Optimization of Accuracy to Autism Spectrum Disorder Identification for Children Using Support Vector Machine and Correlation-based Feature Selection

Lika Alaika [1,*], Alamsyah [1]

[1] Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia
*Corresponding author: likaalaika@students.unnes.ac.id

ARTICLE INFO

Article history
Received
Revised
Accepted

Keywords
Autisme
GSA
SVM
CFS
Data Mining

ABSTRACT

Autism is a developmental disorder that affects the function of the neurological system, which can have a negative impact on the sufferer's quality of life. The ratio of people with autism is relatively high and tends to increase, based on WHO in 2013 the ratio of Indonesian children suffering from autism is 1:160 or more than 112.000. Data mining is the one of data processing techniques that works to find patterns and knowledge from big data. One of the data mining techniques is a classification that works to search for models that reveal and estimate previously unidentified classes. SVM is one of the classification algorithms that use data to find the optimal hyperplane. SVM has the advantage to work well on a dataset that cannot be linearly separated. The disadvantage is that it can be challenging to select parameters that are ideal and to determine which ones have an impact and which ones do not. To reduce attribute dimensions, CFS was provided as a feature selection to improve accuracy based on correlation values. The Autistic Spectrum Disorder Screening for Children Dataset from the UCI machine learning repository was used in this research to compare the accuracy of SVM and CFS. The result of this research is the SVM algorithm yields an accuracy rate of 94.91%. When the SVM algorithm is combined with CFS, the accuracy rate rises to 96.61%, representing an improvement in accuracy of 1.7% by using the 17 selected attributes.

1 Introduction

Autism is one of the types of developmental disorders that can lead to a neurological condition that disturbs brain function and has an impact on a person's growth process, communication skills, and social interaction abilities. In general, autism spectrum disorder can be detected in infants as early as 6 months (Ratama & Munawaroh, 2020). Attention Deficit Hyperactivity Disorder (ADHD), severe mental impairment, and even hearing loss are developmental illnesses that are sometimes mistaken for autism (YPAC, 2010). Indications of autism can generally be identified from the existence of a deviation from normal growth and development. For example, the child may have trouble forming social connections and communicating with other kids in their community, which makes them feel separated from their environment (Sunu, 2012).

Based on data from Badan Pusat Statistik (BPS) and data from the World Health Organization (WHO), there is an increase in autism data in Indonesia with a growing ratio, this can prove that autism in Indonesia is still high and has a tendency to increase (Elsabbagh et al., 2012). Given these facts and supported by the rapid development of information technology, it is necessary to process the data to effectively detect autism so that it may be used. Data mining is one method for turning data into information that finds patterns and information from vast amounts of data (Han et al., 2011). Additionally, data mining can be utilized to facilitate the discovery of previously undiscovered information (Connolly & Begg, 2005).

Various techniques can be used in data mining, classification can be used to predict decisions (Listiana & Muslim, 2017). The goal of classification data is to estimate the unknown class through modeling and grouping data (Yulietha et al., 2017). The classification technique works by taking a systematic approach to building a classification model that is used to predict decisions (Sugiharti & Muslim, 2016). Support Vector Machine (SVM) will be used as the classification algorithm in this research. SVM is frequently used in pattern recognition, text categorization, bioinformatics (Huang & Wang, 2006), microarray gene expression data analysis, object detection, data classification, image classification, handwritten digit recognition, and tone recognition, among other fields (Srivastava et al., 2010). SVM is used because it has a higher level of performance effectiveness than the decision tree, KNN, and linear regression algorithms in predicting disease (Charleonnan et al., 2016). The concept of the SVM algorithm is described as an effort to find the best hyperplane that functions as a separator of two data classes in the input space (Maharani et al., 2017), which has the aim of classifying with a function as a separator between two classes using training data and testing data (Kavzoglu & Colkesen, 2009). Based on the application of learning bias derived from statistical learning theory based on optimization theory, SVM utilizes a fictitious space in the form of linear functions in feature space with higher dimensions and is trained using a learning algorithm (Fachrurrozi & Yusliani, 2015).

The advantage of SVM is that it has good performance on data sets that cannot be separated linearly (Widiastuti et al., 2017). The weakness of SVM is that it does not work well on large data with many attributes, it can reduce the performance of the classifier so that it provides a low level of accuracy (Larasati et al., 2019). To optimize the work of the classifier needs to be done by selecting relevant features using feature selection. By eliminating unnecessary features that will have an impact on the accuracy of the results, feature selection attempts to lower the dimensions of attributes (Utomo & Mesran, 2020). Correlation-based Feature Selection (CFS) is used because it has the most stable performance and a higher precision value (Djatna & Morimoto, 2011).

2 The Proposed Method

2.1 Support Vector Machine

Support Vector Machine (SVM) is one of the algorithms that can perform the classification process in data mining and was first introduced by Vladimir Vapnik, Bernhard Boser, and Isabelle Guyon in 1992 (Han et al., 2011). SVM is a relatively new algorithm (Vijayarani & Dhayanand, 2015) and is classified as the latest supervised machine learning technique that works by minimizing generalization errors (Shrivastri & Deshmukh., 2014). SVM is also one of the ten most effective and accurate algorithms used in data mining techniques (Wu & Kumar, 2009).

SVM employs a learning system that uses a fictitious space represented by linear functions in higher-dimensional feature space (Cristianini & Shawe-Taylor, 2000). While linear data will be separated to achieve the best results, non-linear data is mapped into a dimensional space with a higher level (Kumar & Chezian, 2014). The best linear search will be identified in the newly created dimension as a decision boundary or hyperplane, which is helpful for the process of dividing tuples from one class to another (Han et al., 2012). While this is happening, non-linear mapping is categorized into groups that are constrained by hyperplane utilizing crucial training tuples and margins (Wijaya & Muslim., 2016). SVM has helped solve various problems such as text categorization, handwritten digit recognition, tone recognition, image classification, object detection, microarray gene expression data analysis, and data classification (Srivastava et al., 2010).

2.2 Correlation-based Feature Selection

Correlation-based Feature Selection (CFS) is a feature selection that works by evaluating the value of a subset of attributes by considering the individual predictive ability of each attribute based on the value of the level of redundancy between attributes. For estimations CFS relies on correlation coefficients between attributes and classes as well as correlations between attributes, (Karegowda et al., 2010). Additionally, this feature selection method can increase data comprehensiveness, which will enhance prediction performance (Jain & Singh, 2018).

The value of the correlation coefficient is expressed in numbers with a range of -1 to 1, if the value is close to 0 it means the correlation is weaker or there is no correlation, and a result close to 1 means the positive correlation is stronger, and if the result is close to -1, the negative correlation is stronger. This correlation value will help the selection of attributes in CFS by taking into account the relationship of a certain attribute with other attributes and an attribute with a class. The selected attribute is an attribute that has a high correlation value with the class, but the attribute linkage with other attributes is low (Wirawan & Eksistyanto, 2015).

3 Method

In this research, CFS and SVM will be applied to detect autism spectrum disorder for children. The desired results in this research are the accuracy of the proposed method. The flowchart of this research method can be seen in Figure 1.

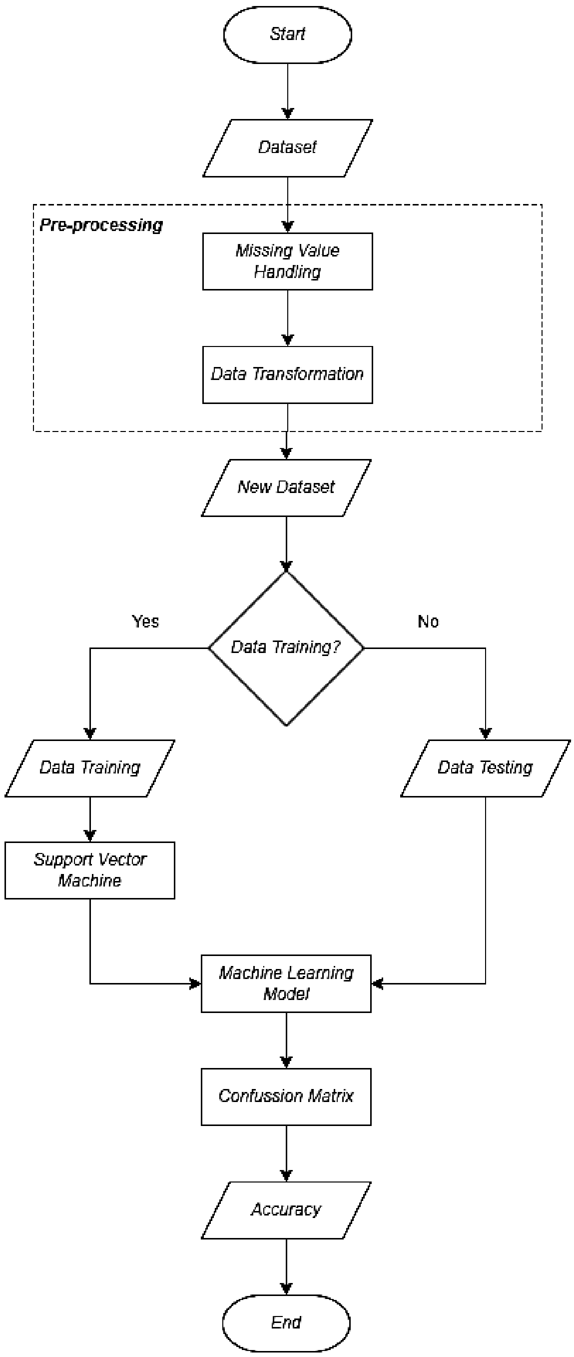

Figure 1. *Flowchart* SVM

Based on the SVM classification stages without applying the CFS, the first process is to retrieve the dataset. After the data is obtained, the data will be preprocessed by replacing the missing value to fill in the blank data and the data transformation process that aims to simplify the process that will be carried out by the Python programming language. After the preprocessing stage is complete, a new dataset will be generated. From the dataset, it will be divided into two parts, with details of 80% of the training data that will be used for SVM classification and 20% of the testing data which will be continued for testing the model to be used. After testing the model, a confusion matrix will be generated, from the confusion matrix it can be calculated and the accuracy value is obtained. After the classification process is complete, it will be followed by merging with feature selection techniques. The combined process of SVM and CFS is shown in Figure 2.

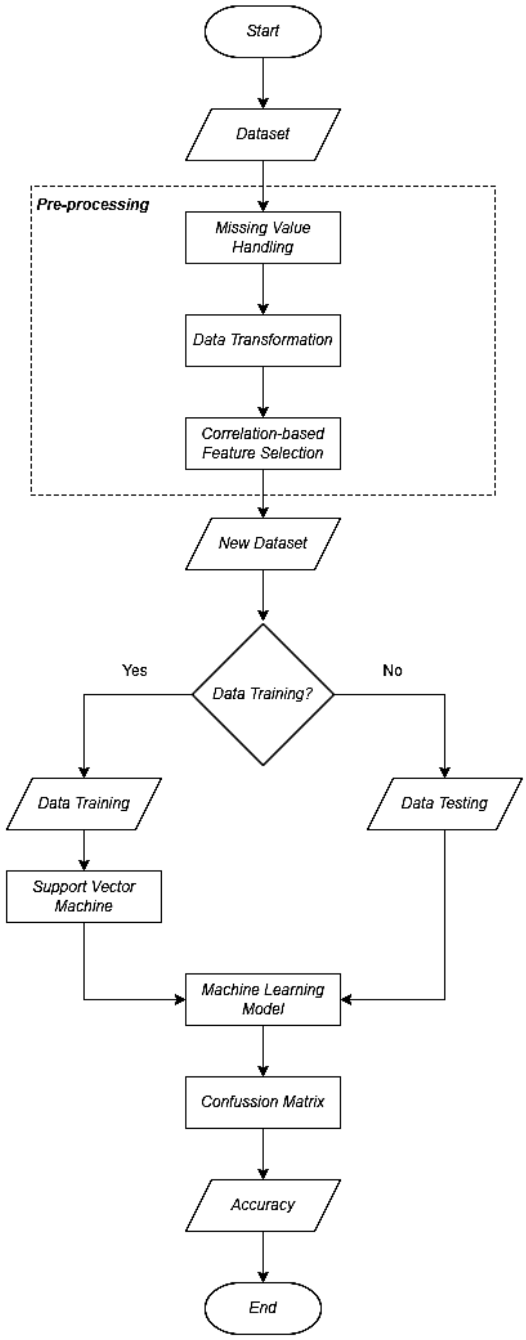

Figure 2. *Flowchart* SVM and CFS

The steps taken in the process of merging SVM and CFS are the same as the stages in the classification process, but the difference is that in the preprocessing stage, CFS feature selection is added so that the new dataset processed in the classification process is a dataset that has been reduced by considering the level of correlation value between attributes and class and between an attribute with another attribute. This process is used to obtain results with a higher accuracy.

4 Results and Discussion

4.1 Result

4.1.1 *Data Collection*

Data collection in this research was carried out by studying literature related to existing problems such as the objects used in this research and the SVM and CFS methods that will be used to help solve existing problems (Adi et al., 2018). The dataset used in this research was uploaded on December 24, 2017 obtained from the UCI machine learning repository entitled Autistic Spectrum Disorder Screening Data for Children. This dataset was uploaded by Fadi Fayez Thabtah from the Department of Digital Technology at the Manukau Institute of Technology, Auckland, New Zealand. This dataset about autism has 21 attributes and has 292 data samples. The data collection was carried out by Fadi Fayez Thabtah using a mobile application called ASD Tests which had been implemented and published for the Android and iOS operating systems in order to facilitate accessibility for users (Thabtah et al., 2018). The Autistic Spectrum Disorder Screening Data for Children dataset attribute is described in detail in Table 1.

Table 1. Attribute Dataset

Attribute	Type	Description
Age	Number	Years
Gender	String	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean (yes or no)	Whether the case was born with jaundice
Family member with PDD	Boolean (yes or no)	Whether any immediate family member has a PDD
Attribute	Type	Description
Who is completing the test	String	Parent, self, caregiver, medical staff, clinician, etc.
Country of residence	String	List of countries in text format
Used the screening app before	Boolean (yes or no)	Whether the user has used a screening app
Screening Method Type	Integer (0,1,2,3)	The type of screening methods chosen based on age category (0=toddler, 1=child, 2= adolescent, 3= adult)
Question 1 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 2 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 3 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 4 Answer	Binary (0, 1)	The answer code of the question based on the screening method used

Attribute	Type	Description
Question 5 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 6 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 7 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 8 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 9 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 10 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Screening Score	Integer	The final score obtained based on the scoring algorithm of the screening method used. This was computed in an automated manner
Class/ASD	Boolean (yes or no)	The final result for screening autism

According to Thabtah, et al (2018), the list of questions related to the process is described in Table 2.

Table 2. List of Questions

Variable	Question
Question 1 Answer	She/he often notices small sounds when others do not
Question 2 Answer	She/he usually concentrates more on the whole picture rather than the small details
Question 3 Answer	In social group, she/he can easily keep track of several different people’s conversations
Question 4 Answer	She/he finds it easy to go back and forth between different activities
Question 5 Answer	She/he doesn’t know how to keep a conversation going with his/her peers
Question 6 Answer	She/he is good at social chit-chat
Question 7 Answer	When she/he is read a story, she/he finds it difficult to work out the character’s intentions or feelings
Question 8 Answer	When she/he was in preschool, she/he used to enjoy playing pretending games with other children
Question 9 Answer	She/he finds it easy to work out what someone is thinking or feeling just by looking at their face
Question 10 Answer	She/he finds it hard to make new friends

4.1.2 *Missing Value Handling*

In this dataset there are missing or incomplete data which are marked with a question mark symbol (?). This missing value will affect the dataset, which may have an impact on the process of defining the target for resolving data processing issues (Alamsyah et al., 2021). The incomplete data will be handled by missing values using the most frequent method. This method works by filling in the data using the most frequency in an attribute. The values obtained from the highest number of frequencies for each attribute to fill in the missing values are shown in Table 3.

Table 3. Attribute Missing Value

No.	Attribute	Value
1.	Age	6
2.	ethnicity	White-Europeean
3.	relation	Parent

4.1.3 *Data Transformation*

The purpose of data transformation is to speed up dataset processing. gender, ethnicity, jaundice, autism, country_of_res, used_app_before, relation, age_desc, and Class/ASD are among the attributes whose data were changed. In order to simplify the process, the Class/ASD attribute has been converted to Class at this point. The list of modified attribute is displayed in Table 4.

Table 4. Data Transformation Attribute

Attribute	Transformation
Gender	M=1; F=0
Ethnicity	White-European=1; Middle Eastern=2; South Asian=3; Asian=4; Black=5; Others=6; Pasifika=7; Hispanic=8; Turkish=9; Latino=10
jaundice	Yes=1; No=0
autisme	Yes=1; No=0
country_of_res	Isle of man=0; South Africa, Nigeria, Ghana, Libya, Egypt=1; India, Afghanistan, Georgia, Pakistan, Bangladesh, Oman, Qatar, South Korea, Philippines, Malaysia, Bhutan, Saudi Arabia, Japan, China, Nepal, Lebanon, Armenia, Jordan, Bahrain, Kuwait, United Arab Emirates, Syria, Iraq=2; United Kingdom, Austria, Europe, Malta, Bulgaria, Turkey, Sweden, Ireland, Netherlands, Romania, Germany, Italy, Latvia, Russia=3; United States, U.S Outlying Islands, Canada, Brazil, Mexico, Costa Rica, Argentina=4; New Zealand, Australia=5
used_app_before	Yes=1; No=0
relation	Parent=1; Self=2; Relative=3; Health care professional=4
age_desc	4-11 years=1
Class	Yes=1;No=0

4.1.4 *Feature Selection*

CFS is an automatic feature selection technique, which means that the user is not required to indicate how many features should be chosen. CFS operates by calculating the correlation between each attribute and the class, as well as between each attribute and other attributes. The selected attribute has a low correlation value with other attributes but a high correlation value with the class. The list of attributes and correlation weights of the CFS process is shown in Table 5.

Table 5. CFS Score

No.	Attribute	CFS Score
1.	A1_Score	0,393534
2.	A2_Score	0,229045
3.	A3_Score	0,395545
4.	A4_Score	0,568502
5.	A5_Score	0,379859
6.	A6_Score	0,417271
7.	A7_Score	0,273949
8.	A8_Score	0,438389
9.	A9_Score	0,486163
10.	A10_Score	0,439942
11.	age	0,07523
12.	gender	0,038782
13.	ethnicity	0,034092
14.	jaundice	-0,025049
15.	autisme	-0,0488
16.	country_of_res	0,119663
17.	used_app_before	-0,047212
18.	result	0,835869
19.	age_desc	-
20.	relation	-0,041998
21.	Class	1

Based on the CFS process, 17 selected attributes were obtained, namely A2_Score, A3_Score, A4_Score, A5_Score, A6_Score, A7_Score, A8_Score, A9_Score, A10_Score, age, gender, ethnicity, jaundice, autism, contry_of_res, result, relation. There is also an attribute that does not have a correlation value, namely age_desc, this is because in the dataset the age_desc attribute has the same data, namely '4-11 years' making it impossible to calculate the correlation value.

4.1.5 *Data Mining*

The first process is classification using the SVM algorithm on the Autistic Spectrum Disorder Screening data for Children Dataset with evaluation using a confusion matrix. Table 6 shows the results of the classification calculated using the confusion matrix.

Table 6. Confusion Matrix SVM

	True Positive	True Negative	Jumlah
False Positive	30	1	31

False Negative	2	26	28
Jumlah	32	27	59

$$Accuracy = \frac{TP+TN}{P+N}\ x\ 100\%$$

$$Accuracy = \frac{30+26}{31+28}\ x\ 100\% = 94,91\%$$

The accuracy results obtained from the SVM algorithm are 94.91%. Because the accuracy results are higher than the error rate, it can be concluded from these results that the SVM algorithm can detect autism spectrum disorders in children effectively.

The second process is SVM algorithm classification by applying CFS as feature selection. From the CFS process, 17 selected attributes were obtained. The new attributes will then be processed in the classification process using the SVM algorithm. The resulting classification results are then calculated using a confusion matrix as shown in Table 7.

Table 7. Confusion Matrix SVM and CFS

	True Positive	True Negative	Jumlah
False Positive	25	2	27
False Negative	0	32	32
Jumlah	25	34	59

$$Accuracy = \frac{TP+TN}{P+N}\ x\ 100\%$$

$$Accuracy = \frac{25+32}{27+32}\ x\ 100\% = 96,61\%$$

From the CFS process, 17 attributes were obtained which are shown in Table 8.

Table 8. Selected Attribute

No.	Selected Attribute
1.	A2_Score
2.	A3_Score
3.	A4_Score
4.	A5_Score
5.	A6_Score
6.	A7_Score
7.	A8_Score
8.	A9_Score
9.	A10_Score
10.	age
11.	gender
12.	ethnicity
13.	jaundice
14.	autism

15.	contry_of_res
16.	result
17.	relation

The accuracy results obtained from the application of CFS with 17 selected attributes in the SVM algorithm is 96.61%. These results show that the selection of CFS features in the Autistic Spectrum Disorder Screening Data for Children Dataset can improve the accuracy of the SVM algorithm classification.

4.2 Discussion

In this research, a comparison of accuracy using the SVM algorithm before and after implementing the feature selection process using CFS was carried out to increase accuracy in the diagnosis of autism spectrum disorders in children using the Autistic Spectrum Disorder Screening Data for Children Dataset. Based on the research conducted, there is an increase in accuracy when the SVM classification algorithm is applied using CFS feature selection. There is also a comparison of the accuracy shown in Table 9.

Table 9. Accuracy Comparison

Method	Accuracy
SVM	94,91%
SVM + CFS	96,61%

Based on this research, the accuracy of SVM has increased by 1.7% when compared to SVM when CFS has been applied, the increase is calculated from the original accuracy of 94.91% to 96.61%. The increase in accuracy is supported by the handling of missing values using the most frequent mode and data transformation with 17 attributes generated in the CFS process are A2_Score, A3_Score, A4_Score, A5_Score, A6_Score, A7_Score, A8_Score, A9_Score, A10_Score, age, gender, ethnicity, jaundice, autism, contry_of_res, result, relation.

5 Conclusion

Based on the results of research and discussion related to the implementation of SVM and CFS methods for the identification of autism spectrum disorders for children, it is proven that the SVM and CFS algorithms can be applied properly. The research process begins with retrieving datasets from the UCI machine learning repository. There are 292 data samples with 21 attributes used in this research. After the data is collected, the data will be preprocessed by using the missing value handling using the most frequent and data transformation. After the data preprocessing stage is complete, the feature selection process is continued using CFS which produces 17 selected attributes. Following the procedure, a brand-new dataset will be created, which will be split into 80% training data and 20% testing data for the SVM classification process. The categorization results are then assessed using a confusion matrix in order to determine the accuracy value.

The accuracy of the SVM application in this research was 94.91% and the combined application of SVM and CFS obtained an accuracy rate of 96.61% using the 17 best attributes. The research results is demonstrate that the accuracy of the SVM classification when combined with CFS provides results that are better than the accuracy of the SVM classification without using CFS, with an increase in accuracy of 1.7%.

References

Adi, P. T. K., Sugiharti, E., & Alamsyah, A. (2018). Comparison between SAW and TOPSIS methods in selection of broiler chicken meat quality. *Scientific Journal of Informatics*, *5*(1), 90.

Alamsyah, A., Prasetiyo, B., Al Hakim, M. F., & Pradana, F. D. (2021). Prediction of COVID-19 using recurrent neural network model. *Scientific Journal of Informatics*, *8*(1), 98-103.

Charleonnan, A., Fufaung, T., Niyomwong, T., Chokchueypattanakit, W., Suwannawach, S., & Ninchawee, N. (2016). Predictive analytics for chronic kidney disease using machine learning techniques. *Proceedings 2016 management and innovation technology international conference (MITicon),* 80.

Connolly, T. M., & Begg, C. E. (2005). *Database systems: a practical approach to design, implementation, and management*. Scotland: Pearson Education.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Essex: Cambridge University press.

Djatna, T., & Morimoto, Y. (2011). Pembandingan stabilitas algoritma seleksi fitur menggunakan transformasi ranking normal. *Jurnal Ilmiah Ilmu Komputer*, *6*(2), 245006.

Elsabbagh, M., Divan, G., Koh, Y. J., Kim, Y. S., Kauchali, S., Marcín, C., Montiel-Nava, C., Patel, V., Paula, C. S., Wang, C., Yasamy, M. T., & Fombonne, E. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism research*, *5*(3), 160-179.

Fachrurrozi, M., & Yusliani, N. (2015). Analisis sentimen pengguna jejaring sosial menggunakan metode support vector machine. *Sriwijaya Journal of Information System (JSI)*.

Han, J., Jian, P., & Micheline, K. (2012). *Data mining: Concepts and Techniques* (3th ed.). Massachusetts: Elsevier/Morgan Kaufmann, Waltham.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and. Techniques* (3th ed.). Massachusetts: Morgan Kauffman.

Huang, C. L., & Wang, C. J. (2006). A GA-based feature selection and parameters optimizationfor support vector machines. *Expert Systems with applications*, *31*(2), 231-240.

Jain, D., & Singh, V. (2018). Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal*, *19*(3), 179-189.

Karegowda, A. G., Manjunath, A. S., & Jayaram, M. A. (2010). Comparative study of attribute selection using gain ratio and correlation based feature selection. *International Journal of Information Technology and Knowledge Management*, *2*(2), 271-277.

Kavzoglu, T., & Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, *11*(5), 352-359.

Kumar, K. S., & Chezian, D. R. M. (2014). Support vector machine and k-nearest neighbor based analysis for the prediction of hypothyroid. *International journal of pharma and bio sciences*, *5*(4), 447-453.

Larasati, U. I., Muslim, M. A., Arifudin, R., & Alamsyah, A. (2019). Improve the accuracy of support vector machine using chi square statistic and term frequency inverse document frequency on movie review sentiment analysis. *Scientific Journal of Informatics*, *6*(1), 138-149.

Listiana, E., & Muslim, M. A. (2017). Penerapan adaboost untuk klasifikasi support vector machine guna meningkatkan akurasi pada diagnosa chronic kidney disease. *Prosiding SNATIF*, 875-881.

Maharani, M., Hasibuan, N. A., Silalahi, N., Nasution, S. D., Mesran, M., Suginam, S., Sutiksno, D. U., Nurdiyanto, H., Buulolo, E., & Yuhandri, Y. (2017). Implementasi data mining untuk pengaturan layout minimarket dengan menerapkan association rule. *JURIKOM (Jurnal Riset Komputer)*, *4*(4).

Ratama, N., & Munawaroh, M. (2020). Implementasi metode fuzzy tsukamoto untuk deteksi dini autisme pada balita berbasis android. *Jurnal Informatika dan Rekayasa Elektronik*, *3*(2), 129-139.

Shrivastri, S., & Deshmukh, R. (2014). Data classification particle swarm optimization and gravitational search algorithm. *International Journal of Innovative Research in Science, Engineering and Technology*, *3*(2), 9734-9741.

Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. *International Journal of Computer Applications*, *17*(8), 43-48.

Srivastava, D. K., Patnaik, K. S., & Bhambhu, L. (2010). Data classification: A rough-SVM approach. *Contemporary Engineering Sciences*, *3*(2), 77-86.

Sugiharti, E., & Muslim, M. (2016). On-line clustering of lecturers performance of computer science Department of Semarang State University using k-means algorithm. *Journal of Theoretical & Applied Information Technology*, *83*(1).

Sunu, C. (2012). *Panduan memecahkan masalah autism unlocking*. Yogyakarta: Lintang Terbit.

Thabtah, F., Kamalov, F., & Rajab, K. (2018). A new computational intelligence approach to detect autistic features for autism screening. *International journal of medical informatics*, *117*, 112-124.

Utomo, D. P., & Mesran, M. (2020). Analisis komparasi metode klasifikasi data mining dan reduksi atribut pada data set penyakit jantung. *Jurnal Media Informatika Budidarma*, *4*(2), 437-444.

Vijayarani, S., & Dhayanand, S. (2015). Liver disease prediction using svm and naïve bayes algorithms. *International Journal of Science, Engineering and Technology Research (IJSETR)*, *4*(4), 816-820.

Widiastuti, N. I., Rainarli, E., & Dewi, K. E. (2017). Peringkasan dan support vector machine pada klasifikasi dokumen. *Jurnal Infotel*, *9*(4), 416-421.

Wijaya, K. P., & Muslim, M. A. (2016). Peningkatan akurasi pada algoritma support vector machine dengan penerapan information gain untuk mendiagnosa chronic kidney disease. *Prosiding 3rd Seminar Nasional Ilmu Komputer (SNIK) Universitas Negeri Semarang,* 22-27.

Wirawan, I. N. T., & Eksistyanto, I. (2015). Penerapan naive bayes pada intrusion detection system dengan diskritisasi variabel. *Jurnal Ilmiah Teknologi Informasi*, *13*(2), 182-189.

Wu, X., & Kumar, V. (Eds.). (2009). *The top ten algorithms in data mining*. London: CRC press.

Yulietha, I., Faraby, S., & Adiwijaya, A. (2017). Klasifikasi sentimen review film menggunakan algoritma support vector machine. *eProceedings of Engineering*, *4*(3).

YPAC. 2010. *Buku pedoman penanganan dan pendidikan autisme*. Jakarta: YPAC Press