



Pemodelan Klasifikasi Pasien Autisme berbasis Artificial Neural Network (ANN)

Presentation by :

KELOMPOK 1





Abstract



Autism Spectrum Disorder (ASD) merupakan kondisi *neurodevelopmental* yang ditandai dengan kesulitan dalam interaksi sosial, komunikasi, dan perilaku repetitif. Penelitian ini bertujuan untuk mengembangkan model klasifikasi ASD menggunakan *Artificial Neural Network* (ANN) dan membandingkannya dengan model *K-Nearest Neighbors* (K-NN) dan *Support Vector Classification* (SVC). Dataset yang digunakan diambil dari situs Kaggle dengan judul "Autism Spectrum Disorder (ASD) Data Set", yang merupakan data dari proses *screening* autisme menggunakan AQ-10. *Hyperparameter tuning* pada model ANN dengan pendekatan bayesian atau *Bayesian Optimization* menaikkan nilai akurasi model ANN dimana menunjukkan performa yang paling baik dalam klasifikasi ASD dengan akurasi sebesar 92%. Model K-NN mencapai akurasi sebesar 90% dalam klasifikasi ASD. Model SVC mencapai akurasi sebesar 79% dalam klasifikasi ASD. Model ANN berpotensi untuk menjadi alternatif metode skrining autisme yang lebih akurat dan efisien.

Kata kunci: *Autism Spectrum Disorder* (ASD), Klasifikasi, *Artificial Neural Network* (ANN), *K-Nearest Neighbors* (K-NN), *Support Vector Classification* (SVC), *Bayesian Optimization*, *Hyperparameter tuning*, *Screening*





Latar Belakang

Menurut WHO (2023), sekitar **1 dari 100** orang di dunia diperkirakan memiliki Autism Spectrum Disorder (ASD), kondisi neurodevelopmental yang menyebabkan kesulitan sosial, perilaku repetitif, dan komunikasi (Erkan & Thanh, 2019).

Tanda-tanda autisme dapat dideteksi dengan screening awal yang penting untuk penanganan dini. Screening melibatkan tes seperti Autism Spectrum Quotient (AQ). AQ dirancang untuk membantu orang dewasa mendeteksi ciri-ciri autistik melalui kuesioner. Versi singkat dari tes ini, disebut AQ-10-Adult, terdiri dari 10 pertanyaan

Analisis ini akan memodelkan klasifikasi autisme menggunakan Jaringan Saraf Tiruan (ANN) dan membandingkannya dengan model machine learning lainnya untuk meningkatkan akurasi diagnosis. Selain itu, penelitian ini juga akan menggunakan Bayesian Optimization, sebuah metode untuk mengoptimalkan hyperparameters model machine learning secara efisien, guna meningkatkan performa model dalam klasifikasi autisme.



Metode Penelitian

ARTIFICIAL NEURAL NETWORK

01

Artificial Neural Network (ANN) adalah model yang terdiri dari jaringan neuron buatan yang terhubung dan mampu memproses serta belajar dari data. Dalam analisis ini, ANN digunakan untuk klasifikasi ASD melalui beberapa tahap: pre-processing data, eksplorasi data, pembentukan model menggunakan Keras, pelatihan model dengan membagi data menjadi set pelatihan dan pengujian, serta evaluasi model menggunakan metrik seperti akurasi dan confusion matrix. Validasi dilakukan dengan K-Fold Cross-Validation untuk memastikan kemampuan generalisasi model.

MACHINE LEARNING

02

Digunakan beberapa metode machine learning untuk mengklasifikasikan autisme dan membandingkannya dengan model ANN. Metode pertama adalah Support Vector Machine (SVM), yang mencari hyperplane optimal untuk memisahkan kelas-kelas dalam data berdimensi N. Model SVM diinisialisasi, dilatih, dan dievaluasi menggunakan precision, recall, dan f1-score. Metode kedua adalah K-Nearest Neighbors (KNN), yang mengklasifikasikan data berdasarkan mayoritas dari k tetangga terdekat. Model KNN diinisialisasi, dilatih, dan dievaluasi dengan metrik yang sama seperti SVM.

BAYESIAN OPTIMIZATION

03

Bayesian Optimization memprediksi performa model untuk memilih hyperparameter terbaik. Dalam klasifikasi ASD, hyperparameter yang dioptimalkan meliputi jumlah lapisan, unit per lapisan, dan learning rate. Prosesnya mencakup mencoba kombinasi awal, memprediksi performa, dan memilih kombinasi berikutnya untuk meningkatkan hasil. Callback 'EarlyStopping' menghentikan pelatihan jika performa tidak membaik. Kombinasi terbaik digunakan untuk model akhir guna memastikan deteksi ASD optimal.



Hyperparameter Tuning: Bayesian Approach

Mampu secara efisien menemukan hiperparameter yang optimal dengan iterasi yang lebih sedikit dengan langkah untuk membangun model probabilitas menggunakan fungsi objektif

PERSAMAAN BAYES

$$p(m|n) = \frac{p(m|n) * p(n)}{p(m)}$$

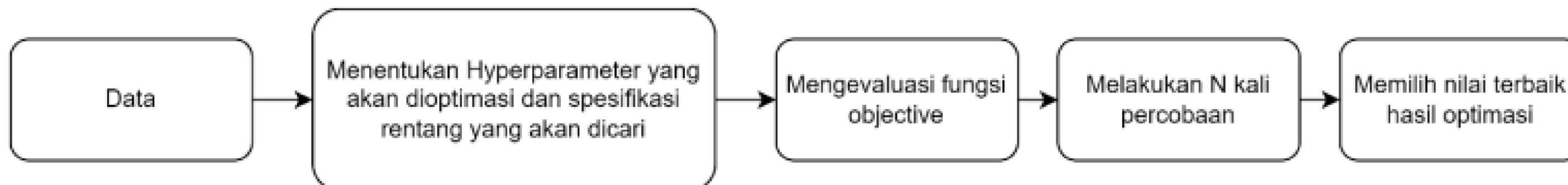
Dengan $p(m|n)$ = hiperparameter probabilitas yang menilai fungsi objektif.

OUTPUT FUNGSI OBJEKTIF

$$p(a|b) = \begin{cases} l(a), & \text{jika } b < b^* \\ g(a), & \text{jika } b \geq b^* \end{cases}$$

Nilai $b < b^*$ menunjukkan biaya yang lebih rendah dari fungsi objektif dari ambang batas, diberi label $l(a)$, dan jika lebih besar, maka diberi label $g(a)$

ALUR HYPERPARAMETER TUNING



Preprocessing Data

OUTLIERS DETECTION

Outliers dideteksi menggunakan boxplot, kemudian interquartile range (IQR) ditentukan menggunakan batas atas dan batas bawah dari kolom tersebut. Nilai-nilai yang berada di luar batas dihapus. Setelah dideteksi, terdapat outliers pada variabel ‘age’. Outliers tersebut dihapus menggunakan teknik ini.

MISSING VALUE IMPUTATION

Karena missing value hanya terdapat pada satu variabel yaitu variabel ‘age’ dengan dua nilai yang missing, maka data dengan missing value ditangani menggunakan imputasi dengan nilai rata-rata.

FEATURE SELECTION

Dalam proses ini, fitur-fitur yang dianggap tidak relevan atau tidak berguna untuk model dihapus untuk meningkatkan kinerja model dan mengurangi kompleksitas. Dalam hal ini variabel ‘age_desc’ dihapus.

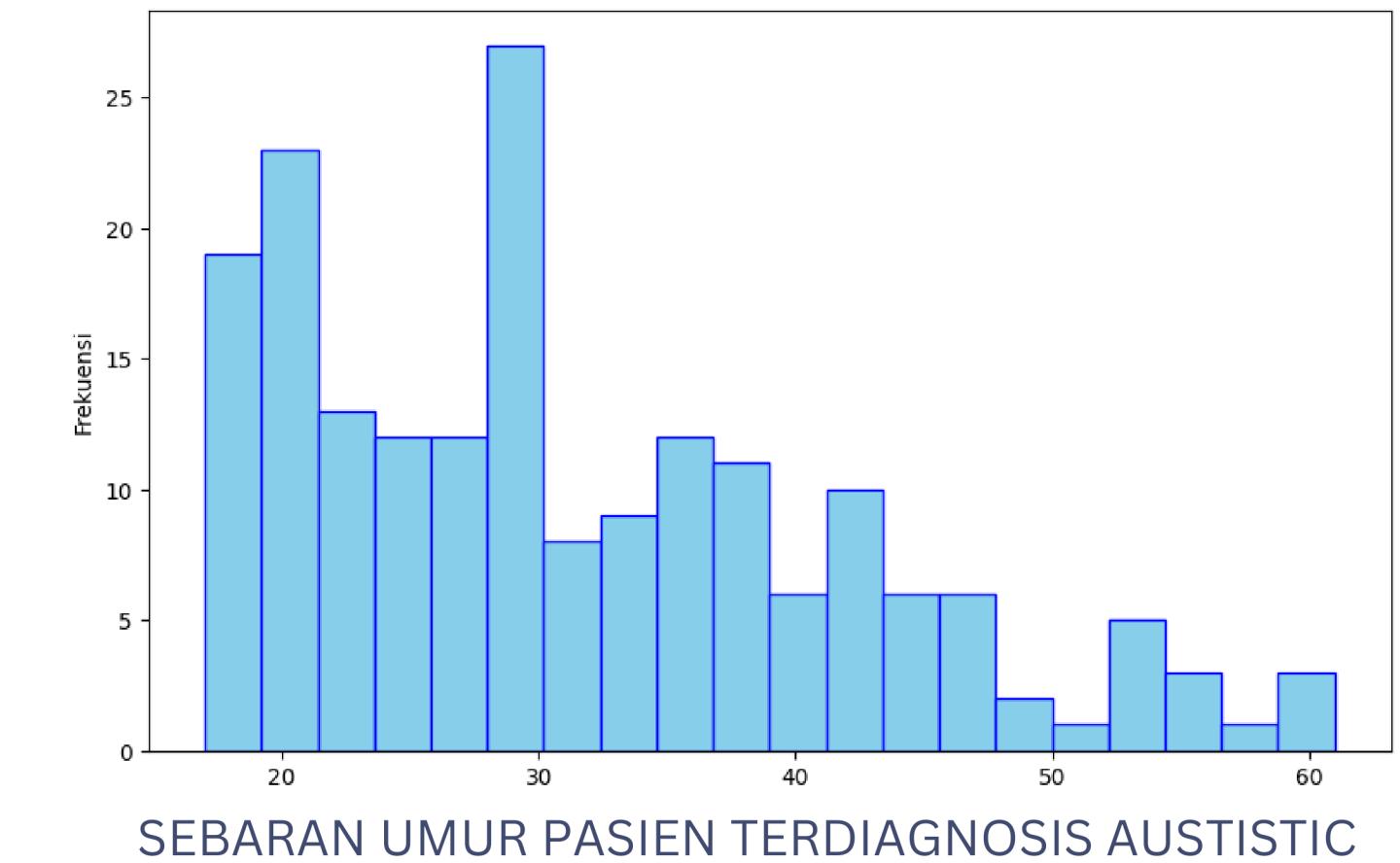
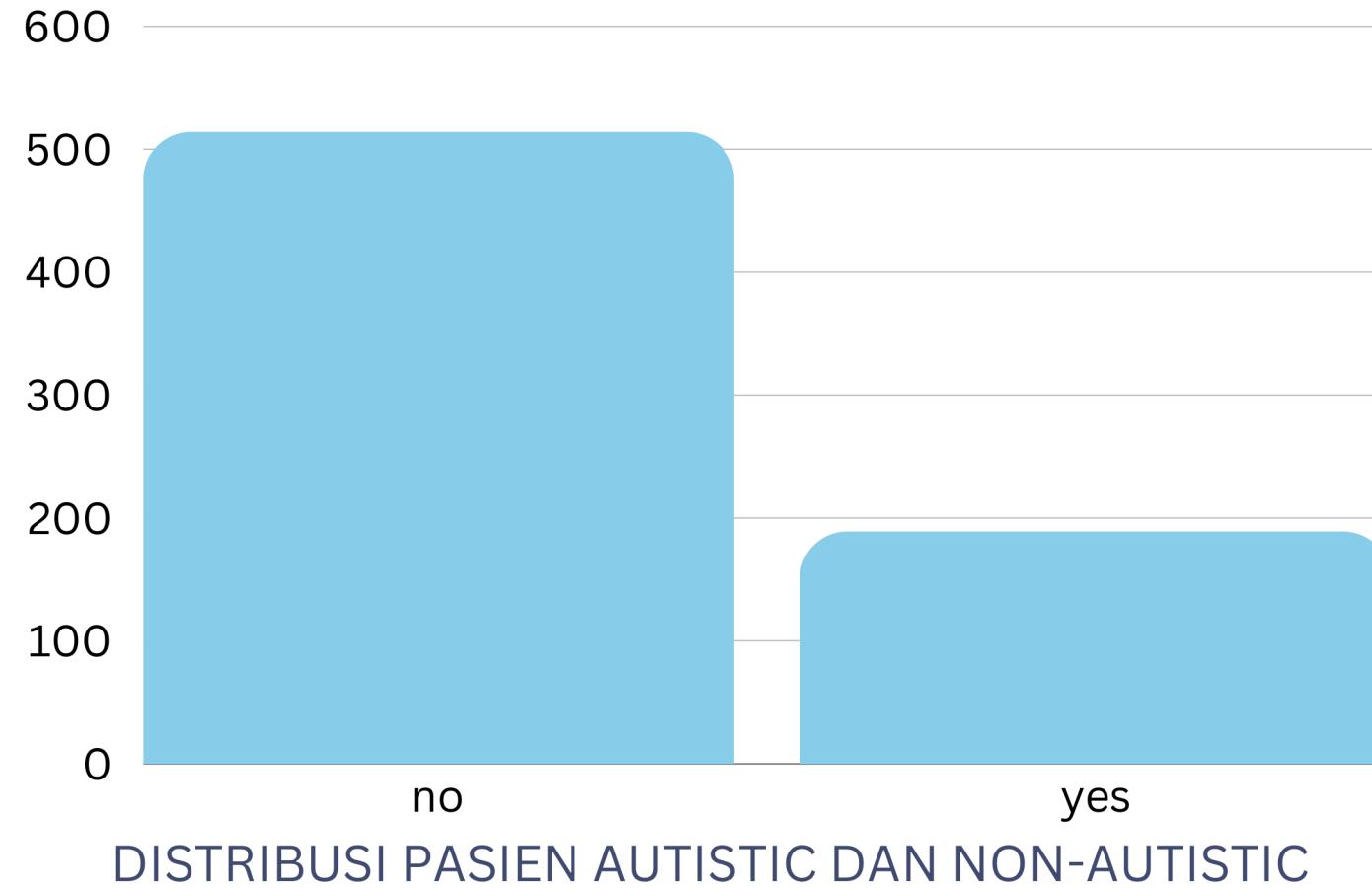
HANDLING CATEGORICAL FEATURES

Data awal yang digunakan masih terdapat fitur kategorikal. Oleh karena itu, fitur kategorikal perlu diubah menjadi format numerik agar bisa digunakan oleh model.

DATASET SPLITTING

Pada proses ini, total jumlah dataset pasien autisme dibagi menjadi dua bagian untuk pelatihan dan pengujian. Berdasarkan model yang diusulkan, bagian data latih berisi 80% data sedangkan 20% data sisanya digunakan untuk tujuan pengujian. Dari total 703 baris, 562 baris pelatihan digunakan untuk membangun model klasifikasi dan 141 baris sisanya digunakan untuk bagian pengujian untuk mengevaluasi model yang telah dibangun.

Exploratory Data Analysis (EDA)

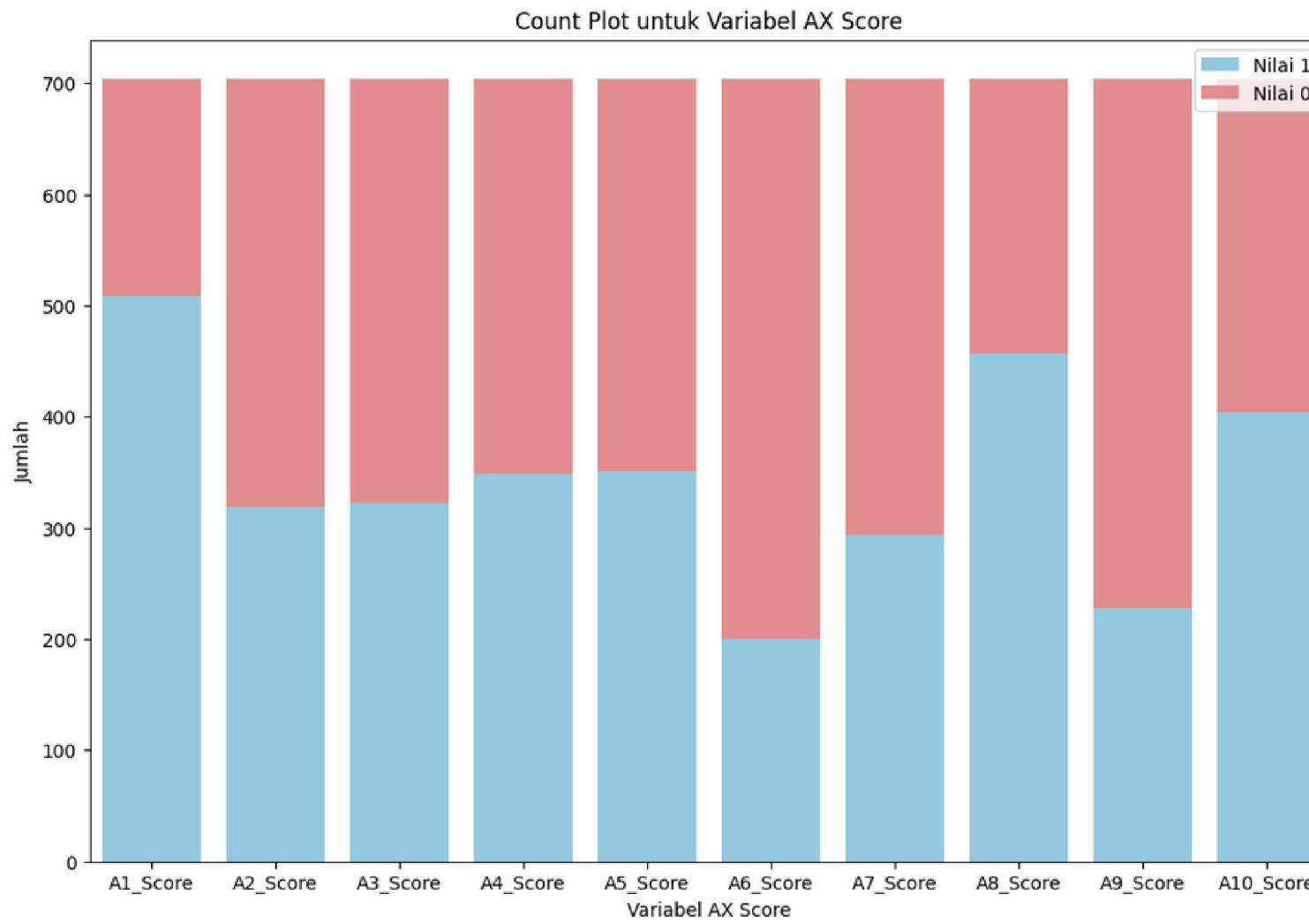


Gambar tersebut menggambarkan distribusi jumlah pasien yang didiagnosis dengan ASD dan yang tidak didiagnosis dengan ASD dalam dataset penelitian ini dimana menunjukkan jumlah pasien non-ASD lebih dominan.

Gambar tersebut menggambarkan distribusi umur partisipan penelitian yang terdeteksi mengalami ASD dimana mayoritas partisipan yang terdeteksi ASD berada di kelompok umur remaja dan dewasa.



Exploratory Data Analysis (EDA)



Gambar plot ini menunjukkan distribusi jumlah partisipan berdasarkan skor AX mereka. Skor AX adalah skor yang diberikan untuk mengukur beberapa aspek dalam penelitian.



Hasil dan Diskusi

Model klasifikasi dengan Artificial Neural Network

ANALISIS KLASIFIKASI UNTUK AUTISME PADA ORANG DEWASA DILAKUKAN PEMODELAN DENGAN MEMBUAT LAYER ANN. ARSITEKTUR YANG DIGUNAKAN PADA PEMODELAN INI ADALAH SEBAGAI BERIKUT :

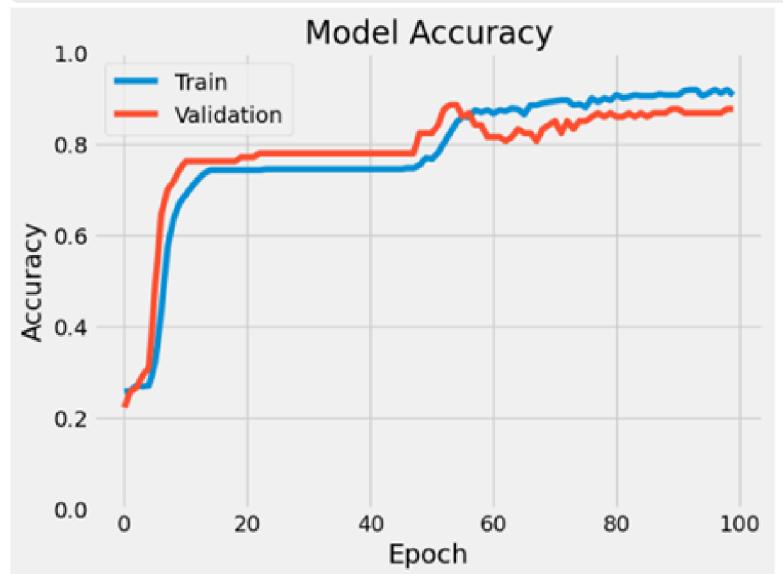
Nama Layer	Bentuk Input	Bentuk Output	Parameter
Input Layer	(None, 19)	(None, 19)	-
dense	(None, 19)	(None, 10)	200
dense_1	(None, 10)	(None, 8)	88
dense_2	(None, 8)	(None, 1)	9

MODELLING

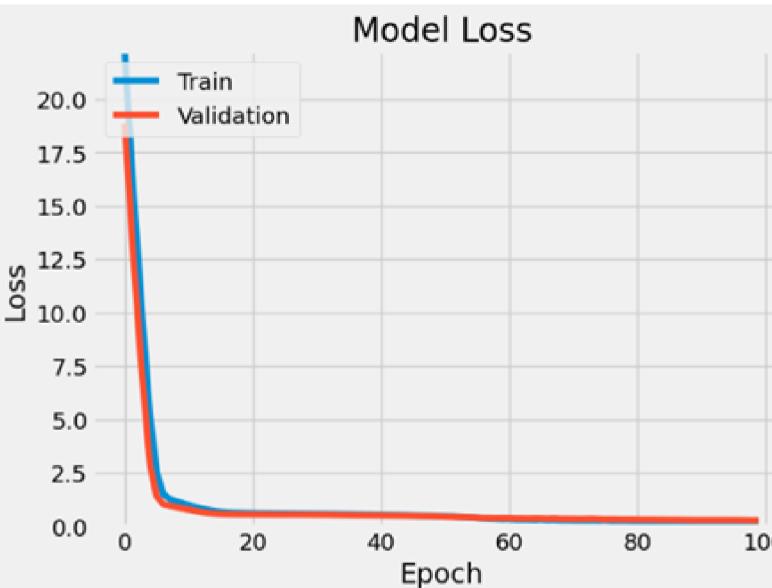
OPTIMIZER: ADAM

LEARNING RATE: 0,001

PLOT AKURASI



PLOT ERROR (LOSS)



AKURASI: 87%

TUNING

Proses Hyperparameter tuning dengan Bayesian Optimization

Hyperparameter	Jangkauan Pencarian
n_layers	(1,10)
n_units	(16,512)
Learning_rate	(1e-6, 1e-2)

Jumlah Iterasi : 15

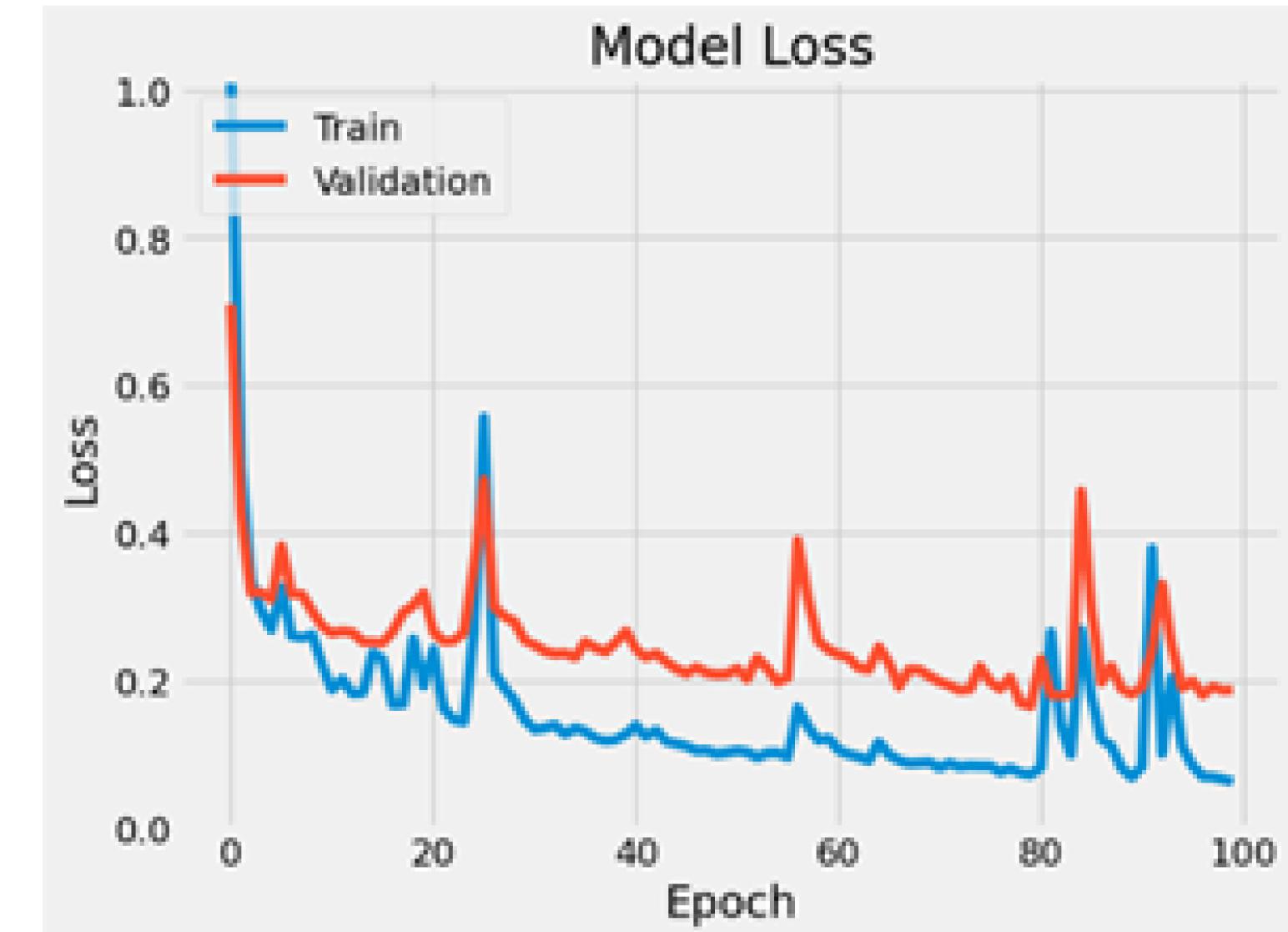
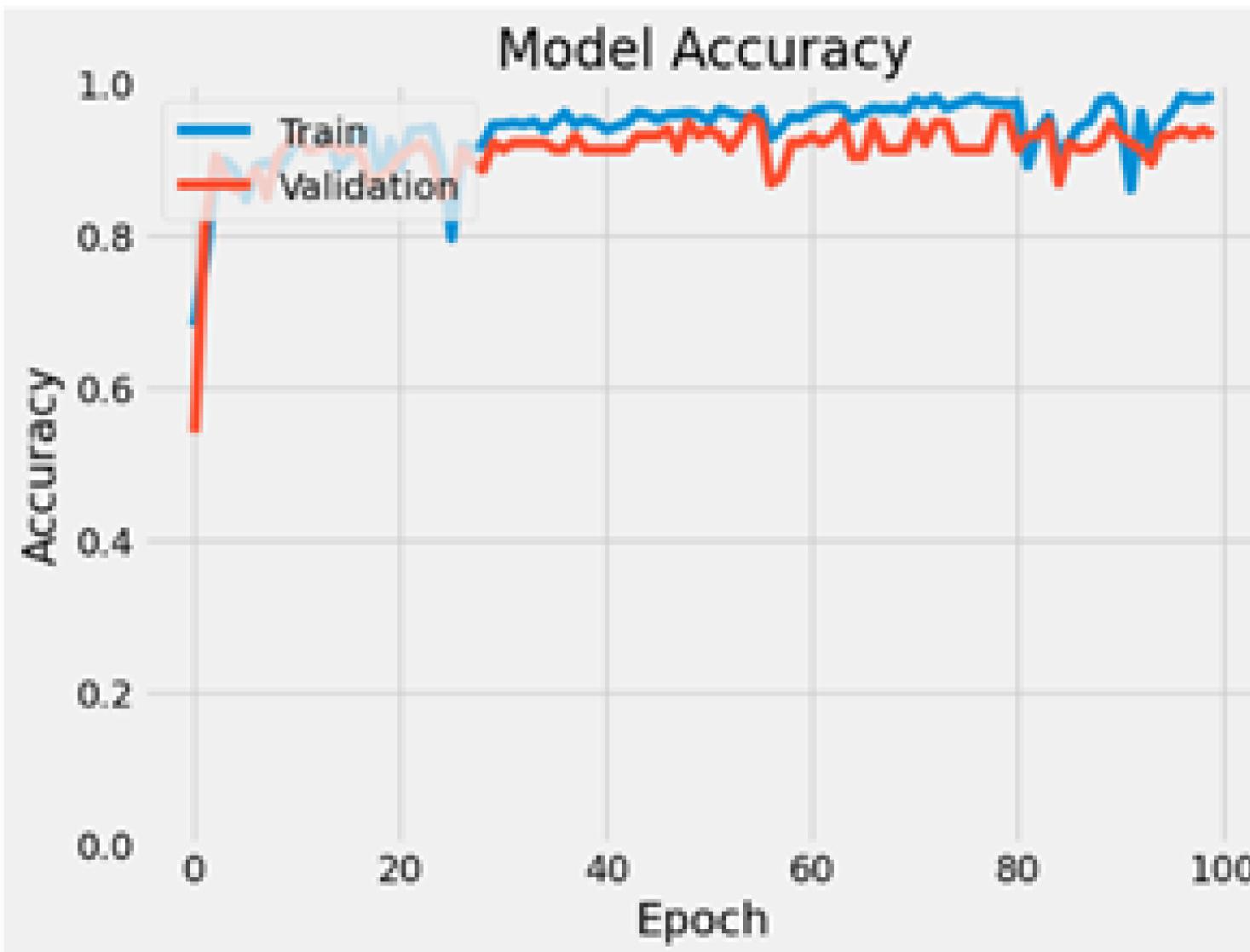
Callback : Early Stopping

Iter	Target	Learning_rate	n_layers	N_units
1	0.8794	0.004171	7.483	16.06
2	0.9007	0.003024	2.321	61.8
3	0.8936	0.001863	4.11	212.8
4	0.8865	0.005389	4.773	355.9
5	0.8865	0.002045	8.903	29.58
6	0.8936	0.002928	2.371	60.78
7	0.9149	0.003375	1.624	62.39
8	0.8865	0.00382	1.099	63.12
9	0.8794	0.002276	2.4	485.9
10	0.6525	0.007263	7.558	112.6
11	0.8156	0.001907	6.944	208.7
12	0.8936	0.002746	1.654	216.4
13	0.766	0.004839	6.232	217.1
14	0.8652	0.003723	1.23	210.8
15	0.8652	0.006581	8.46	25.2

Model Terbaik:

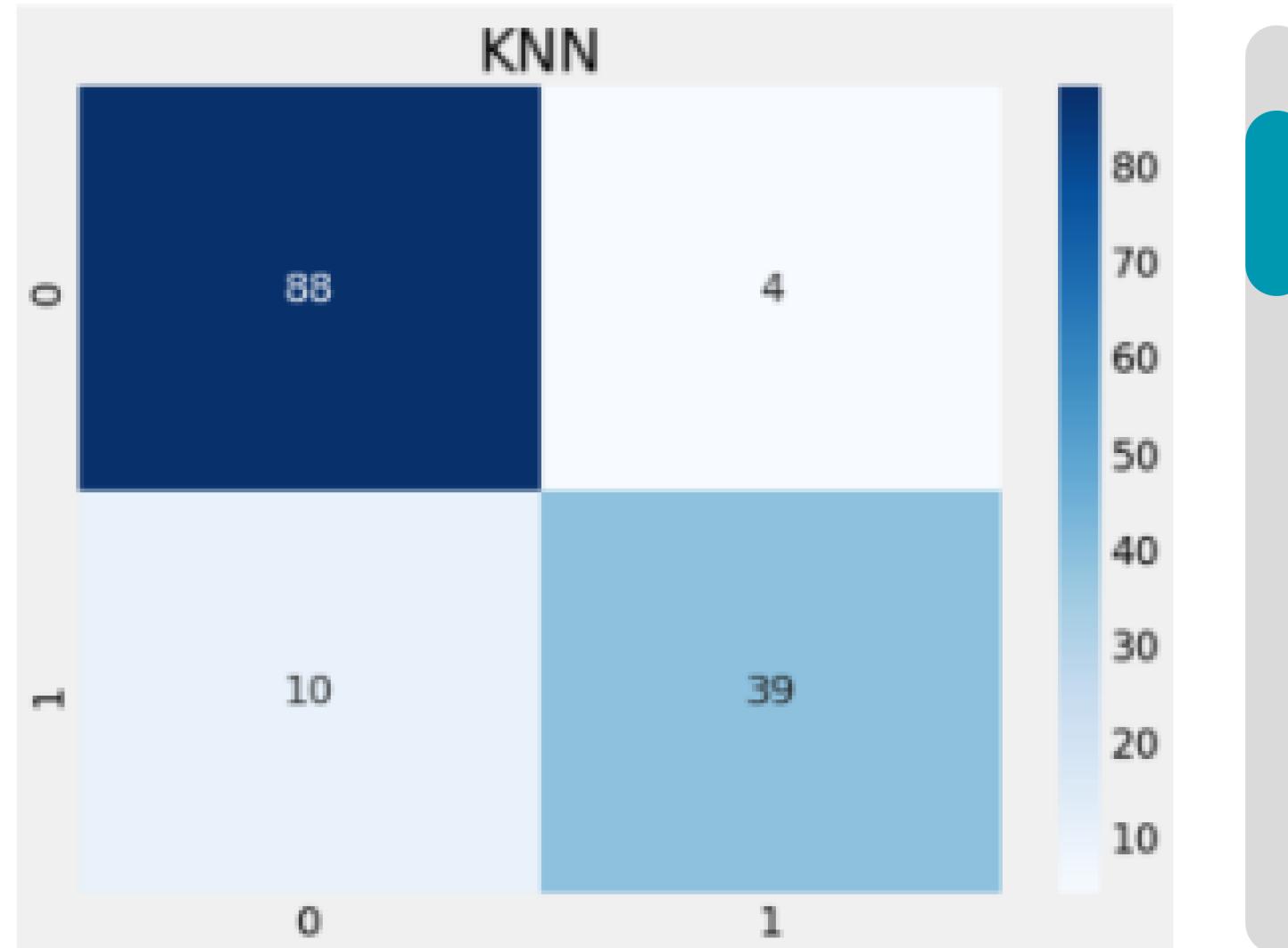
Nama Layer	Bentuk Input	Bentuk Output	Parameter
Dense_499_input	(None, 19)	(None, 19)	-
dense_499	(None, 19)	(None, 62)	1240
dense_500	(None, 62)	(None, 1)	63

Hasil training dengan model terbaik



AKURASI: 92%





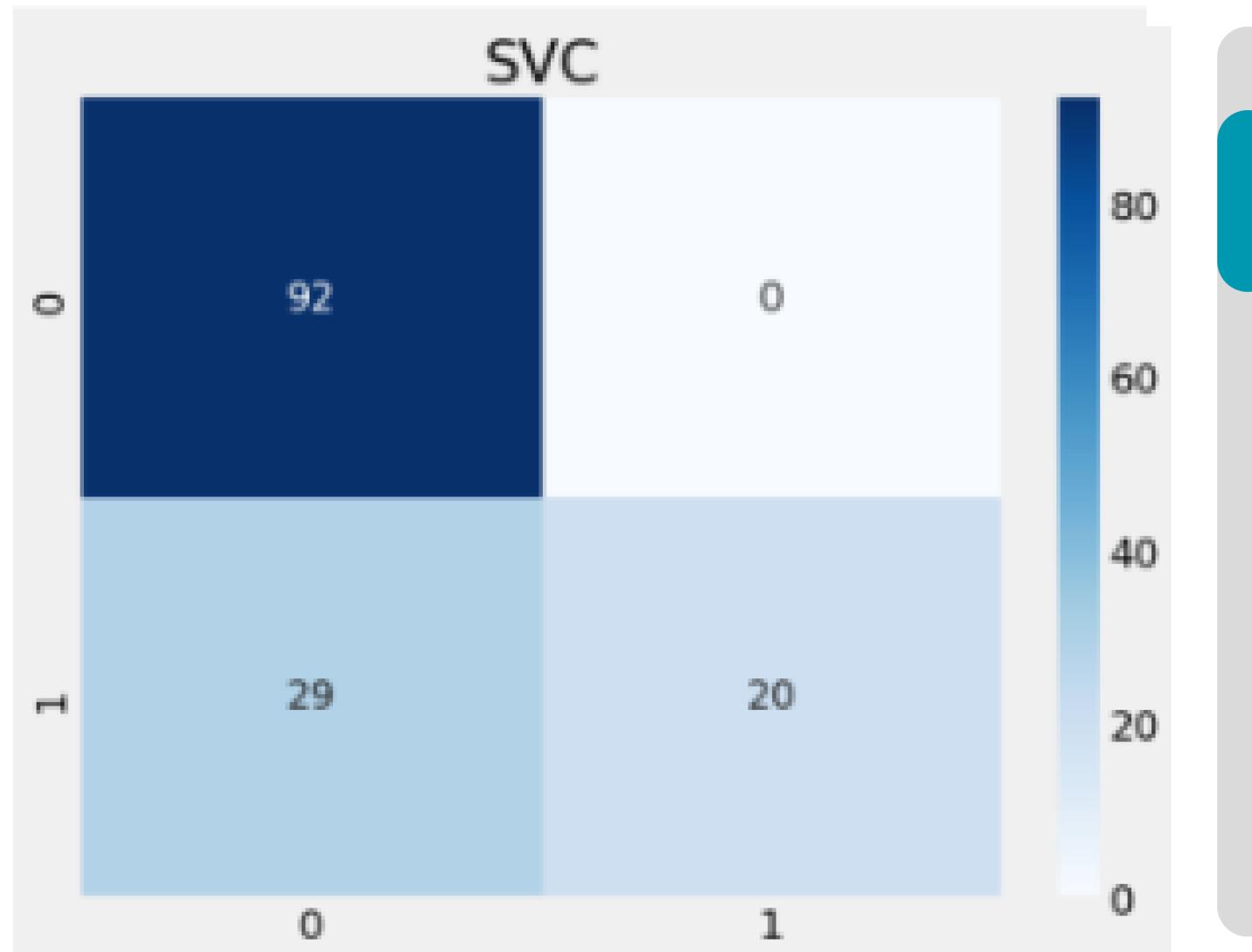
K-Nearest Neighbors

PERMODELAN

Permodelan *K-Nearest Neighbors* dilakukan dengan paket program Python menggunakan library dari “scikit learn” dan tanpa menggunakan pengaturan pada parameter dari model.

AKURASI

Tingkat akurasi dari model dalam melakukan klasifikasi autisme pada orang dewasa sebesar **90%** dengan **rata-rata ketepatan pada setiap kelasnya adalah 89%**. Selain itu didapat dari visualisasi *confusio' matrix* bahwa model dapat dengan sangat baik memprediksi kelas 0 dan cukup baik dalam memprediksi kelas 1.



Support Vector Classification

PERMODELAN

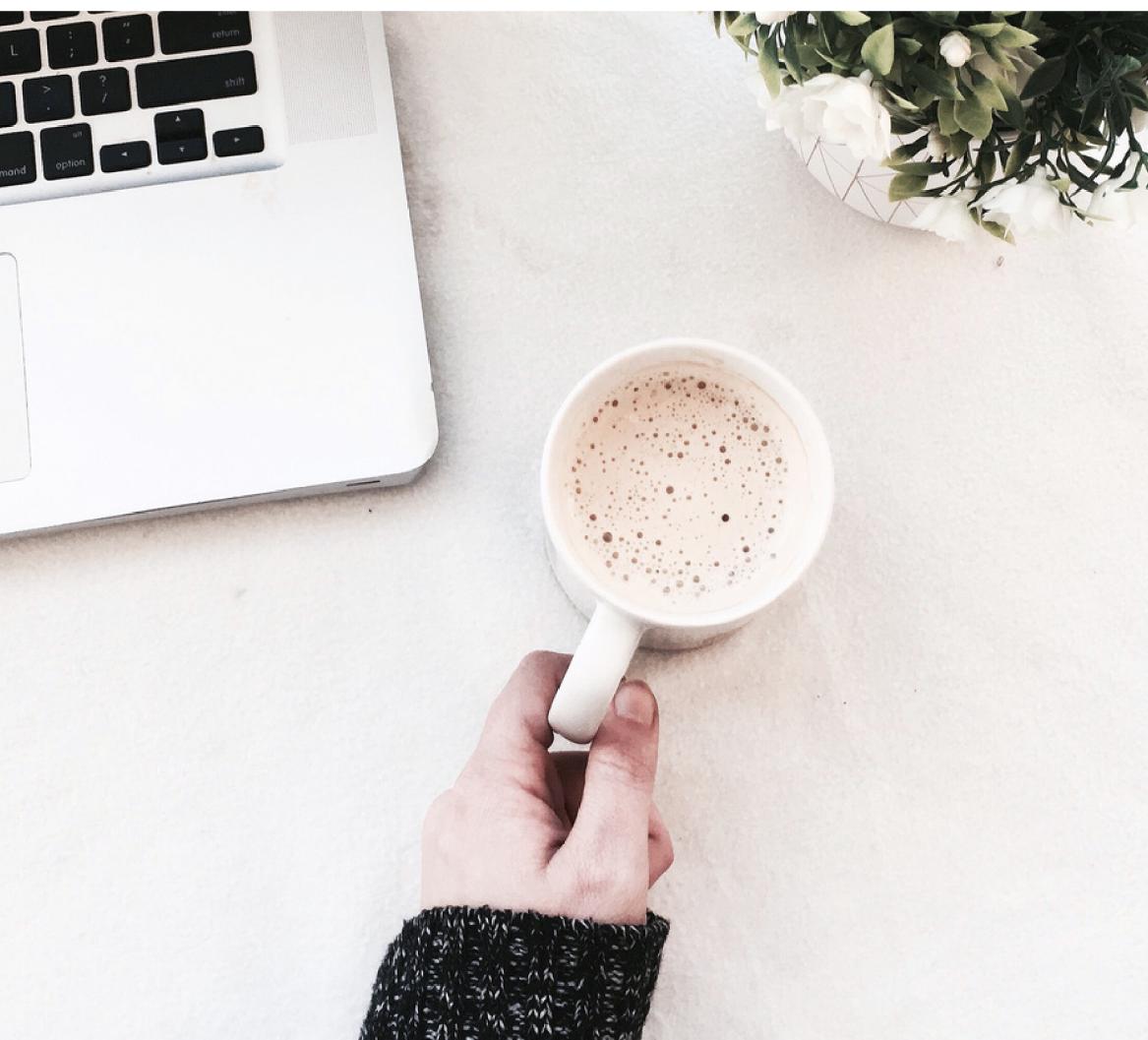
Permodelan *Support Vector Classification* (sVC) dilakukan dengan paket program Python menggunakan library dari “scikit learn” dan tanpa menggunakan pengaturan pada parameter dari model.

AKURASI

Tingkat akurasi dari model dalam melakukan klasifikasi autisme pada orang dewasa sebesar **79%** dengan **rata-rata ketepatan pada setiap kelasnya adalah 72%**. Selain itu didapat dari visualisasi *confusio' matrix* bahwa model dapat dengan sangat baik memprediksi kelas 0, tetapi **masih kurang baik dalam memprediksi kelas 1**.



Kesimpulan & Saran



Dalam melakukan klasifikasi autisme pada orang dewasa dengan menggunakan **data yang kecil dan tidak seimbang antar kelas 0 dan 1**, diperoleh bahwa model ANN memperoleh akurasi yang lebih besar jika dibandingkan dengan algoritma *machine learning* konvensional seperti K-NN dan SVC.

Pada kasus ini, diperoleh tingkat akurasi **ANN sebesar 92%**, sedangkan tingkat akurasi algoritma **K-NN dan SVC berturut-turut sebesar 90% dan 79%** dan **hanya dapat dengan baik memprediksi pada kelas 1 saja**

Saran yang dapat diberikan dari penelitian ini adalah dalam kasus pengklasifikasian autisme orang dewasa lebih baik jika menggunakan algoritma ANN karena memiliki tingkat akurasi yang tinggi baik secara umum, maupun pada setiap kelasnya.





UNIVERSITAS GADJAH MADA

Thank You So Much!



Presentation by :

KELOMPOK 1

