Towards Socially Intelligent Automated Tutors: Predicting Learning Style Dimensions from Conversational Dialogue

Naeemeh Adel, Annabel Latham, Keeley A Crockett
The Intelligent Systems Group, School of Computing, Mathematics and Digital Technology,
The Manchester Metropolitan University, Chester Street,
Manchester M1 5GD, UK
N.Adel@mmu.ac.uk, A.Latham@mmu.ac.uk, K.Crockett@mmu.ac.uk

Abstract— **Conversational Intelligent Tutoring Systems (CITS) that automatically adapt to learning styles (LS) can improve learning, however current modelling of LS has ignored Neutral learners. This paper presents research examining the ability of data mining algorithms to predict LS dimensions from behaviour captured during natural language tutorials with Oscar CITS. Two datasets, 2ClassBDS and 3ClassBDS, were cleaned and prepared for the data mining task of predicting student LS. Each dataset comprised four sub-datasets representing the four Felder-Silverman LS dimensions. 3ClassBDS included a third Neutral class describing individuals with a balance of LS preferences. Naïve Bayes, Decision Trees, Lazy Learning and Neural Networks algorithms were applied to each dataset and parameters adjusted to improve prediction accuracies. The 2ClassBDS dataset results show good prediction, with decision trees (Simple CART) achieving accuracies of 81.33-86.66%. For 3ClassBDS results were mixed, with the J48 algorithm achieving 56-73% accuracy, indicating that further work and data is needed.**

Keywords— *Student modelling; data mining; conversational agents; learning styles*

I. INTRODUCTION

The concept of personal tutoring has been around for a long time, from the days when the rich and the elite were home schooled, to the modern day when some still choose to home school their children. Providing private tutoring to an individual focuses the tutor's attention solely on that one individual rather than have the attention of the tutor split between a class full of students. Individual tutors spend more time with the sole individual and pick up cues to move according to the learner's pace, allowing for optimum learning. However, individual personalized tutoring is costly and limited in flexibility: learners must be available at the same time as tutors, and one-to-one learning experiences are not scalable. If it were possible to develop an intelligent machine that could mimic a human tutor, allowing learners to question and discuss topics whilst providing a tutorial adapted to a learner's knowledge and learning preferences alongside intelligent solution analysis and problem solving support (hints), the problems of flexibility and scalability could be overcome. By providing each student with their own individual tutor the learning experience for that student should improve as the tutor focuses only on them, progressing at their pace and answering questions. This goal of developing an automated, socially intelligent personal tutor, which is available to learners 24/7 and not bound by location or circumstance, motivates this research.

Intelligent Tutoring Systems (ITS) extend the traditional e-learning systems, which were merely content delivery systems, by incorporating computational intelligence techniques to adapt a tutorial to the learner. Adaptation is based on knowledge, but more recent ITS aim to be socially intelligent by adapting to affect (such as learner emotion, mood or learning styles). Conversational Intelligent Tutoring Systems aim to mimic a human face-to-face tutorial by additionally incorporating a natural language interface which enables them to communicate with a learner using conversation.

Oscar is a Conversational Intelligent Tutoring System (CITS) that mimics the role of a real life tutor by holding conversations with the learner, and being able to answer their questions and guide them through a tutorial [1]. In tailoring the tutorial to best fit a student's knowledge and learning preferences, the learning experience is enhanced and students are left with a better personal learning experience [2]. Oscar CITS creates a student model for each learner by noting their knowledge and behavior during interactions, and continually adjusting their learning style scores. By analysing the interactions and results of these tutorials, possible patterns can be picked up, to allow for improvements in the CITS. Oscar adapts the learning experience of a participant following diagnosis of their learning style using the Felder and Silverman (FS) model [3]. The FS model categorises an individual's learning style over four aspects, known as dimensions. Originally, Oscar CITS automatic learning style prediction was based on analyzing the best-predictor behaviour variable exhibited by an individual to determine their preference for its associated dimension of learning style. Whilst this obtained good prediction accuracy, significantly better results were obtained using alternative machine learning approaches, specifically fuzzy decision trees [1] and multilayer perceptron neural networks [4] due to their ability to discover patterns in the combined behavior aspects of a participant in association with their preferred learning style.

There is a balance to be struck between obtaining better prediction accuracy and the transparency of the classification decision. Understanding how the classification of learning style was made by a machine learning algorithm is useful when deciding on suitable adaptation strategies within the tutorial. Machine learning approaches have been shown to be suitable to predict learning styles. For example, Garcia et al. [5] used Bayesian networks to infer learning styles from historical behaviour of participants during interaction with an intelligent tutoring system. In this study, precisions of 58-77% were achieved for three of the four possible dimensions of the Felder and Silverman model. Further tuning of the model improved predictions to 66-80% [5]. Naïve Bayes was used in EDUCE [6] to predict which resources students would prefer based on past choices. A detailed review of state of the art in the automatic detection of learning styles can be found in Feldman, Monteserin & Amandi [7].

The FS model measures an individual's learning style using a self-assessment questionnaire, the Index of Learning Styles, that assigns a score for each of the four learning style dimensions. The score represents the learner's strength of preference for the learning style, and Felder and Silverman describe individuals whose score indicates a very weak preference as *Neutral* learners [1]. Neutral learners are balanced at the centre of a learning style dimension and prefer a mix of teaching styles, however state of the art automatic prediction techniques are binary classifiers, unable to distinguish between learners with a strong preference and neutral learners. Therefore a neutral learner with a slight verbal preference is classified the same as a verbal learner with a strong preference, despite preferring very different teaching styles.

A more sophisticated learner model which distinguishes neutral learners from learners with a strong preference for a learning style should lead to better adaptation of tutoring. Published adaptation algorithms [8] rely on precise models of learning style (i.e. a style and a preference score) to determine the best adaptation for an individual. Oscar CITS incorporates three adaptations for each learning style dimension, as Neutral learners have different needs to those with a strong learning style preference [8, 3].

The contribution of the research presented in this paper is in the improvements to classification accuracy and investigating the modeling of neutral learners as a separate class. Behaviour data generated through the natural language interaction between the participants and Oscar CITS [8] was used to construct two datasets. These datasets, referred to in this work as 2ClassBDS and 3ClassBDS, were then cleaned and prepared for the purpose of data mining. The data mining task was to predict learning style preferences along four dimensions. Thus each dataset comprised four sub-datasets to represent the four learning style dimensions (Active or Reflective, Sequential or Global, Sensing or Intuitive and Verbal or Visual) in the Felder and Silverman model [9]. In addition, the 3ClassBDS dataset identified a *Neutral* class where students exhibited behavior indicating a balance of preferences for a particular dimension (i.e. no strong preference for either learning style on that dimension).

This paper will examine a number of traditional data mining algorithms and their ability to predict learning styles, including Naïve Bayes, Decision Trees, Lazy Learning and Neural Networks. The software used to test these algorithms was Weka [10], an open-source tool which implements a collection of machine learning algorithms for data mining tasks. Algorithms were applied using default parameters then manually tuned to investigate if the predictive accuracy could be improved.

The overall results have shown that high level prediction is possible for 2ClassBDS (the 2-class dataset), where best results in terms of accuracy for each dimension were achieved using decision trees (Simple CART). The Active or Reflective dimension (ACT-REF) had the best result taking both the dimension accuracy and the class accuracy into account. Further work and data is needed to improve prediction for the 3ClassBDS which additionally incorporated a Neutral class.

This research seeks to address the following research questions:

1. Is it possible to model learning style dimensions from conversational dialogue using machine learning techniques?

2. Is it possible to model neutral learners from conversational dialogue using machine learning techniques?

The rest of the paper is organized as follows: Section II provides a brief overview of data mining and the classification techniques adopted in this research. Section III describes the preparation of the two datasets. Section IV outlines the experimental methodology, with results and discussion in section V. Finally, section VI describes the conclusion and future work.

II.   DATA MINING

There are numerous data mining algorithms, but for this work the focus is on classification algorithms. Classification algorithms predict one or more discrete variables which are based on the other attributes in the dataset [11]. They are frequently more accurate than human-created rules since humans are often incapable of articulating what they know and searching for hypothesis to explain the data [12]. The classification algorithms most commonly used to predict a discrete attribute are Decision Trees, Naïve Bayes, Clustering and Neural Network algorithms [11]. Seven algorithms were applied to the two datasets: they are summarised below.

A.   Naïve Bayes

Naïve Bayes (NB) algorithm uses Bayes Theorem that calculates the probability of an event occurring given the probability of another event. A Naïve Bayes model is easy to build with no complicated iterative parameter estimations; this makes it very useful for large datasets [13]. Naïve Bayes provides an adaptive classifier that can improve initial knowledge based predictions for the class of a new instance by refining the model based on the evidence provided by the whole history of processed cases [14]. The Naïve Bayes

classifier is often used for supervised learning; it is flexible and can deal with any number of attributes or classes [13].

B. Bayesian Networks

Bayesian Networks are graphical representations of probabilistic relationships among a set of random variables that display relationships clearly [15]. They are popular in machine learning applications, because they are simple yet efficient and allow each attribute to independently contribute towards the end result making it suitable for many domains [16].

C. C4.5 Algorithm

C4.5 is an extension of the ID3 algorithm, a branch of decision trees and handles the over fitting of data and how deep to grow a decision tree. Over fitting occurs when an algorithm generates very accurate results on a dataset but fails to generalise well from what it has learned [17]. C4.5 reduces error pruning, handles continuous attributes, and improves the overall computational efficiency of the algorithm [18]. J48 is the C4.5 algorithm that is used in Weka. For J48 to classify a new item it must create a decision tree based on the attribute values of the available training data, and is able to identify the attributes that discriminate the various instances the most. This allows maximum information to be collected from the data in order to have the highest information gain [19]. If there are any values present for which there is no ambiguity, and the data instances falling in this category have the same values as the target variable, then this branch is terminated and is assigned the target value that has been obtained [19].

D. Classification And Regression Trees

Classification And Regression Trees (CART) is a classification method that builds decision trees using historical data and can be used on both numerical and categorical variables. The decision trees are then used to classify new data. The decision trees are made up of a set of questions that splits the learning sample into smaller and smaller chunks by only asking yes and no questions. CART searches all possible variables and values to find the best split. The answer to each question determines what the next question will be [20]. Once a branch cannot be split any further, it is referred to as being terminal [21]. The basic idea of CART is to ensure each node is purer than that of its parent node using an impurity measure to measure this pureness in each of the nodes and splits [22].

E. Random Forest Algorithms

Random Forest algorithms grow many classification trees to the largest extent possible with no pruning. Random Forest algorithms are highly accurate and run well on large databases. This algorithm can handle missing data well and will maintain a good accuracy level despite missing data [23]. This algorithm also provides accuracy and variable importance information with the results [24].

F. Lazy Learning

Lazy Learning is an instance based learning algorithm, which is an extension of the nearest neighbour classification algorithm. Lazy algorithms are good to use as they are supervised learning algorithms where instances are used to classify objects. The cost for updating object instances is low and the learning rate of the algorithm is fast. It is also possible to extend the algorithm and be able to obtain concept descriptions. Concept descriptions are the primary output of instance-based learning algorithms. They are the functions that map instances to categories [25].

G. Neural Networks

Neural Networks are organised in a series of layers, on the left side of the vector the input vector enters; this is then projected to a hidden layer. Each of the units in the hidden layers are a weighted sum of the values in that layer. This layer will then project to an outer layer and this is where the desired answer appears. Weights at both the hidden layer and output layer are adjusted so that the actual output corresponds to the desired output. Once it has been trained the neural network will take a new data point and will output either a zero or a one. If the classifier is uncertain of the output it, will produce a value somewhere in between [26].

III. DATASETS

A. Capturing the Dataset

The datasets for this research were collected from a study conducted using Oscar CITS [8]. The participants were all under- or post-graduate students from Manchester Metropolitan University (MMU), who had little or no knowledge of Structured Query Language (SQL). In total, there were 114 participants who completed an SQL tutorial with Oscar CITS under different studies. In the first study, 10 participants were chosen who had English as their first language and had previous experience of an undergraduate ORACLE SQL course at varying levels. The second study involved 43 participants who had previous experience and different levels of expertise of an undergraduate SQL course. The participants were a selection of undergraduate and post graduate students who either volunteered for this experiment or were referred by their lecturers to complete the Oscar CITS SQL revision tutorial. In the final study, 61 students, who were enrolled on the undergraduate Advanced Database Application unit, took part in the study. For these students, Oscar SQL revision tutorial was integrated into the teaching and learning environments and they used the system during official scheduled timetabled labs [8].

All 114 participants initially completed the Index of Learning Styles questionnaire to formally diagnose their FS learning styles: this established the ground truth used to test the classification accuracy. The results were recorded in the Oscar CITS student model, which holds information about behavior and the results of tutorial questions and tests for each participant according to an assigned user ID. Next, participants completed a 12 question Multiple Choice Question (MCQ) test to establish their existing SQL knowledge, and the results were recorded in the Oscar CITS student model. Participants then went on to engage in an SQL revision tutorial conversation with Oscar CITS. At the end of the tutorial participants completed the same MCQ test to assess their learning gain, with the results again being stored in the student model [8]. Of

the 114 participant who took part in these three studies, only 75 completed the SQL revision tutorial, and only the completed studies were used to evaluate Oscar CITS [8].

B. Preparing the Dataset

The learner behaviour captured by Oscar CITS throughout the tutorial conversation were encoded into 41 variables, making up the dataset for each participant [8].

Initial preparation has been done on this dataset to look at the different ways of classifying the four learning dimensions of Felder and Silverman.

1- Global or Sequential (GLO-SEQ) – learners will fall under either global scale or sequential scale. Sequential learners learn in small chunks and tend to gain an understanding of the problem in logical steps. Global learners learn in large chunks and learn the material randomly without seeing connections.

2- Visual or Verbal (VIS-VRB) – learners will fall under either visual scale or verbal scale. Visual learners prefer representations of presented material such as diagrams, pictures or films. Verbal learners prefer written or spoken explanations such as handouts or presentations.

3- Active or Reflective (ACT-REF) – learners will fall under either active scale or reflective scale. Active learners like to try things out and work with others in groups, while reflective learners like to work alone and think things through.

4- Sensing or Intuitive (SNS-INT) – learners will fall ether under either sensing scale or intuitive scale. Sensing learners are drawn towards facts and procedures and like problem solving. They do not like surprises and complications in learning. They tend to be good at memorizing facts and enjoy hands on work. Intuitive learners on the other hand are drawn towards theories and meaning. They like innovative ideas and do not like repetition. They are good at understanding new material and working through mathematical formulas.

For the two-class classification, each of the participants was described as one of the two possible learning style classes for each of the four dimensions of Felder and Silverman's learning styles [9]. This dataset has been labelled 2ClassBDS and will be referred to by this name from this point forward.

For the three-class classification, each participant was described as one of three classes for each of the four dimensions, being placed at either end or somewhere in the middle of the scale. Learners placed in the middle of the scale are known as *Neutral* for that dimension.

Therefore for the three-class dataset the dimensions have the following classes:

1- Global, Sequential or Neutral (GLO-SEQ-NEU)

2- Visual, Verbal or Neutral (VIS-VRB-NEU)

3- Active, Reflective or Neutral (ACT-REF-NEU)

4- Sensing, Intuitive or Neutral (SNS-INT-NEU)

This dataset has been labelled 3ClassBDS and will be referred to by this name from this point forward.

TABLE 1. DISTRIBUTION OF CLASSES IN THE DATASETS

Var.[a]	Description	Type	2ClassBDS	3ClassBDS
Class	Contains the learner's learning style class given by the ILS results	Nominal	ACT: 43 REF: 32	NEU: 32 ACT: 26 REF: 17
			GLO: 30 SEQ: 45	NEU: 28 GLO: 16 SEQ: 31
			VIS: 65 VRB: 10	NEU: 17 VIS: 53 VRB: 5
			INT: 30 SNS: 45	NEU: 27 INT: 15 SNS: 33

a. Variable

Table 1 shows the distribution of the learners in the 2ClassBDS and 3ClassBDS datasets. As can be seen from Table 1, in the 2ClassBDS there is a fair distribution between the classes in each of the learning dimensions except for the VIS-VRB dimension, where the distribution is tilted more towards the Visual class. For the 3ClassBDS, the numbers are again quite similar in the dimensions except for the VIS-VRB-NEU dimension, where most of the distribution is tilted towards the Visual class. This is in line with Felder and Silverman's study of engineering students [3] which found that most engineering students are Visual learners.

IV. EXPERIMENTAL METHODOLOGY

This research explores the accuracy obtained on the two datasets using the seven algorithms of Naïve Bayes, Bayes Network, J48, Simple CART, Random Forest, IBK and RBF, described in section II.

The classification algorithms used in this study were applied directly to both datasets. For each dataset a number of experiments were run for each of the seven chosen algorithms using Weka. 10-fold cross validation was used in all experiments. Cross-validation is a systematic way of doing repeated holdout that actually improves upon it by reducing the variance of the estimate [12].

Initially each algorithm was tested using its default parameters assigned by Weka [10]. Then the algorithm parameters were tuned empirically; in some instances this resulted in improved classification. The best accuracy for each algorithm was then recorded and a comparison of results over all algorithms indicated the best algorithm for classifying the learning style dimension within each dataset.

V. RESULTS AND DISCUSSION

Table 2 shows the results for the 2ClassBDS dataset, listing the highest accuracy obtained for each of the seven algorithms tested. From Table 2, it can be seen that the Simple CART algorithm generated the highest classification accuracy, followed closely by the J48 algorithm. The Simple CART algorithm is very robust when it comes to outliers and can isolate them easily. It also does not need the variables to be

Algorithm / Dataset	Naïve Bayes	Bayes Network	J48	Simple CART	Random Forest	IBK	RBF
ACT-REF	78.66	78.66	81.33	82.66	84.00	66.66	73.33
SEQ-GLO	84.00	84.00	85.33	82.66	81.33	50.66	58.66
SNS-INT	70.66	72.00	81.33	81.33	73.33	62.66	69.33
VIS-VRB	85.33	84.00	85.33	86.66	88.00	72.00	78.66

Algorithm / Dataset	Naïve Bayes	Bayes Network	J48	Simple CART	Random Forest	IBK	RBF
ACT-REF-NEU	66.66	66.66	73.33	72.00	61.33	52.00	62.66
SEQ-GLO-NEU	60.00	61.33	65.33	56.00	56.00	41.33	52.00
SNS-INT-NEU	57.33	57.33	56.00	64.00	54.66	48.00	53.33
VIS-VRB-NEU	72.00	72.00	68.00	65.33	70.66	54.66	61.33

selected in advance and can identify the most significant variables and eliminate the others. This all helps in the algorithm gaining a better accuracy of the data and thus being the algorithm with the best possible results for this dataset.

Table 3 shows the results of the seven algorithms that have been tested for the 3ClassBDS and the highest accuracy obtained from each one. From Table 3 it can be seen that the first highest best algorithm for this dataset is the J48 algorithm, followed closely by the Simple CART algorithm.

Comparing the results for the two datasets, shown graphically in Fig. 1 and Fig. 2, it can be seen that the accuracy levels have dropped significantly in the 3ClassBDS compared to the 2ClassBDS. This is for a number of reasons; possible reasons could be that there are three scales in the 3ClassBDS as opposed to two in the 2ClassBDS.

This, coupled with the low number of records (75) brings the accuracy for each class down because by reclassifying some learners as Neutral, and thus reducing the number of records for each class, the dataset is now less representative of each class. This reduction of the number of class examples for the algorithms to learn from leads to lower accuracy. However,
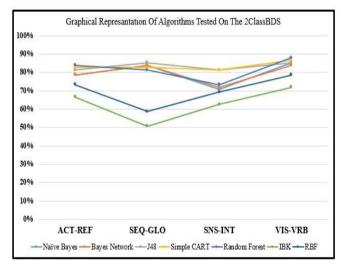
this is also a good indication of the number of Neutral learners whose needs have previously not been met as they have been incorrectly classified under the existing binary classification of learning styles.

VI. CONCLUSION AND FURTHER WORK

In conclusion, this research examined the behaviour datasets captured during the live tutorials of Oscar CITS. The datasets were divided into two sets, one containing a two class dataset, referred to as 2ClassBDS, and the other containing a three class dataset, referred to as 3ClassBDS, where the third class was a Neutral class that highlighted the fuzziness linked to each learning dimension. These datasets were analysed using seven data mining algorithms. With the 2ClassBDS, high level prediction was possible and the Simple CART algorithm gave the best accuracy for this dataset, of 81.33-86.66%. The 3ClassBDS gave a much lower prediction level than the 2ClassBDS and the best result for this dataset was achieved using the J48 algorithm, with an accuracy range of 56.00-73.33%.

Further work that could be done to help with the progress

Fig. 1. Graph showing dataset 2CLASSBDS best results for all algorithms
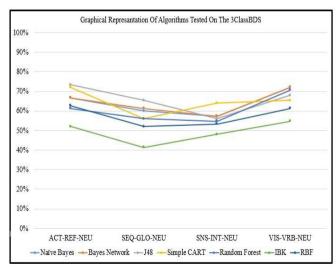

Fig. 2. Graph showing dataset 3CLASSBDS best results for all algorithms

of this research could be to collect more data. As the dataset is very small currently (75 records), it is not sufficient to be able to produce an accurate reading. The larger the dataset the more the algorithms will be able to learn about the dataset. This was noticeably visible in the 3ClassBDS, as it could be seen by the very low or in some cases 0 class% that were present in the dimensions, how little the algorithm was able to learn about each class. By increasing the number of records to be tested, the learning of the algorithms for each dimension is also increased and the results will be a better reflection of the findings.

Furthermore, due to the uncertain nature of the data, fuzzy decision trees could be used to model the Neutral class, as currently this class in not conclusive of anything. In the 2ClassBDS, it can be seen that the accuracy predictions are very high values, but this accuracy is lowered significantly in the 3ClassBDS. This is due to the introduction of the Neutral class in the 3ClassBDS, representing the fuzziness of the data. For the 2ClassBDS, the experiments have shown that prediction is possible, but in the 3ClassBDS this prediction is not as clear, due to the Neutral class.

Finally, to improve the research, the best results for the 2ClassBDS can be taken and the rules can be implemented in a CITS to monitor the outcome and see if it produces the desired result of determining the learning dimensions of each student. Since the datasets used in this research was obtained by Oscar CITS which was tested for engineering students, the results favours the learning habits for these types of students. It will be interesting to compare these results with those of a CITS designed for students of other courses, such as business or art students who have different learning habits than those of engineering students.

REFERENCES

[1] K. Crockett, A. Latham, D. McLean and J. O'Shea, "A fuzzy model for predicting learning styles using behavioural cues in a conversational intelligent tutoring system," *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on,* 2013, pp. 1-8.

[2] A. Latham, K. Crockett, D. McLean and B. Edmonds, "A conversational intelligent tutoring system to automatically predict learning styles," Computers and Education, 59(1), 2012, pp. 95-109.

[3] R. M. Felder and L. K. Silverman, "Learning and teaching styles in engineering education," Engineering Education, 78(7), 1998, pp. 674-681.

[4] A. Latham, K. Crockett and D. McLean, "Profiling student learning styles with multilayer perceptron neural networks," Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, 2013, pp. 2510-2515.

[5] P. Garcia, A. Amandi, S. Schiaffino and M. Campo, "Evaluating bayesian networks' precision for detecting students' learning styles, Computers & Education, 49(3), 2007, pp. 794-808.

[6] D. Kelly and B. Tangney, "Predicting learning characteristics in a multiple intelligence based tutoring system" In Intelligent Tutoring Systems, Lester C. James, Rosa M. Vicari and Fabio Paraguacu (eds.). Springer Berlin Heidelberg, 2004, pp. 678-688.

[7] J. Feldman, A. Monteserin and A. Amandi, "Automatic detection of learning styles: state of the art," Artificial Intelligence Review, 2014, pp. 1-30.

[8] A. M. Latham, "Personalising learning with dynamic prediction and adaptation to learning styles in a conversational intelligent tutoring system," Ph.D. Dissertation, Manchester Metropolitan University, Manchester, 2011.

[9] International Centre for Educators' Learning Styles, "The Felder-Silverman learning and teaching styles model," 2014, Retrieved June 18, 2015 from http://www.icels-educators-for-learning.ca/index.php?option=com_content&view=article&id=47&Itemid=64

[10] The University of Waikato, "Data mining with Weka, Transcripts," 2014, Retrieved October 13, 2015 from http://www.cs.waikato.ac.nz/ml/weka/mooc/dataminingwithweka/transcripts/Transcript2-5.txt

[11] Microsoft, Data mining algorithms (analysis services – data mining)," 2015, Retrieved July 9, 2015 from https://msdn.microsoft.com/en-us/library/ms175595.aspx

[12] R. Schapire, "Machine learning algorithms for classification," 92-page hand out, Algorithms and Data Structures, New Jersey: Princeton University, 2011.

[13] C. A. Ratanamahatana and D. Gunopulos. "Scaling up the naive Bayesian classifier: using decision trees for feature selection," 2002, Retrieved July 11, 2015 from http://alumni.cs.ucr.edu/~ratana/DCAP02.pdf

[14] K. Gibert, M. Sanches-Marre and V. Codina, "Choosing the right data mining technique: classification of methods and intelligent recommendation," Ph.D. Dissertation, International Environmental Modelling and Software Society, 2010, Retrieved August 3. 2015 from http://iemss.logismi.co:8080/xmlui/handle/iemss/10283

[15] Bayes Nets, "Bayes nets," 2007, Retrieved July 9, 2015 from http://www.bayesnets.com/

[16] D. Xhemali, C. J. Hinde and R. G. Stone, "Naïve bayes vs. decision trees vs. neural networks in the classification of training web pages," International Journal of Computer Science, 4(1), 2009, pp. 16-23

[17] H. Garner, "Clojure for data science," Packt Publishing Ltd, 2015

[18] H. J. Hamilton, "C4.5 Tutorial - knowledge discovery in databases," Saskatchewan: University of Regina, 2012, Retrieved July 12, 2015 from http://www2.cs.uregina.ca/~hamilton/courses/831/notes/ml/dtrees/4_dtrees2.html

[19] A. Padhye, "Chapter 5 – classification methods," 2006, Retrieved July 11, 2015 from http://www.d.umn.edu/~padhy005/Chapter5.html

[20] R. Timofeev, Theory and applications, Master Thesis, Centre of Applied Statistics and Economics, Humboldt University, Berlin, 2004.

[21] V. Rao, "Introduction to classification and regression trees (CART)," 2013, Retrieved July 11, 2015 from http://www.datasciencecentral.com/profiles/blogs/introduction-to-classification-regression-trees-cart

[22] B. Li, "Predictive modelling: decision trees: the cart algorithm," 18-page hand out, Data mining in business, China: Southwest University of Finance and Economics, 2006.

[23] L. Breiman and Adele Cutler, "Random forests," Salford Systems, 2004, Retrieved July 12, 2015 from https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

[24] N. Horning, "Introduction to decision trees and random forests," 27-page hand out, Indonesia: Woods Hole Research Centre, 2010.

[25] D. W. Aha, Dennis Kibler and Marc K. Albert, "Instance-based learning algorithms," Machine Learning, 1991, 6(1), pp. 37-66.

[26] D. Benyamin, "A gentle introduction to random forests, ensembles, and performance metrics in a commercial system," 2012, Retrieved July 12, 2015 from https://citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics/

[27] N. Adel, "Prediction of learning style dimensions using natural language dialogue from a conversational intelligent tutoring system," MSc Dissertation, Manchester Metropolitan University, Manchester, 2015.