

Analisis Perbandingan Akurasi dalam Identifikasi Autism dengan SVM dan Naive Bayes

Ferawaty¹, Muhammad Zarlis², Erna Budhiarti Nababan³

^{1,2,3}Universitas Sumatera Utara, Jl. Dr. T. Manshur No. 9. Kampus USU, Medan 20155

¹STMIK Mikroskil, Jl. Thamrin No. 112, 124, 140, Telp. (061) 4573767, Fax. (061) 4567789

¹gold3n27@gmail.com, ²m.zarlis@yahoo.com, ³ernabr@usu.ac.id

Abstrak

Gangguan autisme banyak ditemukan pada anak yang berumur 3 tahun ke bawah. Pendiagnosaan gangguan penyakit ini telah dilakukan dengan menggunakan berbagai metode, terutama metode dalam dunia psikologis. Peneliti mengambil data dari data Survey of Pathways to Diagnosis and Services yang dilakukan pada tahun 2011[1]. Peneliti menggunakan metode Support Vector Machine (SVM) dan metode Naive Bayes untuk menyelesaikan kasus gangguan autisme yang mengalami kesalahan diagnosa. Dalam hasil penelitian ini dilakukan perbandingan metode Support Vector Machine (SVM) dengan metode Naive Bayes. Metode Support Vector Machine (SVM) menghasilkan rata – rata klasifikasi 93,12%, sedangkan metode Naive Bayes menghasilkan rata – rata klasifikasi 73,34%.

Kata kunci— Autism, Support Vector Machine, Naive Bayes, keakuratan

Abstract

Autism disorders found in children aged 3 years and under. Diagnosis disorders has been carried out using various methods, especially methods in the psychological world. The researchers took data from the Survey data of Pathways to Diagnosis and Services done in 2011. The researchers used a method of Support Vector Machine (SVM) and Naive Bayes method to resolve cases of autism disorder that is experiencing the error diagnosis. In the results of the research carried out a comparison of the methods of Support Vector Machine (SVM) with Naive Bayes method. The method of Support Vector Machine (SVM) result in a mean median classification – 93,12%, whereas the and Naive Bayes method produce align the classification of 73,34%

Keywords— Autism, Support Vector Machine, Naive Bayes, Accuracy

1. PENDAHULUAN

Autism merupakan penyakit yang berhubungan dengan otak. Banyak anak yang terdeteksi menderita penyakit autisme. Autism sering didapati pada anak yang berumur 3 tahun ke bawah. Penyakit ini sering terjadi karena faktor kehidupan dari si penderita. Anak yang mengalami gejala autis biasanya dikarenakan orang tua yang bekerja, kurang interaksi dengan orang lain, dan lain sebagainya. Untuk mengetahui gejala autisme terhadap si penderita, maka akan dilakukan penelitian terhadap si penderita dengan cara mencari psikiater atau dokter penyakit jiwa.

Salah satu metode penelitian yang dilakukan oleh psikiater adalah dengan menggunakan metode *Childhood Autism Rating Scale* (CARS). Metode CARS ini menggunakan pertanyaan – pertanyaan yang harus dijawab oleh orang tua dalam membantu mengidentifikasi autisme terhadap si penderita. Selain menggunakan metode – metode psikiater, para peneliti juga melakukan perkembangan ke dalam bidang teknologi, yaitu dengan membuat sistem pakar. Pembuatan sistem pakar yang dilakukan oleh Rahmawati Rizki Fitri pada tahun 2012 di Jawa Timur dengan menggunakan metode *Forward Chaining*. Penelitian yang dilakukan Rahmawati adalah dengan menggunakan cara – cara manual yaitu dengan cara pengamatan langsung, bertanya kepada orang tua penderita dan berbagai metode lainnya.

Selain sistem pakar yang dilakukan oleh Rahmawati, ada juga sistem pakar yang dibuat oleh Apila dan teman – temannya pada tahun 2014 dengan menggunakan metode *and Statistical Manual of Mental* (DSM IV) yang dipadukan menggunakan ICD 10 dan *Checklist Autism in Toddlers* (CHAT). Ada juga penelitian yang dilakukan oleh Sista Pradjna Paramitha pada tahun 2016. Sista menggunakan metode *Naive Bayes* untuk mendiagnosa jenis gangguan autis pada balita.

Penelitian menggunakan metode *Support Vector Machine* dapat dilihat pada penelitian yang dilakukan oleh Rachman pada tahun 2012. Rachman menggunakan metode SVM untuk melakukan penelitian terhadap tingkat keganasan kanker payudara. Selain Rachman, ada juga peneliti lain yaitu Sari. Sari melakukan penelitian pada tahun 2014 dengan menggunakan metode SVM untuk mengklasifikasi perbedaan karakteristik di dua daerah yang berlainan di Indonesia.

Dengan adanya penelitian yang dilakukan oleh beberapa peneliti terdahulu, penulis melakukan perbandingan dalam identifikasi autisme dengan menggunakan metode *Support Vector Machine*(SVM) dengan metode *Naive Bayes*.

2. TINJAUAN PUSTAKA

2.1. Autisme

Kata autis pertama kali digunakan oleh Leo Kanner yang merupakan seorang *psychiatrist* anak di Universitas Johns Hopkins di Baltimore[2]. Kata Autis berasal dari kata *autos* dan *isme*, yang berarti diri sendiri dan suatu aliran, sehingga anak autis dapat diartikan sebagai suatu paham yang tertarik pada dunianya sendiri. Leo Kanner menemukan autisme pada tahun 1943. Gangguan ini dideskripsikan oleh Leo Kanner sebagai ketidakmampuan berinteraksi dengan orang lain, penguasaan bahasa menjadi tertunda, echolalia, mutism, pembalikan kalimat, aktivitas bermain repetitive, stereotype, ingatan yang kuat, keinginan obsesif untuk mempertahankan keteraturan di dalam lingkungannya.

Menurut Kanner dan Asperger (1943), orang yang mengalami gangguan perkembangan mempunyai banyak istilah, yaitu:

1. *Autism* (autisme)
Nama untuk gangguan perkembangan komunikasi, sosial dan perilaku pada anak.
2. *Autis* (autis)
Anak yang mengalami gangguan autisme.
3. *Autistic Child* (anak autistik)
Keadaan anak yang mengalami gangguan autis.

Menurut *Pervasive Development Disorder* (PDD) bahwa gangguan perkembangan terbagi atas beberapa kelompok, yaitu :

1. *Autistic Disorder* (Autism)
Keadaan ini terjadi sebelum usia 3 tahun. Anak tersebut akan mengalami hambatan dalam sosial, komunikasi dan mempunyai kemampuan untuk berimajinasi serta adanya perilaku stereotype pada minat dan aktivitas.
2. *Asperger's Syndrome*
Hambatan perkembangan dalam interaksi dengan orang lain. aktivitas dan minatnya terbatas, tetapi tidak menunjukkan keterlambatan dalam berbicara, dan tingkat kecerdasan rata – rata hingga diatas rata – rata.
3. *Pervasive Developmental Disorder – Not Otherwise Specified* (PDD-NOS)
PDD-NOS merujuk pada istilah *atypical autism*. Diagnosa ini berlaku bila anak tidak menunjukkan keseluruhan kriteria pada diagnosa tertentu (*Autisme, Asperger*, atau *Rett Syndrome*).
4. *Rett's Syndrome*
Syndrome ini lebih banyak terjadi pada anak perempuan. Mengalami perkembangan yang normal sebelumnya, kemudian terjadi kemunduran / kehilangan kemampuan yang dimilikinya. Kemampuan pergerakan tangan digantikan dengan gerakan – gerakan yang berulang – ulang. *Rett's Syndrome* kebanyakan terjadi pada umur 1 – 4 tahun.

5. Childhood Disintegrative Disorder (CDD)

Selama 2 tahun pertama menunjukkan perkembangan yang normal, kemudian semua kemampuan yang telah dipelajari tiba – tiba menghilang.

Dari apa yang telah disebutkan diatas, maka disimpulkan bahwa anak – anak autis mengalami kesulitan dalam perkembangan otak yang kompleks, sehingga mempengaruhi banyak fungsi – fungsi, seperti : imajinasi, perasaan dan persepsi yang terjadi sebelum usia menanjak 3 tahun.

2.2. Support Vector Machine

Pada tahun 1992, Boser, Guyon, Vapnik mempresentasikan *Support Vector Machine* (SVM) di *Annual Workshop on Computational Learning Theory*. Usaha pencarian *hyperplane* terbaik dapat dijelaskan secara sederhana. Pencarian *hyperplane* terbaik berfungsi sebagai pemisah dua buah kelas pada *input space*. +1 dan -1 adalah anggota pola dari dua buah kelas, dan juga merupakan *discrimination boundaries* atau disebut juga sebagai alternatif garis pemisah. Jarak *hyperplane* dengan pola terdekat dari masing – masing kelas disebut sebagai margin. Usaha untuk mencari lokasi *hyperplane* merupakan inti dari proses pembelajaran SVM [3].

Data yang tersedia dinotasikan sebagai sedangan label masing-masing dinotasikan untuk $i = 1, 2, \dots, l$, yang mana l adalah banyaknya data. Diasumsikan kedua kelas -1 dan +1 dapat terpisah secara sempurna oleh *hyperplane* berdimensi d , yang didefinisikan :

$$x_i \cdot w + b = 0 \quad (1)$$

Pattern x_i yang termasuk kelas -1 (sampel negatif) dapat dirumuskan sebagai pattern yang memenuhi pertidaksamaan

$$w \cdot x + b \leq -1 \quad (2)$$

Sedangkan x_i pattern yang termasuk kelas +1 (sampel positif) memenuhi pertidaksamaan :

$$w \cdot x + b \geq +1 \quad (3)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu $\frac{1}{\|w\|}$. Hal ini dapat dirumuskan sebagai Quadratic Programming (QP) Problem, yaitu mencari titik minimal persamaan (4) dengan memperhatikan constraint persamaan (8).

$$\min_w \tau(w) = \frac{1}{2} \|w\|^2 \quad (4)$$

$$y_i (x_i \cdot w + b) - 1 \geq 0, \quad \forall i \quad (5)$$

Problem ini dapat dipecahkan dengan berbagi teknik komputasi, diantaranya dengan *Lagrange Multiplier*.

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (y_i (x_i \cdot w + b) - 1) \quad \text{dengan } i = 1, 2, \dots, l \quad (6)$$

α_i adalah *Lagrange multipliers*, yang bernilai nol atau positif ($\alpha_i \geq 0$). Nilai optimal dari persamaan (6) dapat dihitung dengan meminimalkan L terhadap w dan b , dan memaksimalkan L terhadap α_i . Dengan memperhatikan sifat bahwa pada titik optimal gradient $L = 0$, persamaan (6) dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung α_i , sebagaimana persamaan dibawah ini:
Maksimasi :

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (7)$$

Dengan *Constraint*:

$$\alpha_i \geq 0 (i=1, 2, \dots, l) \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (8)$$

Dari hasil perhitungan ini diperoleh α_i yang kebanyakan bernilai positif. Data yang berkorelasi dengan α_i yang positif inilah yang disebut sebagai *support vector* [4].

2.3. Naive Bayes

Naive Bayes merupakan algoritma klasifikasi yang sederhana dimana setiap atribut bersifat *independent* dan memungkinkan berkontribusi terhadap keputusan akhir[5].

Rumus Bayes yang digunakan dalam pemrograman adalah sebagai berikut :

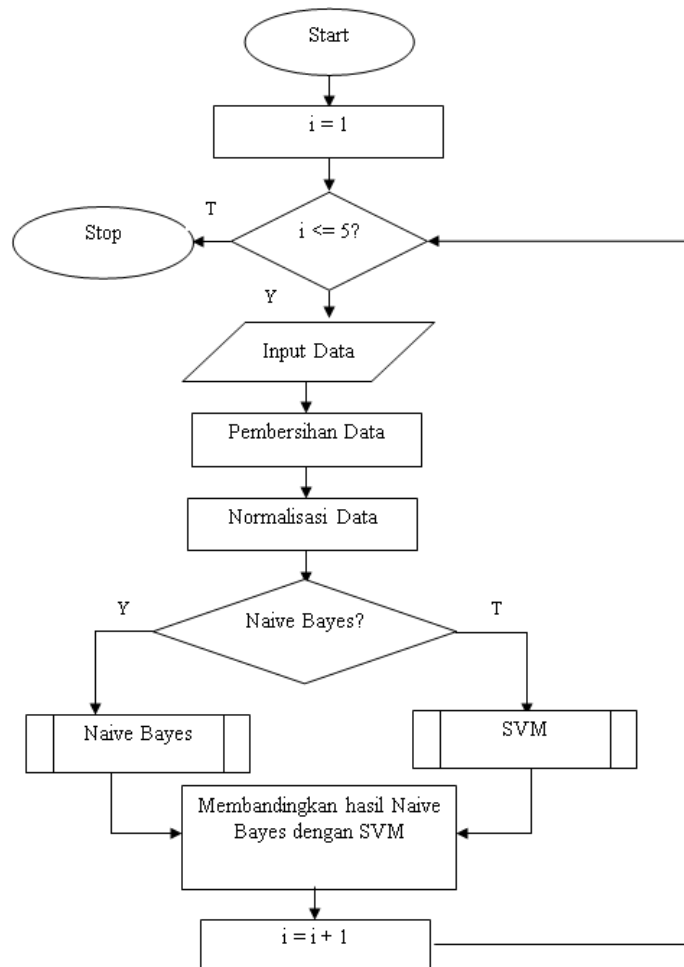
$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (9)$$

Dimana $P(H|X)$ merupakan probabilitas H di dalam X atau persentase banyaknya H di dalam X. $P(X|H)$ merupakan probabilitas X di dalam H, $P(H)$ merupakan probabilitas prior dari H dan $P(X)$ merupakan probabilitas prior dari X.

3. METODOLOGI PENELITIAN

Data autism yang digunakan berasal dari data *Survey of Pathways to Diagnosis and Services* yang dilakukan pada tahun 2011 oleh *The Data Resources Center for Child and Adolescent Health, Child and Adolescent Health Measurement Initiative*[1].

Pada gambar 3.1 dibawah ini merupakan flowchart jalannya penelitian yang dilakukan.



Gambar 1. Flowchart Penelitian

Pada percobaan pertama pada metode SVM dan metode *Naive Bayes* menggunakan data sebanyak 66 data. Pada percobaan kedua, menggunakan data sebanyak 166 data. Pada percobaan ketiga menggunakan data sebanyak 383 data. Pada percobaan keempat, menggunakan data sebanyak 682. Pada percobaan kelima menggunakan data sebanyak 1078. Pada setiap percobaan dilakukan pelatihan dengan pengambilan data secara random sebanyak 70%. Kemudian 30% dari data digunakan sebagai data *testing*.

Data yang sudah diinput akan dibersihkan dengan menghilangkan baris yang mengandung data yang hilang. Kemudian dilakukan normalisasi data. Normalisasi yang dilakukan dengan menggunakan cara normalisasi min – max. Setelah melakukan normalisasi, maka akan dilakukan pencarian nilai *output*(y). Nilai *output* (y) dicari dengan cara menggunakan metode IF. Jika hasil > 0 maka y = 1, jika tidak, maka y = -1.

Kemudian dilakukan pencarian nilai *weight*(w) dan bias(b). Untuk menghitung nilai *weight*(w) dan bias(b), maka digunakan rumus $y_i(w \cdot x_i + b) \geq 1$ dimana $i = 1, 2, 3, \dots, N$. Dengan menggunakan rumus yang digunakan, maka akan menghasilkan beberapa persamaan. Persamaan – persamaan yang diperoleh akan dijumlahkan untuk memperoleh nilai angka untuk *weight* dan bias. Setelah diperoleh hasil *weight* dan bias, maka ditentukan nilai *hyperplane*. Nilai *hyperplane* ditentukan dengan menggunakan rumus

$$f(x) = w_1x_1 + w_2x_2 + w_3x_3 + b = 0 \quad (10)$$

Dari nilai *hyperplane* yang diperoleh, akan dilakukan pengujian dengan menggunakan rumus $\text{sign}(f(x))$ dengan nilai $\alpha = 1$. Nilai $\alpha \geq 0$. Fungsi SIGN yang digunakan merupakan pengecekan pada hasil perhitungan yang dilakukan pada data uji. Jika hasil > 0 maka nilai = 1, sedangkan nilai hasil < 0 akan memperoleh nilai = -1

Sedangkan untuk mencari hasil nilai Naive Bayes, maka dilakukan pencarian hasil dengan menggunakan rumus:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (11)$$

Setelah memperoleh hasil untuk metode SVM dan metode *Naive Bayes*, maka dilakukan perbandingan terhadap akurasi, *classification error*, *weighted mean recall*.

4. HASIL DAN PEMBAHASAN

4.1. Hasil

Pada Tabel 1 dibawah ini akan menampilkan hasil klasifikasi menggunakan metode SVM.

Tabel 1. Hasil Klasifikasi Metode SVM

	SVM (autis)		
	Accuracy	classification_error	weighted_mean_recall
percobaan 1	80,00%	20,00%	33,33%
percobaan 2	96,00%	4,00%	33,33%
percobaan 3	93,91%	6,09%	33,33%
percobaan 4	97,56%	2,44%	33,33%
percobaan 5	98,14%	1,86%	33,33%
	93,12%	6,88%	33,33%

Pada tabel 1 menampilkan hasil klasifikasi metode SVM, dimana diperoleh setiap jumlah data yang diproses menampilkan yang berbeda – beda. Semakin besar jumlah data yang diproses, maka semakin kecil rata – rata yang diperoleh. Seperti yang terlihat pada tabel 1 menampilkan rata – rata

keakuratan sebesar 93,12%, sedangkan klasifikasi error sebesar 6,88%, dan *weighted Mean Recall* sebesar 33,33%.

Tabel 2. Hasil Klasifikasi Metode Naive Bayes

	Naive Bayes (autis)		
	Accuracy	classification_error	weighted_mean_recall
percobaan 1	80,00%	20,00%	64,58%
percobaan 2	86,00%	14,00%	62,50%
percobaan 3	59,13%	40,87%	64,51%
percobaan 4	95,12%	4,88%	39,00%
percobaan 5	46,44%	53,56%	24,00%
	73,34%	26,66%	50,92%

Pada tabel 2 menampilkan hasil klasifikasi metode Naive Bayes, dimana diperoleh setiap jumlah data yang diproses menampilkan yang berbeda – beda. Semakin besar jumlah data yang diproses, maka semakin kecil rata – rata yang diperoleh. Seperti yang terlihat pada tabel 2 menampilkan rata – rata keakuratan menggunakan Naive Bayes adalah 73,34%, sedangkan klasifikasi error sebesar 26,66%, dan *weighted Mean Recall* sebesar 50,92%.

4.2. Pembahasan

Dari banyaknya percobaan yang dilakukan pada kedua metode, 70% digunakan sebagai data *training*, sedangkan 30% dari data digunakan sebagai data *testing*. Pengambilan data *training* yang sebesar 70% dilakukan secara random.

Pada percobaan pertama, dengan menggunakan metode SVM, menghasilkan keakuratan sebesar 80%, *classification error* sebesar 20%, dan *weighted mean recall* sebesar 33,33%. Sedangkan pada metode *Naive Bayes* menghasilkan keakuratan sebesar 80%, *classification error* sebesar 20%, dan *weighted mean recall* sebesar 64,58%.

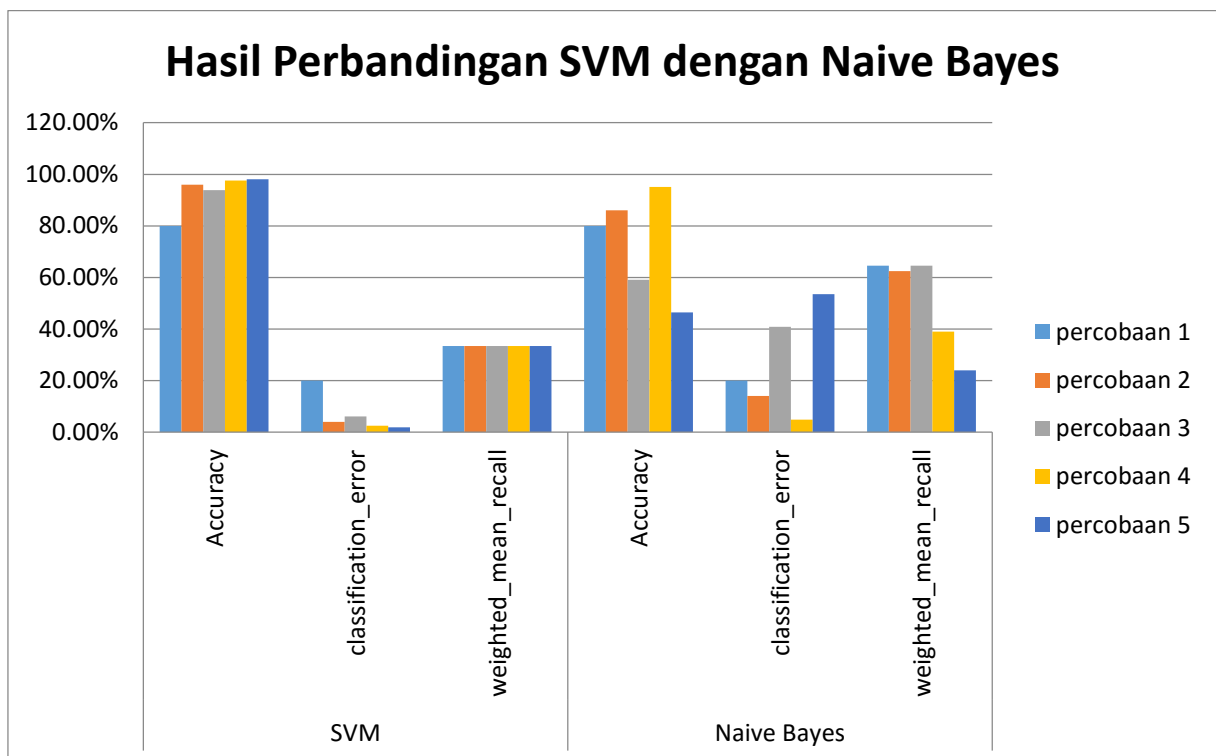
Pada percobaan kedua, dengan menggunakan metode SVM, menghasilkan keakuratan sebesar 96%, *classification error* sebesar 4%, dan *weighted mean recall* sebesar 33,33%. Sedangkan pada metode *Naive Bayes* menghasilkan keakuratan sebesar 86%, *classification error* sebesar 14%, dan *weighted mean recall* sebesar 62,50%.

Pada percobaan ketiga, dengan menggunakan metode SVM, menghasilkan keakuratan sebesar 93,91%, *classification error* sebesar 6,09%, dan *weighted mean recall* sebesar 33,33%. Sedangkan pada metode *Naive Bayes* menghasilkan keakuratan sebesar 59,13%, *classification error* sebesar 40,87%, dan *weighted mean recall* sebesar 64,51%.

Pada percobaan keempat, dengan menggunakan metode SVM, menghasilkan keakuratan sebesar 97,56%, *classification error* sebesar 2,44%, dan *weighted mean recall* sebesar 33,33%. Sedangkan pada metode *Naive Bayes* menghasilkan keakuratan sebesar 95,12%, *classification error* sebesar 4,88%, dan *weighted mean recall* sebesar 39%.

Pada percobaan kelima, dengan menggunakan metode SVM, menghasilkan keakuratan sebesar 98,14%, *classification error* sebesar 1,86%, dan *weighted mean recall* sebesar 33,33%. Sedangkan pada metode *Naive Bayes* menghasilkan keakuratan sebesar 46,44%, *classification error* sebesar 53,56%, dan *weighted mean recall* sebesar 24%.

Untuk memudahkan membaca penelitian ini, dapat dilihat gambar grafik perbandingan hasil klasifikasi SVM dengan *Naive Bayes*.



Gambar 2. Grafik Hasil Perbandingan Metode SVM dengan Metode Naive Bayes

Pada gambar Gambar 2 menampilkan hasil perbandingan metode SVM dengan metode *Naive Bayes*, dimana batang berwarna biru merupakan hasil percobaan 1. Batang berwarna merah merupakan hasil percobaan kedua. Batang berwarna hijau merupakan percobaan ketiga. Batang berwarna ungu merupakan hasil percobaan keempat. Batang berwarna biru muda merupakan hasil percobaan kelima.

Dari Gambar 2, dapat dilihat bahwa keakuratan pada metode svm lebih akurat, *classification error* yang dihasilkan semakin kecil, dan *weight mean recall* yang dihasilkan selalu sama. Sedangkan pada metode *Naive Bayes* keakuratan yang dihasilkan tidak teratur, begitu juga dengan *classification error* yang dihasilkan, sedangkan pada *weight mean recall* yang dihasilkan semakin menurun.

4. KESIMPULAN

Berdasarkan pengujian dengan menggunakan banyak data yang berbeda – beda saat percobaan dapat diambil kesimpulan:

1. Metode SVM menghasilkan tingkat akurasi rata - rata klasifikasi 93,12%. Hasil 93,12% ini merupakan total dari hasil rata – rata dari beberapa percobaan yang telah dilakukan dengan menggunakan metode SVM.
2. Metode *Naive Bayes* menghasilkan rata – rata klasifikasi 73,34%. Hasil 73,34% ini merupakan total dari hasil rata – rata dari beberapa percobaan yang telah dilakukan dengan menggunakan metode *Naive Bayes*.

5. SARAN

Pada penelitian ini terdapat beberapa saran untuk pengembangan penelitian ini, antara lain:

1. Penambahan kernel ke dalam metode SVM yang digunakan.
2. Membandingkan metode SVM dengan metode lain.

DAFTAR PUSTAKA

- [1] 2011 Survey of Pathways to Diagnosis and Services. Maternal and Child Health Bureau in collaboration with the National Center for Health Statistics. 2011 Pathways [Insert SPSS/SAS/Stata] Indicator Data Set prepared by the Data Resource Center for Child and Adolescent Health, Child and Adolescent Health Measurement Initiative. www.childhealthdata.org
- [2] Ozonoff, Sally.,dkk, 2002, *A Parent Guide to Asperger Syndrome & High-Functioning Autism*. New York : Guilford Publications.
- [3] Christianini, N and Sahwe Taylor, J., 2000, *AN INTRODUCTION TO SUPPORT VECTOR MACHINE AND OTHER KERNEL BASED LEARNING MEATHODS*, Cambridge University Press.
- [4] Abram, Kristin. 2014. *Exploring The Impact of Music Theraphy On Children With Complex Communication Needs And Autism Spectrum Disoders: A Focus Group Study*. The Honor Tutorial Collage : Ohio University.
- [5] Aprilia D, Johar A, Hartuti P. 2014. *Sistem Pakar Diagnosa Autisme pada Anak*. Jurnal Rekursif Vol 2.