Received Oktober 20th, 2024; accepted December 4th, 2024. Date of publication December 30th, 2024
Digital Object Identifier: https://doi.org/10/25047/jtit.v11i2.5517

A Stacking Approach to Enhance K-Nearest Neighbors Performance for Autism Screening

MUHAMMAD NAUFAL[1,2], HARUN AL AZIES [1,3]

[1]Study Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia
[2]Research Center for Intelligent Distributed Surveillance and Security, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia
[3]Research Center for Quantum Computing and Material Informatics, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia

CORESPONDING AUTHOR: Harun Al Azies (email: harun.alazies@dsn.dinus.ac.id)

ABSTRACT The increasing prevalence of autism spectrum disorders necessitates improved early screening methods for children to ensure timely intervention and support. While existing screening techniques play a vital role, they often face challenges regarding accuracy, accessibility, and scalability. This research addresses these gaps by enhancing the K-Nearest Neighbors (K-NN) algorithm by implementing a stacking model that integrates multiple distance metrics—Manhattan and Minkowski—to improve predictive performance. Utilizing a public dataset, the study employed K-Fold Cross-Validation with K=5 to ensure a robust evaluation of the models. The results demonstrated that the stacking model achieved an average accuracy of 86.67%, significantly surpassing the traditional K-NN approaches, which reported accuracies of 82.67% for Manhattan and 81.33% for Minkowski. A user-friendly web interface was also developed to facilitate real-world application, allowing users to input data and receive immediate predictive outcomes regarding autism risk. These findings confirm the effectiveness of the stacking method in enhancing K-NN performance and highlight its potential for practical use in autism screening. Future research may explore alternative machine learning algorithms and additional features to refine the predictive capabilities and user experience further.

KEYWORDS: Autism screening, K-Nearest Neighbors, stacking model, machine learning, predictive performance.

I.INTRODUCTION

Autism Spectrum Disorder (ASD) is a developmental condition that affects communication, social interaction, and behaviour. According to the World Health Organization (WHO), as of November 2023, about 1 in 100 children worldwide are diagnosed with ASD (WHO, n.d.). Early detection of ASD is essential because it can significantly influence various aspects of a child's life, such as their education, socialization, and access to appropriate therapeutic interventions [1]. Without timely identification, children with ASD may face more significant challenges as early behavioural patterns become more entrenched and complex to change over time. Common early signs of autism include difficulties with social interaction, communication, and emotional regulation, as well as repetitive behaviours and challenges in processing sensory information. Recognizing these signs early can lead to effective interventions, ultimately improving outcomes for children diagnosed with ASD [2].

Detecting Autism Spectrum Disorder (ASD) at an early stage can significantly influence various aspects of a child's life, including their educational strategies, social skills, and the therapeutic interventions that are designed to meet their specific needs. Research shows that early intervention can have a transformative impact on the lives of children with ASD, often leading to better long-term outcomes compared to treatments initiated later [3]. When parents are able to recognize the early signs of ASD and seek professional help promptly, they can consult with specialists such as paediatricians or child psychiatrists. This timely intervention allows for the development of tailored treatment plans and therapies that can enhance the child's learning experiences and social interactions, ultimately fostering a supportive environment for their growth and development. The earlier the diagnosis, the more influential the interventions can be, helping to mitigate potential challenges that might arise as the child matures [4].

Numerous studies on the early detection of Autism Spectrum Disorder (ASD) have employed various approaches, including image analysis, video observations, audio assessments, brain signal monitoring, and questionnaires [5][6]. Among these methods, questionnaires provide a straightforward and accessible means of data collection, making them particularly popular in both research and clinical settings. This study focuses specifically on a questionnaire-based approach for ASD detection, primarily because of its ease of implementation and efficiency in data acquisition. Furthermore, methods utilized in previous research span a wide range, from rule-based systems and basic statistical techniques to advanced machine learning and deep learning algorithms. This diversity in methodologies underscores the adaptability and flexibility required in developing effective detection strategies for ASD, allowing researchers to tailor their approaches based on available resources and specific research goals.

Machine learning-based screening approaches utilizing questionnaires have employed algorithms like Support Vector Machine (SVM), Logistic Regression (LR) [7] modified SVM [8] and K-Nearest Neighbor (KNN) [9]. These approaches have performed well with tabular data from questionnaires with relatively simple models.

One widely used algorithm in this context is K-Nearest Neighbors (KNN), known for its simplicity, non-parametric nature, and ability to handle non-linear data. KNN does not require complex training processes, making it suitable for small datasets [10]. However, KNN is sensitive to outliers and performs poorly with high-dimensional data [11]. Previous studies have compared distance metrics used in KNN, finding significant variations in performance depending on the dataset [12], [13]. Choosing a distance metric such as Manhattan or Minkowski can improve KNN's performance, particularly in datasets with outliers or high-dimensional data [14].

Ensemble techniques have been employed to address KNN's limitations. As recent research shows, these techniques combine multiple classifiers to create a more accurate model [15]. By integrating KNN with other models, ensemble methods reduce KNN's sensitivity to outliers and high-dimensional data, resulting in improved predictive performance [16]. This study contributes to the field by comparing different distance metrics and implementing stacking techniques in KNN to enhance the prediction accuracy of ASD in children.

II.METHOD

This research consists of several stages designed to clearly identify each process in addressing the research problems and producing new contributions. The research steps are divided into four main sections as follows. The stages of the research can be seen in the diagram presented in Figure 1 below.

This flowchart illustrates the research process for detecting Autism Spectrum Disorder (ASD) using a machine-learning approach. It begins with dataset collection, followed by a preprocessing stage to ensure data quality by handling missing values, normalization, and feature encoding. The data is then split into training and testing sets to ensure unbiased model evaluation.

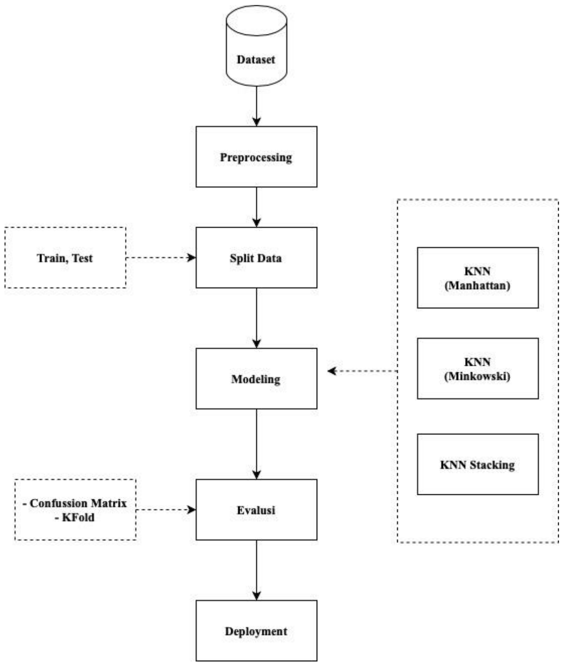


FIGURE 1. Flowchart of the Research Process in Detecting Autism Spectrum Disorder.

The modeling phase involves building three models: KNN using Manhattan distance, KNN using Minkowski distance, and KNN Stacking, which combines multiple models to improve accuracy. The models are evaluated using a confusion matrix to assess performance and K-Fold Cross-Validation to minimize overfitting. The best-performing model is then deployed, making it ready for practical use in detecting ASD in new data. This process is designed to be iterative, ensuring optimal and applicable results.

A. Data Collection

The research began by collecting a publicly available dataset titled "Autistic Spectrum Disorder Screening Data for Children" from the UCI Machine Learning Repository. This dataset, which is in CSV format, contains 292 records focused on predicting Autism Spectrum Disorder (ASD) in children aged 4 to 11 years [17]. It consists of 20 features and two class labels, representing various demographic and behavioural factors. Key demographic variables include the child's age, gender, ethnicity, history of jaundice, family history of ASD, and the relationship of the respondent to the child (e.g., parent or guardian). In addition to these demographic details, the dataset includes responses to 10 behavioural screening questions (Q1 to Q10), which are answered with a binary "Yes" or "No." These questions are designed to assess specific behavioural traits that may be indicative of ASD. The class label, also binary, identifies whether the child is likely to be diagnosed with ASD based on the combination of demographic and behavioural data. This dataset serves as the foundation for building predictive models to assess the risk of ASD in children.

B. Preprocessing

In the preprocessing stage, several key steps were undertaken to improve the dataset's suitability for predictive modelling. First, irrelevant or redundant features were manually removed to avoid noise in the model. These excluded features were:

- age_desc: A descriptive feature that does not contribute to the prediction task.
- country_of_res: The country of residence was deemed irrelevant to the ASD diagnosis.
- Ethnicity: Although demographic, ethnicity was not directly linked to ASD prediction in this study.
- used_app_before: This feature, indicating prior use of the screening app, was not essential for the current task.
- Result: The actual result of the screening could lead to data leakage if included.
- Autism: Another result-based feature that might cause bias in prediction.

After feature selection, the dataset faced missing values in the relation feature, with 44 records missing information on the respondent's relationship to the child. Rather than discarding these rows, appropriate imputation methods were applied to retain as much data as possible, yielding a final dataset with 248 complete records. The dataset was pretty balanced with respect to class labels, containing 122 negative (no ASD) and 126 positive (ASD) cases.

One-hot encoding was applied to handle categorical variables such as gender, jaundice status, and relationship to the child. This process transformed these categorical features into numerical binary vectors, which allowed machine learning algorithms to process them effectively. This transformation was essential as many algorithms, such as KNN and stacking models, require numerical inputs for proper operation.

Thus, after careful feature selection, imputation of missing values, and encoding, the dataset was cleaned and prepped with 248 instances, and features transformed into a format suitable for machine learning model, and also performed a 70:30 train-test data split with an equal proportion of labels.

C. Model Development

Model development involves training and evaluating the K-Nearest Neighbor (KNN) algorithm using different distance metrics, specifically Manhattan and Minkowski distances. The research utilizes K=7 as the KNN parameter. The Manhattan distance between two points *x* and *y* is calculated as follows:

$$d(x,y) = \sum_{i=1}^{n} |x_i - y_i| \quad (1)$$

The Minkowski distance generalizes both Euclidean and Manhattan distances and is given by

$$d(x,y) = \left(\sum_{i=1}^{n} (x_i - y_i)^p\right)^{\frac{1}{p}} \quad (2)$$

where $p = 1$ corresponds to Manhattan distance and $p = 2$ corresponds to Euclidean distance.

The computational steps of KNN include setting the K value, calculating distances using the specified metrics, sorting data based on ascending distances, and making predictions based on the majority class of the nearest neighbors. Additionally, the study employs a Stacking Ensemble Classifier, which combines multiple classifiers for improved accuracy. The base learners are KNN with Minkowski distance, while the meta learner is KNN with Manhattan distance. The outputs from the base learners are utilized to generate new features for the meta-learner to make final predictions.

Several metrics are used to evaluate the models' performance. The confusion matrix is constructed to compare predicted and actual labels, allowing for calculating key evaluation metrics. Accuracy is assessed by measuring the ratio of correct predictions to total predictions, while precision evaluates the proportion of accurate positive predictions among all optimistic predictions. Recall, on the other hand, calculates the proportion of actual positives that were correctly identified. The F1-Score combines precision and recall into a single metric, providing a comprehensive view of the model's performance. K-Fold Cross Validation is also employed, where the dataset is divided into K subsets to ensure that the model is trained and tested effectively across different data segments [18]. The stacking technique applied in this study is illustrated in the diagram shown in Figure 2 below.

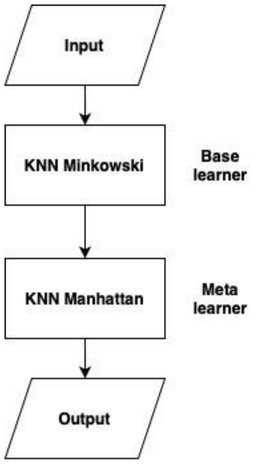


FIGURE 2. Stacking Ensemble Technique for K-Nearest Neighbor Classification.

The stacking technique illustrated in Figure 2 enhances the performance of K-Nearest Neighbors (K-NN) classification by combining multiple models to improve predictive accuracy. In this ensemble approach, various base models are trained on the same dataset, capturing different aspects of the data. The outputs of these base models are then fed into a meta-learner, which learns to make final predictions based on the projections of the base models. This layered structure allows the stacking ensemble to leverage the strengths of different algorithms, ultimately resulting in improved classification performance, particularly in complex tasks such as autism screening.

D. Deployment

In the deployment stage, the trained machine learning model was integrated into a web application to provide a user-friendly interface for making predictions on Autism Spectrum Disorder (ASD). The web application was developed using standard web technologies—HTML for structure, CSS for styling, and JavaScript for dynamic interactivity. The primary goal of the deployment was to allow users to enter questionnaire responses via the website, and based on their input, the model would deliver real-time predictions about the likelihood of ASD in children.

To achieve this, the machine learning model, which had been trained and optimized, was converted into the ONNX (Open Neural Network Exchange) format. ONNX, or Open Neural Network Exchange, is an open-source format that allows machine learning models to be easily shared and deployed across different platforms and frameworks. Converting a model into the ONNX format makes it highly portable. It can be used in various environments, from web browsers to mobile devices and IoT systems, without significant reconfiguration or retraining. In the context of this research, adopting ONNX has a transformative impact. It ensures the model can run efficiently and in real-time directly in a browser using ONNX.js, eliminating the need for server-based computations. This reduces latency and makes the system more accessible and scalable for a broader audience. Additionally, ONNX optimizes model performance, enabling faster processing while maintaining accuracy, crucial for real-time applications like Autism Spectrum Disorder detection. By leveraging ONNX, the research output becomes adaptable to future technologies and platforms, ensuring long-term usability and more comprehensive practical implementation.

This conversion was necessary to ensure compatibility across different platforms and enable efficient execution within the browser environment. ONNX.js, a JavaScript runtime for ONNX models, was then used to load the model within the web application. By leveraging ONNX.js, the model was seamlessly integrated into the web interface, enabling real-time processing of user inputs directly from the browser without needing to rely on a server for predictions. The deployment workflow is straightforward: users access the questionnaire form, provide the required inputs (such as demographic details and responses to screening questions), and then submit the form. The ONNX-based model processes the data and instantly returns a prediction indicating the likelihood of ASD. The results are accompanied by tailored recommendations or next steps based on the model's output, making the application both informative and actionable.

This deployment provides an accessible tool for early screening, allowing users to interact directly with a predictive model and receive real-time feedback, as shown in the simplified illustration in Figure 3.

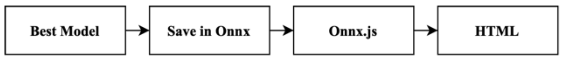


FIGURE 3. Illustration of Model Deployment.

This deployment offers an intuitive and user-friendly interface for early screening of Autism Spectrum Disorder (ASD) in children. Users input relevant information through a web-based questionnaire, and the predictive model—integrated into the system using the ONNX format—analyzes the data in real time. The illustration in Figure 3 demonstrates the seamless interaction between the user input and the model's processing, showcasing how the system generates predictions and provides immediate feedback on potential ASD risk based on the responses. This approach enables accessible and efficient early screening.

III. RESULT AND DISCUSSION

This research aimed to evaluate the K-Nearest Neighbors (K-NN) model's effectiveness in diagnosing autism in children by employing various parameters and methods. The Neighbor parameter was set to its default value of 7. To mitigate bias in testing, the evaluation utilized K-Fold Cross-Validation with K=5, partitioning the dataset into five segments. This approach ensured that each model's performance could be reliably assessed across diverse subsets of data, reflecting its applicability in real-world scenarios. The performance analysis employed a confusion matrix to provide a comprehensive overview of each model's classification outcomes. Table 1 summarizes the training accuracy results across the five folds for each model, reflecting the accuracy measurements obtained from each fold.

TABLE 1. K-Fold Training Results

Fold	K-NN Manhattan	K-NN Minkowski	K-NN Stacking
1	91.43	91.43	94.29
2	88.57	88.57	88.57
3	85.71	85.71	91.43
4	82.35	85.29	85.29
5	88.24	88.24	94.12
Mean	87.26	87.85	90.74

The results in Table 1 illustrate the performance of three models—K-NN Manhattan, K-NN Minkowski, and K-NN Stacking—evaluated using K-Fold cross-validation. The K-NN Manhattan and Minkowski models showed similar performance, with mean accuracies of 87.26% and 87.85%, respectively, indicating that the distance metric had minimal impact on their results. Despite this, their performance was inconsistent across the folds, and neither model achieved the highest accuracy in any particular fold. On the other hand, the K-NN Stacking model demonstrated a clear advantage, achieving the highest accuracy of 94.29% in Fold 1 and a mean accuracy of 90.74% across all folds. This superior performance highlights the power of the stacking technique, which aggregates predictions from multiple base models to improve accuracy and generalization. By combining different K-NN variants, the stacking model overcame the individual limitations of each model, resulting in a more robust and reliable prediction system.

These results emphasize the potential of K-NN Stacking in addressing the complexities of autism detection, where slight variations in input features can significantly impact outcomes. Its higher consistency and accuracy make it an ideal candidate for developing a reliable and scalable screening tool, aligning perfectly with the research goal of creating a precise diagnostic system. The findings confirm that stacking is valuable for enhancing machine learning models in medical diagnostics, offering increased reliability and practical applications in real-world scenarios. The results of the confusion matrix for binary classification are presented in Table 2.

TABLE 2. Confusion Matrix Results

Model	TP	TN	FP	FN
KNN Manhattan	38	24	13	0
KNN Minkowski	37	24	13	1
Stacking	37	28	9	1

Table 2 summarizes the confusion matrix results for each model, presenting values for True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The stacking model demonstrated the highest True Negative (TN) value at 28, indicating its superior ability to correctly identify non-autistic cases compared to the other models, which both recorded a TN of 24. This is a crucial finding, as a higher TN count contributes to reducing the number of misclassifications for children who do not have autism, thereby improving the overall reliability of the screening tool.

Regarding True Positives (TP), the stacking model scored 37, just 1 point lower than the KNN Manhattan model, which had a TP of 38. This suggests that while the stacking model may slightly lag in its ability to classify autistic children compared to the KNN Manhattan model correctly, it still maintains a robust performance. The stacking model's higher TN and slightly lower TP demonstrate that while it may slightly sacrifice sensitivity in identifying autistic cases, it significantly improves specificity (correctly identifying non-autistic cases). This balance is crucial for developing a robust screening tool, as it helps minimize false positives and negatives, leading to a more reliable diagnostic tool for autism detection.

The False Positives (FP) analysis reveals that the stacking model outperforms both KNN Manhattan and KNN Minkowski models, achieving the lowest FP count at 9. This indicates that the stacking model is more effective in minimizing incorrect classifications of children as autistic when they are not, which is crucial for avoiding unnecessary stress and intervention for families. To provide a more comprehensive view of each model's performance, calculations for accuracy, precision, recall, and F1 Score were performed, as shown in Table 3.

TABLE 3. Model Performance

Model	Accuracy	Precision	Recall	F1-Score
KNN Manhattan	82.67	87.25	82.43	82.04
KNN Minkowski	81.33	85.00	81.12	80.76
Stacking	86.67	88.49	86.52	86.47

Table 3 presents a comprehensive overview of each model's performance based on several key metrics: accuracy, precision, recall, and F1 Score. The stacking model outperforms the other models in all evaluated metrics, achieving an accuracy of 86.67%. This high accuracy indicates that the stacking model correctly classifies a significant majority of both autistic and non-autistic cases, which is critical for ensuring a reliable screening tool for autism in children.

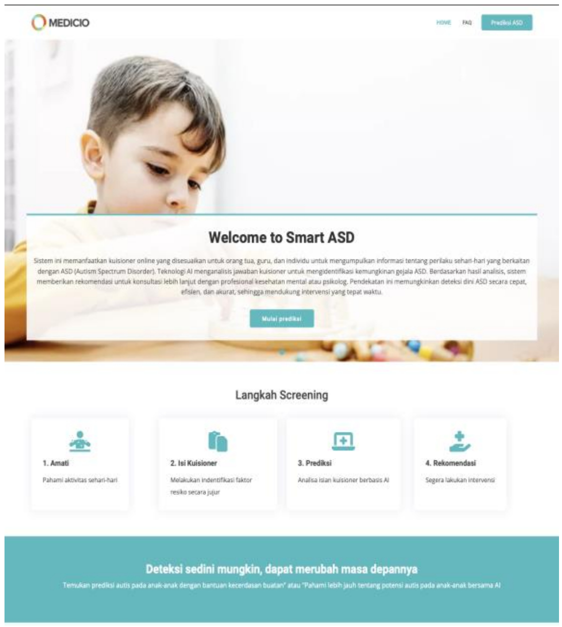
Precision, which measures the proportion of positive results among all optimistic predictions, is highest for the stacking model at 88.49%. This suggests that when the stacking model predicts a child as autistic, it is more likely to be correct than the other models, thus reducing the chances of false positives.

Recall, or sensitivity, reflects the model's ability to identify actual autistic cases correctly. The stacking model's recall of 86.52% indicates that it successfully identifies a substantial proportion of children with autism, which is vital for timely diagnosis and intervention.

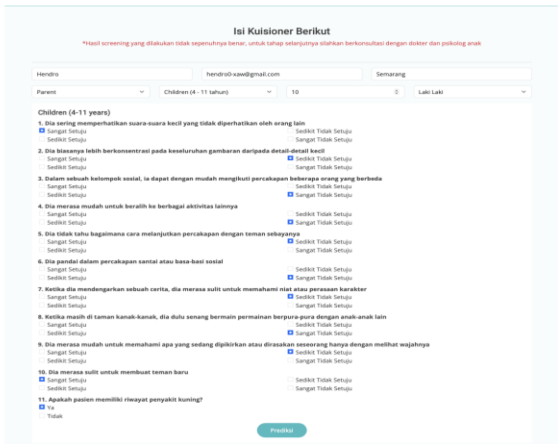
The F1 Score, which balances precision and recall, is also the highest for the stacking model at 86.47. This metric is particularly important in this context because it reflects the model's ability to perform well across both precision and recall, ensuring that it not only effectively identifies children with autism but also minimizes misclassification.

Using macro averages in this evaluation is advantageous as it treats each class equally, irrespective of their representation in the dataset. This approach is especially relevant in medical screening applications like this one, where class imbalance may exist between autistic and non-autistic children. By employing macro averages, the evaluation accurately reflects the model's overall performance across both categories, thereby ensuring that the results are not skewed by the prevalence of one class over the other.

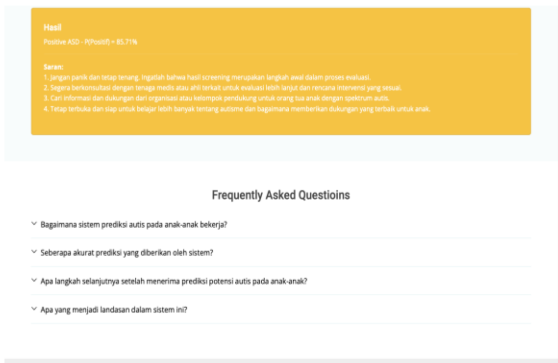
Upon determining that the KNN-stacking model was the best-performing algorithm, we proceeded to deploy it into a web-based application aimed at facilitating autism screening. The user interface of this application is designed for clarity and usability, making it accessible for users without technical expertise. A detailed overview of the interface components is provided below, as illustrated in Figure 4



(a)



(b)



(c)

FIGURE 4. Snapshot of the generated website (a) Header display, (b) Questionnaire form, and (c) Prediction results and recommendations.

(a) Header Display

This section prominently features the application title, "Autism Screening Tool," along with a brief description of its purpose. The header includes instructional text to guide users on using the tool effectively. The layout is visually appealing, utilizing appropriate colors and fonts to enhance readability and engagement.

(b) Questionnaire Form

The questionnaire form is a pivotal interface element where users input their responses. It consists of various fields, including multiple-choice questions, checkboxes, and rating scales, designed to capture comprehensive data regarding children's developmental milestones, social interactions, and behavioral patterns. Each question is carefully crafted to ensure clarity, enabling users to provide accurate and relevant information. For instance, questions might assess social behaviors (e.g., "Does your child make eye contact?") and developmental milestones (e.g., "At what age did your child begin speaking?"). The form also features validation checks to ensure all necessary fields are completed before submission, preventing incomplete data entry.

(c) Prediction Results and Recommendations:

After users complete the questionnaire and click the "Predict" button, the application processes the input data through the KNN-stacking model. The model utilizes the data to generate results indicating the likelihood of autism spectrum disorder (ASD) in the child. The results are presented in an easily digestible format, often with a confidence score (e.g., "There is a 78% likelihood that your child may exhibit signs of autism."). This score helps convey the seriousness of the results while also allowing for nuance in interpretation.

The results are accompanied by tailored recommendations, including suggestions for further assessment by a healthcare professional, support group resources, or early intervention strategies. This information empowers users by providing actionable steps based on the predictive analysis.

The deployment of the KNN-stacking model was achieved using the ONNX (Open Neural Network Exchange) format, which was chosen for its compatibility across various platforms. This format ensures seamless integration into web environments that utilize HTML, CSS, and JavaScript, facilitating a smooth user experience. By leveraging the ONNX format, the KNN-stacking model can be efficiently loaded within the web application, allowing for real-time predictions based on user input. The model's architecture combines multiple KNN approaches to enhance predictive accuracy while operating behind the scenes. This setup enables the system to respond quickly to users without noticeable delays, ensuring an efficient and responsive interaction.

The application is designed with user engagement in mind, guiding users through a straightforward interaction process. Upon entering the application, users are greeted by a welcoming header display that provides an overview of the tool's purpose. Following this introduction, users are prompted to complete a questionnaire, which is essential for gathering the necessary data for the model's predictions.Once all questions have been answered and the questionnaire submitted, the system will trigger the KNN-stacking model to process the data. Users receive instant feedback in the results and recommendations sections within moments, providing valuable insights based on their responses. This efficient flow enhances user experience and encourages timely interventions for those seeking early autism screening.

IV.CONCLUSION

This study shows that the stacking model significantly enhances the KNN algorithm for autism detection, achieving the highest accuracy of 86.67%, compared to 82.67% for the Manhattan distance and 81.33% for the Minkowski distance. The stacking model's ability to combine multiple KNN variants makes it the most effective method for accurate and reliable early autism screening.

The model has been successfully deployed as a simple website, offering a practical tool for parents and healthcare providers to screen children for autism. For future work, exploring alternative algorithms like SVM, deep learning, and hyperparameter tuning could improve performance further. Additionally, improving the user interface and considering mobile app development would enhance accessibility and user experience, making the tool more versatile and widely adopted for early detection of autism.

ACKNOWLEDGMENT

This article is the research result funded by Universitas Dian Nuswantoro under Hibah Penelitian Pemula Perguruan Tinggi (Grant: 076/A.38-04/UDN-09/VII/2024). Furthermore, the authors would also like to thank Universitas Dian Nuswantoro for the continuous support in completing this study.

REFERENCE

[1] S. Horiuchi, O. Nomura, L. Curry, M. Kushima, L. Arnold, and Z. Yamagata, "'Caregivers' awareness matters': Call for a child-centred approach to early detection of autism spectrum disorder through well-child visits in Japan," *Child*, vol. 49, no. 6, pp. 985–994, Nov. 2023, doi: 10.1111/cch.13107.

[2] L. Francés *et al.*, "An approach for prevention planning based on the prevalence and comorbidity of neurodevelopmental disorders in 6-year-old children receiving primary care consultations on the island of Menorca," *BMC Pediatr*, vol. 23, no. 1, p. 32, Jan. 2023, doi: 10.1186/s12887-023-03844-2.

[3] M. Elbattah, R. Carette, F. Cilia, J.-L. Guérin, and G. Dequen, "Applications of machine learning methods to assist the diagnosis of autism spectrum disorder," in *Neural Engineering Techniques for Autism Spectrum Disorder, Volume 2*, Elsevier, 2023, pp. 99–119. doi: 10.1016/B978-0-12-824421-0.00013-8.

[4] H. Manohar, P. Kandasamy, V. Chandrasekaran, and R. P. Rajkumar, "Early Diagnosis and Intervention for Autism Spectrum Disorder: Need for Pediatrician-Child Psychiatrist Liaison," *Indian Journal of Psychological Medicine*, vol. 41, no. 1, pp. 87–90, Jan. 2019, doi: 10.4103/IJPSYM.IJPSYM_154_18.
[5] B. Zhang, P. You, J. Gai, J. Kang, and D. Cheng, "Multimodal Approaches for Autism Spectrum Disorder Prediction," in *Proceedings of the 2023 4th International Symposium on Artificial Intelligence for Medicine Science*, Chengdu China: ACM, Oct. 2023, pp. 769–774. doi: 10.1145/3644116.3644245.
[6] H. Abbas, F. Garberson, E. Glover, and D. P. Wall, "Machine learning approach for early detection of autism by combining questionnaire and home video screening," *Journal of the American Medical Informatics Association*, vol. 25, no. 8, pp. 1000–1007, Aug. 2018, doi: 10.1093/jamia/ocy039.
[7] M. S. Farooq, R. Tehseen, M. Sabir, and Z. Atal, "Detection of autism spectrum disorder (ASD) in children and adults using machine learning," *Sci Rep*, vol. 13, no. 1, p. 9605, Jun. 2023, doi: 10.1038/s41598-023-35910-1.
[8] A. Novianto and M. D. Anasanti, "Autism Spectrum Disorder (ASD) Identification Using Feature-Based Machine Learning Classification Model," *Indonesian J. Comput. Cybern. Syst.*, vol. 17, no. 3, p. 259, Jul. 2023, doi: 10.22146/ijccs.83585.
[9] T. Shrivastava, V. Singh, and A. Agrawal, "Autism spectrum disorder detection with kNN imputer and machine learning classifiers via questionnaire mode of screening," *Health Inf Sci Syst*, vol. 12, no. 1, p. 18, Mar. 2024, doi: 10.1007/s13755-024-00277-8.
[10] D. Prasad, S. K. Goyal, A. Sharma, A. Bindal, and V. S. Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *j comput theor nanosci*, vol. 16, no. 10, pp. 4425–4430, Oct. 2019, doi: 10.1166/jctn.2019.8536.
[11] S. Zhang, "Challenges in KNN Classification," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 10, pp. 4663–4675, Oct. 2022, doi: 10.1109/TKDE.2021.3049250.
[12] H. A. Abu Alfeilat *et al.*, "Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review," *Big Data*, vol. 7, no. 4, pp. 221–248, Dec. 2019, doi: 10.1089/big.2018.0175.
[13] A. Setiawan, "Perbandingan Penggunaan Jarak Manhattan, Jarak Euclid, dan Jarak Minkowski dalam Klasifikasi Menggunakan Metode KNN pada Data Iris," *juses*, vol. 5, no. 1, pp. 28–37, May 2022, doi: 10.24246/juses.v5i1p28-37.
[14] S. Nayak, M. Bhat, N. V. Subba Reddy, and B. Ashwath Rao, "Study of distance metrics on k - nearest neighbor algorithm for star categorization," *J. Phys.: Conf. Ser.*, vol. 2161, no. 1, p. 012004, Jan. 2022, doi: 10.1088/1742-6596/2161/1/012004.
[15] S. Mehta and K. S. Patnaik, "Stacking Based Ensemble Learning for Improved Software Defect Prediction," in *Proceeding of Fifth International Conference on Microelectronics, Computing and Communication Systems*, vol. 748, V. Nath and J. K. Mandal, Eds., in Lecture Notes in Electrical Engineering, vol. 748. , Singapore: Springer Singapore, 2021, pp. 167–178. doi: 10.1007/978-981-16-0275-7_14.
[16] M. A. Mahfouz, A. Shoukry, and M. A. Ismail, "EKNN: Ensemble classifier incorporating connectivity and density into kNN with application to cancer diagnosis," *Artificial Intelligence in Medicine*, vol. 111, p. 101985, Jan. 2021, doi: 10.1016/j.artmed.2020.101985.
[17] F. Thabtah, "Autistic Spectrum Disorder Screening Data for Children." UCI Machine Learning Repository, 2017. doi: 10.24432/C5659W.
[18] Muljono, S. A. Wulandari, H. A. Azies, M. Naufal, W. A. Prasetyanto, and F. A. Zahra, "Breaking Boundaries in Diagnosis: Non-Invasive Anemia Detection Empowered by AI," *IEEE Access*, vol. 12, pp. 9292–9307, 2024, doi: 10.1109/ACCESS.2024.3353788.

MUHAMMAD NAUFAL holds a Bachelor's degree (S.Tr.T) in Informatics Engineering at Harapan Bersama Polytechnic, Tegal, Indonesia, in 2017. He was continuing his Master in Informatics Engineering at Dian Nuswantoro University, Semarang, Indonesia, with a Magister of Computer (M.Kom) degree in 2022.

He is currently a junior lecturer in informatics engineering at Dian Nuswantoro University, Semarang, Indonesia and is affiliated with the Research Center for Intelligent Distributed Surveillance and Security. His research interest is in computer vision, applied Artificial Intelligence and he has published several of them in scientific journals and conference proceedings.

HARUN AL AZIES was born in Surabaya, Indonesia, on April 26, 1996. He received an Associate Degree (A.Md) in Business Statistics from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, in 2017. He earned his Bachelor of Science (B.Sc.) degree in statistics from PGRI Adi Buana University, Surabaya, in 2021, and completed his Master of Statistics (M.Stat) degree at the Faculty of Science and Data Analytics at ITS in 2022. He is a Junior Lecturer in Informatics Engineering at the Faculty of Computer Science, Dian Nuswantoro University in Semarang, and is affiliated with the Research Center for Quantum Computing and Materials Informatics. His research interests include computational statistics, data science, and machine learning.