	c b a a b c b c c a a b b a c	90	90	98
2	15	f	a b c a b c c a c a a c a a c b a c	40	60	90

Whole-brain dominance (WB), Slight preference toward the right dominance (SR), Moderate preference for the right (MR), Right brain dominant (RB) which can be predicting the quality skill of students data based on rule in data mining techniques.

From Table. 2, shows the training set of 1000 instances, each recording the values of seven attributes as well as classification. The domain values are collected from the Higher Secondary School, Department of Computer Science Computer Technology, BCA, Commerce and Information Technology in Rathinam College of Arts and Science and Computer Science and Engineering, Electronic and Communication and Information technology are taken from Rathinam Technical Campus. Here the score marks of logical reasoning, Numerical Ability, perceptual Speed and Brain Dominance Hemisphere are handled with respect to the age.

V. PROPOSED METHODOLOGY

A. *Classification*

In Classification process, the derive model is to predict the class of objects whose class label is unknown. The derived model is based on the analysis an asset of training data. In educational data mining, the work of data was predicted by logical rule using Classification algorithms [5] with respect of common domain values to identifying the qualitative performance of required details.

It can be focused to predicting the cognitive skill of students relaets with brain dominant hemisphere. In this technique, it can be classified the functioning of cognitive style such as logical reasoning, Numerical ability, perceptual speed for analyzing the skill for the students.

In Naïve Bayes algorithm [1], to reduce computation in evaluating P $(X|C_i)$, the naive assumption of class conditional is made. The attributes are conditionally independent to one another by given the class label of the tuple which predicts the data in tuple where X belongs to the class Ci [13].

$$P(C_i|X) > P(C_j|X) \quad \text{for } 1 \le j \le m, j \ne i \qquad (1)$$

By Bayes' theorem, the classic for which P (Ci |X) represents maximum posterior hypothesis.

$$P(Ci|X) = \frac{P(X|Ci)P(Ci)}{P(X)} \qquad (2)$$

The classic for which P $(C_i |X)$ is maximized is called the maximum posteriori hypothesis. It can easily estimate the probabilities $P(x_1|C_i)\times P(x_2|C_i)\times\cdots \times P(xn|C_i)$ from the training tuples by the following relationship.

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i) \qquad (3)$$

Rule Base to classify the skill data as follows

If logical_Reasoning > 95 = Very High
{>76 and <=95} = High
{>56 and<=76} = middle
{>36 and <=56} =Low
<=36 = Very Low

If Numerical_ability > 95 = Very High
{>76 and <=95} = High
{>56 and <=76} = middle
{>36 and <=56} = Low
<=36 = Very Low

If Personality > 95 = Very High
{>76 and <=95} = High
{>56 and <=76} = middle
{>36 and <=56} =Low
<=36 = Very Low

Identify Brain Dominant Hemisphere

An algebraic sum of option "a" and "b" scores based on the condition to check for analyzing the Dominar of students

If select "a" option then (-) minus sign in front of "a" and
If select "b" option then (+) plus sign in front of "b"
Do not consider "C" Option

Rule Base classifier for analysing the brain dominant students

If score= -15 to -13 and -12 to -9 then dominant = leftbrain dominant
If score= -8 to -5 then dominant = moderate preference for the left
If score= -4 to -1 then dominar = slight preference toward the left
If score= 0 then dominar = whole-brain dominance
If score= +1 to +4 then dominar = slight preference toward the right dominance
If score= +5 to +8 then dominar= moderate preference for the right
If score= +9 to +12 and +13 to +15 then dominar = rightbrain dominant.

B. Association of Skill relates with brain dominant Hemisphere

FP Growth is an algortihm [5] that generates frequent itemsets from an FP-Tree [6] [7] [8] [9] by exploring the tree in a bottom- up approach. It finds all the frequent items [19] endings with a particular suffix by divide and conquer stategy to split the problem into smaller sub problem. to discover the frequent itemset [17] [18] without candidate item generation which is constructed using 2 passes over the dataset, In pass 1 to scan the data and find support for each item and sort in decreasing order based on their support. In pass 2, FP growth read transaction at a time and maps it to a path. Fixed order is used so path can overlap when transaction share items when the counters are incremented.

In table 4, table 5, table 6 and table 7 shows the FP Tree growth of table view with respect of size, support, item1, item2 item3, item4 set values for analysing the rule of Logical Reasoning (LR), Numerical Ability (NA) and Personality (P) and dominar of RBD, WBD, SR, MR, LD based on the minimum support=0.1.

Finally, it shows the frequent items are analysed in 4 item set with support of 0.1 by size manner. For the decision tree building the confidence level can be set as 0.5 which can be relate with brain dominant of skill students, an association rule can be shown in fig 4, 5, 6 experimented in rapid miner [8].

From fig.2, represents the rule of association for analyse the relation based on Brain dominar with minimum criteria of confidence value and range. Here confidence value denotes 0.5 and it analysed above the confidence value of item set in dataset and support value can denoted as 0.1.

From the result the right brain dominar(RBD) of students attains Numerical Ability (NA)is high, Logical Reasoning(LR) is Medium and Perceptual (P) is High.

VI. Experiment and Result Analysis

In classification techniques, it can be experimented with training data by the given attributes like logical reasoning, numerical ability and personality for analyse the skill level of students based on braindominar (class label).

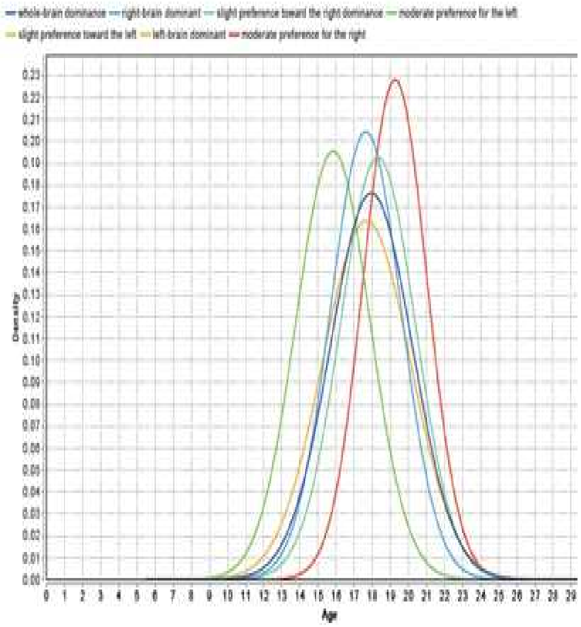

Fig.5 Analysing Barin Dominant Student based on Age

From fig.5, shows the distribution of Class such as
whole-brain dominance in 0.137
moderate preference for the left in 0.017,
slight preference toward the left in 0.019
slight preference toward the right dominance in 0.422,
moderate preference for the right in0.089,
left-brain dominant in 0.054,
right-brain dominant in 0.258

From fig. 6, illustrates all dominar students most attained High and Vey High score in perceptual Speed Accuracy and fig. 7, represents moderate preference in right dominar, right brain dominar, slight preference in left and moderate preference in left dominar attained medium score, Whole Brain dominar, left brain dominar attained High score, Right brain dominar also attained low score, Slight preference of right dominar attained very low score in logical reasoning.

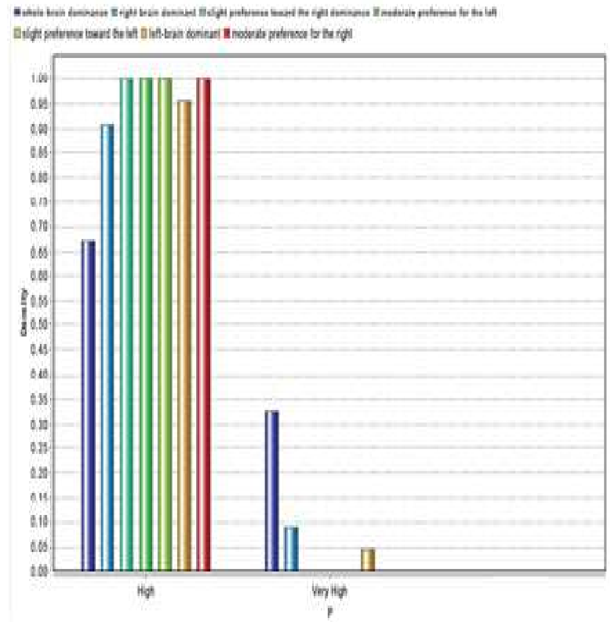

Fig.6 Analysing Perceptual Speed Accuarcy (P) of Student based on Brain Dominar

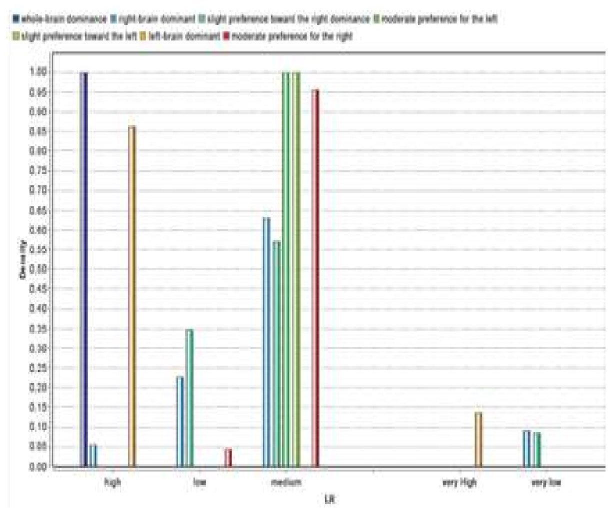

Fig.7 Analysing Logical Reasoning (LR) of Student based on Brain Dominar

From fig. 8, represents moderate preference in right dominar, right brain dominar, slight preference in right and moderate preference in left dominar attained medium score, Whole Brain dominar , Right brain dominar attained High score. Left brain dominar also attained low score, Slight preference of left dominar attained very low score and Right brain dominar also attained Very High score in Numerical ability.

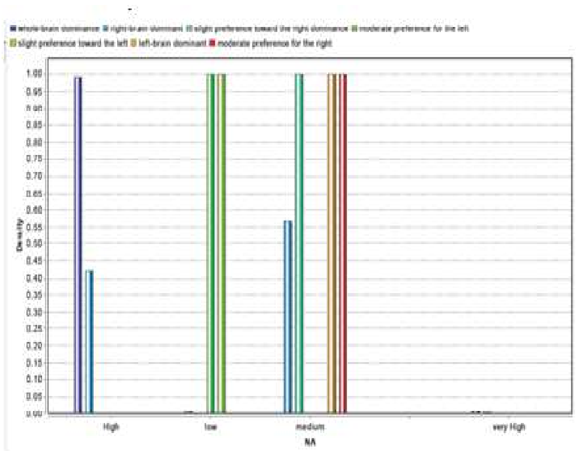

Fig.8 Analysing Numerical Abiity (NA) of Student based on Brain Dominar

A. *Performance Measure*

Naive Bayes classification can be measured 1000 instance of attributes are classified in 0.01 seconds.88.7% of instance are classifed from training data which are predicted by the class of brain dominar and 11.3% instances are incorrect classified under the classification as shown in table 8. True positive rate and False positive rate can be executed in the performance chart, under this evaluation precision, Recall, Accuracy, F- measure and confusion matrix can be computed based on the equation 4, 5 and 6. Precision is a measure of the accuracy provided that a specific class has been retrieved from predicting. It is defined by

$$\text{Precision} = \frac{\text{diagonal element}}{\text{sum of relevant column}}$$

i.e., $$\text{Precision} = \frac{\text{tp}}{(\text{tp} + \text{fp})} \quad (4)$$

where tp and fp are the numbers of true positive from the prediction p and false positive predictions for the considered class when the actual value is n as shown in table 2

$$\text{F-measures} = \frac{2{*}\text{precision}{*}\text{recall}}{(\text{precision} + \text{recall})} \quad (5)$$

$$\text{Accuracy} = \text{tp} + \text{tn} / \text{p} + \text{n} \quad (6)$$

where precision can be seen as a measure of exactness, whereas recall is a measure of completeness or quantity. Recall is nothing but the true positive rate for the class. From Table.8 and table 9, shows the average weight rate of the classification of true positive rate, false positive rate, precision and F-measure.

VII. Conclusion

In this research, it can be concluded that cognitive skill of students can be predicted based on the category of brain dominance hemisphere. This models can be gathered by using problem solving using data mining techniques. In this research, 1000 instance of training data set, can be used to analysing the predicting by Naïve Bayes Classification algorithm which can be produced their efficiency of rule base to classify by execution time of accuracy is 0.01 second. 88.7% are correctly classified for classify the brain dominar, 87.2 % are correctly classified for classify the perceptual speed scored students, 74.6 % are are correctly classified for classify the Logical Reasoning and 99.2 % are correctly. Finally, frequent pattern mining evolves the result of this analysis to produced Right brain dominar attained high and medium score in Numerical Ability and medium and low or very low score in logical reasoning score. Left brain dominar attained high and medium in logical reasoning and low or very low in Numerical ability. Whole brain dominar attained high score in both Numerical ability and logical reasoning.

References

[1] Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, 2nd ed., Morgan Kaufmann Publishers, 2006.
[2] Arun K Pujari, "Data mining techniques", University Press (India) Private Limited.
[3] David Hand, Heikki Mannila & adhraic Smyth, "Principles of Data Mining", MIT Press, 2001.
[4] Arbi Ghazarian, "Pauses in man-machine interactions: a clue to users " Skill levels and their user interface requirements", Int. J. Cognitive Performance Support, Vol. 1, No.1, 2013.
[5] R.Agrawal, T.Imielinski, and A.Swami, "Mining Association Rules Between Sets Of Items In Large Databases", In proceedings of the ACM SIGMOD International Conference on Management of data, pp. 207-216,1993.
[6] M. J. Zaki and C.J. Hsiao, "CHARM: An efficient algorithm for closed association rule mining", Technical Report 99-10, Computer Science Dept., Rensselaer Polytechnic Institute,1999.
[7] Anurag Choubey, Ravindra Patel,J. L.Rana, "A Survey Of Efficient Algorithms And New Approach For Fast Discovery Of Frequent Item Set For Association Rule Mining", International Journal of Soft Computing and Engineering,2011.
[8] Rakesh Agrawal and Ramakrishnan Srikant, "Fast Algorithms For Mining Association Rules In Large Databases", In Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pp 487-499, Santiago, Chile,1994.
[9] M. Houtsma, and Arun Swami, "Set-Oriented Mining for Association Rules in Relational Databases", IEEE International Conference on Data Engineering, pp. 25–33, 1995.
[10] Heikki, Mannila, "Data mining: machine learning, statistics, and databases", IEEE, 1996.
[11] Mayilvaganan M and Kalpanadevi D, "Designing a Human Computer Interface System based on Cognitive Model", IEEE Xplore. Advancing Technology for Humanity. Catalog Number: CFP1420J-
[12] PRT. ISBN: 978-1-4799-3974-9. pp: 283-286.
[13] Mayilvaganan M, Kalpanadevi D. "Designing a Human Computer Interface Systems Tasks and Techniques", International Journal of Engineering Research and General Science 2014, pp: 425-428.
[14] Chai, K.; H. T. Hn, H. L. Chieu; "Bayesian Online Classifiers for Text Classification and
[15] Filtering",Proceedings of the 25th annual international ACM SIGIR conference on Research and Development in Information Retrieval, August 2002, pp 97-104
[16] Nguyen Thai Nghe , Paul Janecek , and Peter Haddawy "A comparative Analysis of Techniques for Predicting Academic Performance" 37th SEE/IEEE Frontiers in EducationConference.
[17] Fabrizio Sebastiani. Machine Learning in Automated Text Categorization. ACM Computing Surveys, 34(1):1-47, 2002.
[18] Yin, X., and Han, J. (2003). CPAR: Classification based on predictive association rule. In SDM 2003, San Francisco, CA.
[19] Agrawal, R., Amielinski, T., and Swami, A. (1993). Mining association rule between sets of items in large databases. In Proceeding of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216, Washington, DC, May 26-28.
[20] Srikant, R., Vu, Q., and Agrawal, R. 1997. Mining association rules with item constraints. In Proc. 1997 Int. Conf. Knowledge Discovery and Data Mining (KDD'97), Newport Beach, CA, pp. 67–73.

[21] Pei, J., Han, J., and Mao, R. 2000. CLOSET: An efficient algorithm for mining frequent closed itemsets. In Proc. 2000 ACM-SIGMOD Int. Workshop Data Mining and Knowledge Discovery (DMKD'00), Dallas, TX, pp. 11–20

[22] Kamber, M., Han, J., and Chiang, J.Y. 1997. Metarule-guided mining of multi-dimensional association rules using data cubes. In Proc. 1997 Int. Conf. Knowledge Discovery and Data Mining (KDD'97), Newport Beach, CA, pp. 207–210.