Journal of Natural Sciences and Engineering, Vol. 2, (2020)
DOI number: 10.14706/JONSAE2020215

Machine Learning in Autism Spectrum Disorder Diagnosis

Naida Nalo[1], **Jasmin Kevrić**[1]

[1]International Burch University, Sarajevo, Bosnia and Herzegovina

naida.nalo@stu.ibu.edu.ba

jasmin.kevric@ibu.edu.ba

***Abstract*— This paper represents an overview of Machine Learning techniques used in Autism Spectrum Disorder - ASD diagnosis. ASD is detected based on behavioral screening which is time consuming and can only be taken by a medical professional. The idea is to find a smaller number of features that are still able to equally well provide satisfying results and not lose the accuracy, sensitivity nor specificity. Some of the algorithms mostly used in recent studies were Artificial Neural Network - ANN and Alternating Decision Trees - ADTrees. The researches usually use WEKA software package for applying the algorithm and obtaining results.**

***Keywords—Machine Learning, Autism Spectrum Disorder, diagnosis, features, ANN, ADTree, WEKA*.**

1. Introduction

Autism Spectrum Disorder (ASD) is defined as a developmental disorder that reflects in difficulty to communicate and interact with people, to have minimal interests and to generate pattern like behaviors. There are three separate conditions that are combined into the Autism Spectrum Disorder and those are: Autistic disorder, Asperger's syndrome and Pervasive developmental disorder not otherwise specified (PDD-NOS). Even though ASD can be a lifelong disorder, an early diagnosis and proper treatment can help improve communications skills and overall ability to function. However, the diagnosis sometimes takes a lot of time, which prolongs the appropriate treatment [1].

Two widely used clinical diagnosis tools for diagnosing autism are The Autism Diagnostic Interview-Revised (ADI-R) [2] and Autism Diagnostic Observation Schedule (ADOS) [3]. ADI-R consists of ninety-three questions that are to be answered by a clinician. This process can take up to two and a half hours to conduct. The ADOS contains four modules, that can be used to test children and adults, according to behavioral and language levels of the person to be tested. This tool uses an algorithm that results with a diagnosis based on the scores of responses. Each module has its own scores [4]. It has become of great importance to find a faster but reliable method of diagnosing ASD, since the earliest treatment gives a greater chance for improvement.

This paper gives insight on the studies conducted in the past on the subject of Machine Learning in ASD diagnosis. A brief review on differences between two versions of Diagnostics and Statistical Manual of Mental Disorders (DSM-IV [5] and DSM-5 [6]) is made.

2. Literature review

The ADI-R is one of two most widely used instruments for behavioral diagnosis of ASD [7]. It is structured in a form of an exam containing ninety-three questions and can be applied to individuals from the age of eighteen months and above. The questions are answered by a trained professional but still take up to two and a half hours to finish. And additionally, the gap between the initial screening and the resulting diagnosis can be around thirteen months, depending on the socioeconomic status of the family [8]. This introduces an additional delay in the early treatment crucial for proper development of the person, especially children. In [9], it was proposed to create an exam that can be conducted in minutes, rather than hours and receive satisfying results.

Machine learning was used to select the right amount of questions, out of initial ninety-three, that would be able to classify the person in either autism or non-spectrum class. In total, fifteen algorithms were tested, and the one that performed the best with the given data was found to be the Alternating Decision Tree (ADTree). This classification algorithm managed to successfully classify all individuals diagnosed with ASD using only eight questions, that were previously tested with a complete set of ninety-three questions of ADI-R and misclassified only one.

However, this research was proved to be unreliable by [10]. This paper brought to attention the importance of understanding both the computational and clinical area before giving any conclusions. The research conducted in [9], was limited by the imbalance of data as well as excluding a big part of it due to missing values. This resulted in only a two-class diagnosis, when originally it should have been three. The middle class, which is the most difficult to identify, that was removed was the ASD Spectrum, leaving only the ASD and Non-spectrum cases. A recommendation from [10] is to use the Unweighted Average Recall – UAR, which is a measure of performance that works better for such unbalanced data and that was used in this paper, when they tried to replicate the work done in [9]. Their results were algorithm dependent and if another algorithm was to be applied to the same data, the number of features would vary.

Another important issue that was discussed in [4] were the differences between results of DSM-IV and the new DSM-5. From [4] we learn that although both of the screening methods abovementioned have shown good sensitivity, specificity and high reliability in experiments, the majority of those studies were based on the DSM-IV rather than the new criteria for diagnosing ASD, the DSM-5. Several studies that were mentioned in [4] had conflicting results when using the two different versions of the manual. This introduces the need to reevaluate the current tools for diagnosis, and to adjust them to the new criteria of diagnosing ASD.

Combining all the research and experiments from the past, and critically analyzing the results, suggestions and advices for the future projects are provided in [11]. The author highlights that none of the screening tools currently in use, have incorporated the machine learning algorithm for diagnosing ASD from the recent studies. Along with the problem of unbalanced data, the overlap of ASD, ADHD and Asperger Syndrome as well as different forms of ASD, represents another obstacle in diagnosing, mentioned in [11]. Most studies simplified the classification process by removing these classes and leaving just the 'Severe Autism' and 'Non-spectrum' as a possible outcome. This of course leads to unreliable classifiers with questionable sensitivity, specificity and accuracy.

3. Problem formulation

For the purpose of this project, in total three datasets have been downloaded from UCI Machine Learning Repository [12]. They deal with data related to ASD screening of three different sets of population: toddlers, adolescents and adults. The data was collected through an application in a form of a quiz [13]. The data sets consist of twenty questions, out of which ten are behavioral features, while the other ten are individual characteristics. The application offers four modules representing the age category for individuals from the age: 12-36 months, 4-11 years, 12-16 years and 17 years and older. Ten questions, that differed depending on the age of the individual, from the application are provided in Table 1, in Appendix 1. Description of the data set is provided in Table 2.

Table 2. Data set characteristics

ASD Screening Data			
MODULE	CHILDREN	ADOLESCENTS	ADULTS
Number of instances	292	104	704
Number of attributes		21	
Missing values		Yes	

The first module is based on current parent-assisted ASD screening tool, the Quantitative Checklist for Autism Toddlers (Q-CHAT), while the remaining three are based on appropriate versions of Autism Spectrum Quotient (AQ), which are considered to be good candidates of diagnosing and were somewhat referred to as a 'red flag'. These screening tools were discussed in [4], where it has been noted that the ten questions can only be used for acknowledging if additional clinical testing is required and is not a definite diagnosis. An analysis that studied these tools is [14]. Next step of this paper will be to investigate the dataset and determine the best way to make the most of it.

4. Machine learning methods

Seven algorithms that were chosen for the process of attribute selection with their brief description are provided in this chapter. Bayes Net's function is to learn the Bayesian networks. This algorithm assumes nominal attributes and no missing values. Search process is done using K2 or TAN algorithm. More sophisticated search methods, used for search, are built on genetic algorithm, hill-climbing, simulated annealing, etc. Search speed can additionally be enhanced by ADTrees [15].

Simple Logistic is an algorithm that builds regression models and fits them with use of LogitBoost and simple regression functions as base learners. Number of iterations are calculated using cross-validation, supporting attribute selection [15].

Decision Stump's function is to build one-level decision trees for sets with a categorical or numeric value. Missing values in this algorithm are dealt with by seeing them as a separate value and creating a third branch from the stump [15].

J48 is an algorithm that creates a pruned or unpruned C4.5 decision tree. C4.5 This algorithm produces a classifier in a form of a decision tree, which can be either a leaf or a decision node. A leaf indicates a class, a decision node specifies a test with one branch and a subtree for every possible outcome of the test [16].

Logistic Model Tree, or LMT, combines two most popular methods of classification: linear logistic regression and tree induction. This algorithm results in not only classification but also in explicit probability estimates of the class. Another advantage of LMT is that it results in a single tree which makes it easier to interpret [17].

Random Forest is an algorithm that combines tree predictors. Each tree is dependent on values of a random vector, which is sampled independently and with same distribution for all trees of the forest. Generalization error of this algorithm depends on strength of individual trees of thee forest and their correlation. As the number of trees grows, the generalizatioon error converges to the limit [18].

REPTree algorithm represents a fast decision tree learner. This algorithm uses information regarding gain or variance and prunes it with reduced-error pruning to build a ecision or regression tree. Values for numeric attributes are only sorted once, which optimized its speed [15].

5. Results

Classification procedure of this research paper was split in two main parts. First part was applying a 10-fold cross-validation to all attributes and all three datasets. Cross-validation of n-folds splits the original dataset into n parts where n-1 parts are used as a train test, while the nth part is used as a test set [19]. Another method used in this part was applying a percentage split of three different values: 50%, 70% and 90%. Percentage split separates the original dataset into train and test according to the chosen percentage.

Both methods were tested using sixteen algorithms, giving in total 64 results for each dataset. The second part of classification involved attribute selection. Algorithms chosen for this part resulted from the first part, since only those that gave 100% accuracy for all three datasets were again used in attribute selection part. A more detailed 10-fold cross-validation results and algorithm performance are presented in Table 3, and the results of percentage-split and algorithm performance is shown in Table 4.

Table 3. 10-fold cross-validation accuracy results

Method	**Cross-validation 10**		
Algorithm	Dataset		
Bayes	**Child**	**Adolescent**	**Adult**
BayesNet	100%	100%	100%
NaiveBayes	98.9726%	98.0769%	97.017%
MultinomialText	51.7123%	60.5769%	73.1534%
BayesUpdateable	98.9726%	98.0769%	97.017%
Functions			
Logistic	95.2055%	95.1923%	97.017%
MultilayerPerceptron	99,6575%	89.4231%	100%
SimpleLogistic	100%	100%	100%
SMO	100%	89.4231%	100%
Lazy			
IBk	88.3562%	90.3846%	94.8864%
Trees			
DecisionStump	100%	100%	100%
HoeffdingTree	100%	99.0385%	99.858%
J48	100%	100%	100%
LMT	100%	100%	100%
RandomForest	100%	100%	100%
RandomTree	93.1507%	80.7692%	96.1648%
REPTree	100%	100%	100%

Table 4. Percentage split accuracy

Method	Percentage Split (50%-50%, 70%-30%, 90%-10%)								
Algorithm	Dataset								
Bayes	**Child**			**Adolescent**			**Adult**		
BayesNet	100%	100%	100%	100%	100%	100 %	100%	100%	100%
NaiveBayes	98.6%	96.5%	96.5%	98.07 %	100%	100 %	98.01 %	98.5%	97%
MultinomialText	51.36 %	55.68 %	41.37 %	61.53 %	54.83 %	50%	74.14 %	74.88 %	80%
BayesUpdateable	98.6%	96.5%	96.5%	98.07 %	100%	100 %	98.01 %	98.57 %	97.14 %
Functions									
Logistic	93.1%	89.7%	93.1%	84.61 %	87.09 %	90%	96.59 %	94.78 %	95.71 %
MultilayerPerceptron	97.2%	97.7%	100%	94.23 %	93.54 %	80%	100%	100%	100%
SimpleLogistic	100%	100%	100%	100%	100%	100 %	100%	100%	100%
SMO	96.5%	95.4%	100%	92.3%	93.54 %	80%	100%	100%	100%
Lazy									
IBk	89.04 %	89.7%	86.2%	88.46 %	90.32 %	100 %	95.73 %	94.31 %	94.28 %
Trees									
DecisionStump	100%	100%	100%	100%	100%	100 %	100%	100%	100%
HoeffdingTree	98.6%	100%	100%	98.07 %	100%	100 %	100%	100%	100%
J48	100%	100%	100%	100%	100%	100 %	100%	100%	100%
LMT	100%	100%	100%	100%	100%	100 %	100%	100%	100%
RandomForest	100%	100%	100%	100%	100%	100 %	100%	100%	100%
RandomTree	93.8%	94.3%	82.7%	67.3%	74.19 %	100 %	100%	90.99 %	100%
REPTree	100%	100%	100%	100%	100%	100 %	100%	100%	100%

Algorithms used for the second part of classifying process of this research paper were chosen according to the percentage of accuracy of Table 3. Out of four Bayes algorithms, only BayesNet gave 100%, SimpleLogistic is the only one out of four Function algorithms that proved the best, and lastly, Tree algorithms shown good results with DecisionStump, J48, LMT, RandomForest and REPTree performing in 100% accuracy for all three datasets. These seven algorithms were used in attribute selection part of classification. All three datasets originally had 21 attributes, and the previous two methods mentioned above included all attributes in the process. Attribute selection method [20] is a process of selecting the most relevant attributes and by doing so, reducing the processing time.

In total, five attribute evaluators have been used in attribute selection process. Those were: ClassifierAttributeEval, CorrelationAttributeEval, ReliefAttributeEval, CfsSubsetEval and WrapperSubsetEval. In a combination of these evaluators, three search methods were used: BestFirst, GreedyStepwise and Ranker [15]. ClassifierAttributeEval evaluates the worth of an attribute with use of a user-specified classifier [21]. CorrelationAttributeEval evaluates the worth of an attribute by measuring the correlation between the attribute and the class [21]. ReliefFAttributeEval sampling of instances happens randomly, and the neighboring instances of the same or different class is checked on [15]. CfsSubsetEval evaluates the worth of a subset of attributes by considering the individual predictive ability of each attribute along with the degree of redundancy between them. Missing values can be seen as a separate value or, with proportion to their frequency, its counts can be distributed among other values [15]. WrapperSubsetEval evaluates attribute sets by using a learning scheme. Cross-validation estimates the accuracy of the learning scheme for a set of attributes [15] .

ClassifierAttributeEval, CorrelationAttEval and ReliefAttributeEval required Ranker as a search method. In all three datasets, number of attributes chosen for the Ranker was five. CfsSubsetEval and WrapperSubsetEval work using either BestFirst or GreedySetpwise search method. Combining the evaluators with search methods, we obtained 56 results for each dataset. After the attribute selection was performed on the complete set of 21 attributes, all evaluators resulted with 100% accuracy, regardless of the algorithm used. The attributes of all three datasets are presented in Table 5.

Table 5. Attributes by number with description

Attribute	Description
1 - 10	Score of 10 questions
11	Age (number)
12	Gender (male or female)
13	Ethnicity (list provided)
14	Born with jaundice
15	Autism in family
16	Country of residence
17	Used app before
18	Result of app (automated calculation)
19	Age description (toddler, child, adolescent, adult)
20	Relation (who is completing the test)
21	Class ASD/NoASD

Attribute that was present in all three datasets and that showed extremely high correlation was the 18th attribute. This attribute represents the score of ten questions of the application [13]. Therefore, a new approach was used. The 18th attribute was removed completely, and the process of selection was repeated for all three datasets. Results of selection are shown in Table 6, 7 and 8, in Appendix 2. Results of accuracy are shown in Table 9, 10 and 11, in Appendix 3.

The lowest performance for child, adolescent and adult dataset was achieved by DecisionStump, resulting in 78.082%, 70.192% and 82.822% respectively, as can be observed from the results. The lowest number of attributes selected is 1, and the highest is 14. However, the best results required less than that. The algorithms that showed best performance for child dataset were SimpleLogistic and LMT. These algorithms, with applied CfsSubsetEvaluator, resulted in accuracy of 98.973%, and used 10 attributes. BayesNet showed best results, with applied ClassifierAttributeEvaluator, it performed in 90.385% accuracy for adolescent dataset and used only 5 attributes. Simple Logistic successfully classified the adult dataset, with impressive accuracy of 99.432% and used 11 attributes in the process.

6. Conclusion

The conclusion is split into two parts, one regarding actions taken to review already written papers and discuss their results, and second which deals with actions taken to derive our own conclusion through processing datasets. This research paper involved three datasets: child, adolescent and adult, with each having 21 attributes. Original datasets were processed using two methods for splitting the dataset into train and test and used sixteen algorithms for both. The obtained results from the first test helped choose algorithms for the second part of testing which involved attribute selection. According to the results, seven algorithms stood out. Attribute selection was performed on all three datasets using seven evaluators. All results had 100% accuracy, despite using different number of attributes. This leaded to another approach which included removing the 18th attribute and reapplying the selection process. Number of attributes for best performances were dependent on the dataset and therefore are different. One should keep in mind that classification process using five attributes can only be used as an indicator of whether further medical testing should be conducted.

The main lesson learned, reading papers written so far on this topic, is that integrating ML in ASD diagnosis and its screening tools is a much harder process than it seems. Finding the appropriate number of features and managing to reduce the time of diagnosis depends on many parameters. Many experiments, in an attempt to reduce the time required for the diagnosis process, have discarded some important issues for the sake of simplicity. Their initially admiring results could not be taken for granted, due to misbalanced data and questionable reliability. The algorithm should not be dependent on data. The issue of distinguishing between ASD and PDD related disorders (ADHD, Asperger Syndrome) represents a big obstacle for proper diagnosis of ASD. The algorithm should be provided with a similar number of all possible outcomes in order to learn to better distinguish between categories. The need to reevaluate the current diagnosis tools and adjust them to fit the new criteria from DSM-5 arises.

APPENDIX 1

Table 1. Questions from the ASDQuiz application [13]

	13-36 months - TODDLER	4-11 years – CHILD	12-16 years – ADOLESCENT	17 & older – ADULT
1.	Does your child look at you when you call his/her name?	He/she often notices small sounds when other do not?	He/she notices patterns in things all the time?	I often notice small sounds when others do not?
2.	How easy is it for you to get eye contact with your child?	He/she usually concentrates more on the whole picture rather than the small details?	He/she usually concentrates more on the whole picture rather than the small details?	I usually concentrate more on the whole picture, rather than the small details?
3.	Does your child point to indicate that he/she wants something (e.g. toy out of reach)?	In a social group, he/she can easily keep track of several different people's conversation?	In a social group, he/she can easily keep track of several different people's conversation?	I find it easy to do more than one thing at once?
4.	Does your child point to share interest with you? (e.g. pointing at an interesting sight)	He/she finds it easy to go back and forth between different activities?	If there is an interruption, he/she can switch back to what he/she was doing very quickly?	If there is an interruption, I can switch back to what I was doing easily?
5.	Does your child pretend? (e.g. care for dolls, talk on a toy phone)	He/she doesn't know how to keep a conversation going with his/her peers?	He/she frequently finds that he/she doesn't know how to keep a conversation going?	I find it easy to read between the lines when someone is talking to me?
6.	Does your child follow where you are looking?	He/she is good at social chit-chat?	He/she is good at social chit-chat?	I know how to tell if someone listening to me is getting bored?
7.	If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort you/them? (e.g. gives a hug)	When he/she reads a story, he/she finds it hard to work out the character's intentions or feelings?	When he/she was younger, he/she used to enjoy playing games involving pretending with other children?	When I'm reading the story, I find it difficult to work out the character's intentions?
8.	Would you describe your child's first words as: typical, unusual, the child doesn't speak?	When he/she was on preschool, he/she used to enjoy playing games involving pretending with other children?	She/he finds it difficult to imagine what it would be like to be someone else?	I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plants, etc.)
9.	Does your child use simple gestures (e.g. waves goodbye)?	He/she finds it difficult to work out what someone is thinking or feeling just by looking at their face?	He/she finds social situations easy?	I find it easy to work out what someone is thinking just by looking at their face
10.	Does your child stare at nothing with no apparent purpose?	He/she finds it hard to make new friends?	He/she finds it hard to make new friends?	I find it difficult to work out people's intentions?

APPENDIX 2

Table 6. Attribute selection results - child dataset

<table>
<tr><th>Algorithms vs. selected attributes</th><th>ClassifierAttributeEval</th><th>CorrelationAttributeEval</th><th>ReliefAttributeEval</th><th>CfsSubsetEval (BF & Greedy)</th><th>WrapperSubsetEval (BF</th><th>Wrapper, Greedy</th></tr>
<tr><td>BayesNet</td><td rowspan="7">4,9,8,10,1</td><td rowspan="7">4,9,10,8,6</td><td rowspan="7">4,1,10,8,9</td><td rowspan="7">1-10</td><td>1-10,17</td><td>4,10</td></tr>
<tr><td>SimpleLogistic</td><td>3,4,6,7,10,12,16</td><td>4,6,10,12,16</td></tr>
<tr><td>DecisionStump</td><td>4</td><td>4</td></tr>
<tr><td>J48</td><td>1,3,4,5,7,8,10,14</td><td>4,10</td></tr>
<tr><td>LMT</td><td colspan="2">4,6,10,12,16</td></tr>
<tr><td>RandomForest</td><td>1-5,7,8,9,10,15,17</td><td>4,10</td></tr>
<tr><td>REPTree</td><td>1,4,8,10</td><td>4,10</td></tr>
</table>

Table 7. Attribute selection results - adolescent dataset

<table>
<tr><th>Algorithms vs. selected attributes</th><th>ClassifierAttributeEval</th><th>CorrelationAttributeEval</th><th>ReliefAttributeEval</th><th>CfsSubsetEval (BF & Greedy)</th><th>WrapperSubsetEval (BF</th><th>Greedy</th></tr>
<tr><td>BayesNet</td><td rowspan="2">5,4,3,10,6</td><td rowspan="7">5,4,10,3,6</td><td rowspan="7">5,3,10,4,8</td><td rowspan="7">3,4,5,6,7,8, 10,17</td><td>1,2,3,4,5,7,10,12</td><td>1,3,4,5,10,12</td></tr>
<tr><td>SimpleLogistic</td><td colspan="2">3,4,5,10,14</td></tr>
<tr><td>DecisionStump</td><td rowspan="3">5,4,10,3,6</td><td colspan="2">5</td></tr>
<tr><td>J48</td><td colspan="2">2,5,9,14</td></tr>
<tr><td>LMT</td><td colspan="2">3,4,5,10,14</td></tr>
<tr><td>RandomForest</td><td>5,4,3,10,6</td><td>15,7,8,10,17</td><td>2,3,5,8</td></tr>
<tr><td>REPTree</td><td>5,4,10,3,6</td><td colspan="2">4,5,7,10</td></tr>
</table>

Table 8 Attribute selection results - adult dataset

<table>
<tr><th>Algorithms vs. selected attributes</th><th>ClassifierAttributeEval</th><th>CorrelationAttributeEval</th><th>ReliefAttributeEval</th><th>CfsSubsetEval (BF & Greedy)</th><th>WrapperSubsetEval (BF</th><th>Greedy</th></tr>
<tr><td>BayesNet</td><td>9,6,16,19,7</td><td rowspan="7">9,6,5,4,3</td><td rowspan="7">5,9,6,4,7</td><td rowspan="7">1-10,16</td><td colspan="2">1-10,15,16</td></tr>
<tr><td>SimpleLogistic</td><td rowspan="6">9,6,8,7,19</td><td>1,3,5,6,9,12,15</td><td>1,3,5,6,9</td></tr>
<tr><td>DecisionStump</td><td>9</td><td>9</td></tr>
<tr><td>J48</td><td>1-5,7-10,16,15</td><td>1,5,9</td></tr>
<tr><td>LMT</td><td>1,3,5,6,9,12,15</td><td>1,3,5,6,9</td></tr>
<tr><td>RandomForest</td><td colspan="2">1-10,12,15,18,19</td></tr>
<tr><td>REPTree</td><td colspan="2">1,2,3,5,6,7,9,15</td></tr>
</table>

APPENDIX 3

Table 9. Attribute selection accuracy - child dataset

<table>
<tr><th>Attribute evaluator</th><th>BayesNet</th><th>SimpleLogistic</th><th>DecisionStump</th><th>J48</th><th>LMT</th><th>RandomForest</th><th>REPTree</th></tr>
<tr><td>ClassifierAttributeEval</td><td>86.644%</td><td>86.301%</td><td rowspan="6">78.082%</td><td>85.959%</td><td>85.959%</td><td>85.274%</td><td>85.616%</td></tr>
<tr><td>CorrelationAttEval</td><td>84.932%</td><td>84.589%</td><td>84.932%</td><td>84.589%</td><td>81.849%</td><td>85.959%</td></tr>
<tr><td>ReliefAttributeEval</td><td>85.616%</td><td>88.014%</td><td>83.904%</td><td>87.329%</td><td>85.274%</td><td>83.219%</td></tr>
<tr><td>CfsSubsetEval</td><td>95.206%</td><td>98.973%</td><td>91.438%</td><td>98.973%</td><td>93.151%</td><td>83.562%</td></tr>
<tr><td>WrapperSubsetEval (BF)</td><td>91.438%</td><td>95.890%</td><td>85.274%</td><td>95.890%</td><td>89.384%</td><td>84.589%</td></tr>
<tr><td>WrapperSubsetEval (Greedy)</td><td>82.192%</td><td>82.192%</td><td>82.192%</td><td>83.904%</td><td>84.247%</td><td>83.562%</td></tr>
</table>

Table 10. Attribute selection accuracy - adolescent dataset

<table>
<tr><th>Attribute evaluator</th><th>BayesNet</th><th>SimpleLogistic</th><th>DecisionStump</th><th>J48</th><th>LMT</th><th>RandomForest</th><th>REPTree</th></tr>
<tr><td>ClassifierAttributeEval</td><td>90.385%</td><td rowspan="2">89.423%</td><td rowspan="4">70.192%</td><td rowspan="3">80.769%</td><td rowspan="2">89.423%</td><td>86.539%</td><td rowspan="2">75%</td></tr>
<tr><td>CorrelationAttEval</td><td>89.423%</td><td>85.577%</td></tr>
<tr><td>ReliefAttributeEval</td><td>82.692%</td><td>81.731%</td><td>81.731%</td><td>83.654%</td><td>76.923%</td></tr>
<tr><td>CfsSubsetEval</td><td>87.5%</td><td>89.423%</td><td>84.615%</td><td>88.462%</td><td>86.539%</td><td>68.269%</td></tr>
<tr><td>WrapperSubsetEval (BF)</td><td>88.462%</td><td>86.539%</td><td rowspan="2">68.269%</td><td>72.115%</td><td>85.577%</td><td rowspan="2">78.846%</td><td>72.115%</td></tr>
<tr><td>WrapperSubsetEval (Greedy)</td><td>77.885%</td><td>78.846%</td><td>71.154%</td><td>78.846%</td><td>71.154%</td></tr>
</table>

Table 11. Attribute selection accuracy - adult dataset

<table>
<tr><th>Attribute evaluator</th><th>BayesNet</th><th>SimpleLogistic</th><th>DecisionStump</th><th>J48</th><th>LMT</th><th>RandomForest</th><th>REPTree</th></tr>
<tr><td>ClassifierAttributeEval</td><td>90.341%</td><td>89.347%</td><td rowspan="4">84.517%</td><td>88.21%</td><td>89.357%</td><td>89.347%</td><td>87.5%</td></tr>
<tr><td>CorrelationAttEval</td><td>91.051%</td><td>90.909%</td><td rowspan="2">89.063%</td><td>90.909%</td><td>91.761%</td><td>90.199%</td></tr>
<tr><td>ReliefAttributeEval</td><td>90.767%</td><td>90.625%</td><td>90.625%</td><td>90.057%</td><td>89.205%</td></tr>
<tr><td>CfsSubsetEval</td><td>96.307%</td><td>99.432%</td><td>92.046%</td><td>99.006%</td><td>91.193%</td><td>86.364%</td></tr>
<tr><td>WrapperSubsetEval (BF)</td><td>95.313%</td><td>95.881%</td><td rowspan="2">82.822%</td><td>89.347%</td><td>95.028%</td><td>94.46%</td><td>89.063%</td></tr>
<tr><td>WrapperSubsetEval (Greedy)</td><td>92.046%</td><td>93.04%</td><td>89.063%</td><td>93.892%</td><td>91.761%</td><td>89.921%</td></tr>
</table>

7. References

[1] National Institute of Mental Health, Available at: https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd/index.shtml

[2] Lord, C., Rutter, M., & Le Couteur, A. 1994. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism and Developmental Disorders, 24, 659–685

[3] Lord, C, Risi, S, Lambrecht, L et al, 2000. The Autism Diagnostic Observation Schedule-Generic: a standard measure of social and communication deficits associated with the spectrum of autism J Autism Dev Disord. 2000; 30:205–223.7.

[4] F. Thabtah, 2017. Autism Spectrum Disorder Screening: Machine Learning Approach and DSM-5 Fulfillment, ICMHI '17, May 20-22, 2017, Taichung City, Taiwan © 2017 Association for Computing Machinery. ACM ISBN 978-1-4503-5224-6/17/05

[5] American Psychiatric Association 2000. Diagnostic and statistical manual for mental disorder (4th edn), Text revision. Washington, DC: American Psychiatric Association

[6] American Psychiatric Association. 2013. Diagnostic and statistical manual of mental disorders: DSM-5. Washington, D.C: American Psychiatric Association

[7] C. Lord et al., 1994. Autism Diagnostic Interview-Revised: A Revised Version of a Diagnostic Interview for Caregivers of Individuals with Possible Pervasive Developmental Disorders, Journal of Autism and Developmental Disorders, Vol. 24, No. 5, 1994

[8] L. D. Wiggins, J. Baio, C. Rice, 2006, Examination of the Time Between First Evaluation and First Autism Spectrum Diagnosis in a Population-based Sample, Developmental and Behavioral Pediatrics, Vol. 27, No.2, April 2006.

[9] Wall DP, Dally R, Luyster R, Jung J-Y, DeLuca TF (2012) Use of Artificial Intelligence to Shorten the Behavioral Diagnosis of Autism. PLoS ONE 7(8): e43855. doi: 10.1371/journal.pone.0043855

[10] D. Bone, M. S. Goodwin, M. P. Black, et al, 2014. Applying Machine Learning to Facilitate Autism Diagnostics: Pifalls and Promises, Springer Science + Business Media New York 2014

[11] F. Thabtah (2018): Machine learning in autistic spectrum disorder behavioral research: A review and ways forward, Informatics for Health and Social Care, DOI: 10.1080/17538157.2017.1399132

[12] UCI Machine Learning Repository, Available at: https://archive.ics.uci.edu/ml/index.php

[13] Thabtah, F. (2017). ASDTests. A mobile app for ASD screening. www.asdtests.com

[14] C.Allison, B. Auyeung, S. Baron-Cohen, 2012, Toward Brief "Red Flags" for Autism Screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist in 1000 Cases and 3000 Controls, Journal of the American Academy of Child & Adolescent Psychiatry, DOI: 10.1016/j.jaac.2011.11.003

[15] I.H. Witten, E. Frank (2005), Data Mining: Practical Machine Learning Tools and Techniques, Elsevier

[16] J. Ross Quinlan, C4.5: Programs for Machine Learning, Avilable at: https://link.springer.com/article/10.1007/BF00993309

[17] University of Waikato, Logistic Model Trees, Available at:https://www.cs.waikato.ac.nz/ml/publications/2005/LMT.pdf

[18] L. Breiman (2001), Random Forests, Available at: https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf

[19] Towards Data Science, Cross-Validation in Machine Learning, K-Fold Cross Validation, Available at: https://towardsdatascience.com/cross-validation-in-machine-learning-72924a69872f

[20] Machine Learning Mastery, An Introduction to Feature Selection, Available at: https://machinelearningmastery.com/an-introduction-to-feature-selection/

[21] Weka – Version 3.8.3, Waikato Environment for Knowledge Analysis, ClassifierAttributeEval