MATHunesa
Jurnal Ilmiah Matematika *Volume 09 No 01*
e-ISSN : 2716-506X | p-ISSN : 2301-9115 *Tahun 2021*

KLASIFIKASI *AUTISM SPECTRUM DISORDER* MENGGUNAKAN ALGORITMA *NAÏVE BAYES*

Erina Seviyanti Dewi
Jurusan Matematika, FMIPA, Universitas Negeri Surabaya
email: erina.17030214021@mhs.unesa.ac.id

Abstrak

Autism Spectrum Disorders (ASD) merupakan suatu kelainan pada otak manusia yang mengakibatkan seseorang mengalami gangguan dalam melakukan komunikasi dan interaksi terhadap sosial. Meningkatnya penderita *Autism Spectrum Disorders* di dunia memerlukan deteksi dini terhadap kelainan tersebut untuk mengurangi risiko negatif yang ditimbulkan dan memberikan perawatan yang tepat untuk penderitanya. Beberapa peneliti telah melakukan klasifikasi *Autism Spectrum Disorders*, namun belum ada yang mengklasifikasikan *Autism Spectrum Disorders* dengan algoritma *Naive Bayes*. Sehingga pada penelitian ini klasifikasi *Autism Spectrum Disorders* dilakukan berdasarkan *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition* (DSM-5) dengan algoritma *Naive Bayes*. Dalam dataset terdapat 292 anak dengan dua kelas yaitu normal dan *Autism Spectrum Disorders* yang diperoleh dari *UCI Machine Learning Repository*. Terdapat 20 atribut yang digunakan untuk mengklasifikasikan *Autism Spectrum Disorders* pada anak. Data yang ada dibagi menjadi data latih dan data uji berdasarkan *hold out validation*. Berdasarkan hasil klasifikasi, *Naive Bayes* dengan rasio data latih dan uji 1:1 menghasilkan akurasi tertinggi sebesar 98.6301%.
Kata Kunci: *Autism Spectrum Disorders*, DSM-5, *Naive Bayes*

Abstract

Autism Spectrum Disorders (ASD) is an abnormality in the human brain that causes a person to experience problems in communicating and interacting with society. The increase in Autism Spectrum Disorders sufferers in the world requires early detection of these disorders to reduce negative risks posed then provide appropriate care for sufferers. Several researchers have done Autism Spectrum Disorders classification, but no one has classified Autism Spectrum Disorders using the Naive Bayes algorithm. So that in this study the Autism Spectrum Disorders classification was carried out based on the Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-5) using the Naive Bayes algorithm. The dataset consisted of 292 children with two classes, namely normal and Autism Spectrum Disorders, which were obtained from UCI Machine Learning Repository. There are 20 attributes used to classify Autism Spectrum Disorders in children. Existing dataset is divided into train data and test data based on hold out validation. Based on the classification results, Naive Bayes with a ratio of training and testing data of 1:1 produces the highest accuracy of 98.6301%.
Keywords: *Autism Spectrum Disorders*, DSM-5, *Naive Bayes*

PENDAHULUAN

Autism Spectrum Disorders (ASD) adalah suatu kelainan yang mengakibatkan seseorang mengalami gangguan dalam komunikasi dan interaksi terhadap sosial (Lord dkk., 2020). Gangguan pada perkembangan otak ini disebabkan oleh kecacatan genetik pada janin dalam kandungan (Pardo dkk., 2009). Dalam dua tahun terakhir ini, kasus *Autism Spectrum Disorders* (ASD) terutama pada anak-anak menunjukkan kenaikan kasus yang cukup tinggi. Pada 2018, penelitian yang dilakukan Baio telah mengidentifikasi jumlah anak *Autism Spectrum Disorders* (ASD) adalah 1,47% dari total populasi anak di dunia (Baio dkk., 2018). Sedangkan *World Health Organization* (WHO) telah mengumumkan bahwa setiap 66 anak di dunia, seorang diantaranya mengalami *Autism Spectrum Disorders* (ASD) (WHO, 2019). Berdasarkan jumlah kasus anak *Autism Spectrum Disorders* (ASD) yang selalu bertambah, diperlukan deteksi dini dan melakukan perawatan yang sesuai (Parmeggiani dkk., 2019).

Klasifikasi individu dengan *Autism Spectrum Disorders* (ASD) saat ini merupakan topik yang menarik bagi peneliti. Penelitian yang dilakukan oleh Ibrahim mengklasifikasikan *Autism Spectrum Disorders* (ASD) berdasarkan *Electroencephalography* (EEG) menjadi dua dan tiga kelas. Klasifikasi *Electroencephalography* (EEG) menjadi dua kelas yaitu normal dan *Autism Spectrum Disorders* (ASD), sedangkan klasifikasi menjadi tiga kelas yaitu normal, *Autism Spectrum Disorders* (ASD), dan *epilepsy*. Dalam penelitian ini, Ibrahim menggunakan kombinasi *Discrete Wavelet Transform* (DWT) dan beberapa metode non linier yaitu *Shannon entropy*,

largest Lyapunov exponent, *standard deviation* dan *band power*. Untuk mengklasifikasikan *Electroencephalography* (EEG) tersebut, digunakan algoritma *Artificial Neural Network* (ANN), *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), dan *Linear Discriminant Analysis* (LDA). *Channel* yang digunakan dalam penelitian ini yaitu *single-channel* dan *multi-channel*. Akurasi tertinggi dicapai pada kombinasi metode *Discrete Wavelet Transform* (DWT) dan *Shannon entropy* dengan algoritma *K-Nearest Neighbor* (KNN), yaitu 94,6% untuk klasifikasi menjadi tiga kelas dengan multi-channel (Ibrahim dkk., 2018).

Beberapa penelitian telah menggunakan *machine learning* dalam klasifikasi *Autism Spectrum Disorders* (ASD). *Machine learning* adalah metode yang dapat menghasilkan prediksi hasil suatu sistem berdasarkan model yang telah dibentuk sebelumnya (Parikesit dkk., 2019). Berbagai algoritma dalam *machine learning* seperti *Naive Bayes, Random Forest*, dan *Support Vector Machine* (SVM) adalah contoh algoritma yang sering digunakan dan tervalidasi (Jamal dan Scaria, 2013). Menurut penelitian Kukreja dalam membandingkan beberapa metode *machine learning*, disimpulkan bahwa *Naive Bayes* ialah metode yang tercepat dan efektif (Kukreja, M., dkk., 2012).

Berdasarkan penelitian yang telah dilakukan untuk mengklasifikasikan *Autism Spectrum Disorders* (ASD), belum adanya klasifikasi berdasarkan data DSM-5 atau *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition* menggunakan metode *machine learning*. Selain itu metode *Naive Bayes* belum banyak digunakan dalam klasifikasi *Autism Spectrum Disorders* (ASD). Sehingga dalam penelitian ini dilakukan klasifikasi *Autism Spectrum Disorders* (ASD) berdasarkan data *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition* (DSM-5) menggunakan algoritma *Naive Bayes*. Variasi beberapa rasio data pelatihan dan data pengujian dilakukan pada penelitian ini untuk mengetahui rasio terbaik dalam menghasilkan akurasi tertinggi pada klasifikasi *Autism Spectrum Disorders* (ASD). Selain itu dilakukan variasi nilai *batchsize* pada algoritma *Naive Bayes*.

KAJIAN PUSTAKA

AUTISM SPECTRUM DISORDERS (ASD)

Gangguan yang sering disebut *Autism Spectrum Disorders* (ASD) ialah suatu kelainan dimana dapat mengakibatkan seseorang mengalami gangguan dalam komunikasi dan interaksi terhadap sosial (Lord dkk., 2020). Beberapa gejala yang sering dialami seseorang dengan *Autism Spectrum Disorders* (ASD) adalah perilaku yang hiperaktif, cenderung agresif, sering melakukan tindakan untuk menyakiti diri, dan sering mengalami perubahan suasana hati (Kazdoba dkk., 2016). Faktor penyebab seseorang mengalami *Autism Spectrum Disorders* (ASD) dapat berasal dari faktor genetik dan faktor lingkungan (Lyall, K dkk., 2014). Menurut Beata dalam penelitiannya, faktor lingkungan yang mempengaruhi terjadinya *Autism Spectrum Disorders* (ASD) ialah bayi yang memiliki berat badan terlalu rendah atau disebut BBLR atau Berat Badan Lahir Rendah, usia ibu pada waktu mengandung, dan infeksi yang terjadi saat usia kehamilan (Beata dkk., 2016).

DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS FIFTH EDITION (DSM-5)

Suatu alat yang diciptakan oleh *American Psychiatric Association* (APA) untuk mendiagnosa penyakit psikis pada manusia yaitu *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition* (DSM-5) (APA, 2013). Dalam *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition* (DSM-5) terdapat pedoman mengenai pengetahuan beberapa gangguan jiwa yang dapat digunakan oleh praktisi untuk menganalisa kejiwaan seseorang. *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition* (DSM-5) menyajikan pendekatan dimensional yang terdiri dari beberapa pertanyaan mengenai kebiasaan seseorang dan informasi individu. Pendekatan dimensional ini mampu menilai dan memetakan gangguan kejiwaan (Vahia, 2013). Metode ini merupakan pembahuruan dari *Diagnostic and Statistical Manual of Mental Disorders Fourth Edition* (DSM IV) dan telah dikenalkan pada masyarakat sejak tahun 2013. Kriteria yang digunakan dalam *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition* (DSM-5) mampu mengenali gangguan secara spesifik. Pembaharuan metode ini salah satunya terdapat pada rekonseptualisasi *Asperger*

Syndrome dengan gangguan yang lebih spesifik menjadi *Autism Spectrum Disorders* (APA, 2013).

NAIVE BAYES

Naive Bayes adalah suatu algoritma dimana dapat mengklasifikasikan data berdasarkan teori peluang serta Teorema Bayesian dengan memberikan asumsi bahwa variabel X merupakan variable yang independen (Dwi dkk., 2018). Dalam klasifikasi dengan menggunakan *Naive Bayes*, frekuensi dan kombinasi nilai yang ada dalam dataset dijumlahkan satu sama lain sehingga menghasilkan probabilitas. *Naive Bayes* termasuk metode *supervised*, dimana dalam melakukan klasifikasi, algoritma ini memerlukan data pelatihan. Menurut penelitian yang dilakukan oleh Syarifah dan Muslim terdapat banyak keuntungan dalam menggunakan algoritma *Naive Bayes* untuk klasifikasi data (Syarifah dan Muslim, 2015). Hal ini dikarenakan algoritma *Naive Bayes* tidak membutuhkan data latih yang banyak untuk menentukan mean dan varian dari variabel yang diperlukan pada proses klasifikasi. Selain itu, *Naive Bayes* telah terbukti dapat mengklasifikasikan data secara mudah, cepat, dan memiliki akurasi yang tinggi (Supriyanto dkk., 2013).

METODE

PRA-PEMROSESAN DATA

Pada tahap pra-pemrosesan data, data *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition* (DSM-5) yang telah diinputkan kemudian dibagi menjadi data latih dan data uji. Penentuan data latih dan data uji ini berdasarkan metode *hold out validation*. Pada metode *hold out validation*, data yang telah ada dipartisi menjadi dua sesuai dengan rasio yang diinginkan. Pada penelitian ini terdapat 9 rasio data latih dan uji yaitu 9:1, 4:1, 7:3, 3:2, 1:1, 2:3, 3:7, 1:4, dan 1:9. Rasio data 9:1 berarti pada klasifikasi digunakan data latih sebanyak 9/10 dari total dataset dan 1/10 digunakan sebagai data uji, hal tersebut berlaku untuk rasio yang lainnya.

PROSES KLASIFIKASI NAIVE BAYES

Tahapan klasifikasi data dalam peneltian ini menggunakan algoritma *Naive Bayes* dengan variasi nilai *batchsize*. Nilai *batchsize* yang digunakan pada penelitian ini yaitu *batchsize* 50, 100, dan 200. Variasi ini dilakukan untuk mengetahui performa sistem, waktu untuk membangun model, dan waktu untuk melakukan pengujian data.

Algoritma *Naive Bayes* didasarkan pada *Conditional Probability* (Peluang Bersyarat). Peluang bersyarat adalah peluang kemunculan suatu kejadian bergantung pada kejadian yang telah muncul sebelumnya. Peeluang bersyarat ini biasanya dilambangkan dengan P(Y|X) dimana peluang kejadian Y terjadi dengan syarat kejadian X. Peluang bersyarat terdapat pada persamaan (1) dan (2).

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)} \quad (1)$$

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} \quad (2)$$

dimana $P(Y|X) \neq P(X|Y)$

Dari peluang bersyarat, kemudian dibentuk *Bayes Rule*. Pada *Bayes Rule*, P(X|Y) digunakan untuk menentukan nilai P(Y|X). Berdasarkan persamaan (2), maka:

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$$

$$P(X \cap Y) = P(X|Y) \,.\, P(Y) \quad (3)$$

Subtitusi persamaan (3) pada persamaan (1), sehingga:

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)}$$

$$P(Y|X) = \frac{P(X|Y) \,.\, P(Y)}{P(X)} \quad (4)$$

Persamaan (4) tersebut yang dinamakan *Bayes Rule* dimana P(Y|X) adalah peluang posterior, P(X|Y) adalah peluang X terjadi dengan syarat kejadian Y, P(Y) adalah peluang prior, dan P(X) adalah peluang kemunculan *evidence* (atribut).

Dalam klasifikasi, variabel Y menyatakan kelas dataset dan variabel X menyatakan atribut atau fitur dari dataset. Sehingga variabel Y dikontruksi menjadi K dan variabel X menjadi At.

Pada kehidupan nyata, dalam mengklasifikasikan data atribut yang digunakan umumnya lebih dari satu. Sehingga terdapat multi variabel At. Apabila At tersebut independen (saling bebas), maka:

$$P(Y|X) = \frac{P(X|Y) \,.\, P(Y)}{P(X)}$$

$$P(K|At) = \frac{P(At|K) \,.\, P(K)}{P(At)}$$

$$P(K|At_{i=1}^{n}) = \frac{P(At_1 \ldots At_n|K) \,.\, P(K)}{P(At_1,\ldots,At_n)}$$

$$P(K|At_{i=1}^{n}) = \frac{P(At_2 \ldots At_n|K, At_1).P(At_1|K).P(K)}{P(At_1,\ldots,At_n)}$$

$$P(K|At_{i=1}^{n}) = \frac{P(At_n|K, At_1 \dots At_{n-1}) \dots P(At_1|K).P(K)}{P(At_1,\dots,At_n)} \quad (5)$$

Berdasarkan rumus kejadian saling bebas (independen)

$$P(X \cap Y) = P(X).P(Y) \quad (6)$$

dan peluang bersyarat pada persamaan (2), maka:

$$P(At_i|At_j) = \frac{P(At_i \cap At_j)}{P(At_j)} = \frac{P(At_i)P(At_j)}{P(At_j)} = P(At_i) \quad (7)$$

untuk $i \neq j$.

Sehingga

$$P(At_i|K, At_j) = P(At_i|K) \quad (8)$$

Kemudian subtitusi persamaan (8) ke persamaan (5).

$$P(K|At_{i=1}^{n}) = \frac{P(At_n|K, At_1 \dots At_{n-1}) \dots P(At_1|K).P(K)}{P(At_1,\dots,At_n)}$$

$$P(K|At_{i=1}^{n}) = \frac{P(At_1|K).P(At_2|K) \dots p(At_n|K).P(K)}{P(At_1) \dots P(At_n)}$$

$$P(K|At_{i=1}^{n}) = \frac{P(K)\prod_{i=1}^{n} P(At_i|K)}{\prod_{i=1}^{n} P(At_i)}$$

Karena nilai $P(At_i)$ bernilai sama pada setiap kelasnya maka persamaan tersebut menjadi:

$$P(K|At_{i=1}^{n}) = P(K)\prod_{i=1}^{n} P(At_i|K) \quad (9)$$

Jadi persamaan (9) yang disebut rumus *Naive Bayes*.

Dalam menghitung data kontinu atau numerik, digunakan *Gaussian Naive Bayes*. Pada *Gaussian Naive Bayes* data diasumsikan berdistribusi normal, digunakan sehingga *Probability Density Function* (PDF) dari distribusi normal. Rumus *Gaussian Naive Bayes* terdapat pada persamaan (10).

$$P(At_i|K = k) = \frac{1}{\sqrt{2\pi\sigma_k}} . e^{\frac{-(At_i - \mu_k)^2}{2{\sigma_k}^2}} \quad (10)$$

dengan μ adalah mean dan σ adalah standar deviasi.

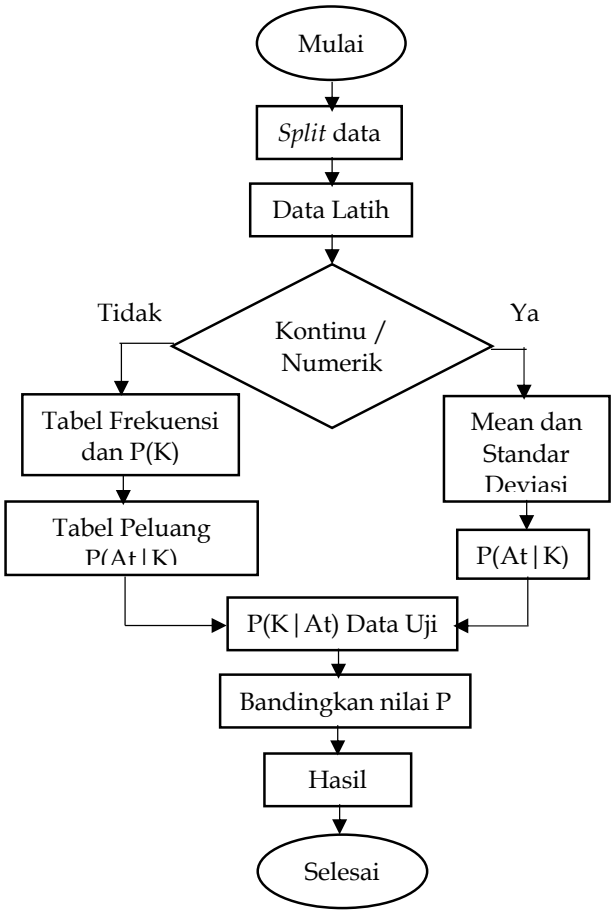
Apabila terjadi peluang memiliki nilai 0, digunakan teknik perhitungan *Laplacian Correction*. Teknik ini menambahkan nilai 1 pada atribut yang memiliki nilai peluang 0 pada peluang prior dan *conditional probability*.

$$P(Y = Class_k) = \frac{\sum_{i=1}^{n}(y_i = Class_k) + 1}{N + k} \quad (11)$$

$$P(X = At_j|Y = Class_k) = \frac{\sum_{i=1}^{n}(x_i = At_j, y_i = Class_k) + 1}{N \sum_{i=1}^{n}(y_i = Class_k) + A} \quad (12)$$

dimana k adalah banyak kelas, dan A adalah banyaknya nilai yang berbeda pada satu atribut.

Diagram alir algoritma *Naive Bayes* terdapat pada Gambar 1.



Gambar 1. Diagram Alir Algoritma *Naive Bayes*

Berdasarkan Gambar 1, tahapan algoritma *Naive Bayes* adalah:

Tahap 1: Input dataset dan *split* data menjadi data latih dan data uji berdasarkan rasio tertentu. Misalkan terdapat dataset dengan jumlah data 292 data. Apabila *split* data dengan rasio 9:1, berarti 9/10 dari total dataset digunakan sebagai data latih dan 1/10 dari total dataset digunakan sebagai data uji. Sehingga diperoleh:

$$Data\ Latih = \frac{9}{10} x292 = 262.8\ data$$

$$Data\ Uji = \frac{1}{10} x292 = 29.2\ data$$

Karena hasil perhitungan jumlah data latih dan data uji dalam bentuk desimal, maka dilakukan pembulatan hasil ke nilai terdekat. Sehingga dari contoh tersebut dihasilkan jumlah data latih yaitu 263 data dan jumlah data uji yaitu 29 data. Data latih dipilih secara

acak sebanyak 263 data, dan dataset yang tidak terpilih sebagai data latih digunakan sebagai data uji.

Tahap 2: Berdasarkan data latih, bila data tersebut numerik atau kontinu, hitung mean dan standar deviasi (std) dari masing-masing kelas dengan rumus sebagai berikut:

$$Mean = \mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (13)$$

$$Std = \sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2} \quad (14)$$

Apabila data bukan numerik, membuat tabel frekuensi dengan $[\boldsymbol{n}(At_i \cap K)]$ dan hitung peluang kelas prior $P(K)$.

Tahap 3: Menghitung peluang $P(At_i|K)$, jika data nominal (bukan numerik) maka dihitung menggunakan peluang bersyarat $P(X|Y)$ pada rumus (2), dimana X adalah atribut dan Y adalah kelas data. Namun apabila data bertipe numerik maka dihitung menggunakan rumus (10).

Tahap 4: Menghitung $P(K|At_{i=1}^{n})$ data uji dengan rumus (9). Apabila terdapat $P(At_i|K)$ yang bernilai 0, maka dilakukan proses *Laplacian Correction* dengan menggunakan rumus (11) dan (12).

Tahap 5: Membandingkan nilai $P(K|At_{i=1}^{n})$ dari tiap kelas.

Tahap 6: Mengklasifikasikan data uji pada kelas tertentu berdasarkan nilai $P(K|At_{i=1}^{n})$ terbesar.

HASIL DAN PEMBAHASAN

Pada penelitian ini, dataset diperoleh dari *UCI Machine Learning repository* (UCI Machine Learning, 2017). Dalam dataset tersebut terdapat 292 subjek anak-anak yang terdiri dari 141 subjek *Autism Spectrum Disorders* (ASD) dan 151 subjek normal. Data diperoleh dari anak usia 4 hingga 11 tahun dengan 208 pria dan 84 wanita. Terdapat 20 atribut dalam dataset ini, yaitu 10 pertanyaan mengenai perilaku anak dan 10 karakter individu yang dapat digunakan sebagai acuan dalam mendeteksi *Autism Spectrum Disorders* (ASD) pada anak. Karakterter individu tersebut meliputi umur, jenis kelamin, suku, penyakit kuning, keluarga yang mengalami *Autism Spectrum Disorders* (ASD), siapa yang menyelesaikan tugas yang diberikan, negara, penggunaan *screening app*, dan *screening score*.

Dataset yang ada dilakukan uji klasifikasi dengan menggunakan algoritma *Naive Bayes* dan menerapkan beberapa variasi nilai *batchsize* dan rasio data latih terhadap data uji. Dari pengujian tersebut didapatkan *confusion matrix* yang terbentuk setelah proses klasifikasi. *Confusion matrix* terdiri dari TP (*True Positive*), FP (*False Positive*), FN (*False Negative*), dan TN (*True Negative*). Bentuk dari *Confussion Matrix* terdapat pada Tabel 1.

Tabel 1. *Confussion Matrix*

		Kelas Prediksi	
		No	Yes
Kelas Aktual	No	TP	FN
	Yes	FP	TN

dengan TP (*True Positive*) merupakan jumlah kelas 'No' pada data uji yang diklasifikasikan sebagai kelas 'No', FP (*False Positive*) adalah jumlah kelas 'Yes' pada data uji yang diklasifikasikan sebagai kelas 'No', FN (*False Negative*) adalah jumlah kelas 'No' pada data uji yang diklasifikasikan sebagai kelas 'Yes', dan TN (*True Negative*) merupakan jumlah kelas 'Yes' pada data uji yang diklasifikasikan sebagai kelas 'Yes'. Hasil klasifikasi tersebut kemudian diukur performa sistem dengan menghitung akurasi, *precison*, *recall*, dan *specificity*.

Akurasi adalah jumlah data hasil pengujian yang diklasifikasikan benar terhadap total data yang diuji. Nilai akurasi dihitung menggunakan persamaan (15).

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (15)$$

Pecision merupakan tingkat efektivitas sistem dalam memberikan informasi yang relevan dari total prediksi oleh sistem. Perhitungan *precision* dirumuskan pada persamaan (16).

$$Precision = \frac{TP}{TP+FP} \quad (16)$$

Selain dievalusi menggunakan perhitungan akurasi dan precision, performa sistem juga dapat diukur melalui *recall* (*sensitivity*). *Recall* merupakan tingkat kemampuan model untuk menemukan ulang informasi yang bermakna dalam dataset. *Recall* dihitung dengan persamaan (17).

$$Recall = \frac{TP}{TP+FN} \quad (17)$$

Specificity merupakan tingkat efektivitas sistem dalam memprediksi kelas negatif. Rumus *specificity* terdapat pada persamaan (18).

$$Specificity = \frac{TN}{TN+FP} \quad (18)$$

Berdasarkan hasil klasifikasi menggunakan algoritma *Naive Bayes* dengan melakukan nilai *batchsize* dan rasio data latih, didapatkan hasil akurasi dari setiap percobaan. Hasil akurasi terdapat pada Tabel 2.

Tabel 2. Tabel Akurasi Algoritma *Naive Bayes*

Rasio	Batch Size		
	50	100	200
9:1	96.5517	96.5517	96.5517
4:1	96.5517	96.5517	96.5517
7:3	96.5909	96.5909	96.5909
3:2	98.2906	98.2906	98.2906
1:1	**98.6301**	**98.6301**	**98.6301**
2:3	97.7143	97.7143	97.7143
3:7	98.0392	98.0392	98.0392
2:4	97.8632	97.8632	97.8632
1:9	95.4373	95.4373	95.4373

Variasi dilakukan dengan menerapkan nilai *batchsize* yang berbeda untuk setiap persentase data latih yaitu 50, 100, dan 200. Pada penelitian ini, nilai *batchsize* dipilih secara acak. *Batchsize* merupakan subset dari data latih. Misalkan terdapat rasio 9:1, maka diperoleh jumlah data latih sebesar 263 data dan 29 data uji. Apabila digunakan nilai *batchsize* 50, maka dilakukan proses pelatihan data dengan mengambil 50 data pertama, selanjutnya 50 data kedua, dan seterusnya hingga data latih terakhir. Karena jumlah data latih yaitu 263 data, maka proses pelatihan terakhir hanya digunakan 13 data. Apabila nilai *batchsize* lebih besar dari jumlah data latih yang digunakan, maka dalam proses membangun model digunakan data latih tanpa menerapkan nilai *batchsize*.

Menurut Kotsiantis, konsep dari algoritma *Naive Bayes* ialah menggunakan *batch mode* dalam melakukan pelatihan data, dimana data latih dikumpulkan dan dihitung peluang setiap kejadian (Kotsiantis, 2013). Hal ini menyebabkan data yang telah dipartisi menjadi beberapa *batch* dikumpulkan kembali untuk dihitung peluangnya. Sehingga diperoleh data latih yang sama dengan menerapkan beberapa nilai *batchsize* pada persentase data latih. Data latih digunakan untuk membangun model dalam klasifikasi. Model yang diperoleh berdasarkan Tabel 2 ialah berupa peluang kejadian yang nantinya akan digunakan pada klasifikasi data uji.

Berdasarkan Tabel 2, hasil akurasi terbaik terdapat pada rasio 1:1 dengan akurasi 98,6301% disetiap nilai *batchsize*. Hal ini disebabkan rasio 1:1 antara data latih dan data uji merupakan rasio yang ideal dalam melakukan klasifikasi data. Sedangkan hasil akurasi terendah terdapat pada rasio 1:9 dengan akurasi 95.4373%. Jumlah data latih yang sangat kecil mengakibatkan model kurang melakukan pembelajaran terhadap data yang ada.

Berdasarkan analisis hasil klasifikasi pada Tabel 2, nilai *batchsize* pada algoritma *Naive Bayes* tidak mempengaruhi hasil akurasi dalam klasifikasi. Secara konsep pemilihan nilai *batchsize* yang kecil maka akurasi sistem menjadi lebih baik. Namun pada eksperimen yang telah dilakukan, diperoleh hasil bahwa nilai *batchsize* tidak berpengaruh signifikan pada akurasi. Hal ini dikarenakan data latih yang digunakan dalam setiap nilai *batchsize* adalah sama. Perubahan nilai *batchsize* berpengaruh terhadap waktu latih dan uji data yang ditunjukkan pada Gambar 2 dan Gambar 3. Nilai *batchsize* yang kecil maka waktu pelatihan data semakin lama, karena sistem memerlukan waktu lebih untuk membagi data latih dalam beberapa *batch*. Semakin kecil nilai *batchsize* maka semakin banyak *batch* yang terbentuk sehingga waktu yg diperlukan semakin lama

Selain hasil akurasi pada klasifikasi data *Autism Spectrum Disorders* (ASD), terdapat nilai *precision*, *recall*, dan *specificity* untuk mengukur performa sistem. Sebelum data diklasifikasikan dengan menggunakan algoritma *Naive Bayes*, data dibagi menjadi data latih dan data uji. Data latih digunakan untuk menentukan peluang $P(K)$ dan $P(At_i|K)$ dari setiap atribut pada suatu kelas, dan hasil perhitungan peluang tersebut digunakan untuk menghitung $P(K|At_{i=1}^{n})$ pada data uji, sehingga data uji dapat diklasifikasikan dalam kelas tertentu. Data latih dan data uji tersebut berpengaruh terhadap hasil klasifikasi kelas data uji yang direpresentasikan dengan *confussion matrix*.

Hasil *Confussion Matrix* tersebut digunakan untuk menghitung nilai *precision*, *recall,* dan *specificity* pada setiap kelas. Akurasi yang sama pada setiap nilai *batchsize*, mengakibatkan nilai *precision*, *recall,* dan

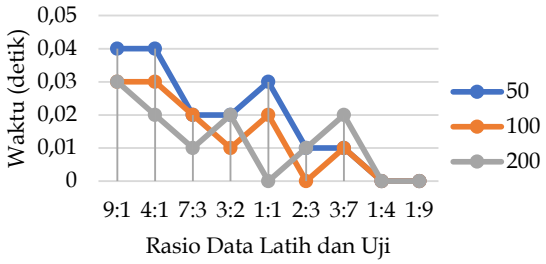
specificity yang dihasilkan juga bernilai sama. Hasil *precision*, *recall*, dan *specificity* terdapat pada Tabel 3.

Tabel 3. Tabel Hasil *Precision*, *Recall*, dan *Specificity* Algoritma *Naive Bayes*

Rasio		Precision	Recall	Specificity
9:1	No	0.923	1	0.99187
	Yes	1	0.941	
	Avg.	0.968	0.966	
4:1	No	0.966	0.966	0.98182
	Yes	0.966	0.966	
	Avg.	0.966	0.966	
7:3	No	0.979	0.959	**1**
	Yes	0.950	0.974	
	Avg.	0.966	0.966	
3:2	No	1	0.969	0.98765
	Yes	0.963	1	
	Avg.	0.984	0.983	
1:1	No	0.987	0.987	0.98592
	Yes	0.986	0.986	
	Avg.	**0.986**	**0.986**	
2:3	No	0.989	0.968	**1**
	Yes	0.964	0.988	
	Avg.	0.977	0.977	
3:7	No	1	0.963	0.97436
	Yes	0.960	1	
	Avg.	0.981	0.980	
1:4	No	0.984	0.976	0.96552
	Yes	0.973	0.982	
	Avg.	0.979	0.979	
1:9	No	0.992	0.921	0.94118
	Yes	0.917	0.992	
	Avg.	0.957	0.954	

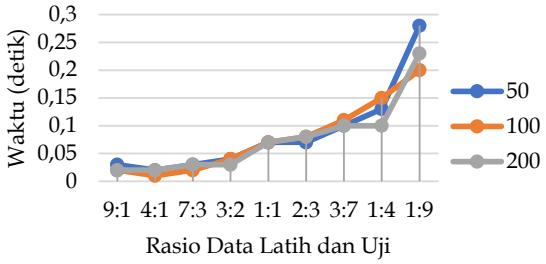
Berdasarkan Tabel 3, Algoritma *Naive Bayes* dengan menggunakan rasio data 1:1 menghasilkan *precision* tertinggi dibandingkan rasio lainnya yaitu 0.986. Hal ini merepresentasikan bahwa dengan membagi dataset menjadi dua bagian yang sama besar untuk digunakan sebagai data latih dan data uji, memberikan tingkat ketepatan yang tinggi dalam memberikan informasi yang diinginkan pengguna dengan hasil prediksi oleh sistem yang ada. Selain itu rasio tersebut juga menghasilkan *recall* terbaik dengan nilai 0.986 yang berarti dapat menghasilkan akurasi yang tinggi dalam mengklasifikasikan data secara benar. Dalam memprediksi kelas negatif (*Autism Spectrum Disorders*), rasio data 7:3 dan 2:3 adalah yang terbaik, hal ini diketahui berdasarkan nilai *specificity* yaitu 1. Dan dari Tabel 3, diketahui bahwa dengan menggunakan rasio data lebih dari 1:4 menghasilkan nilai *precision*, *recall*, dan *specificity* yang baik. Sedangkan penggunaan rasio data 1:9 memiliki performa sistem yang kurang baik, karena data yang digunakan untuk melatih model jumlahnya sangat kecil.

Klasifikasi data *Autism Spectrum Disorders* (ASD) pada anak dengan menggunakan algoritma *Naive Bayes* juga menghasilkan *training* dan *testing time*. *Training time* adalah waktu dimana sebuah sistem membangun model berdasarkan data latih pada algoritma klasifikasi yang digunakan. Sedangkan *testing time* adalah waktu yang dibutuhkan sebuah sistem untuk mengklasifikasi data uji menggunakan algoritma tertentu. *Training* dan *testing time* klasifikasi data terdapat pada Gambar 2 dan Gambar 3.



Gambar 2. *Training Time* Algoritma *Naive Bayes*

Berdasarkan Gambar 2, waktu yang dibutuhkan sistem untuk membangun model dengan berbagai variasi nilai *batchsize* dan rasio data relatif berbanding lurus dengan rasio data yang digunakan. Waktu pelatihan tertinggi terdapat pada penggunaan rasio data 9:1 dan nilai *batchsize* 50 dengan waktu 0.04 detik. Sedangkan untuk waktu pelatihan terendah terdapat pada rasio data 4:2 dan 1:9 dengan waktu 0 detik. Hal ini dikarenakan jumlah data latih yang sedikit menyebabkan waktu pelatihan juga membutuhkan waktu yang sangat rendah. Untuk variasi jumlah *batch* yang digunakan memberikan pengaruh terhadap waktu pelatihan data. Semakin besar jumlah atribut yang digunakan pada setiap *batch* menghasilkan waktu pelatihan yang rendah.



Gambar 3. *Testing Time* Algoritma *Naive Bayes*

Pada Gambar 3, waktu yang dibutuhkan sistem untuk melakukan klasifikasi data berdasarkan data yang telah dilatih berbanding terbalik dengan rasio data yang digunakan. Apabila rasio data yang digunakan semakin besar maka waktu yang dibutuhkan untuk menguji data juga akan semakin rendah. Hal ini dikarenakan apabila rasio data tinggi maka persentase data uji menjadi kecil sehingga hal ini mempengaruhi proses pengujian data. Berdasarkan Gambar 3 rasio data 1:9 menghasilkan waktu tertinggi dari setiap *batch*-nya yaitu antara 0.2 hingga 0.3 detik.

Hasil klasifikasi *Autism Spectrum Disorders* (ASD) pada anak dengan menggunakan algoritma *Naive Bayes* dengan berbagai variasi nilai *batchsize* dalam pengujian menunjukkan nilai akurasi terbaik diperoleh dari rasio data 1:1 di berbagai nilai *batchsize*. Berdasarkan waktu yang dibutuhkan untuk membangun model, rasio data 1:9 dan 1:4 menghasilkan waktu terendah dibandingkan dengan persentase data latih lainnya. Namun rasi data 1:9 menghasilkan waktu tes terbesar. Berdasarkan hasil klasifikasi, dengan melakukan *split* data menjadi data latih dan data uji menggunakan rasio dapat diketahui komposisi terbaik dalam mengklasifikasikan data berdasarkan hasil akurasi. Sedangkan membagi data latih dalam *batch* tertentu bertujuan untuk meminimumkan waktu pelatihan data, namun karena algoritma *Naive Bayes* melakukan pelatihan dengan menghitung data latih secara langsung, maka tujuan tersebut tidak terpenuhi.

Pada penelitian sebelumnya, dataset *Autism Spectrum Disorders* (ASD) pada anak telah digunakan oleh Fadi Thabtah untuk diklasifikasikan dengan berbagai metode dalam *machine learning* (Thabtah, 2017). Namun pada penelitian tersebut belum dijelaskan algoritma *machine learning* yang digunakan beserta hasil klasifikasinya dan lebih banyak membahas mengenai metode *screening* ASD yang terbaru yaitu *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition* (DSM-5). Sehingga pada penelitian ini ingin melengkapi penelitian Fadi Thabtah dengan membahas lebih spesifik salah satu metode *machine learning* yaitu *Naive Bayes*.

Penelitian mengenai klasifikasi *Autism Spectrum Disorders* (ASD) dengan berbagai metode *machine learning* salah satunya *Naive Bayes* telah dilakukan oleh Usta (Usta, M. B dkk. 2019). Namun penelitian tersebut mengklasifikasikan faktor penyebab *Autism Spectrum Disorders* (ASD) berdasarkan penilaian Autism Behavior Checklist, Aberrant Behavior Checklist, Clinical Global Impression pada usia 0, 1, 2, dan 3 tahun. Sehingga pada penelitian ini ingin menambahkan penelitian tersebut dengan menggunakan algoritma *Naive Bayes* dengan berbagai rasio data latih dan uji untuk mengklasifikasikan anak autis atau normal berdasarkan data DSM-5.

SIMPULAN

Pada penelitian ini, dilakukan klasifikasi *Autism Spectrum Disorders* (ASD) pada anak menggunakan data *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition* (DSM-5). Algoritma yang digunakan adalah *Naive Bayes* dengan menerapkan beberapa nilai *batchsize* yaitu 50, 100, dan 200. Selain itu variasi rasio data juga digunakan untuk mengetahui nilai akurasi terbaik pada proses klasifikasi. Rasio data latih dan uji yang digunakan yaitu 9:1, 4:1, 7:3, 3:2, 1:1, 2:3, 3:7, 1:4, dan 1:9

Berdasarkan hasil akurasi, *precision*, dan *recall* diketahui bahwa penerapan algoritma *Naive Bayes* dengan rasio data 1:1 merupakan rasio terbaik dalam mengklasifikasikan *Autism Spectrum Disorders* (ASD) pada anak berdasarkan *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition* (DSM-5). Nilai akurasi pada rasio data 1:1 adalah 98.6301% dengan *precision* dan *recall* sebesar 0.986. Sedangkan berdasarkan *specificity*, algoritma *Naive Bayes* dengan rasio 7:3 dan 2:3 merupakan hasil terbaik dengan nilai 1. Dalam waktu pelatihan, rasio data 1:9 merupakan yang terendah. Sedangkan dalam waktu pengujian, rasio data 9:1 merupakan yang terendah dibandingkan lainnya. Dalam penelitian ini, nilai *batch size* tidak berpengaruh dalam hasil akurasi.

DAFTAR PUSTAKA

Association, A. P. (2013). Diagnostic and Statistical Manual of Mental Disorders (DSM-5)-5th Edition. Washington, D.C.: American Psychiatric Publishing.

Baio, J., Wiggins, L., & and L. Christensen, D. (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. MMWR Surveill Summ, 67(6), 1-23.

Beata, T., Bolton, P., Happe, F., Rutter, M., & Rijsdijk, F. (2016). Heritability of Autism Spectrum

Disorder: A Meta Analysis of Twin Studies. Journal of Child Psychology and Psychiatry, 57(5), 585-595.

Dwi, P., Saptono, R., & Anggrainigsih. (2018). Academic Articles Classification Using Naive Bayes Classifier (NBC) Method. Jurnal Ilmiah Teknologi dan Informasi, 7(2), 74-81.

Ibrahim, S., Djemal, R., & Alsuwailem, A. (2018). Electroenchepalography (EEG) Signal Processing for Epilepsy and Autrism Spectrum Disorder. Biocybernetics and Biomedical Engineering, 38, 16-26.

Jamal, S., & Scaria, V. (2013). Cheminformatic Models Based On Machine Learning for Pyruvate Kinase Inhibitors of Leishmania Mexicana. BMC Bioinformatics, Biomed Central, 14(1), 329-335.

Kazdoba, T. M., Leach, P. T., & Crawley, J. N. (2016). Behavioral Phenotypes of Genetic Mouse Model of Autism. Genes Brain Behavior, 15, 7-26.

Kotsiantis, S. (2013). Increasing the accuracy of incremental naive bayes classifier using instance based learning. International Journal of Control, Automation and Systems, 11(1), 159–166.

Kukreja, M., Johnston, S. A., & Stanfford, P. (2012). Comparative Study of Classification Algorithms for Immunosignaturing Data. BMC Bionformatics, 13, 139-145.

Lord, C., Brugha, T. S., & Charman, T. (2020). Autism Spectrum Disorder. Nature Reviews Disease Primers, 6(5).

Lyall, K., Schmidt, R. J., & Hertz-Picciotto, I. (2014). Maternal Lifestyle and Environmental Risk Factors for Autism Spectrum Disorders. International Journal of Epidemiology, 43(2), 443-464.

Pardo, C. A., Vargas, D. L., & Zimmerman, A. W. (2009). Immunity, Neuroglia, and Neuroin Flammation in Autism. International Review of Psychiatry, 17(6), 485-495.

Parikesit, A. A., Nurdiansyah, R., & Agustriawan, D. (2019). Penerapan Pendekatan Machine Learning Pada Pengembangan Basis Data Herbal Sebagai Sumber Informasi Kandidat Obat Kanker. Jurnal Teknologi Industri Pangan, 29(2), 175-182.

Parmeggiani, A., Corinaldesi, A., & Posar, A. (2019). Early Feature of Autism Spectrum Disorder: A Cross-Sectional Study. Italian Journal Pediatrics, 45(144).

Supriyanto, C., & Parida, P. (2013). Deteksi Penyakit Diabetes Type II dengan Naives Bayes Berbasis Particle Swarm Optimization. Jurnal Teknologi Informasi, 9(2).

Syarifah, A., & Muslim, M. A. (2015). Pemanfaatan Naive Bayes Untuk Merespon Emosi dari Kalimat Berbahasa Indonesia. UNNES Journal of Mathematics, 4(2), 147-156.

Thabtah, F. (2017). Autism spectrum disorder screening: Machine learning adaptation and DSM-5 fulfillment. ACM International Conference Proceeding Series, Part F129311, 1–6.

UCI Machine Learning. (2017). Retrieved from https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++

Usta, M. B., Karabekiroglu, K., Sahin, B., Aydin, M., Bozkurt, A., Karaosman, T., Aral, A., Cobanoglu, C., Kurt, A. D., Kesim, N., Sahin, İ., & Ürer, E. (2019). Use of machine learning methods in prediction of short-term outcome in autism spectrum disorders. Psychiatry and Clinical Psychopharmacology, 29(3), 320–325.

Vahia V. N. (2013). Diagnostic and Statistical Manual of Mental Disorders 5: A quick glance. Indian Journal of Psychiatry, 55(3), 220–223.

World Health Organization. (2019, November 7). (World Health Organization) Retrieved November 10, 2020, from Autism spectrum disorders: https://www.who.int