Data Mining Techniques to Study Therapy Success with Autistic Children

Gondy Leroy, Annika Irmscher, Marjorie H. Charlop-Christy, Sarah Kuriakose, Alefiyah Pishori, Lorryana Wurzman, Aria Ash-Rafzadeh, Michelle Seffrood, Stephanie Buxton

***Abstract*— Autism spectrum disorder has become one of the most prevalent developmental disorders, characterized by a wide variety of symptoms. Many children need extensive therapy for years to improve their behavior and facilitate integration in society. However, few systematic evaluations are done on a large scale that can provide insights into when, where, and how therapy has an impact. We describe how data mining techniques can be used to provide insights into behavioral therapy as well as its effect on participants. To this end, we are developing a digital library of coded video segments that contains data on appropriate and inappropriate behavior of autistic children in different social settings during different stages of therapy. In general, we found that therapy increased appropriate behavior and decreased inappropriate behavior. Decision trees and association rules provided more detailed insights for high and low levels of appropriate and inappropriate behavior. In particular, we found that a child's interaction with a parent or therapist led to especially high levels of appropriate behavior and behavior is most predictable while therapy is in progress.**

Keywords: Autism, data mining, decision trees, association rules

I. Introduction

AUTISM spectrum disorder (ASD) has become one of the most prevalent developmental disorders over the last few years and its prevalence is still growing at a rate of 10-17% per year (Autism Society of America)[1]. It is a serious mental disorder that afflicts more than 500,000 children in the United States and is more common than childhood cancer or Down's syndrome [2]. The disorder is characterized by a wide variety of possible symptoms such as developmental disabilities, extreme withdrawal, lack of social behavior, severe language and attention deficits, and repetitive behaviors. The symptoms' intensity ranges from almost unnoticeable to very severe. Because of this wide variety of symptoms and intensity, therapy needs to be individualized for every person.

Therapeutic interventions that leverage information technology (IT) are still in their infancy in this area. Therapists and researchers alike point to the need for more research and applications to help integrate autistic persons in society and help them and their families achieve a higher quality of life. To address this issue Moore, McGrath, and Thorpe present a framework for research and development to facilitate computer-aided learning for people with autism [3]. They argue that it would be most beneficial if computer-aided learning software would focus on three main impairments: social and interpersonal skills, communication difficulties, and rigidity of thinking. Most of the current research and development can be classified into these three groups.

Social and interpersonal skill training refers to helping people with autism understand why other people behave as they do. Bishop [4] is working on a tool that can help an autistic person better understand social situations by using a mobile phone to provide translations for idioms such as "cat got your tongue?". This phrase does not make sense when taken literally, which many autistic persons would do. In a small study they evaluated how the system would be received and found that people with autism would find it useful.

Others focus on the second impairment: communication. The use of information technology varies from approaches that use low complexity technology, from the therapist's point of view, to high complexity. For example, Dauphin et al [5] used PowerPoint slides with video segments to teach sociodramatic play. Others trained autistic people to point to pictures to aid communication [6]. This approach has been taken a step further by Miller et al [7] who developed a communication system for use with personal digital assistants. Even more technology-intensive are simulations used to teach verbal communication [8] or social robots for diagnosis and treatment [9].

Finally, there are also attempts to address the third major impairment: rigidity of thinking. Rajendran and Mitchell [10] tried to improve interpersonal understanding and understanding of mental states of

G. Leroy is with the School of Information Systems and Technology, Claremont Graduate University, Claremont, CA, 91711 (909-607-3270; fax: 90-621-8564; e-mail: gondy.leroy@ cgu.edu).

A. Irmscher is with the School of Information Systems and Technology, Claremont Graduate University, Claremont, CA, 91711 (e-mail: annikairmscher@gmail.com).

M.H. Charlop-Christy is with the Department of Psychology, Claremont McKenna College, Claremont, CA 91711 (e-mail: marjorie.charlop-christy@claremontmckenna.edu).

others with the Bubble Dialogue program. They trained children with Asperger's syndrome, i.e., high-functioning individuals with autism. Unfortunately, this training was not successful.

The most popular types of interventions use information technology to interact directly with autistic children. However, compared to fields such as biology or commerce, information technology is not used as much to capture and analyze the impact of interventions on a large scale, i.e. the impact of therapy on behaviors. Large-scale systematic studies could help specify characteristics of the syndrome and explain the wide variety of these characteristics; it could also help provide profiles for autistic children and identify the therapies that are most suitable for them. To our knowledge, few researchers have used advanced data mining to study systematic differences, or the lack thereof, in therapy outcome for autistic children.

II. Data mining For Therapy data

A. Problem Description

Most of the current research that leverages information technology focuses specifically on the use of IT to train autistic children or to improve communication with them. This research is of tremendous value to caregivers and autistic children to improve their quality of life. However, it focuses on children already assigned to a particular therapy.

Additional research is needed to find the cause and optimal matching of children to therapies. Data mining is optimal for this. Two approaches are possible. The first is the use of data mining to find biological associations, such as specific genes related to autism [11]. This type of data mining research is similar to data mining in molecular biology and genetics. Data mining can also help fine-tune diagnosing and find underlying causes for drug related treatments. However, a third approach, mining the actual behaviors, may provide more specific insights into when specific therapies work, which behavioral profiles exist, and how these two factors may interact. It may also help fine-tune the autism diagnosis. Later, these should be combined.

Our work focuses on the use of technology to study the effects of behavioral therapy. Data mining techniques can be used on behavioral therapy data from two perspectives: to find what is characteristic for one or a group of children (behavioral profile) and to discover what is characteristic for a specific therapy approach (therapy profile). Ideally, the two will be combined. Sherer and Schreibman [12] followed the first approach (behavioral profile) and developed profiles of responders and non-responders based on average frequencies of specific behaviors, such as toy play. We will use data mining techniques to examine how behavior changes for a group of children receiving a specific therapy (therapy profile). Our goal is to shed light on the effect that therapy has on appropriate and inappropriate behavior observed in autistic children.

B. Data Mining Techniques

Data mining techniques can be grouped into several categories depending on the goal of the application (description, classification/prediction, or optimization) or the paradigms used (connectionist, evolutionary, symbolic). Not all approaches are suitable for our problem since it is our goal is to provide insights on the effect of therapy. Both therapists and other caregivers should be able to interpret the results. The goal is not to predict on-the-fly when a child will behave appropriately or not. Instead, our goal is to prove human-understandable knowledge. This eliminates the use of black box approaches, such as a feetforward backpropagation network. In addition, datasets are small because they require correctly hand-coded video segments, a labor-intense and time-consuming process. We limit ourselves therefore to symbolic machine learning, which results in clear, human understandable output. In particular, we selected decision trees and association rules as our approach.

We focused on the increase of appropriate behavior and the decrease of inappropriate behavior by the autistic children in different conditions. Two general data mining characteristics can contribute to better insight in the therapy's effect. The first characteristic is the actual picture provided. For example, if we present a set of rules of conditions that lead to mostly inappropriate behavior, therapist get more detailed information, which may lead to additional focus and caregivers may get a better idea of when to expect inappropriate behavior. Moreover, comparing such rules over different stages during therapy is informative. The second characteristic is the accuracy of the prediction. It does not only tell the quality of the prediction, but increased accuracy with ongoing therapy indicates that the children's behavior becomes more regular.

We chose to use both decision trees and association rules because they provide human-understandable output and because they provide easy to understand accuracy measures for therapists.

Decision trees contain a set of rules that are hierarchically combined. The rules are used to categorize data elements into pre-specified groups. The rules consist of a combination of attributes that lead to a specific outcome. When a tree has been learned, the final decision tree can be rewritten as a set of conditional rules [13]. We selected ID3, originally developed by Quinlan [14, 15] as our approach. ID3 is a greedy algorithm that chooses attributes to become part of a decision tree depending on how well they reduce uncertainty (entropy reduction) of labeling an instance with a particular class value. At each node in the tree, remaining attributes are tested to find the one that reduced uncertainty the most. This attribute is used to continue growing the decision tree. By dividing the dataset into training and test sets, the accuracy of a tree can be measured.

Association rule mining [16] can also provide understandable predictions of outcome. The goal of association rule mining is to discover if-then rules that describe associated data points. However, different measures are used to evaluate the rules: support and confidence. The support of a rule is the strength of the rule compared to all other discovered rules (how often does it apply in the dataset). The confidence of a rule is a measure of certainty that the right hand side element of the rule is associated with the left hand side (how accurate is the rule).

III. PREDICTING AUTISTIC BEHAVIOR

A. Data Set

The Claremont Autism Center provides behavioral treatment to children diagnosed with ASD. This is done through direct treatment of the children and through training of their parents. The therapy focuses on development of appropriate behaviors and the elimination of inappropriate behaviors.

To evaluate the effectiveness of therapy as well as for research purposes, the Center maintains a library of video recordings of all children in different social settings, during different stages of therapy. All data is collected with parental consent. Children are videotaped before therapy to provide a baseline, during their therapy, and after finishing therapy. Four different types of conditions are videotaped. The child can be 1) on his or her own in the playroom or can be accompanied by 2) a parent, 3) a therapist, or 4) a stranger (a person working at the Center whom the child does not know). When someone accompanies the child, this person can either leave the child alone or engage him or her interactively. Therapists are not present before treatment has started. Table 1 provides an overview of the conditions and treatment stages.

The development of this digital library of behavioral data has started recently. For each condition, several sessions were videotaped. Two therapists evaluate each session for the presence of eight behaviors. All therapists are trained in advance to achieve sufficient co-rater reliability. The individual sessions are randomized to avoid bias as much as possible and the raters do not have knowledge of the therapy stage they are watching. The raters count the presence of eight specific behaviors as they appear in each setting. There are four appropriate (or desired) behaviors: social behavior, appropriate play, appropriate speech, and compliance. There are four inappropriate (or unwanted) behaviors: non-compliance, inappropriate speech, self-stimulation, and disruptive behavior. These individual behaviors are tallied.

Therapists and parents can clearly see that appropriate behaviors increase and inappropriate behaviors decrease as a result of therapy. However, until now, few other details were available. Our goal here is not only to provide general overview data (e.g., how much appropriate behavior occurs), but also to evaluate the conditions in which this occurs and with what kind of regularity. Although there exists a wide variety of individual differences that make it difficult to generalize, our goal is to find trends in behavior or interesting regularities not previously known. This will help fine-tune therapy, e.g., answer questions such as where can improvements be made or when is therapy currently most effective. It will also be beneficial for parents, to help them evaluate what they get out of therapy and see how their child compares against general expectations.

For each video session, therapists record the presence of each of the four appropriate and four inappropriate behaviors. This is recorded as a present (1) or not present (0) value per 10-second interval. There is no more detailed count per such interval. For example, if a child displays non-compliance twice during the 10-second interval, this is only recorded once as "present." All information is captured during video playbacks. Personal, identifying information is removed from the data before it is transferred to a centralized database.

B. Preprocessing

Data mining requires substantial preprocessing of data. This was especially the case for the behavioral therapy data. Autistic children present very different profiles. To make the data comparable between children, all data needs to be normalized.

For the work presented here, we combined the data of 8 children. The number of data elements per child differs since coding of the therapy sessions is ongoing. Although behaviors differ enormously between children, we decided to combine their data for two reasons. The first reason is theoretical: we want to be able to generalize findings for all children with ASD for this therapy if at all possible. The second reason is practical: by combining data, we increase the number of data points, which makes data mining possible.

To prepare the data, we needed to preprocess the counts to make them comparable between children. First, we combined the behavior counts per one-minute intervals for each child. This provides us with a total count of occurrences per minute instead of only the presence/absence per 10-second interval. Table 1 provides an overview of the total number of video minutes (almost 9 hours) coded and integrated in the central database for the different conditions and stages of therapy.

TABLE I
AVAILABLE VIDEO MINUTES PER EXPERIMENTAL CONDITION

		Treatment Stage		
Exp. Conditions		Pre	In-Process	Post
Alone		35	30	20
Therapist	Non-Inter.	-	40	15
	Inter.	-	40	20
Parent	Non-Inter.	40	40	20
	Inter.	35	35	20
Stranger	Non-Inter.	35	35	20
	Inter.	30	35	15
	TOTAL:	175	225	130

During the second preprocessing step, we combined the counts of the different types of appropriate and inappropriate behavior. For example, if a child displayed appropriate speech 3 times in a minute and compliance 2 times in that minute, this score would be combined into 5 counts of appropriate behavior. This provides us with overview data of the children's behavior for the different conditions. Later, we will look at the eight behaviors individually.

To make the data comparable between children, we normalized the counts. For each child we looked up the maximum number of appropriate or inappropriate behaviors. Then, each count for that child was divided by this maximum number, resulting in scores for appropriate and inappropriate behavior between 0 and 1. For example, if a child displayed appropriate behavior at most 10 times in any one-minute interval, then we divided each count of appropriate behavior per 1-minute interval by 10. We labeled each score as either a high or low level of (in)appropriate behavior. A low score was given to normalized scores below the average of normalized (in)appropriate behavior for that child. This label can be thought of as an indication of when a child displayed more or less (in)appropriate behavior than it did on average.

C. General Results

We first looked at the different levels of (in)appropriate behavior during the different treatment stages. Figure 1 provides an overview of how the proportion of behavior type changes. Before treatment started (PRE), 28% of all recorded behaviors were inappropriate. However, during therapy (IN PROCESS), only 9% of all recorded behavior was inappropriate. After therapy (POST), only 10% of the behavior was inappropriate.

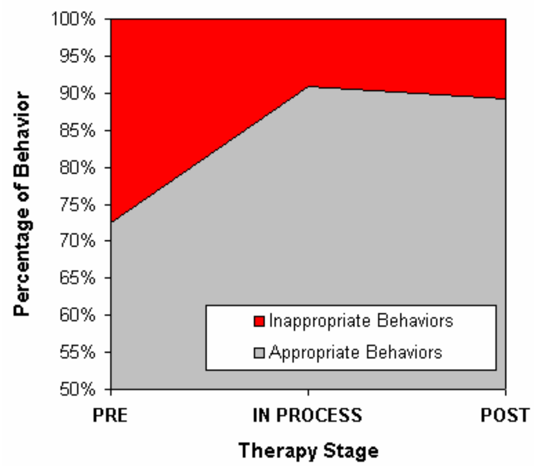

Fig. 1. Overview of the Effect of Therapy on Appropriate and Inappropriate Behavior

D. Decision Trees - Results

The children's improved behavior was not a surprise. The therapists as well as the children's parents witnessed these differences. However few, if any, studies have looked at such behavioral data in detail. We used ID3 to build the decision trees with the Weka open source package [17]. To evaluate the accuracy of the tree, we used 10-fold cross-validation. Our data contains two main attributes: interaction/noninteraction and the person accompanying the child.

1) Predicting Appropriate versus Inappropriate Behavior

A first set of analyses focused on predicting if either appropriate or inappropriate behavior would occur in each of the different settings. Even though we normalized all data and could combine the data for the eight children, no good decision tree could be found. This was probably due to the large variances of this small dataset. Such large differences between children are typical for ASD.

2) Predicting High versus Low Levels of Appropriate or Inappropriate Behavior

For a second set of analyses, we focused on the prediction of high or low (in)appropriate behavior in different conditions. The tree itself can tell us when we can expect low and high levels of (in)appropriate behavior during different stages of therapy. We also provide a baseline evaluation where the most common label is assigned to all conditions. This baseline allows us to evaluate ID3's performance. Table 2 provides an overview of the results. In the baseline condition, we assigned the most common label (high appropriate behavior) to all instances and evaluate that outcome.

TABLE II
DECISION TREE PREDICTION ACCURACY

% Accuracy		Treatment Stage		
		Pre	In-Process	Post
Appropriate Behavior				
	Baseline	52	55	58
	ID3	69	76	59
Inappropriate Behavior				
	Baseline	60	69	83
	ID3	60	69	83

a) Levels of Appropriate Behavior

For appropriate behavior, the decision tree could predict pre-treatment with 69% accuracy when high or low levels of appropriate behavior would occur. The prediction accuracy increased during treatment (76%) but decreased again after treatment. Post treatment appropriate behavior cannot be better predicted with the decision tree than the majority prediction. However, this is not a negative finding from the therapist's perspective, since it is mostly appropriate behavior (see Figure 1).

There is little variance among the trees when comparing the different stages during treatment. Figure 2 shows decision trees for appropriate behavior. Only those nodes that lead to a prediction are displayed. For example, a decision rule based on the treatment IN-PROCESS tree would be that *if there is interaction with the child and the person interacting is the parent, then appropriate behavior is high*. The accuracy is higher for rules from the IN-PROCESS tree, but the tree is almost identical to the PRE-treatment tree. The POST treatment tree is different from the other two and displays the person accompanying the child as the first, most important characteristic (ID3 is a greedy algorithm).

b) Levels of Inappropriate Behavior

We performed a similar analysis for the different levels of inappropriate behavior. However, in this case building a decision tree or simply predicting low inappropriate behavior (baseline) gives the same accuracy in each condition as can be seen in Table 2.

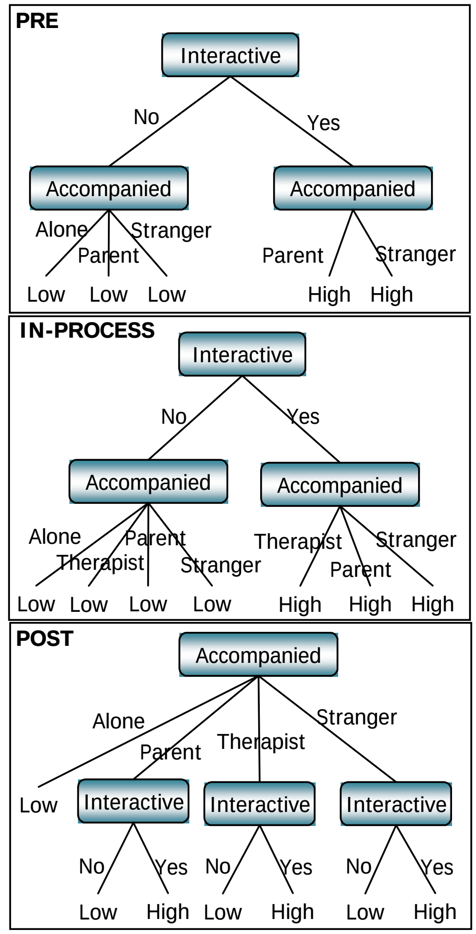

Fig. 2: Decision Trees for Appropriate Behavior Before (Pre), During (In-Process), and after (Post) Therapy.

E. Association Rules - Results

Since the decision trees only provide accurate pictures for two conditions, we had another look at the datasets using association rule mining. For each condition, we considered interesting only those rules where support was minimally 10%, where confidence was at least 90%, and when high or low levels of behavior where predicted. We decided to continue the evaluation of high and low levels of appropriate and inappropriate behavior so we could confirm and fine-tune results found with decision trees.

a) Levels of Appropriate Behavior

We did not find rules with minimal support and confidence (as described above) for the PRE or POST treatment dataset. However, in the IN-PROCESS dataset, we found three rules that met our minimal requirements:

If Parent & Interactive then high appropriate behavior (Confidence = 0.97)

If Alone then low appropriate behavior (Confidence = 0.97)

If Therapist & Interactive then high appropriate behavior (Confidence = 0.95)

These three rules confirm parts of the information found with the decision trees.

b) Levels Inappropriate Behavior

Although decision trees could not shed light on inappropriate behavior, association rule mining found two interesting rules for POST treatment inappropriate behavior:

If Alone then low inappropriate behavior (Confidence = 0.9)

If Stranger & Non-Interactive then low inappropriate behavior (Confidence = 0.9)

No regularities leading to interesting rules could be found in PRE and IN PROCESS conditions.

IV. Discussion

In general, behavior improved during therapy and this change was sustained after treatment. The behavior by autistic children can be very irregular, which was reflected in few classification rules being found. However, behavior was more systematic during treatment. Both decision trees for different levels of appropriate behavior before and during treatment are fairly accurate. From these decision trees, it seems that interacting with the child is a strong force in bringing out good behavior. The post treatment decision tree was based on the smallest data set and this may have caused the lack of predictive power. The rules found with association rule mining confirm the strongest results for predicted levels of appropriate behavior and add details for predicted levels of inappropriate behavior. Being alone brings low levels of appropriate behavior but also low levels of inappropriate behavior. In addition, we found that when a stranger is present who does not interact with the child, inappropriate behavior will be low.

V. Conclusion and Future Directions

The data presented here is based on approximately 9 hours of videotaped sessions of 8 children diagnosed with ASD. For each child, four types of appropriate and 4 types of inappropriate behavior were counted per 10-second sessions. We combined and normalized data for all the children. Therapy was effective in reducing inappropriate behavior and increasing appropriate behavior. In particular, interaction with the child brought out appropriate behavior without a similar increase in inappropriate behavior.

We have only scratched the surface of what is possible with data mining of behavioral datasets. In the near future we will complete the coding of all video sessions. With more data, we believe we will be able to find decision trees that are more accurate in predicting when (in)appropriate behavior is most likely to occur with high or low frequency. We will also look at the eight behaviors individually and how these are affected by the settings. Finally, we will look into different profiles of children to see if some of them respond especially well or not so well to certain conditions and the therapy. This will help evaluate children early on and help adjust their therapy according to their needs at a much earlier stage of treatment. In addition, we will optimize the preprocessing treatments and the data mining algorithms to be fine-tuned for these types of behavioral datasets with such a wide variety of scores per individual.

Additional research into automated coding of video data would be very helpful to gather the necessary data, which is currently a time-consuming and labor-intensive process.

Acknowledgement

We would like to thank the parents of all the children at the Claremont Autism Center who graciously gave their consent to the videotaping and let us use this data for research.

References

[1] L. A. Hogdon, *Visual Strategies for Improving Communication: Practical Support for School and Home*. Troy, Michigan: QuirkRoberts Publishing, 2004.

[2] N. C. o. B. D. a. D. Disabilities, *About Autism, Autism Information Center*: Center for Disease Control, 2005.

[3] D. J. Moore, P. McGrath, and J. Thorpe, "Computer Aided Learning for people with autism - a framework for research and development," *Innovations in Education and Training International*, vol. 37, pp. 218-228, 2000.

[4] J. Bishop, "The Internet for educating individuals with social impairments," *Journal of Computer Assisted Learning*, vol. 19, pp. 546-556, 2003.

[5] M. Dauphin, E. M. Kinney, and R. Stromer, "Using Video-Enhanced Activity Schedules and Matrix Training to Teach Sociodramatic Play to a Child with Autism," *Journal of Positive Behavior Interventions*, vol. 6, pp. 238-250, 2004.

[6] M. L. Israel, L. Ruthel, M. Bates, and N. Smith, "Software to teach nonverbal persons with severe autism and retardation to communicate by pointing to pictures," presented at Johns Hopkins National Search for Computing Applications to Assist Persons with Disabilities, 1992.

[7] T. Miller, G. Leroy, J. Huang, S. Chuang, and M. H. Charlop-Christy, "Using a Digital Library of Images for Communication: Comparison of a Card-Based System to PDA Software," presented at First International Conference on Design Science Research in Information Systems and Technology, Claremont, CA, 2006.

[8] O. E. Hetzroni and J. Tannous, "Effects of a Computer-Based Intervention Program on the Communicative Functions of Children with Autism," *Journal of Autism and Developmental Disorders*, vol. 34, pp. 95-113, 2004.

[9] B. Scassellati, "Quantitative metrics of social response for autism diagnosis," presented at IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN), Nashville, TN, USA, 2005.

[10] G. Rajendran and P. Mitchell, "Computer mediated interaction in Asperger's syndrome: the Bubble Dialogue program," *Computers and Education*, vol. 35, pp. 189-207, 2000.

[11] A. L. Yonan, A. A. Palmer, K. C. Smith, I. Feldman, H. K. Lee, J. M. Yonan, S. G. Fischer, P. Pavlidis, and T. C. Gilliam, "Bioinformatic analysis of autism positional candidate genes using biological databases and computational gene network prediction," *Genes, Brain and Behavior*, vol. 2, pp. 303-320, 2003.

[12] M. R. Sherer and L. Schreibman, "Individual Behavoiarl Profiles and Predictors of Treatment Effectiveness for Children with Autism," *Journal of Consulting and Clinical Psychology*, vol. 73, pp. 525-538, 2005.

[13] P. Langley and H. A. Simon, "Applications of Machine Learning and Rule Induction," *Communications of the ACM*, vol. 38, pp. 55-64, 1995.

[14] J. R. Quinlan, "Discovering Rules by Induction from Large Collections of Examples," in *Expert Systems in the Micro-electronic Age*, D. Michie, Ed. Edingburgh: Edingburgh University Press, 1979, pp. 168-201.

[15] J. R. Quinlan, "Learning Decision Tree Classifiers," *ACM Computing Surveys*, vol. 28, pp. 71-72, 1996.

[16] R. Agrawal and R. Srikant, "Mining Sequential Patterns," presented at Eleventh International Conference on Data Engineering, 1995.

[17] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java*. San Francisco: Morgan Kaufmann, 2000.