

LAPORAN AKHIR
STUDI INDEPENDEN BERSERTIFIKAT
Foundations of AI and Life Skills for Gen-Z
Di Orbit Future Academy

Diajukan untuk memenuhi persyaratan kelulusan
Program MSIB MBKM

oleh :
Fadhilah Nur Amaliah (52419118)



TEKNIK INFORMATIKA
UNIVERSITAS GUNADARMA
2022

Lembar Pengesahan Teknik Informatika Universitas Gunadarma

**Prediksi Tingkat Risiko Penyebaran Covid-19 pada Pulau
Sumatera Menggunakan Metode KNN dan SVM**

Di Orbit Future Academy

oleh :

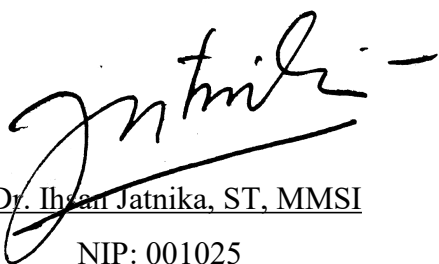
Fadhilah Nur Amaliah (52419118)

disetujui dan disahkan sebagai

Laporan Magang atau Studi Independen Bersertifikat Kampus Merdeka

Bekasi, 16 Juni 2022

Pembimbing Magang atau Studi Independen Teknik Informatika Universitas
Gunadarma



Dr. Ihsan Jatnika, ST, MMSI
NIP: 001025

Lembar Pengesahan

**Prediksi Tingkat Risiko Penyebaran Covid-19 pada Pulau
Sumatera Menggunakan Metode KNN dan SVM**

Di Orbit Future Academy

oleh :

Fadhilah Nur Amaliah (52419118)

disetujui dan disahkan sebagai

Laporan Magang atau Studi Independen Bersertifikat Kampus Merdeka

Bandung, 16 Juni 2022

AI Coach



Angel Metanosa Afinda, S.Kom.

NIP: 2201043

Abstraksi

Orbit Future Academy didirikan dengan tujuan untuk meningkatkan kualitas hidup lewat cara inovasi, Pendidikan, pelatihan keterampilan. Dengan adanya proses pembelajaran melalui Studi Independent melalui program MSIB Batch 2 dengan topik yang diambil yaitu AI. Sebagai Project Akhir untuk syarat kelulusan, penulis mengambil tema Data Science dengan judul “Prediksi Tingkat Risiko Penyebaran Covid-19 pada Pulau Sumatera Menggunakan Metode K-NN dan SVM”. Covid-19 merupakan virus menular yang menyerang saluran pernafasan. Penularan Covid-19 yang sangat cepat ini memiliki dampak yang luas bagi seluruh masyarakat Indonesia dan seluruh dunia. Bidang Data Science ini sangat berperan dalam pemantauan kasus Covid-19. Pada saat ini, forecasting berperan penting untuk memprediksi tingkat risiko Covid-19 di Indonesia. Indonesia melakukan penutupan diseluruh wilayah yang memiliki jumlah kasus positif yang cukup besar, salah satunya adalah wilayah Sumatera. Penelitian ini bertujuan untuk melihat seberapa efektif penerapan model KNN dan SVM untuk prediksi sebuah data penyebaran Covid-19 pada Pulau Sumatera. Pengukuran antara pengujian kedua model tersebut dapat dilihat dari nilai akurasi yang didapat dari hasil confusion matrix. Dari kedua model algoritma yang diuji, dihasilkan bahwa penggunaan model KNN dengan n neighbor 5 menghasilkan akurasi tertinggi 87% dimana hasil lebih efektif dalam memprediksi tingkat penyebaran Covid-19 di Pulau Sumatera.

Kata kunci: Covid-19, KNN, SVM, Prediksi, Kernel.

Kata Pengantar

Puji dan syukur penulis panjatkan kepada Allah SWT. atas segala nikmat dan berkah yang telah dilimpahkan kepada penulis, sehingga dengan ridho-Nya penulis dapat menyelesaikan Laporan Akhir ini.

Laporan Akhir ini disusun guna melengkapi sebagian syarat untuk kelulusan program MSIB Foundations of AI and Life Skills for Gen-Z di Orbit Future Academy. Adapun topik proyek yang dibahas dalam Pembelajaran Studi Independen ini adalah **“Prediksi Tingkat Risiko Penyebaran Covid-19 pada Pulau Sumatera Menggunakan Metode KNN dan SVM”**.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu selama pelaksanaan kegiatan Studi Independen Bersertifikat dan penyusunan Laporan Akhir, antara lain, yaitu:

1. Bapak Dr. -Ing. Ilham A. Habibie, Selaku Co-Founder & President Commissioner Orbit Future Academy.
2. Ibu Prof. Dr. E. S. Margianti, SE., MM., selaku Rektor Universitas Gunadarma.
3. Bapak Dr. Ihsan Jatnika, ST, MMSI., selaku Dosen Pembimbing MSIB Teknik Informatika Universitas Gunadarma.
4. Coach Angel Metanosa Afinda, S.Kom., selaku homeroom coach yang telah memberikan segala saran, kritik, waktu dan bimbingannya yang membangun, sehingga laporan ini dapat segera terselesaikan.
5. Para Coach Orbit Future Academy, yang telah mengajar dan membimbing penulis selama kegiatan Studi Independen di Orbit Future Academy yang tidak dapat penulis sebutkan namanya satu-persatu.

6. Teman-teman satu tim Project Akhir, yang terdiri dari Naila Ameera Larasati, Ghifary Catur Achmad, Nova Nurul Putri, dan Mohammad Luthfan Faohan yang telah senantiasa menemani dan ikut berperan dalam penyusunan Project Akhir.
7. Teman-teman kelas Jupyter XXI dan kelas 3IA17 yang telah memberi dukungan baik secara langsung maupun tidak langsung yang tidak dapat penulis sebutkan namanya satu-persatu.

Penulis menyadari sepenuhnya bahwa Laporan Akhir ini masih jauh dari kata sempurna, karena keterbatasan kemampuan, pengetahuan dan pengalaman penulis. Oleh karena itu saran-saran dan ide-ide yang mendukung sangat penulis harapkan untuk perbaikan selanjutnya.

Daftar Isi

Pendahuluan	1
Latar belakang	1
Lingkup	1
Tujuan	2
Orbit Future Academy	1
Struktur Organisasi	1
Lingkup Pekerjaan	2
Deskripsi Pekerjaan	3
Jadwal Kerja	4
Prediksi Tingkat Risiko Penyebaran Covid-19 pada Pulau Sumatera Menggunakan Metode KNN dan SVM	1
Latar Belakang Proyek Akhir	1
Proses Pelaksanaan Proyek Akhir	1
Hasil Proyek Akhir	1
3.3.1 Pembahasan Algoritma KNN	6
3.3.2 Pembahasan Algoritma SVM	11
Penutup	1
Kesimpulan	1
Saran	1
Referensi	1
Lampiran A. TOR	2
Lampiran B. Log Activity	2
Lampiran C. Jurnal	2

Daftar Tabel

Tabel 2.1 Tugas Student	4
Tabel 2.2 Agenda Kelas	4
Tabel 2.3 Model Algoritma KNN	3
Tabel 2.4 Model Algoritma SVM	4
Tabel 2.5 Evaluasi Matrix	15
Tabel 2.6 Evaluasi Matrix	8

Daftar Gambar

Gambar 2.1 Logo Orbit Future Academy	1
Gambar 2.2 Struktur Organisasi OFA	2
Gambar 2.3 Flowchart Pengerjaan	2
Gambar 2.4 Nilai Performansi Matriks	6
Gambar 2.5 Visualisasi confusion matrix	7
Gambar 2.6 Nilai Performansi Matriks	8
Gambar 2.7 Visualisasi Confusion Matrix	8
Gambar 2.8 Visualisasi Confusion Matrix	9
Gambar 2.9 Nilai Performansi Matriks	9
Gambar 2.10 Visualisasi Confusion Matrix	10
Gambar 2.11 Nilai Performansi Matriks	10
Gambar 2.12 Nilai Performansi Matriks	11
Gambar 2.13 Visualisasi Confusion Matrix	11
Gambar 2.14 Nilai Performansi Matriks	12
Gambar 2.15 Visualisasi Confusion Matrix	12
Gambar 2.16 Nilai Performansi Matriks	13
Gambar 2.17 Visualisasi Confusion Matrix	14
Gambar 2.18 Visualisasi Confusion Matrix	14
Gambar 2.19 Nilai Performansi Matriks	15

Bab I Pendahuluan

I.1 Latar belakang

Indonesia saat ini telah memasuki era revolusi industri ke 4 atau biasa dikenal dengan nama “Industri 4.0”. Di era ini, penggunaan teknologi sangat berperan dalam peningkatan kehidupan sehari-hari kita sebagai manusia. Kecerdasan Buatan (*Artificial Intelligence*) merupakan salah satu teknologi pendukung yang berperan penting dalam Revolusi Industri 4.0. *Artificial Intelligence* (AI) menurut John McCarthy merupakan suatu ilmu dan teknik dalam menciptakan mesin yang bersifat cerdas, terutama dalam menciptakan program atau aplikasi komputer cerdas[1]. Untuk menghadapi tantangan Industri 4.0 ini, mempelajari dan mendalami *Artificial Intelligence* merupakan suatu hal yang penting.

Dalam hal ini, Orbit Future Academy memberikan dukungan dengan menyediakan program Studi Independen Bersertifikasi. Melalui program Foundations of AI and Life Skills for Gen-Z di Orbit, penulis dapat mempelajari dan mengimplementasikan langsung ilmu yang didapat lewat Proyek Akhir. Adapun topik yang dibahas dalam Proyek Akhir adalah **“Prediksi Tingkat Risiko Penyebaran Covid-19 pada Pulau Sumatera Menggunakan Metode KNN dan SVM”**.

Dalam penelitiannya, penulis menggunakan algoritma KNN dan SVM untuk memprediksi tingkat risiko Covid-19 pada Pulau Sumatera. Dua Metode ini biasanya digunakan untuk pemecahan masalah pada bidang *Data Science*. Tujuan dari penelitian adalah untuk melihat seberapa efektif penerapan metode KNN dan SVM untuk prediksi sebuah data penyebaran COVID-19 pada Pulau Sumatera.

I.2 Lingkup

Project yang dikerjakan terfokus pada:

1. Membuat model untuk prediksi data penyebaran COVID-19 pada Pulau Sumatera.
2. Melakukan validasi model menggunakan *confusion matrix*.

I.3 Tujuan

Mengetahui tentang apa saja yang dipelajari dalam domain AI terutama *Data Science* serta dapat mengimplementasikannya untuk mengatasi masalah di kehidupan sehari-hari. Dalam proyek yang dikerjakan, tujuannya adalah untuk melihat seberapa efektif penerapan algoritma KNN dan SVM untuk prediksi sebuah data penyebaran COVID-19 pada Pulau Sumatera.

Bab II Orbit Future Academy

II.1 Struktur Organisasi



Gambar 2.1 Logo Orbit Future Academy

Orbit Future Academy (OFA) didirikan pada tahun 2016 dengan tujuan untuk meningkatkan kualitas hidup melalui inovasi, edukasi, dan pelatihan keterampilan. Label atau *brand* Orbit merupakan kelanjutan dari warisan mendiang Prof. Dr. Ing. B. J. Habibie (presiden Republik Indonesia ke-3) dan istrinya, Dr. Hasri Ainun Habibie. Mereka berdua telah menjadi penggerak dalam mendukung perkembangan inovasi dan teknologi pendidikan di Indonesia. OFA mengkurasi dan melokalkan program/kursus internasional untuk *upskilling* atau *reskilling* pemuda dan tenaga kerja menuju pekerjaan masa depan. Hal ini sesuai dengan slogan OFA, yakni “*Skills-for-Future-Jobs*”.

Visi:

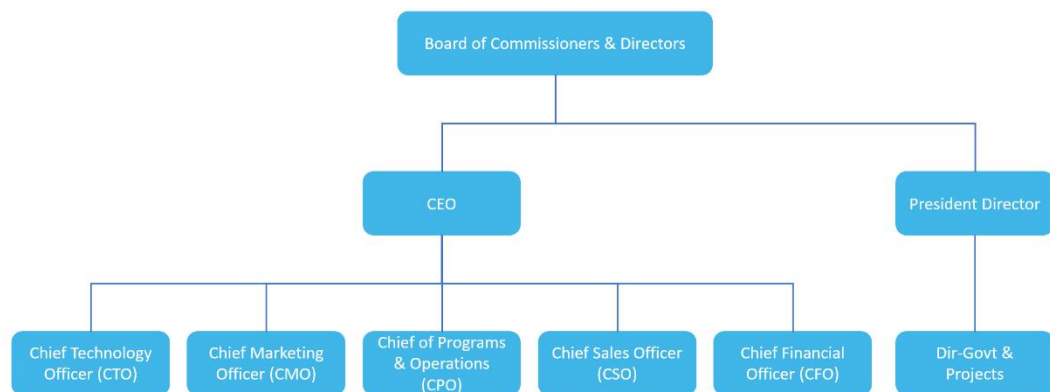
Memberikan pembelajaran berbasis keterampilan transformatif terbaik untuk para pencari kerja & pencipta lapangan kerja.

Misi:

1. Membangun jaringan Orbit Transformation Center (OTC) secara nasional untuk menyampaikan kurikulum keterampilan masa depan berbasis sertifikasi melalui Platform Konten Digital.

2. Secara proaktif bekerja dengan pemerintah & organisasi dengan mengubah tenaga kerja mereka agar sesuai dengan perubahan pekerjaan yang terjadi karena Industri 4.0.
3. Melatih pemuda dengan keterampilan kewirausahaan & mencocokkan mereka dengan peluang masa depan yang muncul di berbagai industri.
4. Menghubungkan jaringan inkubator dan akselerator yang dikurasi ke industri, investor, dan ekosistem start-up global.

Struktur organisasi OFA dapat dilihat pada Gambar 2.2.



Gambar 2.2 Struktur Organisasi OFA

II.2 Lingkup Pekerjaan

Seorang fasilitator akan mendampingi kurang lebih 40 peserta MSIB (student) dalam satu kelas. Terdapat dua jenis fasilitator, yakni:

a. Homeroom Coach

Homeroom coach bertugas menyampaikan materi tentang dasar-dasar AI, memberikan penilaian pada student, dan mengarahkan *student* saat pengerjaan Proyek Akhir (PA).

b. Domain Coach

Domain coach bertugas menyampaikan materi tentang domain AI atau life skills dan memberikan penilaian pada student.

Lingkup pekerjaan student adalah mengikuti kelas bersama homeroom atau domain coach, sesuai agenda kelas, hingga program selesai.

II.3 Deskripsi Pekerjaan

Berikut adalah deskripsi pekerjaan student sebelum pengerjaan PA:

- a. Mengikuti pre-test.
- b. Mengikuti kelas sesi pagi pada pukul 08.00 hingga 11.30 WIB.
- c. Mengikuti kelas sesi siang pada pukul 13.00 hingga 16.30 WIB.
- d. Mengulang materi yang telah disampaikan di kelas sesi pagi dan siang, setelah kelas sesi siang, selama 1 jam (*self-study*).
- e. Mengerjakan latihan individu atau kelompok yang diberikan oleh homeroom atau domain coach saat kelas berlangsung.
- f. Mengerjakan tugas yang diberikan homeroom atau domain coach hingga batas waktu tertentu.
- g. Mengerjakan *mini project* yang diberikan homeroom atau domain coach hingga batas waktu tertentu
- h. Mengikuti post-test.

Student memiliki peran selama pengerjaan , dengan deskripsi pekerjaan sebagai berikut:

Tabel 2.1 Tugas Student

Tugas	Deskripsi tugas
Membuat Project Cycle	Menentukan tema, mencari sumber data dan merapikan data
Membuat Model Data	<ul style="list-style-type: none"> • Membuat model menggunakan Algoritma KNN dengan n-neighbor 3 • Membuat model menggunakan Algoritma SVM dengan jenis kernel Linear
Melakukan Validasi Model	<ul style="list-style-type: none"> • Validasi model menggunakan Confusion matrix • Membuat Visualisasi dari hasil confusion matrix
Menyusun Paper	Menyusun Paper bersama

II.4 Jadwal Kerja

Program ini berlangsung setiap hari kerja (Senin sampai dengan Jumat) selama 8 jam per harinya, dengan rincian sebagai berikut:

Tabel 2.2 Agenda Kelas

Pukul (WIB)	Durasi (jam)	Aktivitas
08.00 s.d. 11.30	3.5	Kelas Sesi Pagi
13.00 s.d. 16.30	3.5	Kelas Sesi Siang
16.30 s.d. 17.30	1	<i>Self-Study</i>

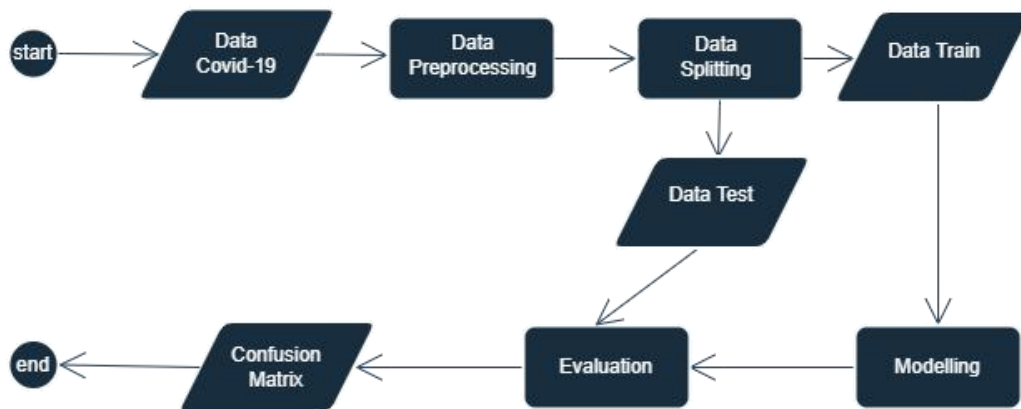
Program ini berlangsung dari bulan Februari 2022 sampai dengan bulan Juli 2022.

Bab III Prediksi Tingkat Risiko Penyebaran Covid-19 pada Pulau Sumatera Menggunakan Metode KNN dan SVM

III.1 Latar Belakang Proyek Akhir

Bidang *Data science* sangat berperan dalam membantu pemantauan COVID-19 di Indonesia. Sebagai salah satu negara yang terkena dampak bahayanya COVID-19, Indonesia melakukan penutupan diseluruh wilayah yang memiliki jumlah kasus positif yang cukup besar, salah satunya adalah wilayah Sumatera. Menurut data 30 April 2022 Pulau Sumatera, terdapat 811.180 orang terkonfirmasi positif, 24.621 orang meninggal, 784.869 orang sembuh total [2]. Dalam hal ini, dibutuhkan informasi prediksi yang akurat mengenai penyebaran COVID-19 di wilayah Sumatera. Oleh karena itu, penulis melakukan prediksi tingkat penyebaran COVID-19 di wilayah Sumatera. *Data Science* merupakan salah satu bidang studi yang mampu membuat prediksi atau *forecasting*. Metode ini digunakan untuk memprediksi masa depan berdasarkan tren masa lalu dan masa sekarang[3]. Dua Metode ini biasanya digunakan untuk pemecahan masalah pada bidang *Data Science*. Tujuan penelitian ini untuk melihat seberapa efektif penerapan metode KNN dan SVM untuk prediksi sebuah data penyebaran COVID-19 pada Pulau Sumatera.

III.2 Proses Pelaksanaan Proyek Akhir



Gambar 2.3 Flowchart Pengerjaan

Penelitian ini merupakan penelitian yang bertujuan untuk melihat seberapa efektif penerapan algoritma KNN dan SVM dalam memprediksi data penyebaran COVID-19 di Pulau Sumatera. Dalam pelaksanaannya dilalui beberapa proses atau alur pengerjaan seperti pada **Gambar 2.3** , berikut penjelasannya

1. Data Acquisition

Dengan menggunakan Open data, dalam penelitian ini data didapat melalui situs informasi dan koordinasi COVID-19 milik web Andra farm. Dataset tersebut berisikan informasi mengenai total kasus COVID-19, total meninggal, total sembuh dan total orang-orang yang masih sakit hingga saat ini di seluruh kota kabupaten di Pulau Sumatera[4].

2. Data Exploration

Pada tahap ini dilakukan seleksi atribut. Penggunaan atribut atau label pada prediksi ini hanya menggunakan 5 atribut yaitu atribut total kasus suspek, total kasus meninggal, masih sakit dan total sembuh sebagai

atribut yang dijadikan indikator prediksi, serta satu atribut identitas yaitu nama kabupaten/kota. Sedangkan untuk atribut lain yang mengandung angka-angka akumulasi penyebaran kasus COVID-19 ditiadakan karena tidak bersinggungan langsung dengan proses selanjutnya. kemudian dataset yang telah diseleksi tersebut di modifikasi dengan penambahan variabel tingkat risiko.

3. Modelling

Tahap selanjutnya adalah membuat model dengan algoritma KNN (*K-Nearest Neighbor*). KNN merupakan salah satu metode yang menerapkan algoritma *supervised* dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Kemudian dilanjutkan dengan membuat model SVM. SVM merupakan algoritma yang bekerja menggunakan pemetaan *nonlinear* untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi. Terdapat 8 model yang dibangun dengan jumlah *n-neighbor* serta kernel yang berbeda. Model yang dibangun pada penelitian ini dapat dituliskan pada **Tabel 2.3** dan **Tabel 2.4**

Tabel 2.3 Model Algoritma KNN

Algoritma	JUMLAH K
KNN	K=3
KNN	K=4

KNN	K=5
KNN	K=6

Tabel 2.4 Model Algoritma SVM

Algoritma	Kernel
SVM	Linear
SVM	Sigmoid
SVM	Polynomial
SVM	RBF

4. Evaluation

Pada Tahapan evaluasi ini KNN dan SVM digunakan untuk membentuk suatu prediksi kelompok suatu masalah dengan akurasi yang tepat. Matriks evaluasi yang digunakan terdiri dari *Accuray*, *Precission*, *Recall* dan *f1-Score*.

Rumus: $Accuracy = (TP+TN) / (TP+FP+FN+TN)$

Rumus : $Precision = (TP) / (TP + FP)$

Rumus : $Recall = TP / (TP + FN)$

Rumus: $F-1\ Score = (2 * Recall * Precision) / (Recall + Precision)$ [5]

III.3 Hasil Proyek Akhir

3.3.1 Pembahasan Algoritma KNN

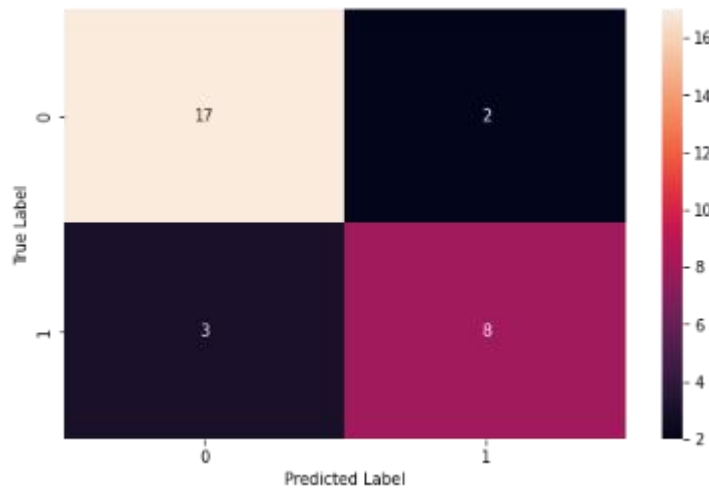
Setelah melakukan proses *split* dan normalisasi data, langkah selanjutnya adalah melakukan pengujian dari hasil pemodelan menggunakan algoritma K-NN dan SVM. Penulis akan melakukan pengujian dengan jumlah *n neighbors* yang berbeda untuk algoritma K-NN dan pengujian terhadap 4 jenis kernel dari algoritma SVM. Maka dari itu hasil dan pembahasan dalam penelitian ini terbagi menjadi ke beberapa skema, yaitu:

3.3.1.1 Hasil Pengujian Model KNN dengan *n neighbor* 3

Pada skenario pertama digunakan algoritma K-NN dengan jumlah *n neighbor* 3. Dari pemodelan ini, melalui performansi matrix didapatkan hasil akurasi sebesar 0.833 atau 83%. Hal ini menunjukkan prediksi yang telah dilakukan memiliki tingkat keakuratan yang baik. Kemudian, diikuti dengan nilai *Precision* sebesar 83%, nilai *recall* sebesar 83% dan *f1_score* sebesar 83%. Hasil dari *confusion matrix* model KNN dengan jumlah *n neighbor* 3 dapat dilihat di **Gambar 2.4**.

	precision	recall	f1-score	support
0	0.85	0.89	0.87	19
1	0.80	0.73	0.76	11
accuracy			0.83	30
macro avg	0.82	0.81	0.82	30
weighted avg	0.83	0.83	0.83	30

Gambar 2.4 Nilai Performansi Matriks



Gambar 2.5 Visualisasi confusion matrix

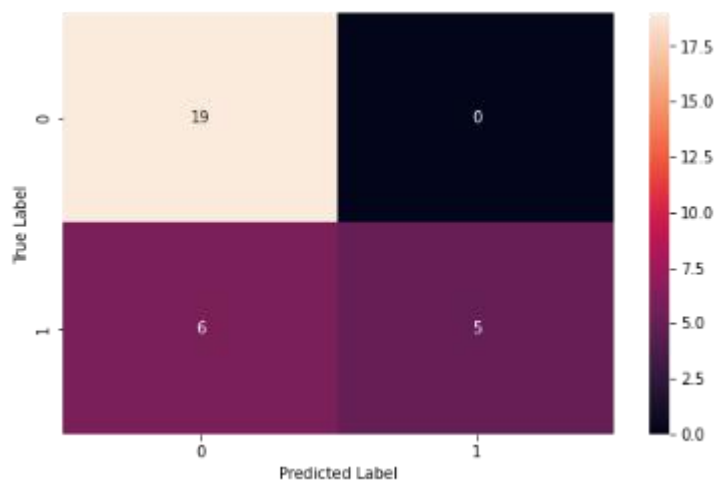
Berdasarkan visualisasi confusion matrix dari **Gambar 2.5**, dari 20 kota tingkat resiko rendah atau 0, model memprediksi terdapat 3 kota yang diprediksi tingkat resiko tinggi (FN). Sedangkan dari 10 kota tingkat resiko tinggi atau 1, model memprediksi terdapat 2 kota yang tingkat resiko rendah.

3.3.1.2 Hasil Pengujian Model KNN dengan n neighbor 4

Pada skenario kedua digunakan algoritma KNN dengan jumlah n Neighbor 4. Tahap selanjutnya adalah melakukan validasi model performansi matrix dan didapatkan hasil akurasi sebesar 0.8 atau 80%. Kemudian, diikuti dengan nilai *Precision* sebesar 85%, nilai *recall* sebesar 80% dan *f1-score* sebesar 78%. Hasil dari confusion matrix model KNN dengan jumlah n neighbor 4 dapat dilihat pada **Gambar 2.6**.

	precision	recall	f1-score	support
0	0.76	1.00	0.86	19
1	1.00	0.45	0.62	11
accuracy			0.80	30
macro avg	0.88	0.73	0.74	30
weighted avg	0.85	0.80	0.78	30

Gambar 2.6 Nilai Performansi Matriks

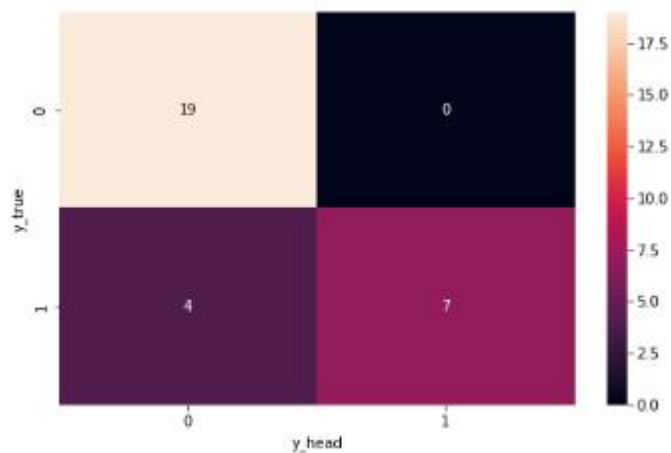


Gambar 2.7 Visualisasi Confusion Matrix

Berdasarkan visualisasi confusion matrix dari **Gambar 2.7**, dari 25 kota tingkat resiko rendah, model memprediksi terdapat 6 kota yang diprediksi masuk ke tingkat resiko tinggi (FN). Sedangkan dari 5 kota tingkat resiko tinggi, model memprediksi terdapat 0 kota yang tingkat resiko rendah dimana prediksi tersebut sesuai dengan *true value*-nya yaitu 5.

3.3.1.3 Hasil Pengujian Model KNN dengan n neighbor 5

Hasil pengujian model KNN dengan n Neighbor berjumlah 5 dapat dilihat di **Gambar 2.8**. Dengan ini dari 25 kota tingkat resiko rendah dan tinggi diprediksi ada 4 kota yang tinggi (FN), sedangkan dari 7 Kota tingkat resiko tinggi, model prediksi terpadat 0 kota serta prediksinya benar.



Gambar 2.8 Visualisasi Confusion Matrix

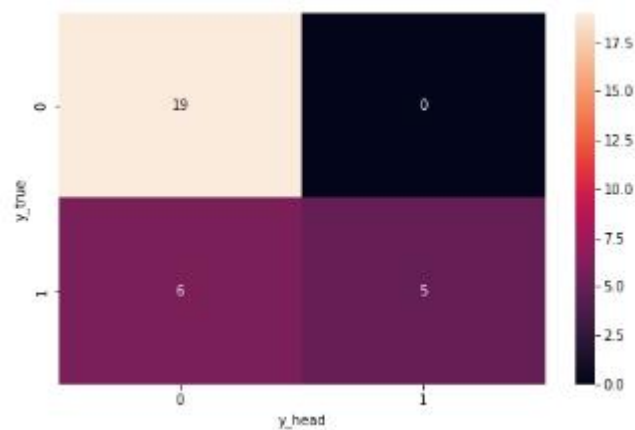
	precision	recall	f1-score	support
0	0.83	1.00	0.90	19
1	1.00	0.64	0.78	11
accuracy			0.87	30
macro avg	0.91	0.82	0.84	30
weighted avg	0.89	0.87	0.86	30

Gambar 2.9 Nilai Performansi Matriks

Berdasarkan hasil dari **Gambar 2.9** menunjukkan bahwa nilai akurasi mencapai 87%. Diikuti dengan nilai *precision* yang cukup tinggi yaitu 89%, nilai *recall* 87% dan *f1-score* bernilai 86%.

3.3.1.4 Hasil Pengujian Model KNN dengan n neighbor 6

Pada skenario keempat dengan menggunakan algoritma K-NN dengan jumlah n Neighbor 6. Jika dilihat hasil dari confusion matrix pada **Gambar 2.10**, dari 25 kota dengan tingkat resiko rendah atau 0, model algoritma memprediksi 6 kota yang masuk ke kategori tingkat resiko tinggi (FN). Sedangkan dari 5 kota dengan tingkat resiko tinggi, model algoritma memprediksi terdapat 0 kota yang tingkat resiko rendah atau prediksinya benar sempurna.



Gambar 2.10 Visualisasi Confusion Matrix

	precision	recall	f1-score	support
0	0.76	1.00	0.86	19
1	1.00	0.45	0.62	11
accuracy			0.80	30
macro avg	0.88	0.73	0.74	30
weighted avg	0.85	0.80	0.78	30

Gambar 2.11 Nilai Performansi Matriks

Berdasarkan nilai ketepatan prediksi **Gambar 2.11** maka dapat diketahui bahwa akurasi menghasilkan nilai 80%, diikuti dengan nilai precision 85%, recall bernilai 80% dan f1-score sebesar 78%.

3.3.2 Pembahasan Algoritma SVM

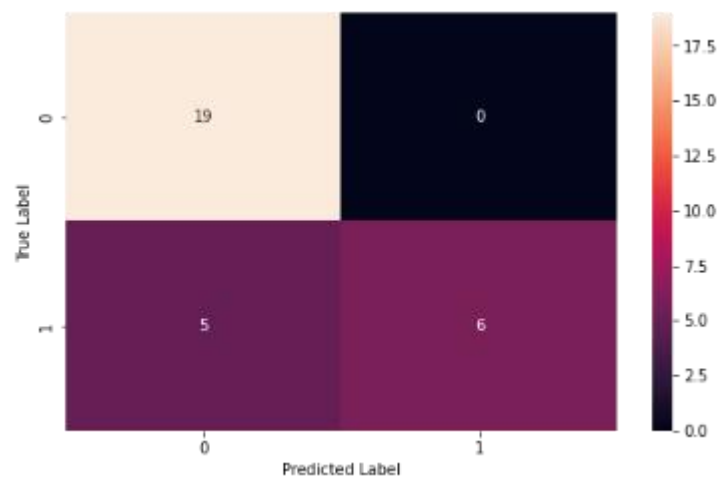
3.3.2.1 Hasil Pengujian Model SVM dengan Jenis Kernel Linear

Skema selanjutnya merupakan pengujian model prediksi menggunakan algoritma SVM. Pada tahap ini jenis kernel yang digunakan merupakan kernel linear. Melalui performa matrik didapatkan hasil akurasi sebesar 0.83 atau 83%. Kemudian, diikuti dengan nilai *Precision* sebesar 87%, nilai *recall* sebesar 83% dan *f1_score* sebesar 82%. Hasil dari confusion matrix model SVM dengan jenis kernel linear dapat dilihat di **Gambar 2.12**.

```
print(classification_report(Y_test, y_pred_svm))
```

	precision	recall	f1-score	support
0	0.79	1.00	0.88	19
1	1.00	0.55	0.71	11
accuracy			0.83	30
macro avg	0.90	0.77	0.79	30
weighted avg	0.87	0.83	0.82	30

Gambar 2.12 Nilai Performansi Matriks



Gambar 2.13 Visualisasi Confusion Matrix

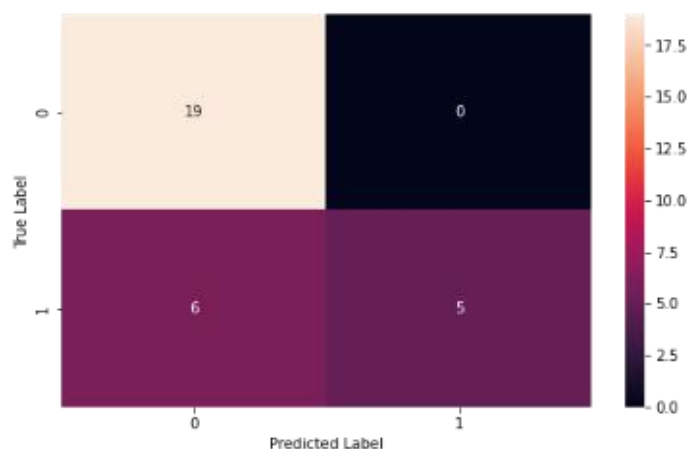
Berdasarkan **Gambar 2.13** Confusion matriks di atas , menunjukkan bahwa dari kota 24 kota dengan tingkat resiko rendah, model svm memprediksi terdapat 5 kota yang diprediksi masuk ke tingkat resiko tinggi (FN). Sedangkan dari 6 kota dengan tingkat resiko tinggi, model memprediksi dengan sempurna bahwa ke 6 kota tersebut termasuk ke tingkat resiko tinggi.

3.3.2.2 Hasil Pengujian Model SVM dengan Jenis Kernel Sigmoid

Pada tahap ini jenis kernel yang digunakan merupakan kernel Sigmoid. Melalui performa matrik didapatkan hasil akurasi sebesar 0.80 atau 80%. Kemudian, diikuti dengan nilai *Precision* sebesar 85%, nilai *recall* sebesar 80% dan *f1_score* sebesar 78%. Hasil dari confusion matrix model SVM dengan jenis kernel Sigmoid dapat dilihat di **Gambar 2.14**

	precision	recall	f1-score	support
0	0.76	1.00	0.86	19
1	1.00	0.45	0.62	11
accuracy			0.80	30
macro avg	0.88	0.73	0.74	30
weighted avg	0.85	0.80	0.78	30

Gambar 2.14 Nilai Performansi Matriks



Gambar 2.15 Visualisasi Confusion Matrix

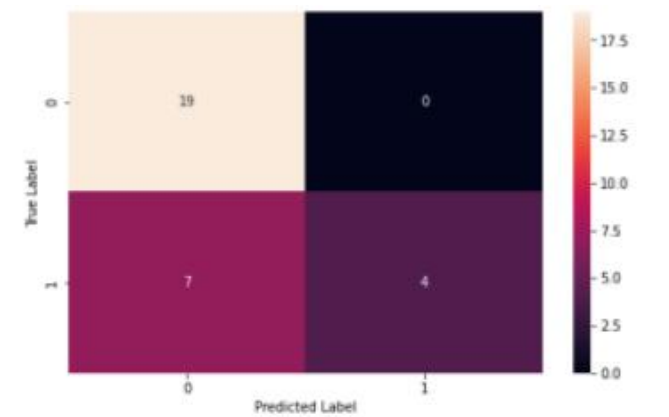
Berdasarkan **Gambar 2.15** confusion matriks di atas, menunjukkan bahwa dari 25 kota dengan tingkat resiko rendah, model svm memprediksi terdapat 6 kota yang diprediksi masuk ke tingkat resiko tinggi (FN). Sedangkan dari 5 kota dengan tingkat resiko tinggi, model memprediksi dengan sempurna bahwa ke 5 kota tersebut termasuk ke tingkat resiko tinggi.

3.3.2.3 Hasil Pengujian Model SVM dengan Jenis Kernel Polynomial

Kernel Polynomial memiliki fungsi kernel digunakan dengan data yang terpisah secara linear. Hasil *confusion matrix* dalam tahap ini menunjukkan bahwa akurasi bernilai sebesar 0.77 atau 77%. Kemudian, diikuti dengan nilai *Precision* sebesar 83%, nilai *recall* sebesar 77% dan *f1_score* sebesar 73%. Hasil dari *confusion matrix* model SVM dengan jenis kernel Sigmoid dapat dilihat di **Gambar 2.16**

	precision	recall	f1-score	support
0	0.73	1.00	0.84	19
1	1.00	0.36	0.53	11
accuracy			0.77	30
macro avg	0.87	0.68	0.69	30
weighted avg	0.83	0.77	0.73	30

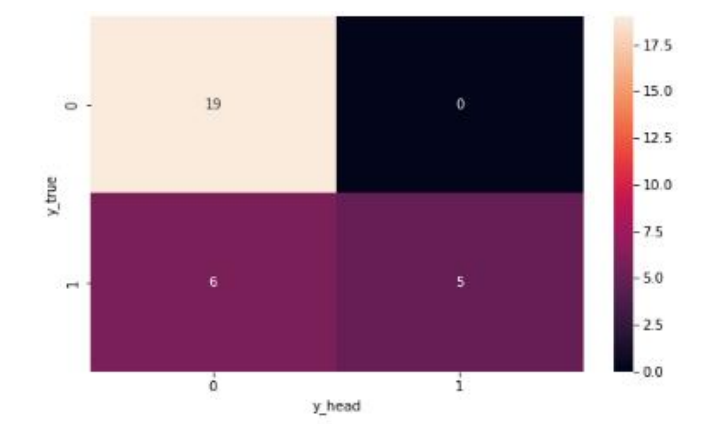
Gambar 2.16 Nilai Performansi Matriks



Gambar 2.17 Visualisasi Confusion Matrix

3.3.2.4 Hasil Pengujian Model SVM dengan Jenis Kernel RBF

Untuk pemodelan algoritma SVM dengan jenis kernel RBF, jika dilihat hasil dari *confusion matrix* pada **Gambar 2.18** dari 25 kota dengan tingkat resiko rendah, model algoritma ini menghasilkan bahwa terdapat 6 kota yang diprediksi masuk ke kategori tingkat risiko tinggi (FN). Sedangkan dari 5 kota tingkat resiko tinggi, model algoritma ini memprediksi dengan sempurna yang dimana prediksinya sesuai dengan nilai aslinya.



Gambar 2.18 Visualisasi Confusion Matrix

	precision	recall	f1-score	support
0	0.76	1.00	0.86	19
1	1.00	0.45	0.62	11
accuracy			0.80	30
macro avg	0.88	0.73	0.74	30
weighted avg	0.85	0.80	0.78	30

Gambar 2.19 Nilai Performansi Matriks

Skema pengujian model prediksi menggunakan algoritma SVM jenis RBF. Melalui performa matrik didapatkan hasil akurasi sebesar 0.80 atau 80%. Kemudian, diikuti dengan nilai Precision sebesar 85%, nilai recall sebesar 80% dan f1_score sebesar 78%.

Tabel 2.5 Evaluasi Matrix

Algoritma	Parameter	Nilai	Akurasi	Precision	Recall	F1-Score
KNN	n-Neighbor	3	83%	83%	83%	83%
		4	80%	85%	80%	78%
		5	87%	89%	87%	86%
		6	80%	85%	80%	78%
SVM	Kernel	Linear	83%	87%	83%	82%
		Sigmoid	80%	85%	80%	78%
		Poly	77%	83%	77%	73%
		RBF	80%	85%	80%	78%

Dari proses pengujian menunjukkan bahwa model prediksi memperlihatkan bahwa tingkat akurasi yang dimiliki oleh KNN model dengan jumlah n neighbor 5 berada pada ambang akurasi yang terbaik yaitu sekitar 87%. Sedangkan nilai akurasi paling rendah dihasilkan oleh SVM model dengan jenis kernel Polynomial.

Bab IV Penutup

IV.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, kesimpulan yang didapat yaitu sebagai berikut:

- a. Hasil dari pengujian model prediksi memperlihatkan bahwa tingkat akurasi yang dimiliki oleh KNN model dengan jumlah n neighbor 5 berada pada ambang akurasi yang terbaik yaitu sekitar 87%.
- b. Hasil dari pengujian model prediksi dengan menggunakan SVM model, menunjukkan bahwa pengujian dengan jenis kernel linear memiliki hasil akurasi terbaik yaitu sekitar 83%.
- c. Dari kedua model yang diuji, dapat disimpulkan bahwa penggunaan model KNN lebih efektif dalam memprediksi tingkat penyebaran COVID-19 di Pulau Sumatera.

IV.2 Saran

- a. Perlu pembahasan yang lebih dalam dan detail dalam proyek.
- b. Disarankan untuk mempelajari dan mendalami teori atau studi literatur terlebih dahulu.

Bab V Referensi

[1]	Gunawan, Kevin, “Rancangan Bangun Sistem Pakar untuk Mendeteksi Penyebab Kerusakan pada Ban Kendaraan dengan Algoritma C4.5”, FTI UMN, 2016.
[2]	Farm Andra, 2022, “Data seputar kasus perkembangan COVID-19 (virus Corona) di Pulau Sumatera per kota / kabupaten pada 30 April 2022,” 2022. https://www.andrafarm.com/_andra.php?_i=daftar-co19-
[3]	algoritma. (2022). “DATA MINING: PROSES PENTING DALAM DATA SCIENCE”[Online]. Tersedia: https://algorit.ma/blog/data-science/data-mining/
[4]	Farm Andra. (2022). “Tabel Harian Korona Sumatera - Sekarang ”[online]. Tersedia: https://www.andrafarm.com/_andra.php?_i=daftar-co19cari&jobs=sumatera#posisiurut
[5]	Dr. Anggreany, Maria. (2020). “Confusion Matrix”.[online]. Tersedia: https://socs.binus.ac.id/2020/11/01/confusion-matrix/

Bab VI Lampiran A. TOR

TERM OF REFERENCE (TOR) STUDI INDEPENDEN BERSERTIFIKAT FOUNDATION OF AI AND LIFE SKILLS FOR GEN-Z DI ORBIT FUTURE ACADEMY

A. Rincian Program

Foundation of AI and Life Skills for Gen-Z adalah program pelatihan *Artificial Intelligence* (AI) daring yang bertujuan untuk memperkenalkan teknologi dan perangkat AI kepada pelajar, sehingga diharapkan mereka dapat mengembangkan produk AI yang memiliki dampak sosial. Program ini berfokus pada komponen utama AI, seperti Data Science (DS), Natural Language Processing (NLP), dan Computer Vision (CV). Selain keterampilan AI, pelajar juga akan mendapat *life skills* yang bermanfaat untuk mencari atau menciptakan lapangan kerja.

B. Tujuan Program

Tujuan yang diharapkan setelah peserta mengikuti program ini:

1. Memiliki wawasan tentang AI dan perkembangannya.
2. Mampu merancang dan mengimplementasikan AI Project Cycle.
3. Mampu menggunakan bahasa pemrograman Python untuk mengembangkan aplikasi AI.
4. Mampu mengembangkan salah satu dari 3 domain AI (DS, NLP, dan CV) hingga tahap *deployment*.
5. Mampu menggunakan *soft skills* dan *hard skills* dalam dunia industri dan lingkungan perusahaan.
6. Mampu mengaplikasikan kiat-kiat yang dibutuhkan seorang wirausahawan yang bergerak di bidang *start-up* dalam mentransformasikan ide ke dalam bentuk produk/jasa sehingga dapat menciptakan peluang bisnis yang terus berinovasi, berevolusi, dan berkelanjutan.

C. Jadwal dan Tempat Pelaksanaan

Jadwal pelaksanaan tertera dalam tabel berikut:

Pukul (WIB)	Durasi (jam)	Aktivitas
08.00 s.d. 11.30	3.5	Kelas Sesi Pagi
13.00 s.d. 16.30	3.5	Kelas Sesi Siang
16.30 s.d. 17.30	1	<i>Self-Study</i>

Kelas akan diselenggarakan secara daring melalui aplikasi *video conference*.

D. Peserta

Peserta program ini adalah mahasiswa yang berasal dari Perguruan Tinggi di bawah Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia.

E. Uraian Tugas Peserta

Selama mengikuti program ini, peserta diharuskan:

1. Mengikuti program dari awal hingga selesai.
2. Mematuhi aturan program.
3. Mematuhi aturan kelas yang dibuat bersama *homeroom* atau *domain coach*.
4. Mengikuti kelas dengan presensi minimal 85%.
5. Membuat laporan harian dan mingguan di *website* Kampus Merdeka.
6. Menyelesaikan Proyek Akhir (PA) beserta laporannya.

Homeroom Coach,



Angel Metanosa Afinda
NIP: 2201043

Bekasi, 21 Februari 2022

Peserta Program,



Fadhilah Nur Amaliah
52419118

Bab VII Lampiran B. Log Activity

Minggu/Tgl	Kegiatan	Hasil
M-10 / 25 Apr2022	Hari ini adalah hari pertama di jadwal untuk mengerjakan project akhir kelompok. Kegiatan hari ini, saya dan masing-masing teman kelompok saya adalah mencari judul untuk Project Akhir.	Judul Project Akhir
M-10 / 26 Apr 2022	Pada kegiatan hari ini, saya beserta anggota kelompok melanjutkan diskusi dan mencari jurnal-jurnal untuk dijadikan referensi.	Menemukan beberapa referensi
M-10 / 27 Apr 2022	Hari ini, saya berdiskusi dengan coach untuk membahas judul project yang saya dan anggota kelompok saya ambil.	ACC judul Project
M-10 / 28 Apr 2022	Tidak ada kegiatan	
M-10 / 29 Apr 2022	Tidak ada kegiatan	
M-11/02 Mei 2022	Libur nasional	
M-11/03 Mei 2022	Libur nasional	
M-11/04 Mei 2022	Libur nasional	
M-11/05 Mei 2022	Libur nasional	
M-11/06 Mei 2022	Libur nasional	
M-12/ 09 Mei 2022	Libur nasional	
M-12/ 10 Mei 2022	Bimbingan dengan Coach	Memutuskan data yang

	membahas tentang data	akan digunakan
M-12/ 11 Mei 2022	Tidak ada kegiatan	
M-12/ 12 Mei 2022	Berdiskusi dengan kelompok membahas dataset	fiksasi data, yaitu dataset pulau Sumatera Barat
M-12/ 13 Mei 2022	Diskusi di grup membahas dan mencoba kodingan	Error dalam kodingan dan masih kendala dengan dataset
M-13/ 16 Mei 2022	Pertemuan kelompok membahas dataset	Masih bermasalah dengan dataset pulau Sumatera Barat
M-13/ 17 Mei 2022	Pertemuan kelompok membahas lanjutan dataset	Dataset masih belum menemukan yang cocok
M-13/ 18 Mei 2022	Tidak ada kegiatan	
M-13/ 19 Mei 2022	Explore dataset	Menemukan dataset yang cocok
M-13/ 20 Mei 2022	Pertemuan kelompok untuk merapikan data	Data siap dipakai
M-14/23 Mei 2022	Pertemuan Kelompok untuk membahas perhitungan tingkat risiko	belum menghasilkan jawaban mengenai rumus perhitungan yang digunakan
M-14/24 Mei 2022	Pertemuan Kelompok untuk membahas lanjutan perihal rumus yang digunakan untuk perhitungan tingkat risiko	belum menghasilkan jawaban mengenai rumus perhitungan yang digunakan
M-14/25 Mei 2022	Melanjutkan diskusi mengenai perhitungan	Menemukan solusi untuk melakukan perhitungan

	data .	data
M-14/26 Mei 2022	Mencari Referensi lain untuk menjadi acuan dan mulai menggarap paper	Referensi paper
M-14/27 Mei 2022	Mencari Referensi tambahan untuk paper	Referensi paper
M-15/ 30 Mei 2022	Pertemuan kelompok untuk membahas perihal kodingan	sudah melakukan pembagian atribut data dan melakukan normalisasi data. Namun, pada tahap split data mengalami error
M-15/ 31 Mei 2022	Pertemuan kelompok melanjutkan pembahasan mengenai kodingan	belum menemukan solusi kodingan, masih mencari referensi lain.
M-15/ 01 Juni 2022	Mengulik dan mencari referensi untuk kodingan	telah mencapai tahap pemodelan
M-15/ 02 Juni 2022	Pertemuan kelompok untuk membahas kodingan dan melanjutkan paper	Pemodelan untuk prediksi menggunakan KNN telah selesai.
M-15/ 03 Juni 2022	Bimbingan dengan coach membahas perihal kodingan	Coach menyarankan untuk menambah dataset , menambah model dengan algoritma SVM dan validasi menggunakan confusion matrix.

M-16/ 06 Juni 2022	Mengulik dan mencari referensi untuk penambahan pemodelan baru menggunakan algoritma SVM	Pemodelan selesai
M-16/ 07 Juni 2022	Mengulik dan mencari referensi untuk melakukan validasi model	Validasi model selesai
M-16/ 08 Juni 2022	Membuat visualisasi confusion matrix	visualisasi confusion matrix
M-16/ 09 Juni 2022	Bimbingan dengan coach membahas kodingan terbaru	Coach menyetujui hasil koding prediksi
M-16/ 10 Juni 2022	Merapikan paper	Progres paper sekitar 89%
M-17/ 13 Juni 2022	Presentasi progress	coach memberikan beberapa revisi untuk ppt
M-17/ 14 Juni 2022	Merapikan PPT dan melanjutkan merapikan paper	paper selesai
M-17/ 15 Juni 2022	Bimbingan dengan coach untuk membahas paper	Masih terdapat beberapa poin yang perlu direvisi
M-17/ 16 Juni 2022	Pengerjaan Laporan Akhir	Laporan Akhir selesai
M-17/ 17 Juni 2022	Pengumpulan Laporan Akhir	

Bab VIII Lampiran C. Jurnal

1. Judul

“Prediksi Tingkat Risiko Penyebaran COVID-19 pada Pulau Sumatera menggunakan Metode KNN dan SVM”

2. Data penulis

- 1) Fadhilah Nur Amaliah
- 2) Ghifary Catur Achmad Dhany
- 3) Mohammad Luthfan Faohan
- 4) Naila Ameera Larasati
- 5) Nova Nurul Putri

3. Abstrak

Abstrak - COVID-19 merupakan virus yang menyebabkan infeksi saluran pernafasan. Penularan COVID-19 yang sangat cepat ini memiliki dampak yang luas bagi seluruh masyarakat dari negara di seluruh dunia. Bidang Data science sangat berperan dalam membantu pemantauan COVID-19 di Indonesia. Pada saat ini, forecasting yang merupakan salah satu penerapan data science berperan penting untuk memprediksi tingkat risiko penyebaran COVID-19. Sebagai salah satu negara yang terkena dampak bahayanya COVID-19, Indonesia melakukan penutupan diseluruh wilayah yang memiliki jumlah kasus positif yang cukup besar, salah satunya adalah wilayah Pulau Sumatera. Dalam hal ini, dibutuhkanannya informasi prediksi yang akurat mengenai penyebaran COVID-19 di wilayah Sumatera. Penelitian ini bertujuan untuk melihat efektivitas penerapan model KNN dan SVM dalam memprediksi tingkat penyebaran COVID-19 di Pulau Sumatera. Pengukuran pengujian kedua model tersebut dapat dilihat melalui confusion matrix. Dimana nilai evaluasi matrix tertinggi yang berupa akurasi, precision, recall dan fl-score

masing-masing menghasilkan 87%, 89%, 87% dan 86% didapat dari penggunaan algoritma KNN dengan n neighbor 5.

Kata Kunci — COVID-19, KNN, SVM, Kernel, Pulau Sumatera, Prediksi.

4. Pendahuluan

Pada Desember tahun 2019, SARS-CoV-2 muncul pertama kali di wilayah Wuhan, Tiongkok. SARS-CoV-2 atau yang biasa dikenal dengan COVID-19 merupakan virus yang menyebabkan infeksi saluran pernafasan. Virus ini memiliki macam gejala, mulai dari gejala yang mirip dengan flu biasa hingga penyakit yang lebih parah seperti Middle East Respiratory Syndrome (MERS). COVID-19 dapat menyebar melalui percikan-percikan cairan tubuh yang keluar saat seseorang yang terinfeksi batuk, bersin atau berbicara. Penularan COVID-19 yang sangat cepat ini memiliki dampak yang luas bagi seluruh masyarakat dari negara di seluruh dunia. Di Indonesia, kasus COVID-19 muncul pertama kali pada awal Maret 2020. Sejak saat itu penyebaran COVID-19 meluas, hingga jumlah angka positif mencapai 4,7 juta jiwa dengan korban meninggal mencapai 132 ribu (Kemenkes RI, 2021)[1].

Saat ini banyak peneliti medis di seluruh dunia telah menemukan obat dan vaksin yang tepat untuk virus COVID-19. Pemerintah Indonesia juga cukup baik dalam menangani pencegahan virus ini, dengan berbagai cara mulai dari diberlakukannya protokol kesehatan, tes deteksi COVID-19 menggunakan PCR dan pemberian vaksin gratis. Bidang Data science juga sangat berperan dalam membantu pemantauan COVID-19 di Indonesia. Data Science merupakan salah satu bidang studi yang mampu membuat prediksi atau *forecasting*. Metode ini digunakan untuk memprediksi masa depan berdasarkan tren masa lalu dan masa sekarang[2]. Beberapa contoh *forecasting* diantaranya adalah peramalan cuaca, prediksi penyakit dan pasar saham. Pada saat ini, *forecasting* berperan penting untuk memprediksi tingkat risiko penyebaran COVID-19.

Sebagai salah satu negara yang terkena dampak bahayanya COVID-19, Indonesia melakukan penutupan diseluruh wilayah yang memiliki jumlah kasus positif yang cukup besar, salah satunya adalah wilayah Sumatera. Menurut data 30 April 2022 Pulau Sumatera, terdapat 811.180 orang terkonfirmasi positif, 24.621 orang meninggal, 784.869 orang sembuh total [2]. Dalam hal ini, dibutuhkannya informasi prediksi yang akurat mengenai penyebaran COVID-19 di wilayah Sumatera. Oleh karena itu, penulis melakukan prediksi tingkat penyebaran COVID-19 di wilayah Sumatera.

Dalam penelitian ini, penulis menggunakan metode algoritma KNN dan SVM untuk memprediksi tingkat risiko COVID-19 pada Pulau Sumatera. *K-Nearest Neighbors* (KNN) digunakan untuk mengklasifikasi sekumpulan data berdasarkan pembelajaran data. Salah satu ciri khusus algoritma ini adalah untuk mencari kasus dengan menghitung kedekatan kasus yang baru dengan kasus yang lama. Sedangkan *Support Vector Machine* (SVM) adalah algoritma klasifikasi yang sering digunakan dalam klasifikasi teks dan dapat menunjukkan penampilan yang baik dibanding algoritma lainnya. Dua Metode ini biasanya digunakan untuk pemecahan masalah pada bidang *Data Science*. Tujuan penelitian ini untuk melihat seberapa efektif penerapan metode KNN dan SVM untuk prediksi sebuah data penyebaran COVID-19 pada Pulau Sumatera. Oleh karena itu, peneliti melakukan komparasi hasil evaluasi dari algoritma KNN dan SVM pada data penyebaran COVID-19 Pulau Sumatera.

5. Studi Literatur

2.1 *K-Nearest Neighbors*

K-Nearest Neighbors atau K-NN melakukan klasifikasi dengan proyeksi data pembelajaran pada ruang berdimensi banyak. Ruang ini dibagi menjadi bagian-bagian yang merepresentasikan kriteria data pembelajaran. Setiap data pembelajaran direpresentasikan menjadi titik-titik c pada ruang dimensi banyak. K-NN adalah algoritma yang menghitung kemiripan pada tiap data uji dan semua data awal untuk menghitung daftar nearest neighbor-nya [3]. Konsep dasar dari K-NN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga terdekatnya dalam data pelatihan [4]. Teknik pencarian tetangga terdekat yang umum dilakukan dengan menggunakan formula jarak euclidean. Berikut beberapa formula yang digunakan dalam algoritma KNN

- *Euclidean Distance*

Jarak *Euclidean* adalah formula untuk mencari jarak antar titik dalam ruang dimensi [5].

- *Hamming Distance*

Jarak *Hamming* adalah cara mencari jarak antara 2 titik yang dihitung dengan panjang vektor biner yang dibentuk oleh dua titik tersebut dalam block kode biner.

- *Manhattan Distance*

Manhattan Distance atau Taxicab Geometri adalah formula untuk mencari jarak d antar 2 vektor p, q pada ruang dimensi n

- *Minkowski Distance*

Minkowski distance adalah formula pengukuran antara 2 titik pada ruang vektor normal yang merupakan hibridisasi yang menggeneralisasi *euclidean distance* dan *manhattan distance*.

Tujuan utama dari algoritma ini yaitu mengklasifikasikan suatu obyek berdasarkan atribut-atribut dan training sample. Algoritma K-Nearest Neighbor (K-NN) menggunakan klasifikasi kedekatan titik sebagai nilai perkiraan dari query instance yang baru. Jarak yang digunakan adalah jarak *Euclidean Distance* [6].

2.2 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) dikenalkan pertama kali oleh Vapnik tahun 1992 [7]. SVM adalah algoritma *machine learning* dengan pendekatan *supervised learning* yang dapat digunakan untuk klasifikasi (SVM *Classification*) dan regresi (SVM *Regression*). SVM dapat mengatasi masalah klasifikasi dan regresi dengan *linear* maupun *non-linear*. Cara kerja dari metode SVM khususnya pada masalah *non-linear* adalah dengan memasukkan konsep kernel ke dalam ruang berdimensi tinggi. Tujuannya adalah untuk mencari *hyperplane* atau pemisah yang dapat memaksimalkan jarak (*margin*) antar data kelas [8].

Dalam SVM terdapat beberapa jenis kernel yang sering digunakan, diantaranya adalah:

1. Kernel *Linear*, umum digunakan untuk dataset linear.
2. Kernel *Polynomial*, umum digunakan untuk data normal.

3. Kernel *Radial Basis Function* (RBF) atau *Gaussian*, umum digunakan untuk data *non-linear*.

4. Kernel *Sigmoid*, merupakan pengembangan dari jaringan syaraf tiruan.

2.3 Confusion Matrix

Confusion Matrix adalah tabel evaluasi informasi yang berisi perbandingan hasil prediksi klasifikasi dari data yang akan dievaluasi. Evaluasi ini dilakukan oleh sistem dengan nilai yang sebenarnya. Contoh *Confusion Matrix* dapat dilihat pada Tabel 1.

Tabel 1

	<u>Actual Values</u>	
	<u>1 (Positive)</u>	<u>0 (Negative)</u>
<u>Predicted Values</u>	TP (True Positive)	FP (False Positive) <u>Type I Error</u>
	FN (False Negative) <u>Type II Error</u>	TN (True Negative)

Confusion Matrix atau *Error Matrix*, membandingkan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya [9]. Untuk mengukur *performance metrics* dari *confusion matrix*, dapat digunakan *Accuracy*, *Precision* dan *Recall*.

- *Accuracy* (Q): Tingkat kedekatan nilai prediksi dengan nilai aktual (sebenarnya). Dari nilai akurasi, kita dapat mengetahui berapa persen kota yang benar diprediksi masuk kategori tingkat resiko rendah maupun tinggi.

Rumus: $Accuracy = (TP+TN) / (TP+FP+FN+TN)$

- *Precision* (PR): rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Dari nilai Precision, kita dapat mengetahui berapa persen kota/kabupaten yang benar masuk dalam tingkat penyebaran dengan resiko tinggi dari seluruh kota/kabupaten yang diprediksi masuk ke kategori resiko tinggi.

Rumus: $Precision = (TP) / (TP + FP)$

- *Recall* atau *Sensitivity* (SE): rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Dengan nilai *Recall*, kita dapat mengetahui berapa persen kota/kabupaten yang diprediksi masuk tingkat risiko tinggi dibandingkan keseluruhan kota/kabupaten yang sebenarnya masuk dalam risiko tinggi.

Rumus: $Recall = TP / (TP + FN)$

- *F1-Score* adalah parameter yang digunakan untuk menghitung matrik kinerja dari kombinasi *Precision* dan *Recall*.

Rumus: $F-1\ Score = (2 * Recall * Precision) / (Recall + Precision)$

6. Hasil dan Analisa

Setelah melakukan proses *split* dan normalisasi data, langkah selanjutnya adalah melakukan pengujian dari hasil pemodelan menggunakan algoritma K-NN dan SVM. Penulis akan melakukan pengujian dengan jumlah *n* neighbors yang berbeda untuk algoritma K-NN dan pengujian terhadap 4 jenis kernel dari algoritma SVM. Maka dari itu hasil dan pembahasan dalam penelitian ini terbagi menjadi ke beberapa skema, yaitu:

Tabel 2.6 Evaluasi Matrix

Algoritma	Parameter	Nilai	Akurasi	Precision	Recall	F1-Score
KNN	n-Neighbor	3	83%	83%	83%	83%
		4	80%	85%	80%	78%
		5	87%	89%	87%	86%
		6	80%	85%	80%	78%
SVM	Kernel	Linear	83%	87%	83%	82%
		Sigmoid	80%	85%	80%	78%
		Poly	77%	83%	77%	73%
		RBF	80%	85%	80%	78%

7. Kesimpulan

Hasil dari pengujian model prediksi memperlihatkan bahwa tingkat akurasi yang dimiliki oleh KNN model dengan jumlah *n* Neighbor 5 berada pada ambang akurasi yang terbaik yaitu 87%.

8. *Author Contributions Statement**

Penulis bertanggung jawab dalam penyusunan bagian Pendahuluan, Metodologi Penelitian dan Hasil Pembahasan.

9. Daftar pustaka

- [1] A. Aditia, “e-ISSN 2715-6885; p-ISSN 2714-9757 <http://jurnal.globalhealthsciencegroup.com/index.php/JPPP>,” *Jurnal Penelitian Perawat Profesiona*, vol. 3, no. 4, hlm. 8, 2021.

- [2] Algoritma. (2022). “DATA MINING: PROSES PENTING DALAM DATA SCIENCE”[Online]. Tersedia: <https://algoritma.blog/data-science/data-mining/>

- [3] T. S. Sabrila, V. R. Sari, dan A. E. Minarno, “Analisis Sentimen Pada Tweet Tentang Penanganan Covid-19 Menggunakan Word Embedding Pada Algoritma Support Vector Machine Dan K-Nearest Neighbor,” *FIJ*, vol. 6, no. 2, hlm. 69, Jul 2021, doi: 10.21111/fij.v6i2.5536.

- [4] M. Lestandy dan L. Syafa’ah, “PREDIKSI KASUS AKTIF COVID-19 MENGGUNAKAN METODE K-NEAREST NEIGHBORS,” *Seminar Nasional Teknologi dan Rekayasa (SENTRA) 2020*, hlm. 45–48, 2020.

- [5] D. M. U. Atmaja, A. R. Hakim, D. Haryadi, dan N. Suwaryo, “PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK PREDIKSI PENGELOMPOKKAN TINGKAT RISIKO PENYEBARAN COVID-19 JAWA BARAT,” vol. 1, hlm. 9, 2021.

- [6] A. Tanggu Mara, E. Sedyono, dan H. Purnomo, “Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba,” *JOINTER - J. of Informatics Engineering*, vol. 2, no. 01, hlm. 24–31, Jun 2021, doi: 10.53682/jointer.v2i01.30.
- [7] I. M. Parapat dan M. T. Furqon, “Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 10, hlm. 3163–3169, 2018.
- [8] I. A. Ropikoh, R. Abdulhakim, U. Enri, dan N. Sulistiyowati, “Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19,” *JAIC*, vol. 5, no. 1, hlm. 64–73, Jul 2021, doi: 10.30871/jaic.v5i1.3167.
- [9] Nugroho. K, “Confusion Matrix untuk Evaluasi Model pada Supervised Learning,” 2022. <https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-pada-supervised-machine-learning-bc4b1ae9ae3f>.
- [10] Farm Andra, 2022, “Data seputar kasus perkembangan COVID-19 (virus Corona) di Pulau Sumatera per kota / kabupaten pada 30 April 2022,” 2022. https://www.andrafarm.com/_andra.php?_i=daftar-co19-
- [11] X. Yayasan Multi Media, “Pahami Lebih Dalam untuk Siap Hadapi Industri Data,” 2020. <https://www.dqlab.id/data-analisis-machine-learning-untuk-proses-pengolahan-data>