

A Novel Approach to Predict the Early Childhood Special Education Learning Skills of Autistic Children Using Ensemble Machine Learning

Dr. Yolvi J. Ocaña-Fernández^{1*}, Dr. Walter Gómez-Gonzales²,
Dr. Luis Alex Valenzuela Fernández³, Dr. Segundo Pio Vásquez Ramos⁴,
Dr. Huguette Fortunata Dueñas Zúñiga⁵, Dr. Jackeline Roxana Huaman Fernandez⁶ and
Dr. Marco Antonio Amapanqui Broncano⁷

^{1*}Universidad Privada San Juan Bautista, Lima, Perú. yolvi.ocana@upsjb.edu.pe,
Orcid: <https://orcid.org/0000-0002-2566-6875>

²Universidad Privada San Juan Bautista, Lima, Perú. walter.gomez@upsjb.edu.pe,
Orcid: <https://orcid.org/0000-0003-0706-7614>

³Universidad Privada San Juan Bautista, Lima, Perú. luisa.valenzuela@upsjb.edu.pe,
Orcid: <https://orcid.org/0000-0002-8743-4092>

⁴Universidad Cesar Vallejo, Lima, Perú. Spivar77@gmail.com,
Orcid: <https://orcid.org/0000-0002-0367-5711>

⁵Universidad Nacional del Altiplano, Puno, Perú. huguet24_2@hotmail.com.pe,
Orcid: <https://orcid.org/0000-0003-2437-4165>

⁶Centro de Altos Estudios Nacionales CAEN-EPG, Lima, Perú. jroxana11713@gmail.com,
Orcid: <https://orcid.org/0000-0001-9391-8205>

⁷Universidad Nacional de Educación “Enrique Guzmán y Valle” Lima, Perú.
20222394@une.edu.pe, Orcid: <https://orcid.org/0000-0001-8646-6642>

Received: March 10, 2023; Accepted: May 15, 2023; Published: June 30, 2023

Abstract

Children with autism spectrum disorder will eventually receive more extensive educational experiences, diverse understanding styles, any distinctive instructional techniques to help all infants achieve. Data mining categorization algorithms in the Weka tool are used to anticipate and forecast infants' performance with Autism Spectrum Disorder (ASD). As a decision-making tool for improving the performance of autistic youngsters, data mining is widely acknowledged. Support Vector Machines (SVMs), Logistic Regression (LR), and Naive Bayes (NB) are some of the techniques that can be used for categorization. The categorization model's outcomes include information on the model's accuracy, error rate, confusion matrices, classifier effectiveness, and execution time.

Keywords: Learning Disabilities, Diagnosis, ASD, Multinomial NB, Logistic Regression (LR) and SVM.

1 Introduction

Learning disability (LD) was initially described in 1963 (Kirk, S.A. (1963)). However, professionals in this field can still not agree on a precise definition of LD (Fletcher, J.M., et al., 2004). Alternately, the techniques frequently rely on scientific study experiments. The "Discrepancy Model" (Schrag, J. (2000).) indicates that there must be a considerable gap between intellectual capacity and academic accomplishment in at least one of the following areas:

- Oral and written expressions
- Normal reading skills
- Listening and reading comprehensions
- Mathematics calculation

Data mining is an interdisciplinary stage of "Knowledge Discovery in Databases" (KDD), and it identifies patterns in large data sets using artificial intelligence, machine learning, and statistical information techniques (Guidelines for Educating Students with Autism Spectrum Disorders", October 2010 Virginia Department of Education, Office of Special Education and Student Services.). Geographical indices are used as information tools (Liloja, 2023). The designs outline the raw input data for further analysis. Data fishing, snooping and dredging relate to using data mining techniques to sample small amount of an extensive population information collection to make statistical judgments about the validity of any patterns identified. These methodologies will test novel ideas against larger data populations ("Guidelines for Identification and Education of Children and Youth with Autism ", July 2005, CONNECTICUT STATE DEPARTMENT OF EDUCATION Division of Teaching and Learning Programs and Services Bureau of Special Education).

Autism presents complex conduct challenges among youngsters with intellectual challenges. Autism symptoms include difficulties with social interaction, anxiety, misunderstanding, empathy, and abnormal constrained, repetitive behaviour. Autism is prevalent among developing children (Felix D.C.C. Beacher et al., 2012). Autism affects the brain's activity and the child's social frame of mind. The high prevalence of scattering among growing children is classified as a mental disorder that affects the cerebral neurons, with side effects observed throughout a child's early development phases. The difficulties in understanding or comprehending non-verbal indications such as outer looks, use of verbal phrases, or irregular eye contact with the public (Volkmar, F.R., et al 2000) discourage any child suffering from persuasive ASD. The complex difficulties reduce the ability to participate in social and communication collaboration successfully. Withdrawal from attention to the facial region may indicate a reduced ability to perceive feelings. The plan is to sift through expository data gathered from people with ASD.

Orly et al. (2016) joined the wellness concepts of open children and its quality data for considering illnesses like Autism. It implements features of children and the prognosis of these characteristics at a young age with the support of data obtained from their quality. Elizabeth et al. (Stevens, E., 2017) suggested a Kmeans cluster to recognize behaviour and learning results. Supply individualized therapy interventions with the greatest effectiveness and minimal time and cost. Chan et al., The ANOVA method evaluates the therapeutic impact of a patented herbal formula, determined by neuropsychological assessments and day-to-day behavioural inspections. This research demonstrates that the executive functioning of ASD children can be enhanced. The emergence of smart analysis techniques will fulfil the DSM's enhanced standards (Thabtah, F. (2017)). In this case study, Dawson et al. hypothesized that a lack of social connection occurs within the first 6 to 24 months of baby autism (Dawson, G., et al.,

2000). Ramya et al. developed a new DRN method for analyzing multiple levels of ASD, DRN concepts introduced in Rapid Miner to identify the efficiency, categorization fault, and operation time with techniques such as Bayesian Boosting, Ada Boost, Vote, Stacking, and Bagging. DRN with 98.6 % (Ramya, R., et al 2017). Geetha Ramaniand and Sivaselvi (2017) suggested a system for identifying ASD and TD (Typically Developing), and yielded effectiveness of 88.46 % in the diagnosis of ASD and TD. The data mining technology is critical for accurately measuring the learning skills increased performance of autistic youngsters. Classification methods are frequently used to classify and evaluate data sets containing autistic youngsters accurately. The primary goal of this study is to anticipate and assess the behavior of autistic children using data mining approaches. Data mining offers a variety of tasks that investigate the effectiveness of autistic youngsters. The categorization task is applied in this research to quantify and estimate the performance of autistic children.

2 Proposed Methodology

Datasets used in the algorithmic categorization rule should be clear, accurate, and capable of being preprocessed to handle missing or redundant attributes. The data must be managed efficiently to get the optimal results from the Data Mining procedure.

Attribute Identification

Dataset acquired from the infant's database is depicted in Table 1.

Table 1: Dataset Acquired from Infant Database

Attributes	Description
Child	Child Autism Stages
Atten	Attention Skills
Memory	Memory Skills
Hand writ	Hand Writing Skills
Write	Writing Skills
Spelling	Spelling Skills
Read	Reading Skills
Language	Language Skills

Ensemble Classifiers for ASD

Three classifiers were chosen based on previous performance (Gambino O.J., et al., 2016). The classifiers and their parameters are defined here:

- **Multinomial NB.** There are no external dependencies. It uses Bayes' theorem. Despite the naive assumption of class conditional independence, this approach produces good results equivalent to more complicated procedures such as neural networks (John G.H., et al., 1995). The classifier's additional soft factor, alpha, was set to 0.5.
- **LR.** Models the likelihood of events occurring as a linear function of a set of predictor factors. Using a recognition approach makes this algorithm a good classifier (Friedman J., et al., 2001). The inverse of regularisation strength, factor C, was set to 1.0.
- **SVM.** This approach transforms the real dataset into a higher dimension—a hyper-plane (Joachims T. et al., 1998). SVMs are used to find the best decision boundary for data classification. The class probabilities were determined using the algorithm given by (Wu T.F., et al., 2004). The kernel coefficient (gamma) was set to 0.00001 and the error term penalty factor (C) 3500.

Ensemble Classification

It diagnoses new cases by combining the outcome of numerous classifiers whose individual conclusions are integrated via weighted or unweighted voting (Dawson, G., et al., 2000). In this study, ensemble categorization categorizes fresh tweets, including in $n = 3$ classifiers denoted by C_1, C_2, \dots, C_n . For each tweet $t^q, q = 1, 2, \dots, t$ classifier $C_i, i = 1, 2, \dots, n$ creates m probabilities.

$P_{i,j}, j = 1, 2, \dots, m$. $P_{i,1}$ denotes the probability created by the classifier C_i that the q -th tweet t^q belongs to class L1, $P_{i,2}$ the probability that it belongs to the class L2 and so on for the m classes

The suggested weighting concept, weights the m probabilities created by the classifier $C_i, i = 1, 2, \dots, n$ with weights $w_i, i = 1, 2, \dots, n$.

The suggested weighting concept weights the probabilities created by the classifier with weights.

Weka Tool Implementation

Weka introduces various ML techniques. It was produced from the above data. WEKA explorer loaded this file. Factors affecting autistic children's learning abilities include the Result of diverse children's handwriting, spelling, language. The implementation used 40 samples. To employ classification algorithms on a dataset, analyze the predictive approach's accuracy, and illustrate the process. Weka imposed Multinomial NB, LR, and SVM. Tenfold cross-validation is set under "Test options".

3 Experimental Results

Categorization analysis and evaluation take time and require a solid understanding of statistics. The method takes a long time and requires skilled analysis to classify and relate data. An unavoidable conclusion emerges from the initial rules: writing skills are significantly related to student achievement. According to the ruleset, communication, handwriting, memory, writing, reading, language, spelling, and the levels of the autistic infant are the multiple aspects that have a high potential variable that affects autistic children's performance.

Table 2: Performance Result of Classifiers

Methods	Time to create the model (sec)	Accurately categorized instances	Wrongly categorized instances	Accuracy (%)	Kappa statistic	Mean absolute error	Root mean squared Error (RMSE)
J48	0.02	36	4	90	0.8462	0.0625	0.2254
Multinomial Naïve Bayes (NB)	0.22	40	4	93	0.8321	0.0541	0.2001
Logistic Regression (LR)	0.25	35	3	92	0.8541	0.0321	0.2124
Support Vector Machine (SVM)	0.28	38	2	95	0.9234	0.0233	0.2154
Proposed Ensemble Classifier	0.01	45	1	98	0.9255	0.0121	0.1999

In Table 2, the performance result of classifiers is depicted. The ensemble classifier is considered as highly effective than other classifiers.

4 Conclusion

Handwriting, spelling, language, reading, writing, memory abilities, and autism school student performance analysis. The work investigates the influence of results, factors like attention on machine learning algorithms. The Ensemble classifier using Multinomial NB, LR and SVM outperforms the other algorithms used in the study. Multiple categorization procedures and association rule mining for the autism infant dataset will be utilized in the future. This study will be very beneficial to educational institutions.

References

- [1] Beacher, F.D., Radulescu, E., Minati, L., Baron-Cohen, S., Lombardo, M.V., Lai, M.C., & Critchley, H.D. (2012). Sex differences and autism: brain function during verbal fluency and mental rotation. *PloS one*, 7(6), 1-11.
- [2] Chan, A.S., Sze, S.L., & Han, Y.M. (2014). An intranasal herbal medicine improves executive functions and activates the underlying neural network in children with autism. *Research in Autism Spectrum Disorders*, 8(6), 681-691.
- [3] Dawson, G., Osterling, J., Meltzoff, A.N., & Kuhl, P. (2000). Case study of the development of an infant with autism from birth to two years of age. *Journal of applied developmental psychology*, 21(3), 299-313.
- [4] Fletcher, J.M., Coulter, W.A., Reschly, D.J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of dyslexia*, 54(2), 304-331.
- [5] Gambino, O.J., & Calvo, H. (2016). A comparison between two spanish sentiment lexicons in the twitter sentiment analysis task. In *Advances in Artificial Intelligence-IBERAMIA: 15th Ibero-American Conference on AI, San José, Costa Rica, Proceedings 15*, 127-138. Springer International Publishing.
- [6] Guidelines for Educating Students with Autism Spectrum Disorders, October 2010 Virginia Department of Education, Office of Special Education and Student Services.
- [7] Guidelines for Identification and Education of Children and Youth with Autism, July 2005, Connecticut State Department of Education Division of Teaching and Learning Programs and Services Bureau of Special Education.
- [8] Hastie, T., Tibshirani, R., & Friedman, J. (2001). The elements of statistical learning. Springer series in statistics. *New York, NY, USA*.
- [9] Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*, 137-142. Berlin, Heidelberg: Springer Berlin Heidelberg.
- [10] John G.H., Langley P. (1995). Estimating continuous distributions in Bayesian classifiers. *Proceedings of the 11th Conference on Uncertainty in artificial intelligence; Montreal, QC, Canada. Burlington, MA, USA: Morgan Kaufmann Publishers Inc*, 338-345.
- [11] Kirk, S.A. (1963). Behavioral diagnoses and remediation of learning disabilities. In *Proceedings of the annual meeting: Conference on exploration into the problems of the perceptually handicapped child Evanston, IL, 1963*, 1-7.
- [12] Liloja, & Ranjana, D. (2023). An Intrusion Detection System Using a Machine Learning Approach in IOT-based Smart Cities. *Journal of Internet Services and Information Security (JISIS)*, 13(1), 11-21.
- [13] Ramani, R.G., & Sivaselvi, K. (2017). Autism spectrum disorder identification using data mining techniques. *International Journal of Pure and Applied Mathematics*, 117(16), 427-436.

- [14] Ramya, R., & Zoraida, B.S.E. (2017). DRN Hybrid Model for Predicting Autism using Rapid Miner Tool. *International Journal of Advanced Research in Computer Science*, 8(8), 111-115.
- [15] Reiner, O., Karzbrun, E., Kshirsagar, A., & Kaibuchi, K. (2016). Regulation of neuronal migration, an emerging topic in autism spectrum disorders. *Journal of neurochemistry*, 136(3), 440-456.
- [16] Schrag, J.A. (2000). Welfare Reform: Issues, Progress, and Implications for Persons with Disabilities. Project FORUM, National Association of State Directors of Special Education, 1800 Diagonal Road, Suite 320, Alexandria, VA 22314.
- [17] Stevens, E., Atchison, A., Stevens, L., Hong, E., Granpeesheh, D., Dixon, D., & Linstead, E. (2017). A cluster analysis of challenging behaviors in autism spectrum disorder. *In 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 661-666.
- [18] Thabtah, F. (2017). Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment. *In Proceedings of the 1st International Conference on Medical and health Informatics*, 1-6.
- [19] Volkmar FR, Klin A (2000) Pervasive Developmental Disorders. In: Kaplan and Sadock's Comprehensive Textbook of Psychiatry CDROM 7th edition. Publisher: Lippincott Williams and Wilkins, Philadelphia, PA.
- [20] Wu T.F., Lin C.J., Weng R.C. (2004). Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5, 975–1005.

Authors Biography



Yolvi J. Ocaña Fernández

University professor, teacher in university teaching and doctor of education from the Universidad Nacional Mayor de San Marcos. He has a master's degree in social communication of science from the International University of Valencia-Spain and a post-doctorate in ethics and research from the Abad Oliva University of Barcelona-Spain. He works as a part-time research professor at the School of Human Medicine of the private San Juan Bautista University.



Walter Gómez-Gonzales

Doctor of Health Sciences, Faculty of Medicine -Universidad Nacional Mayor de San Marcos. Magister in Social Policy, Mention Health and Social Security (UNMSM), Specialist in Public and Community Health (USMP), Diploma in Epidemiology and others. Principal Ordinary Research Professor, Professional School of Human Medicine, San Juan Bautista Private University. Lima Peru.



Luis Alex Valenzuela Fernández

Doctorate in Education and also a license in Natural Sciences and Mathematics from the National University of Education and the title of Biologist - Microbiologist from the National University of San Marcos, in Lima Peru. He currently teaches the specialty of Cellular and Molecular Biology, as well as is a Research Professor at the San Juan Bautista Private University where he has been developing research in the Biomedical and Clinical Laboratory area. He has produced various publications on new technologies, artificial intelligence and its impact on education, among others.



Segundo Pio Vásquez Ramos

Doctor. in Public Management and Governance (Universidad Cesar Vallejo, Peru) Master's Degrees in: University Teaching and Educational Management (Universidad Cesar Vallejo, Peru); Auditing and Public Management, (Catholic University of Trujillo Benedict XVI, Peru); Bachelor of Education in Social Sciences (Catholic University of Trujillo Benedict XVI, Peru), Lawyer with specialization in Administrative Law and Administrative Litigation, (University of Huanuco, Peru) committed to Higher and Regular Basic Education with more than 20 years of teaching experience, with publications in indexed journals.



Huguette Fortunata Dueñas Zúñiga

Bachelor in Psychology and Licentiate in Psychology at the Universidad Particular Andina del Cusco. Master in Public Health, Doctoris Scientiae in Education, Universidad Nacional del Altiplano. Teacher appointed at the National University of the Altiplano. Teaching experience in private universities. Teacher of Master's Degree in Social Sciences, Second Specialty in the Faculty of Education.



Jackeline Roxana Huaman Fernandez

Methodological adviser of the Center for Higher National Studies, researcher with COD ORCID, Postgraduate Professor, Specialist in research project management, reviewer and thesis jury. Speaker and trainer in scientific research topics. PhD in Educational Sciences, Master in Educational Management, Mg in Systems Engineering (c), Graduate in Education.



Marco Antonio Amapanqui Broncano

Systems engineer, graduate in education with a major in Philosophy and Social Sciences, primary education teacher with a master's degree in education, 20 years of experience as a teacher and administrator of educational institutions. Researcher, trainer and specialist in ICT, digital tools, gamifications for teaching and learning, creator of virtual classrooms, web pages and educational software. Entrepreneur of digital literacy projects "Mobile Schools - MUNINET" of the Municipality of Lima that awarded him distinctions as a teacher and administrator of the MUNINET project, he is currently doing his doctoral studies, he works as a teacher in the classroom of pedagogical innovation and alternative basic education.