

Penerapan Teknik Stacking untuk Optimasi Deteksi Dini Anak Autis berbasis Support Vector Machine

Yulianti¹, Teti Desyani², Sri Mulyati³

Teknik Informatika, Universitas Pamulang, Jl. Raya Puspitek No. 46 Buaran, Serpong, Tangerang Selatan, Banten, Indonesia, 15417

e-mail: ¹yulianti@unpam.ac.id, ²dosen00839@unpam.ac.id, ³dosen00391@unpam.ac.id

Submitted Date: October 02nd, 2024

Reviewed Date: October 28th, 2024

Revised Date: October 31st, 2024

Accepted Date: October 31st, 2024

Abstract

Autism is not difficult to detect, but it takes a lot of learning and training for doctors to detect it. Currently ASD is detected by understanding the behavior and intellectual activity of a child. This diagnosis can be subjective, time consuming, inconclusive, does not provide precise insight into genetics and is not suitable for early detection. Machine Learning Methods can make relevant changes to speed up the process. It is known that early intervention is the key to improving children with autism. Obviously speeding up the diagnosis time is even more important in the case of Autism. Big data and machine learning technologies can make major advances in predicting and accelerating the complex and time-consuming process of diagnosis and treatment. Machine learning systems can be developed to take advantage of the vast amount of health and medical data available for predictive modeling and predictive analysis. In this paper, a comparison of several machine learning techniques and models will be tested and analyzed. In this study, it is proposed to apply the stacking technique to predict the presence of ASD. The results showed that the application of the stacking technique could improve the performance of the predictive model in ASD diagnosis.

Keywords: Children; Autism; predictions; Stacking

Abstrak

Autisme tidak sulit dideteksi, tetapi membutuhkan banyak pembelajaran dan pelatihan bagi dokter untuk mendeteksinya. Saat ini ASD dideteksi dengan memahami perilaku dan aktivitas intelektual seorang anak. Diagnosis ini bisa subjektif, memakan waktu, tidak meyakinkan, tidak memberikan wawasan yang tepat tentang genetika dan tidak cocok untuk deteksi dini. Metode Pembelajaran Mesin (Machine Learning) dapat membuat perubahan yang relevan untuk mempercepat proses. Diketahui bahwa intervensi dini merupakan kunci untuk memperbaiki anak autis. Jelas mempercepat waktu diagnosis bahkan lebih penting dalam kasus Autisme. Teknologi big data dan pembelajaran mesin dapat membuat kemajuan besar untuk memprediksi dan mempercepat proses diagnosis dan pengobatan yang kompleks dan memakan waktu. Sistem pembelajaran mesin dapat dikembangkan untuk memanfaatkan sejumlah besar data kesehatan dan medis yang tersedia untuk pemodelan prediktif dan analisis prediktif. Dalam makalah ini, perbandingan dari beberapa teknik dan model pembelajaran mesin akan diuji dan dianalisis. Pada penelitian ini diusulkan penerapan teknik stacking untuk prediksi adanya ASD. Hasil penelitian menunjukkan bahwa penerapan teknik stacking dapat meningkatkan kinerja model prediksi pada diagnosa ASD.

Kata kunci: Anak; Autis; Prediksi; Stacking

1 Pendahuluan

Masalah Autism Spectrum Disorder (ASD) saat ini telah meningkat pesat di antara semua usia

populasi manusia. Deteksi dini penyakit saraf ini dapat sangat membantu dalam pemeliharaan kesehatan mental dan fisik subjek. Dengan



maraknya penerapan model berbasis pembelajaran mesin dalam prediksi berbagai penyakit manusia, deteksi dini mereka berdasarkan berbagai parameter kesehatan dan fisiologis sekarang tampaknya mungkin dilakukan. Faktor ini memotivasi kami untuk meningkatkan minat dalam deteksi dan analisis penyakit ASD untuk meningkatkan metodologi pengobatan yang lebih baik. Deteksi ASD menjadi tantangan karena ada beberapa gangguan mental lain yang sedikit gejalanya sangat mirip dengan gejala ASD, sehingga membuat tugas ini menjadi sulit (Thabtah, 2019).

Gangguan Spektrum Autisme merupakan masalah yang berkaitan dengan perkembangan otak manusia. Seseorang yang mengalami Gangguan Spektrum Autisme umumnya tidak dapat melakukan interaksi sosial dan komunikasi dengan orang lain (Thabtah et al., 2018)(R & R, 2018). Dalam hal ini, kehidupan seseorang biasanya terpengaruh sepanjang hidupnya. Menarik untuk diketahui bahwa faktor lingkungan dan genetik dapat menjadi faktor penyebab penyakit ini. Gejala masalah ini mungkin dimulai pada usia tiga tahun dan dapat berlanjut seumur hidup. Tidak mungkin untuk menyelesaikan pengobatan pasien yang menderita penyakit ini (Constantino et al., 2007), namun efeknya dapat berkurang untuk beberapa waktu jika gejalanya terdeteksi secara dini. Dengan asumsi bahwa gen manusia bertanggung jawab untuk itu, penyebab pasti ASD belum dikenali oleh ilmuwan (Bone et al., 2014). Gen manusia mempengaruhi perkembangan dengan mempengaruhi lingkungan. Ada beberapa faktor risiko yang mempengaruhi GSA seperti berat badan lahir rendah, saudara kandung dengan ASD dan memiliki orang tua, dan lain-lain.

Selain itu, terdapat beberapa masalah interaksi dan komunikasi sosial yang sering terjadi pada individu dengan gangguan perkembangan tertentu. Beberapa contoh dari masalah ini meliputi perilaku seperti tertawa atau cekikikan yang tidak pantas, kurangnya kepekaan terhadap rasa sakit, serta kesulitan dalam melakukan kontak mata yang baik. Individu tersebut mungkin juga menunjukkan respons yang tidak sesuai terhadap suara, enggan untuk menerima pelukan, dan kesulitan dalam mengekspresikan gerak tubuh. Selain itu, mereka mungkin tidak menunjukkan keinginan untuk berinteraksi dengan orang lain, memiliki

keterikatan yang tidak biasa terhadap objek tertentu, dan menunjukkan preferensi untuk hidup sendiri. Penggunaan kata-kata gema atau pengulangan kata juga sering terlihat sebagai bagian dari pola komunikasi yang terganggu. Semua masalah ini menggambarkan tantangan yang signifikan dalam interaksi sosial dan komunikasi yang memengaruhi kehidupan sehari-hari.

Orang dengan ASD (Autism Spectrum Disorder) sering mengalami kesulitan yang berkaitan dengan minat yang terbatas dan pengulangan perilaku yang konsisten. Beberapa contoh spesifik perilaku ini meliputi mengulangi kata atau frasa secara terus-menerus tanpa henti. Mereka juga dapat menunjukkan reaksi emosional yang intens, seperti marah atau gelisah, ketika rutinitas harian berubah, karena perubahan tersebut mengganggu kenyamanan dan pola mereka. Selain itu, mereka mungkin memiliki minat yang sangat sempit pada topik tertentu, misalnya angka atau fakta tertentu, yang ditekuni dengan intensitas tinggi. Dalam beberapa situasi, mereka juga menunjukkan kepekaan sensorik yang berbeda dari orang lain, seperti kurang sensitif terhadap cahaya atau kebisingan, atau sebaliknya, mungkin sangat sensitif terhadap rangsangan sensorik ini. Perilaku-perilaku ini menyoroti karakteristik khas yang mempengaruhi interaksi sehari-hari dan kualitas hidup mereka.

Deteksi dan pengobatan dini merupakan langkah terpenting yang harus dilakukan untuk mengurangi gejala gangguan spektrum autisme dan meningkatkan kualitas hidup penderita GSA. Namun, tidak ada prosedur tes medis untuk mendeteksi autisme (Wall, Kosmicki, et al., 2012). Gejala ASD biasanya dikenali dengan observasi. Pada lansia dan remaja yang bersekolah, gejala ASD biasanya dikenali oleh orang tua dan gurunya. Setelah itu gejala ASD ada dan dievaluasi oleh tim pendidikan khusus sekolah. Tim sekolah ini menyarankan anak-anak ini mengunjungi dokter perawatan kesehatan mereka untuk tes yang diperlukan. Pada orang dewasa, mengidentifikasi gejala ASD sangat sulit daripada anak-anak dan remaja yang lebih tua karena beberapa gejala ASD mungkin tumpang tindih dengan gangguan kesehatan mental lainnya (Wall, Dally, et al., 2012). Perubahan perilaku pada anak dapat dengan mudah diidentifikasi melalui observasi karena dapat dilihat pada usia 6 bulan lebih awal

dibandingkan dengan pencitraan otak spesifik Autisme karena pencitraan otak dapat diidentifikasi setelah usia 2 tahun.

Vaishali R. dkk. (R & R, 2018) telah mengusulkan metode untuk mengidentifikasi Autisme dengan kumpulan perilaku optimal. Dalam pekerjaan ini, dataset diagnosis ASD dengan 21 fitur yang diperoleh dari repositori pembelajaran mesin UCI bereksperimen dengan pembungkus pemilihan fitur biner firefly berbasis kecerdasan swarm. Hipotesis alternatif dari eksperimen mengklaim bahwa model pembelajaran mesin dimungkinkan untuk mencapai akurasi klasifikasi yang lebih baik dengan subset fitur minimum. Dengan menggunakan pembungkus seleksi fitur kunang-kunang biner tujuan tunggal berbasis kecerdasan Swarm, ditemukan bahwa 10 fitur di antara 21 fitur dari dataset ASD cukup untuk membedakan antara pasien ASD dan non-ASD. Hasil yang diperoleh dengan pendekatan ini membenarkan hipotesis dengan menghasilkan akurasi rata-rata pada kisaran 92,12% -97,95% dengan subset fitur optimal yang kira-kira sama dengan akurasi rata-rata yang dihasilkan oleh seluruh dataset diagnosis ASD.

Thabtah dkk. (Thabtah, 2017) telah mengusulkan model penyaringan ASD menggunakan Machine Learning Adaption dan DSM-5. Alat skrining telah digunakan untuk mewujudkan satu atau lebih tujuan dalam skrining ASD. Dalam makalah ini, peneliti membahas klasifikasi ASD Machine Learning dengan pro dan kontranya. Peneliti mencoba untuk menyoroti masalah yang menyertai alat skrining ASD yang ada dan konsistensi dari alat tersebut menggunakan DSM-IV daripada manual DSM-5.

Mythili dan Shanavas (Mythili & Shanavas, 2014) memiliki studi tentang ASD menggunakan Teknik Klasifikasi. Tujuan utama dari makalah ini adalah untuk mendeteksi masalah autisme dan tingkat autisme. Dalam Neural Network ini, teknik SVM dan Fuzzy dengan tools WEKA digunakan untuk menganalisis perilaku dan interaksi sosial siswa. Kosmicki, dkk. (Kosmicki et al., 2015) mengira metode pencarian untuk set paling sedikit sifat untuk deteksi autisme. Dalam hal ini, penulis menggunakan pendekatan pembelajaran mesin untuk mengevaluasi penilaian klinis ASD. ADOS dilakukan pada subset perilaku anak-anak berdasarkan spektrum autisme. ADOS memiliki empat modul. Dalam pekerjaan ini, 8 algoritma

pembelajaran mesin yang berbeda digunakan, yang melibatkan identifikasi fitur mundur bertahap pada lembar skor dari 4.540 individu. Ini menggunakan 9 dari 28 perilaku dari modul 2 dan 12 dari 28 perilaku dari modul 3 untuk mengidentifikasi risiko ASD dengan akurasi keseluruhan masing-masing 98,27% dan 97,66%.

Li dkk. (Li et al., 2017) telah menggunakan pengklasifikasi pembelajaran mesin untuk mendeteksi orang dewasa autisme dengan metode imitasi. Tujuan dari penelitian ini adalah untuk mengetahui permasalahan mendasar terkait dengan kondisi uji diskriminatif dan parameter kinematik. Dataset berisi 16 peserta ASD yang memiliki rangkaian gerakan tangan. Dalam 40 kendala kinematik dari 08 kondisi imitasi telah diekstraksi dengan menggunakan metode pembelajaran mesin. Penelitian ini menunjukkan bahwa untuk sampel kecil, penerapan metode pembelajaran mesin dapat dilakukan untuk menganalisis data berdimensi tinggi dan klasifikasi diagnostik autisme. Tingkat sensitivitas yang dicapai oleh RIPPER yang memiliki fitur V_a (87.30%), CHI (80.95%), IG (80.95%), Correlation (84.13%), CFS (84.13%), dan "no feature selection" (80.00%) pada kumpulan data AQ-Adolescent.

Terbukti dari bagian yang dibahas di atas bahwa ada kebutuhan untuk mengeksplorasi kemungkinan penerapan model berbasis mesin pembelajaran (machine learning) yang dalam untuk mendeteksi ASD pada populasi manusia. Sebagian besar pekerjaan yang dibahas di atas menggunakan pendekatan pembelajaran mesin konvensional dan oleh karena itu kinerjanya terbatas. Dalam pekerjaan ini, performa beberapa model pembelajaran mesin telah dibandingkan dengan model pembelajaran dalam untuk tujuan ini. Model terpisah telah disiapkan untuk kumpulan populasi terpisah (dibahas pada bagian di bawah) dan dibandingkan secara individual.

Pada penelitian ini akan diterapkan algoritma Support Vector Machine (SVM) untuk mengembangkan aplikasi Deteksi Dini Anak Autis. Dengan pengembangan aplikasi ini diharapkan dapat digunakan sebagai alat bantu untuk mendeteksi lebih dini kemungkinan anak autisme, sehingga dapat segera ditangani atau diberikan terapi.

2 Metodologi

Pada penelitian ini digunakan pendekatan kuantitatif. Pada pendekatan penelitian kuantitatif, dilakukan analisa kuantitatif secara teliti terhadap beberapa generasi informasi yang berbentuk kuantitatif. Umumnya pendekatan kuantitatif memiliki tiga bentuk yang berbeda, yaitu pendekatan inferensial, pendekatan eksperimental, dan pendekatan simulasi. Pada pendekatan inferensial, sampel yang diperoleh digunakan untuk membuat dugaan karakteristik populasi, relasinya, dan lain-lain. Pada pendekatan ini, peneliti tidak memiliki kontrol atas karakteristik, variabel, dan responden yang diteliti. Pendekatan eksperimental ditandai dengan adanya kontrol atas lingkungan penelitian oleh peneliti. Eksperimen adalah suatu proses yang sistematis di mana peneliti memiliki kontrol atas variabel berdasarkan pertimbangan agar sesuai dengan tujuan penelitian. Simulasi berarti operasi model numerik yang mewakili struktur proses dinamis. Pada pendekatan simulasi, lingkungan buatan dibuat di mana informasi yang diperlukan dapat dihasilkan.

Tujuan dari metode kuantitatif adalah untuk dapat memahami bagaimana sesuatu dikonstruksi, bagaimana dibangun, dan bagaimana cara kerjanya. Penelitian kuantitatif umumnya didorong oleh hipotesis, kemudian dibuat rumusan dan pengujian secara ketat untuk menunjukkan bahwa hipotesisnya salah. Sehingga usaha yang dilakukan adalah membuktikan bahwa hipotesis yang dibuat adalah salah, jika hipotesisnya tahan uji, maka hipotesis tersebut dianggap benar. Tetapi jika tidak tahan uji, maka hipotesisnya dianggap salah.

Sudut pandang kuantitatif menekankan bahwa pengukuran merupakan dasar yang dapat digunakan untuk menunjukkan hubungan antara observasi dan formalisasi model, teori, dan hipotesis. Penelitian dan metode kuantitatif akan menghasilkan pengembangan model, teori, dan hipotesis yang berkaitan dengan fenomena alam.

Pendekatan yang akan digunakan pada penelitian ini adalah pendekatan eksperimen. Pada penelitian eksperimen dilakukan dengan cara menginvestigasi hubungan sebab-akibat menggunakan pengujian yang dikontrol oleh peneliti. Pada penelitian semiekperimental sering mendapatkan kendala pada tidak cukupnya akses terhadap sampel, masalah etika, dan sebagainya. Untuk pengembangan, evaluasi, dan pemecahan

masalah proyek biasanya dilakukan dengan eksperimen.

Dalam Autisme, tantangan besar yang dihadapi dalam banyak kondisi perawatan kesehatan adalah waktu diagnosis. Diperlukan waktu hingga 6 bulan untuk mendiagnosis anak autis dengan pasti karena proses yang lama, dan seorang anak harus menemui banyak spesialis berbeda untuk mendiagnosis autisme, mulai dari dokter anak perkembangan, ahli saraf, psikiater atau psikolog. Dengan cara tradisional saat ini, waktu yang dibutuhkan untuk menyelesaikan diagnosis Autisme relatif lama.

Oleh karena itu, metode Pembelajaran Mesin dapat membuat perubahan yang relevan untuk mempercepat proses. Diketahui bahwa intervensi dini merupakan kunci untuk memperbaiki anak autis. Jelas mempercepat waktu diagnosis bahkan lebih penting dalam kasus Autisme. Teknologi big data dan pembelajaran mesin dapat membuat kemajuan besar untuk memprediksi dan mempercepat proses diagnosis dan pengobatan yang kompleks dan memakan waktu. Sistem pembelajaran mesin dapat dikembangkan untuk memanfaatkan sejumlah besar data kesehatan dan medis yang tersedia untuk pemodelan prediktif dan analisis prediktif. Dalam makalah ini, perbandingan dari beberapa teknik dan model pembelajaran mesin akan diuji dan dianalisis.

Data diproses sebelumnya untuk membuat prediksi berdasarkan kategori yang berbeda di mana pengujian diklasifikasikan sebagai Autistik. Ada banyak algoritma klasifikasi yang dapat diterapkan. Setiap pengklasifikasi berbeda dalam cara pengumpulan data, pemfilteran data, ekstraksi fitur, dan menggunakan proses ini untuk memasukkan model untuk dipelajari. Dalam pekerjaan ini, efektivitas beberapa algoritma pembelajaran mesin untuk mengevaluasi efektivitas pengobatan atau prediksi hasil pengobatan ASD dinilai.

Stacking adalah cara untuk menggabungkan beberapa klasifikasi atau model regresi. Ada banyak cara untuk model ensemble, model yang banyak dikenal adalah Bagging dan Boosting. Bagging memungkinkan beberapa model serupa dengan varians tinggi yang dirata-ratakan untuk menurunkan varians. Boosting membangun beberapa model inkremental untuk mengurangi bias, sekaligus menjaga varians tetap kecil.

Metode ensemble biasanya digunakan untuk meningkatkan akurasi prediksi dengan menggabungkan prediksi beberapa model pembelajaran mesin. Kearifan tradisional telah menggabungkan apa yang disebut pelajar "lemah". Namun, pendekatan yang lebih modern adalah membuat ansambel dari koleksi model yang kuat namun beragam yang dipilih dengan baik.

Membangun model ansambel yang kuat memiliki banyak kesamaan dengan membangun tim manusia yang sukses dalam bisnis, sains, politik, dan olahraga. Setiap anggota tim memberikan kontribusi yang signifikan dan kelemahan serta bias individu diimbangi oleh kekuatan anggota lainnya.

Jenis ansambel yang paling sederhana adalah prediksi rata-rata tak berbobot dari model yang membentuk perpustakaan model. Misalnya, jika pustaka model menyertakan tiga model untuk target interval (seperti yang ditunjukkan pada gambar berikut), rata-rata tidak tertimbang akan memerlukan pembagian jumlah nilai yang diprediksi dari tiga model kandidat dengan tiga. Dalam rata-rata tanpa bobot, setiap model memiliki bobot yang sama saat model ansambel dibuat.

2.1 Prediksi rata-rata untuk membentuk model ansambel.

Secara lebih umum, Anda bisa memikirkan tentang menggunakan rata-rata tertimbang. Misalnya, Anda mungkin yakin bahwa beberapa model lebih baik atau lebih akurat dan Anda ingin menetapkan bobot yang lebih tinggi secara manual. Namun pendekatan yang lebih baik mungkin memperkirakan bobot ini secara lebih cerdas dengan menggunakan lapisan algoritme pembelajaran lainnya. Pendekatan ini disebut model stacking.

Model Stacking dapat dianalogikan dengan contoh-contoh dunia nyata seperti membangun tim manusia yang sukses dalam bisnis, sains, olahraga, dll. Jika semua anggota tim benar-benar ahli dalam tugas yang sama, maka tim akan menghancurkan tantangan apa pun yang membutuhkan tantangan ini. keterampilan khusus, tetapi akan gagal total saat menangani masalah kehidupan nyata yang kompleks yang membutuhkan banyak keterampilan, pola pikir, dan pendekatan yang beragam. Saya tidak tahu banyak tentang sepak bola Amerika, tetapi bahkan dengan pengetahuan saya yang terbatas, cukup jelas bagi saya bahwa

Anda tidak dapat memenangkan pertandingan sepak bola dengan tim yang hanya terdiri dari quarterback, bahkan jika quarterback itu adalah yang terbaik di liga. Itulah sebabnya tim olahraga yang optimal dan unit bisnis yang sukses terdiri dari kelompok individu yang beragam dengan kekuatan dan kelemahan yang lebih luas.

Model stacking adalah metode ensemble yang efisien di mana prediksi, yang dihasilkan dengan menggunakan berbagai algoritma pembelajaran mesin, digunakan sebagai input dalam algoritma pembelajaran lapis kedua. Algoritme lapisan kedua ini dilatih untuk secara optimal menggabungkan prediksi model untuk membentuk kumpulan prediksi baru. Misalnya, ketika regresi linier digunakan sebagai pemodelan lapisan kedua, ini memperkirakan bobot ini dengan meminimalkan kesalahan kuadrat terkecil. Namun, pemodelan lapis kedua tidak terbatas hanya pada model linier; hubungan antara prediktor bisa lebih kompleks, membuka pintu untuk menggunakan algoritma pembelajaran mesin lainnya.

Stacking (terkadang disebut Generalisasi Stacking) adalah paradigma yang berbeda. Tujuan Stacking adalah untuk mengeksplorasi ruang model yang berbeda untuk masalah yang sama. Idanya adalah bahwa Anda dapat menyerang masalah pembelajaran dengan berbagai jenis model yang mampu mempelajari beberapa bagian dari masalah, tetapi tidak seluruh ruang masalah. Jadi, Anda dapat membangun beberapa pelajar yang berbeda dan Anda menggunakannya untuk membuat prediksi menengah, satu prediksi untuk setiap model yang dipelajari. Kemudian Anda menambahkan model baru yang belajar dari prediksi perantara target yang sama.

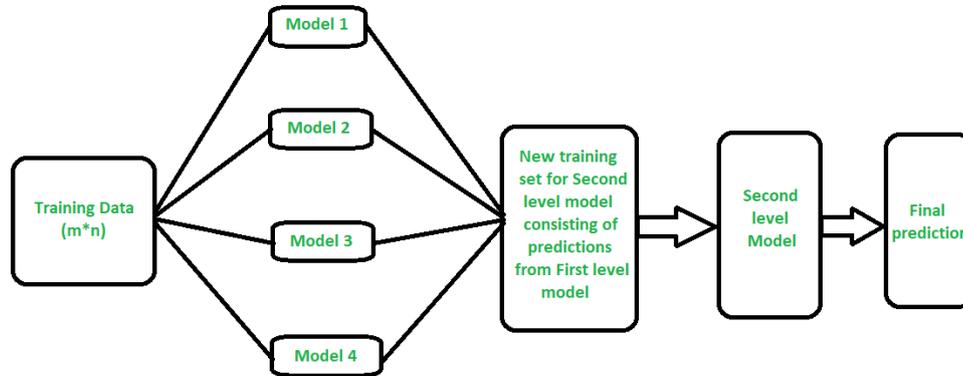
Model Stacking yang ditunjukkan pada Gambar 1 dikatakan ditumpuk di atas yang lain, oleh karena itu dinamai model. Dengan demikian, Anda dapat meningkatkan kinerja Anda secara keseluruhan, dan sering kali Anda mendapatkan model yang lebih baik daripada model perantara individual. Namun perhatikan, itu tidak memberi Anda jaminan apa pun, seperti yang sering terjadi pada teknik pembelajaran mesin apa pun.

Stacking adalah teknik pembelajaran ansambel yang menggabungkan beberapa model klasifikasi atau regresi melalui meta-klasifikasi atau meta-regresor. Model tingkat dasar dilatih berdasarkan set pelatihan yang lengkap, kemudian

meta-model dilatih pada keluaran model tingkat dasar sebagai fitur.

Tingkat dasar sering kali terdiri dari algoritme pembelajaran yang berbeda dan oleh

karena itu ansambel Stacking sering kali heterogen. Algoritme Stacking ditunjukkan pada Gambar 2.



Gambar 1 Model Stacking

Algorithm	Stacking
1:	Input: training data $D = \{x_i, y_i\}_{i=1}^m$
2:	Output: ensemble classifier H
3:	Step 1: learn base-level classifiers
4:	for $t = 1$ to T do
5:	learn h_t based on D
6:	end for
7:	Step 2: construct new data set of predictions
8:	for $i = 1$ to m do
9:	$D_h = \{x'_i, y_i\}$, where $x'_i = \{h_1(x_i), \dots, h_T(x_i)\}$
10:	end for
11:	Step 3: learn a meta-classifier
12:	learn H based on D_h
13:	return H

Gambar 2 Algoritma Stacking

2.2 Data Pendukung

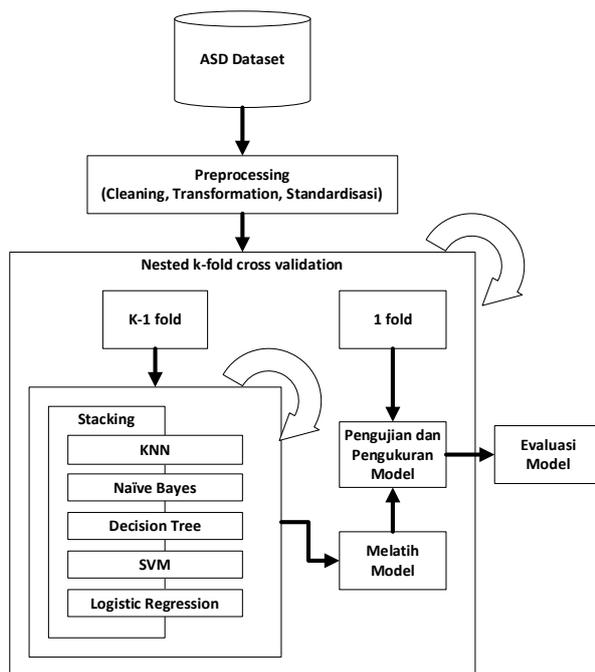
Dataset yang digunakan pada penelitian ini merupakan data sekunder yang diambil dari UCI Dataset (*Autistic Spectrum Disorder Screening Data for Children Data Set*) dengan link <https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++>.

Tabel 1 Spesifikasi Atribut Dataset

Attribute	Type	Description
Age	Number	years
Gender	String	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean (yes or no)	Whether the case was born with jaundice
Family member with PDD	Boolean (yes or no)	Whether any immediate family member has a PDD
Who is completing the test	String	Parent, self, caregiver, medical staff, clinician ,etc.
Country of residence	String	List of countries in text format
Used the screening app before	Boolean (yes or no)	Whether the user has used a screening app
Screening Method Type	Integer (0,1,2,3)	The type of screening methods chosen based on age category (0=toddler, 1=child, 2= adolescent, 3= adult)
Question 1 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 2 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 3 Answer	Binary (0, 1)	The answer code of the question based on the screening method used

Attribute	Type	Description
Question 4 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 5 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 6 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 7 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 8 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 9 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 10 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Screening Score	Integer	The final score obtained based on the scoring algorithm of the screening method used. This was computed in an automated manner

2.3 Model yang Diusulkan

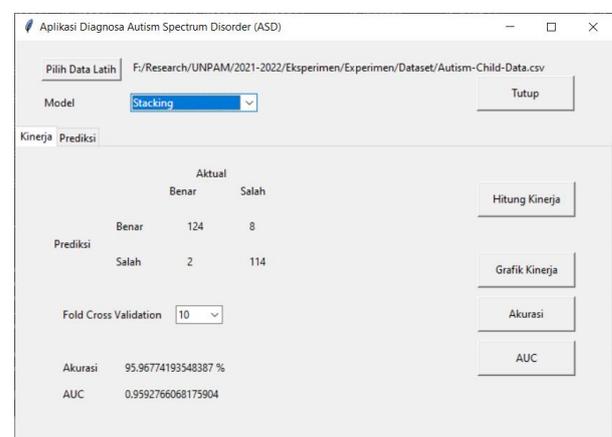


Gambar 3 Model yang Diusulkan

Pada model yang diusulkan digunakan dataset ASD yang telah dikumpulkan. Dataset ASD yang terkumpul dilakukan preprocessing yang mencakup, cleaning, transformasi, dan standarisasi. Dataset akan dibagi menjadi dua bagian menggunakan 10-fold cross validation, yaitu data latih (K-1 fold) dan data uji (1 fold). Data latih digunakan untuk melatih model yang menerapkan KNN, Naive Bayes, Decision Tree, SVM, Logistic Regression, dan stacking dari kelima algoritma tersebut. Model yang telah dilatih kemudian divalidasi menggunakan data uji. Hasil validasi digunakan untuk mengukur kinerja model. Kemudian semua model yang terbentuk dibandingkan untuk mengetahui model terbaiknya.

3 Hasil dan Pembahasan

Pada penelitian ini digunakan bahasa pemrograman Python untuk menerapkan model prediksi/diagnosa Autism Spectrum Disorder (ASD) yang diusulkan. Tampilan aplikasi ditunjukkan pada Gambar 4.



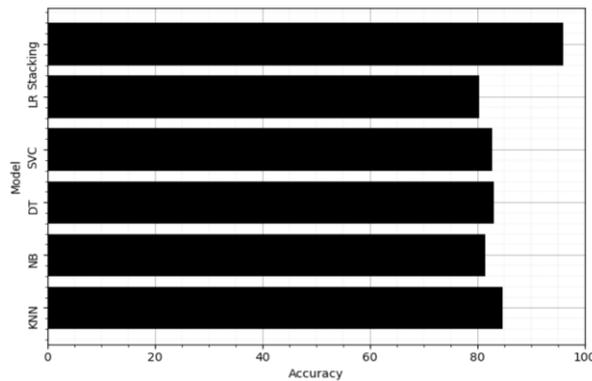
Gambar 4 Tampilan Aplikasi

Hasil penerapan model memberikan hasil kinerja seperti pada Tabel 2.

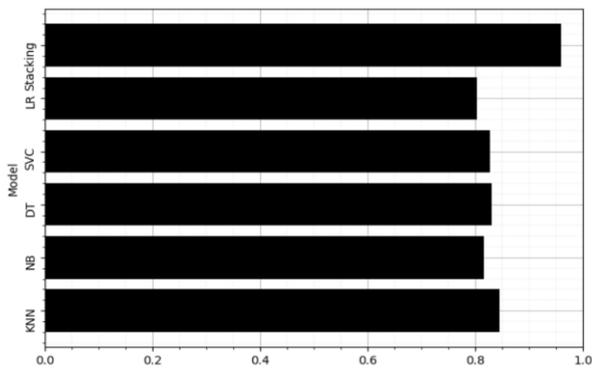
Tabel 2 Kinerja Model

Model	Akurasi	AUC
Stacking	95,97%	0,9593
LR	80,24%	0,8024
SVM (SVC)	82,66%	0,8268
DT	83,06%	0,8301
NB	81,45%	0,8155
KNN	84,67%	0,8456

Kemudian nilai kinerja model divisualisasi menggunakan grafik untuk perbandingannya. Nilai akurasi ditampilkan pada Gambar 5 dan nilai AUC ditunjukkan pada Gambar 6.



Gambar 5 Akurasi model



Gambar 6 AUC model

Berdasarkan nilai yang didapat, menunjukkan bahwa model yang menerapkan teknik stacking memberikan nilai kinerja yang lebih baik. Maka teknik stacking dianggap dapat meningkatkan kinerja model prediksi/diagnosa ASD (Autism Spectrum Disorder).

4 Kesimpulan

Berdasarkan hasil eksperimen yang telah dilakukan dapat disimpulkan sebagai berikut:

- 1) Model prediksi/diagnosa *Autism Spectrum Disorder* (ASD) dapat ditingkatkan menggunakan teknik stacking.
- 2) Model yang menerapkan teknik stacking dapat meningkatkan kinerja model menjadi 95.97% untuk nilai akurasi dan 0,9593 untuk AUC, sehingga kinerjanya masuk kategori sangat baik (*excellent*).

References

Bone, D., Goodwin, M. S., & Lee, M. P. B. C. (2014). *Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and Promises*. 101. <https://doi.org/10.1007/s10803-014-2268-6>

- Constantino, J. N., Lavesser, P. D., Zhang, Y., Abbacchi, A. M., Gray, T., & Todd, R. D. (2007). Rapid Quantitative Assessment of Autistic Social Impairment by Classroom Teachers. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46(12), 1668–1676. <https://doi.org/10.1097/chi.0b013e318157cb23>
- Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a Minimal Set of Behaviors for Autism Detection Through Feature Selection-based Machine Learning. *Translational Psychiatry*, 5(2), e514-7. <https://doi.org/10.1038/tp.2015.7>
- Li, B., Sharma, A., Meng, J., Purushwalkam, S., & Gowen, E. (2017). Applying Machine Learning to Identify Autistic Adults Using Imitation: An Exploratory Study. *PLoS ONE*, 12(8), 1–19. <https://doi.org/10.1371/journal.pone.0182652>
- Mythili, M. S., & Shanavas, A. R. M. (2014). A Study on Autism Spectrum Disorders using Classification Techniques. *International Journal of Soft Computing and Engineering (IJSCE)*, 5, 2231–2307. <http://www.ijscce.org/wp-content/uploads/papers/v4i5/E2433114514.pdf>
- R, V., & R, S. (2018). A Machine Learning based Approach to Classify Autism with Optimum Behavior Sets. *International Journal of Engineering & Technology*, 7(4), 4216–4219.
- Thabtah, F. (2017). Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. *ACM International Conference Proceeding Series, Part F1293*, 1–6. <https://doi.org/10.1145/3107514.3107515>
- Thabtah, F. (2019). Machine Learning in Autistic Spectrum Disorder Behavioral Research: A Review and Ways Forward. *Informatics for Health and Social Care*, 44(3), 278–297. <https://doi.org/10.1080/17538157.2017.1399132>
- Thabtah, F., Kamalov, F., & Rajab, K. (2018). A New Computational Intelligence Approach to Detect Autistic Features for Autism Screening. *International Journal of Medical Informatics*, 117, 112–124. <https://doi.org/10.1016/j.ijmedinf.2018.06.009>
- Wall, D. P., Dally, R., Luyster, R., Jung, J. Y., & DeLuca, T. F. (2012). Use of Artificial Intelligence to Shorten the Behavioral Diagnosis of Autism. *PLoS ONE*, 7(8). <https://doi.org/10.1371/journal.pone.0043855>
- Wall, D. P., Kosmicki, J., Deluca, T. F., Harstad, E., & Fusaro, V. A. (2012). Use of Machine Learning to Shorten Observation-based Screening and Diagnosis of Autism. *Translational Psychiatry*, 2 (December 2011). <https://doi.org/10.1038/tp.2012.10>