

Penerapan Algoritma Naive Bayes Dan Support Vector Machine dalam Memprediksi Autisme

Riki Supriyadi¹, Nurlaelatul Maulidah², Ahmad Fauzi³, Hiya Nalatissifa⁴, Sri Diantika⁵

¹ Universitas Nusa Mandiri
e-mail: riki.rsd@nusamandiri.ac.id

^{2,3,4,5} Universitas Bina Sarana Informatika
e-mail: nurlaelatul.ntl@bsi.ac.id, ahmad.fzx@bsi.ac.id, hiya.hys@bsi.ac.id, sri.szd@bsi.ac.id

Abstrak

Autism Spectrum Disorder (ASD) merupakan gangguan perkembangan saraf yang menyebabkan anak mengalami gangguan dalam kemampuan komunikatif, fungsi sosial, dan perilaku kaku atau berulang. Diagnosis terhadap screening autisme adalah langkah awal untuk mengetahui kondisi anak dalam proses penanganan autisme secara dini. Konsultasi dengan dokter atau tenaga medis menjadi pilihan utama yang dilakukan orangtua, namun keterbatasan tenaga medis yang berfokus terhadap perkembangan otak membuat orang tua sulit menangani anaknya. Machine learning menjadi salah satu alternatif dalam penanganan penyakit terutama autisme untuk membantu meningkatkan kemampuan pendeteksian otomatis. Dalam penelitian yang telah dilakukan dengan menerapkan dua algoritma data mining yaitu algoritma Naive Bayes dan Support Vector Machine (SVM) kedua algoritma kemudian dikomparasi dimana hasil yang ditunjukkan algoritma Naive Bayes mendapatkan hasil akurasi 96,45% sedangkan Support Vector Machine (SVM) mendapatkan hasil akurasi 81,56% dan disimpulkan dalam penelitian ini algoritma yang paling sesuai untuk mendeteksi penyakit autisme yaitu Naive Bayes dimana memiliki akurasi yang lebih baik saat pengujian.

Kata Kunci : Deteksi Autisme , Naive Bayes, Support Vector Machine

Abstract

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that causes children to experience impairments in communicative abilities, social functioning, and rigid or repetitive behaviors. Diagnosis of autism screening is the first step to knowing the condition of children in the process of handling autism early. Consultation with a doctor or medical personnel is the main choice made by oragtua, but the limitations of medical personnel who focus on brain development make it difficult for parents to handle their children. Machine learning is one of the alternatives in the treatment of diseases, especially autism, to help improve automatic detection capabilities. In research that has been done by applying two data mining algorithms, namely naive bayes algorithm and support vector machine (SVM) both algorithms are then compared where the results shown by naive bayes algorithm get an accuracy of 96.45% while support vector machine (SVM) gets an accuracy of 81.56% and concluded in this study the most suitable algorithm to detect autism disease is Naive Bayes which has better accuracy when testing.

Keywords: Autism Detection, Naive Bayes, Support Vector Machine

1. Pendahuluan

Kesehatan menjadi kebahagiaan tersendiri bagi orang tua karena anak merupakan karunia yang diberikan oleh tuhan. Semua orang tua pasti megarapkan

anaknya dapat tumbuh dengan normal dan sehat. Namun tidak semua anak yang

terlahir memiliki kesamaan, karena dalam beberapa kasus ada yang memiliki kelainan atau kebutuhan khusus, seperti mengalami autisme. Autisme merupakan gangguan perkembangan mental pada anak yang menyebabkan anak sulit untuk berinteraksi sosial (Sugara et al., 2018). gangguan *Autism Spectrum Disorder (ASD)* terdiri dari beberapa gangguan perkembangan saraf kompleks yang menyatu disekitar gangguan

dalam kemampuan komunikatif, fungsi sosial, dan perilaku kaku atau berulang. Cacat yang menyebar ini digambarkan sebagai gangguan spektrum karena variasi yang cukup besar dalam bagaimana individu memanifestasikan gejala. (Jennings Dunlap, 2019)

Diagnosis autisme adalah langkah awal untuk mengetahui kondisi anak dalam proses penanganan autisme secara dini. Deteksi dini gangguan autisme merupakan hal yang terpenting dalam proses tumbuh kembang anak (Sugara et al., 2019). Namun keterbatasan pengetahuan dari orang tua mengenai gejala-gejala autisme membuat penanganan yang dilakukan terlambat serta jumlah tenaga medis yang berfokus dengan perkembangan otak khususnya autisme masih terbatas. Selain itu, beberapa bukti menunjukkan bahwa orang dengan gangguan kebutuhan kesehatan seperti spektrum autisme kadang-kadang diabaikan, oleh karena itu, mengakibatkan rendahnya tingkat kepuasan terhadap layanan kesehatan yang semakin diperburuk oleh hambatan dalam mengakses perawatan medis. (Rydzewska et al., 2019)

Perkembangan teknologi tidak bisa dipisahkan dari kehidupan masyarakat saat ini, dari berbagai bidang sudah mengandalkan teknologi termasuk bidang kesehatan. Dalam bidang kesehatan teknologi digunakan untuk mempermudah tenaga medis dalam penanganan penyakit. Diagnosis penyakit terutama autisme secara manual membutuhkan waktu yang cukup lama, maka dari itu diperlukan *tools* untuk membantu untuk pendeteksian secara dini dengan cepat agar dapat segera ditangani lebih lanjut. Berkembangnya teknologi dan informasi yang terus berinovasi terutama bidang kecerdasan buatan. Metode *machine learning* menawarkan otomatis yang efisien dan efektif model klasifikasi untuk masalah autisme karena mereka mengadopsi campuran metode matematika dan pencarian dari ilmu komputer. (Thabtah, 2017), teknik *machine learning* diperkenalkan untuk membantu meningkatkan kemampuan pendeteksian otomatis. (Fauzi et al., 2020). Beberapa macam *tools* yang dapat digunakan untuk melakukan penelitian berbasis data mining yaitu Matlab, Rapid Miner, Weka dan Python (Supriyadi et al., 2020).

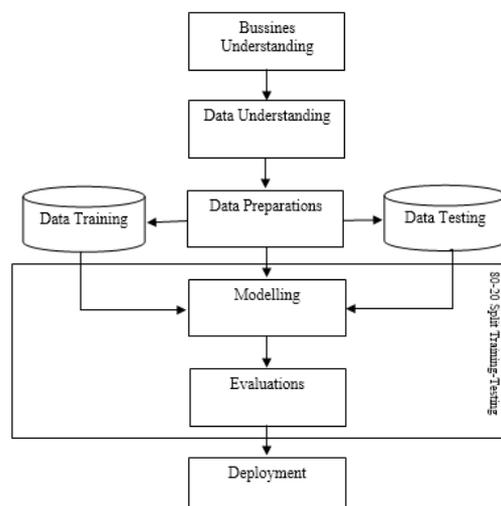
Pada penelitian terkait yang pernah dilakukan oleh (Sugara et al., 2018) deteksi

autism dengan menggunakan aplikasi RapidMiner dan menggunakan metode decision tree C4.5 hasil penelitian ini menunjukkan nilai akurasi sebanyak 72%. Dan penelitian yang dilakukan oleh (Sugara et al., 2019) dalam komparasi Algoritma C4.5 dan Naïve Bayes untuk untuk deteksi dini gangguan autisme pada anak menunjukkan bahwa dengan menggunakan Algoritma C4.5 menghasilkan nilai akurasi sebesar 72% sedangkan menggunakan Algoritma Naïve Bayes menunjukkan hasil 73.33%.

Tujuan dari penelitian yaitu mendapatkan akurasi terbaik dalam hal ini akan menggunakan algoritma Naïve Bayes Dan Support Vector Machine (SVM) untuk menentukan algoritma terbaik dalam memprediksi autisme dengan menggunakan *tools* Weka.

2. Metode Penelitian

Metode yang digunakan dalam penelitian ini yaitu *Cross Industry Standard Process for Data Mining* (CRISP-DM). untuk garis besarnya skema alur penelitian ini dapat dilihat pada Gambar 1. Model Penelitian yang dilakukan dengan perangkat lunak Weka.



Sumber : Penelitian (2021)

Gambar 1. Model Penelitian

2.1. Business Understanding

Tahapan ini merupakan proses dimana penelitian mengarah kepada tujuannya. Perumusan dan definisi masalah diselesaikan pada tahap ini. Penelitian akan menggunakan dataset Autism Screening Adult yang diperoleh dari UCI untuk melakukan prediksi terhadap fitur yang

mempengaruhi terjadinya autis. Hasil dari penelitian diharapkan dapat dijadikan sebagai acuan untuk mendeteksi Autis.

2.2. Data Understanding

Data yang akan diklasifikasikan perlu dipahami sebelum penelitian dilakukan. Dataset yang digunakan merupakan data publik dari dataset *Autism Screening Adult*, merupakan jenis data klasifikasi yang berasal dari UCI *Repository* yang dapat diperoleh di <https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult>. Dataset autisme ini terdiri dari 704 data.

Tabel 1. Atribut Dataset

Atribut	Tipe
Usia	Number
Jenis Kelamin	String
Etnis	String
Sakit kuning saat lahir	Booelan
Keluarga pengidap autis	Booelan
Menggunakan aplikasi sebelumnya	String
Negara tempat tinggal	String
Pernah melakukan tes sebelumnya	Booelan
Tipe tes	Integer
hasil	Integer
Jawaban pertanyaan 1	Binary (0,1)
Jawaban pertanyaan 2	Binary (0,1)
Jawaban pertanyaan 3	Binary (0,1)
Jawaban pertanyaan 4	Binary (0,1)
Jawaban pertanyaan 5	Binary (0,1)
Jawaban pertanyaan 6	Binary (0,1)
Jawaban pertanyaan 7	Binary (0,1)
Jawaban pertanyaan 8	Binary (0,1)
Jawaban pertanyaan 9	Binary (0,1)
Jawaban pertanyaan 10	Binary (0,1)

Sumber : Respository UCI (2017)

2.3. Data Preparation

Pada tahap ini data disiapkan untuk dilakukan proses pelatihan. Datanya sendiri terdiri dari 704 data. Kemudian tahap pengolahan data yang bertujuan untuk membangun dataset akhir yang akan diproses pada tahap pemodelan. Tahap pengolahan data meliputi pemilihan kelas, atribut data, dan transformasi data, dilanjutkan dengan proses pembersihan dan pengaturan data. Proses ini mencoba untuk menghilangkan nilai-nilai yang hilang. Dari 21 atribut yang terdapat dalam dataset, atribut yang dijadikan kelas yaitu *Class/ASD*

yang berisi data Ya dan Tidak.

2.4. Modeling

Algoritma yang digunakan dalam data latih ini ada tiga algoritma yaitu : Algoritma *Naive Bayes* , *Support Vector Machine* (SVM). Data yang akan digunakan untuk peramalan dibagi menjadi dua, yaitu data training dan data testing. Pembagiannya 80% data training dan 20% data testing, artinya *data training* berjumlah 563 dan *data testing* sejumlah 141. Dengan dua algoritma bertujuan untuk mendapatkan model terbaik yang dapat melakukan prediksi dengan tingkat akurasi tertinggi terhadap dataset *Autism Screening Adult* dan memprediksi atribut apa yang paling berpengaruh besar terhadap autis, sehingga memberikan hasil yang optimal.

2.5. Evaluation

Pada tahap ini dilakukan evaluasi berdasarkan dua algoritma yaitu *Naive Bayes* dan *Support Vector Machine* (SVM) kemudian dibuat perbandingan untuk memilih algoritma yang paling tepat berdasarkan nilai dari hasil *correctly classified*, *confusion matrix*, AUC dan *accuracy*.

2.6. Deployment

Setelah tahap evaluasi dilakukan dengan menilai secara detail dari hasil permodelan dan dilakukan pengimplementasian dari keseluruhan Model yang telah di rancang sehingga mendapatkan hasil sesuai yang diharapkan. Prediksi autisme dilakukan berdasarkan atribut yang ada di dalamnya kemudian dapat dikembangkan agar selanjutnya dapat menentukan faktor utama yang paling berpengaruh terhadap autisme sehingga dapat mengambil tindakan penanganan lebih lanjut

3. Hasil dan Pembahasan

Pada penelitian yang telah dilakukan terhadap dataset *Autism Screening Adult* dengan 704 data menggunakan dua algoritma yaitu *Naive Bayes* dan *Support Vector Machine* (SVM), berikut hasil yang telah didapatkan :

3.1. Naive Bayes

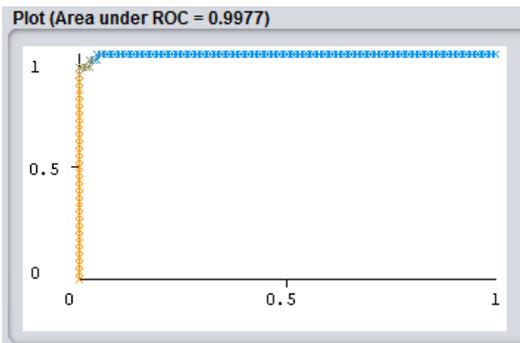
Tabel 2. Hasil Pengujian *Naive Bayes*

	True	True	Precision
--	------	------	-----------

	No	Yes	
Pred.No	106	3	0,981
Pred.Yes	2	30	0,909
Recall	0,972	0,938	

Sumber : Hasil Penelitian (2021)

Pada Tabel 2. diatas menunjukkan *Confusion Matrix* hasil penelitian yang diperoleh dari pengujian dengan algoritma *Naive Bayes* dan mendapatkan hasil akurasi sebesar 96,4539%



Sumber : Hasil Penelitian (2021)

Gambar 2. Hasil AUC Yes *Naive Bayes*

Dapat dilihat pada gambar 2 menunjukkan hasil AUC dengan algoritma *Naive Bayes* menghasilkan 0,9977

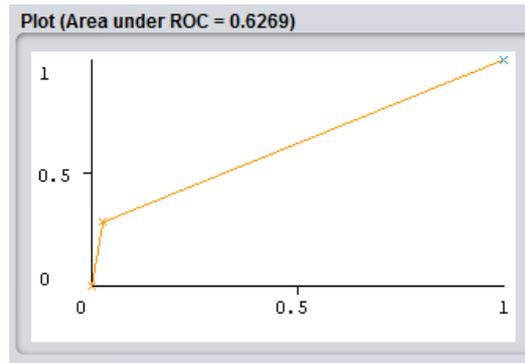
3.2. Support Vector Machine (SVM)

Tabel 3. Hasil Pengujian *Support Vector Machine* (SVM)

	True No	True Yes	Precision
Pred.No	106	3	0,822
Pred.Yes	23	9	0,750
Recall	0,972	0,281	

Sumber : Hasil Penelitian (2021)

Pada Tabel 3. diatas adalah *Confusion Matrix* hasil penelitian yang diperoleh dari pengujian dengan algoritma *Support Vector Machine* (SVM) dan mendapatkan hasil akurasi sebesar 81,5603%



Sumber : Hasil Penelitian (2021)

Gambar 3. Hasil AUC Yes SVM

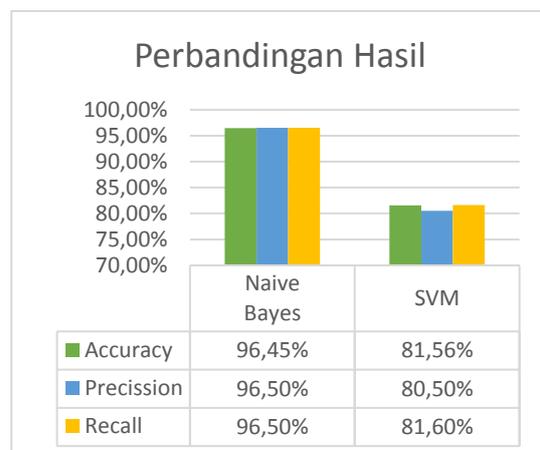
Dapat dilihat pada gambar 3 menunjukkan hasil AUC dengan algoritma SVM menghasilkan 0,6269

Tabel 4. Perbandingan Hasil

	<i>Naive Bayes</i>	SVM
Accuracy	96,4539%	81,5603%
Precision	96,5%	80,5%
Recall	96,5%	81,6%

Sumber : Hasil Penelitian (2021)

Pada tabel 4 menunjukkan hasil perbandingan antara dua algoritma yaitu *Naive Bayes* dan *Support Vector Machine* (SVM) yang dikomparasikan, dimana hasil yang diperoleh oleh *Naive Bayes* secara keseluruhan lebih unggul dibandingkan dengan hasil dari *Support Vector Machine* (SVM).



Sumber : Hasil Penelitian (2021)

Gambar 4. Hasil Perbandingan 2 algoritma

Pada gambar 4 menunjukkan grafik secara visual dari hasil perbandingan dua algoritma yang diterapkan pada penelitian pendeteksi autisme yang menunjukkan hasil masing-masing algoritma

4. Kesimpulan

Penelitian ini dilakukan untuk mendeteksi penyakit autisme dengan tingkat akurasi yang baik. Untuk memprediksi permasalahan tersebut digunakan dua algoritma yaitu *Naive Bayes* dan *Support Vector Machine* (SVM). Yang kemudian dikomparasi hasilnya masing-masing. Dimana hasil evaluasi pengujian tersebut akan dilihat nilai akurasinya. Berdasarkan penelitian yang telah dilakukan tingkat akurasi prediksi *Naive Bayes* mampu memberikan tingkat akurasi yang baik dimana memperoleh hasil akurasi sebesar 96,45% sedangkan pengujian kedua dengan *Support Vector Machine* (SVM) mendapatkan hasil akurasi sebesar 81,56%. Dari hasil penelitian yang sudah dilakukan dapat disimpulkan bahwa penerapan *Naive Bayes* lebih baik dibanding *Support Vector Machine* (SVM) dalam hal mendeteksi penyakit autis.

Dengan hasil penelitian yang telah dilakukan dengan membandingkan dua algoritma hal ini dapat menentukan algoritma yang lebih baik untuk diterapkan dalam pembuatan *tools*, namun hal ini masih dapat dikembangkan lebih lanjut dengan komparasi metode lain atau penambahan fitur lagi agar bisa menghasilkan akurasi yang lebih baik lagi, sehingga dalam pembuatan *tools* pendeteksi autisme dapat lebih optimal nantinya, baik secara hasil dan juga kualitasnya juga kualitasnya lebih optimal nantinya, baik secara hasil dan juga kualitasnya.

Referensi

- Fauzi, A., Supriyadi, R., & Maulidah, N. (2020). *Deteksi Penyakit Kanker Payudara dengan Seleksi Fitur berbasis Principal Component Analysis dan Random Forest*. 2(1).
- Jennings Dunlap, J. (2019). Autism Spectrum Disorder Screening and Early Action. *Journal for Nurse Practitioners*, 15(7), 496–501. <https://doi.org/10.1016/j.nurpra.2019.04.001>
- Rydzewska, E., Hughes-McCormack, L. A., Gillberg, C., Henderson, A., MacIntyre, C., Rintoul, J., & Cooper, S. A. (2019). General health of adults with autism spectrum disorders – A whole country population cross-sectional study. *Research in Autism Spectrum Disorders*, 60(January 2018), 59–66. <https://doi.org/10.1016/j.rasd.2019.01.004>
- Sugara, B., Adidarma, D., & Budilaksono, S. (2019). Perbandingan Akurasi Algoritma C4.5 dan Naive Bayes Untuk Deteksi Dini Gangguan Autisme Pada Anak. *Jurnal IKRA-ITH Informatika*, 3(1), 119–128.
- Sugara, B., Widyatmoko, D., Prakoso, B. S., & Saputro, D. M. (2018). Penerapan Algoritma C4.5 untuk Deteksi Dini Autisme Pada Anak. *Seminar Nasional Teknologi Informasi Dan Komunikasi (SENTIKA)*, 2018(Sentika), 87–96.
- Supriyadi, R., Gata, W., Maulidah, N., & Fauzi, A. (2020). Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah. *E-Bisnis: Jurnal Ilmiah Ekonomi Dan Bisnis*, 13(2), 67–75. <https://doi.org/10.51903/e-bisnis.v13i2.247>
- Thabtah, F. (2017). *Autism Spectrum Disorder Screening : Machine Learning Adaptation and DSM-5 Fulfillment*. 1–6.