

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 :

Ghifary Catur Achmad Dhany / 12119597



SISTEM INFORMASI & UNIVERSITAS GUNADARMA
2022

Lembar Pengesahan Sistem Informasi & Universitas Gunadarma

Foundations of AI and Life Skills for Gen-Z

Di Orbit Future Academy

oleh :

Ghifary Catur Achmad Dhany / 12119597

disetujui dan disahkan sebagai

Laporan Magang atau Studi Independen Bersertifikat Kampus Merdeka

Jakarta, 16 Juni 2022

Pembimbing Magang atau Studi Independen Sistem Informasi & Universitas
Gunadarma



Kuwat Setiyanto, S.Kom., MMSi

NIP: 100930

Lembar Pengesahan
Foundations of AI and Life Skills for Gen-Z
Di Orbit Future Academy

oleh :

Ghifary Catur Achmad Dhany / 12119597

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 melalui inovasi, pendidikan, dan pelatihan keterampilan. Dengan pelaksanaan kegiatan belajar mengajar pada studi independent mengedepankan AI sebagai pembelajaran serta luaran dari MSIB Batch 2 ini dan mengambil tema untuk Project Akhir dengan data science dan judul Prediksi Tingkat Resiko Penyebaran COVID-19 pada Pulau Sumatera menggunakan Metode KKN dan SVM. Jadi dari penelitian Project Akhir kami 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 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 Syukur penulis panjatkan atas kehadiran Allah SWT yang telah memberikan karunia serta rahmat hidayahnya dan memberikan kekuatan serta kemudahan kepada penulis, sehingga penulis dapat menyelesaikan tugas proyek akhir ini yang berjudul “Prediksi Tingkat Risiko Penyebaran COVID-19 pada Pulau Sumatera menggunakan Metode KNN dan SVM”. Maksud yang mendasari penulisan laporan akhir ini adalah selain sebagai salah satu syarat untuk kelulusan program MSIB, laporan akhir ini juga sebagai salah satu persyaratan konversi nilai di kampus. Pada kesempatan ini penulis ingin mengucapkan terimakasih kepada pihak-pihak yang telah membantu penulis baik selama pengerjaan proyek akhir ini. Ucapan terimakasih penulis sampaikan diantaranya kepada:

1. Kedua orang tua penulis.
2. Coach Angel Metanosa Afinda sebagai homeroom coach dan pembimbing.
3. Fadhilah Nur Amaliah, Ghifary Catur Dhany,
Mohammad Luthfan Faohan, Naila Ameera Larasati, Nova Nurul Putri sebagai
partner

Penulis menyadari menyadari bahwa masih ada banyak kekurangan dalam penelitian ini baik secara materi maupun penulisannya. Karena itu kritik dan saran yang membangun sangat dibutuhkan oleh penulis demi membuat laporan ini menjadi lebih baik.

Daftar Isi

Bab I Pendahuluan	1
I.1 Latar belakang	1
I.2 Lingkup	2
I.3 Tujuan	3
Bab II Orbit Future Academy	4
II.1 Struktur Organisasi	4
II.2 Lingkup Pekerjaan	5
II.3 Deskripsi Pekerjaan	6
II.4 Jadwal Kerja	7
Bab III Prediksi Tingkat Risiko Penyebaran COVID-19 pada Pulau Sumatera menggunakan Metode KNN dan SVM	8
III.1 Latar Belakang Proyek Akhir	8
III.2 Proses Pelaksanaan Proyek Akhir	8
III.3 Hasil Proyek Akhir	10
Bab IV Penutup	19
IV.1 Kesimpulan	19
IV.2 Saran	19
Bab V Referensi	20
Bab VI Lampiran A. TOR	21
Bab VII Lampiran B. Log Activity	23
Bab VIII Lampiran C. Jurnal	27

Daftar Tabel

Bab I Pendahuluan	1
I.1 Latar belakang	1
I.2 Lingkup	2
I.3 Tujuan	3
Bab II Orbit Future Academy	4
II.1 Struktur Organisasi	4
Gambar 2.1 Logo Orbit Future Academy	4
Gambar 2.2 Struktur Organisasi OFA	5
II.2 Lingkup Pekerjaan	5
II.3 Deskripsi Pekerjaan	6
II.4 Jadwal Kerja	7
Tabel 2.1 Agenda Kelas	7
Bab III Prediksi Tingkat Risiko Penyebaran COVID-19 pada Pulau Sumatera menggunakan Metode KNN dan SVM	8
III.1 Latar Belakang Proyek Akhir	8
III.2 Proses Pelaksanaan Proyek Akhir	8
Gambar 3.1 Alur Penelitian	8
Gambar 3.2 Data Keseluruhan Pulau Sumatera	9
Gambar 3.3 Dataset Sumatera.csv	9
III.3 Hasil Proyek Akhir	10
Gambar 3.4 Nilai Performansi Matriks n Neighbor 3	11
Gambar 3.5 Visualisasi confusion matrix n Neighbor 3	11
Gambar 3.6 Nilai Performansi Matriks n Neighbor 4	12

Gambar 3.7 Visualisasi Confusion Matrix n Neighbor 4	12
Gambar 3.8 Visualisasi Confusion Matrix n Neighbor 5	13
Gambar 3.9 Nilai Performansi Matriks n Neighbor 5	13
Gambar 3.10 Visualisasi Confusion Matrix n Neighbor 6	14
Gambar 3.11 Nilai Performansi Matriks n Neighbor 6	14
Gambar 3.12 Nilai Performansi Matriks Kernel Linear	14
Gambar 3.13 Visualisasi Confusion Matrix Kernel Linear	15
Gambar 3.14 Nilai Performansi Matriks Kernel Sigmoid	15
Gambar 3.15 Visualisasi Confusion Matrix Kernel Sigmoid	16
Gambar 3.16 Nilai Performansi Matriks Kernel Polynomial	16
Gambar 3.17 Visualisasi Confusion Matrix Kernel Polynomial	17
Gambar 3.18 Visualisasi Confusion Matrix Kernel RBF	17
Gambar 19 Nilai Performansi Matriks Kernel RBF	17
Tabel 3.1 Hasil Algoritma dan Parameter	18
Bab IV Penutup	19
IV.1 Kesimpulan	19
IV.2 Saran	19
Bab V Referensi	20
Bab VI Lampiran A. TOR	21
Bab VII Lampiran B. Log Activity	23
Bab VIII Lampiran C. Jurnal	27

Daftar Gambar

Bab I Pendahuluan	1
I.1 Latar belakang	1
I.2 Lingkup	2
I.3 Tujuan	3
Bab II Orbit Future Academy	4
II.1 Struktur Organisasi	4
Gambar 2.1 Logo Orbit Future Academy	4
Gambar 2.2 Struktur Organisasi OFA	5
II.2 Lingkup Pekerjaan	5
II.3 Deskripsi Pekerjaan	6
II.4 Jadwal Kerja	7
Tabel 2.1 Agenda Kelas	7
Bab III Prediksi Tingkat Risiko Penyebaran COVID-19 pada Pulau Sumatera menggunakan Metode KNN dan SVM	8
III.1 Latar Belakang Proyek Akhir	8
III.2 Proses Pelaksanaan Proyek Akhir	8
Gambar 3.1 Alur Penelitian	8
Gambar 3.2 Data Keseluruhan Pulau Sumatera	9
Gambar 3.3 Dataset Sumatera.csv	9
III.3 Hasil Proyek Akhir	10
Gambar 3.4 Nilai Performansi Matriks n Neighbor 3	11
Gambar 3.5 Visualisasi confusion matrix n Neighbor 3	11
Gambar 3.6 Nilai Performansi Matriks n Neighbor 4	12
Gambar 3.7 Visualisasi Confusion Matrix n Neighbor 4	12

Gambar 3.8 Visualisasi Confusion Matrix n Neighbor 5	13
Gambar 3.9 Nilai Performansi Matriks n Neighbor 5	13
Gambar 3.10 Visualisasi Confusion Matrix n Neighbor 6	14
Gambar 3.11 Nilai Performansi Matriks n Neighbor 6	14
Gambar 3.12 Nilai Performansi Matriks Kernel Linear	14
Gambar 3.13 Visualisasi Confusion Matrix Kernel Linear	15
Gambar 3.14 Nilai Performansi Matriks Kernel Sigmoid	15
Gambar 3.15 Visualisasi Confusion Matrix Kernel Sigmoid	16
Gambar 3.16 Nilai Performansi Matriks Kernel Polynomial	16
Gambar 3.17 Visualisasi Confusion Matrix Kernel Polynomial	17
Gambar 3.18 Visualisasi Confusion Matrix Kernel RBF	17
Gambar 19 Nilai Performansi Matriks Kernel RBF	17
Tabel 3.1 Hasil Algoritma dan Parameter	18
Bab IV Penutup	19
IV.1 Kesimpulan	19
IV.2 Saran	19
Bab V Referensi	20
Bab VI Lampiran A. TOR	21
Bab VII Lampiran B. Log Activity	23
Bab VIII Lampiran C. Jurnal	27

Bab I Pendahuluan

I.1 Latar belakang

Direktorat Jenderal Pendidikan Tinggi (Ditjen Dikti) Kemendikbud Ristek telah mencanangkan Kampus Merdeka Belajar dengan delapan program, salah satunya Magang dan Studi Independen Bersertifikat (MSIB). Direktur Jenderal Pendidikan Tinggi (Dirjen Dikti) Nizam mengatakan program MSIB bertujuan untuk memberikan ruang bagi mahasiswa untuk mendapatkan pengalaman dalam memahami dunia profesional dan menciptakan tenaga kerja yang profesional. Magang Bersertifikat Kampus Independen adalah program magang yang dipercepat dan dipercepat dengan pengalaman belajar yang dirancang dengan baik. Pembelajaran mandiri di Sertifikasi Kampus Mandiri adalah pembelajaran di kelas yang dirancang dan disesuaikan dengan tantangan praktis yang dihadapi mitra/industri. Pada tahun 2021, 160 mitra industri akan terlibat dalam penerapan MSIB. Salah satu mitra industri yang menjadi anggota pelaksana MSIB adalah Orbit Future Academy (OFA), OFA sendiri

Salah satu mitra industri yang terlibat dalam implementasi MSIB adalah Orbit Future Academy (OFA), perusahaan yang didirikan pada tahun 2016 untuk meningkatkan kualitas hidup melalui inovasi, pendidikan, dan pelatihan keterampilan. Label atau merek Orbit merupakan kelanjutan dari warisan mendiagnose profesor. PhD. Inge. B.J. Habibie (Presiden ke-3 Republik Indonesia) dan istrinya Dr. Hasri Ainun Habibie. Mereka semua menjadi motor penggerak yang mendukung inovasi dan perkembangan teknologi pendidikan di Indonesia. OFA mengkurasi dan melokalisasi program/kursus internasional untuk meningkatkan atau melatih kembali kaum muda dan tenaga kerja untuk pekerjaan di masa depan. Hal ini sejalan dengan slogan OFA, "Skills for the Future of Work."

Orbit Future Academy (OFA) juga menawarkan beberapa program kursus bersertifikat industri contohnya seperti Kecerdasan Buatan, Kewirausahaan Startup, Keterampilan Pendidik, Literasi Digital, Kecakapan Hidup, dan banyak lagi. Program yang tersedia saat ini di OFA untuk usia 12-60 tahun dan program kursus ini juga ditujukan untuk pencarian kerja serta pencipta kerja dimasa depan. OFA sendiri merupakan mitra pendidikan pemerintah Indonesia, saat ini OFA juga memberikan program ke 170 lebih universitas dan meluncurkan program keterampilan nasional untuk sektor K12 Sekolah & Madrasah, yang ditujukan untuk menjangkau 10.000 sekolah, 1 juta guru, dan 5 juta siswa dalam jangka waktu 5 tahun.

I.2 Lingkup

Studi Independent Foundations of AI and Life Skills For Gen-Z diselenggarakan oleh Orbit Future Academy yang berfokus pada komponen utama AI seperti Data Science, Natural Language Processing, dan Computer Vision. Program ini dimulai pada 21 Februari 2022 sampai dengan 22 Juli 2022. Setelah menjalani serangkaian proses pembelajaran, seluruh peserta diharapkan dapat menghasilkan luaran berupa projek akhir dengan rincian sebagai berikut :

- 9 - 13 Mei : Tahap Data Acquisition, Reprocessing, dan Pemodelan awal
- 16 Mei - 20 Mei 2022 : Tahap Pemodelan, Evaluasi, dan Perancangan UI/UX
- 23 Mei - 3 Juni 2022 : Tahap Improvisasi, Deployment Project, serta Penyusunan Laporan
- 6 Juni - 10 Juni 2022 : Menyelesaikan laporan
- 13 Juni - 17 Juni 2022 : Submit laporan ke website KM

Selama program berlangsung dibagi menjadi dua tahap yaitu fase individu dan kelompok. Dua fase awal, peserta menjalankan program secara individu. Lalu tiga fase selanjutnya, peserta terlibat secara berkelompok sesuai dengan pilihan masing - masing untuk mulai mengerjakan proyek akhir tiap kelompok.

I.3 Tujuan

Adapun Tujuan dari pembuatan tugas akhir pada kajian ini adalah sebagai berikut

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 atau merancang tentang Data science
5. Mampu memecahkan permasalahan prediksi tingkat risiko pada pulau sumatera menggunakan algoritma KNN dan SVM.

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:

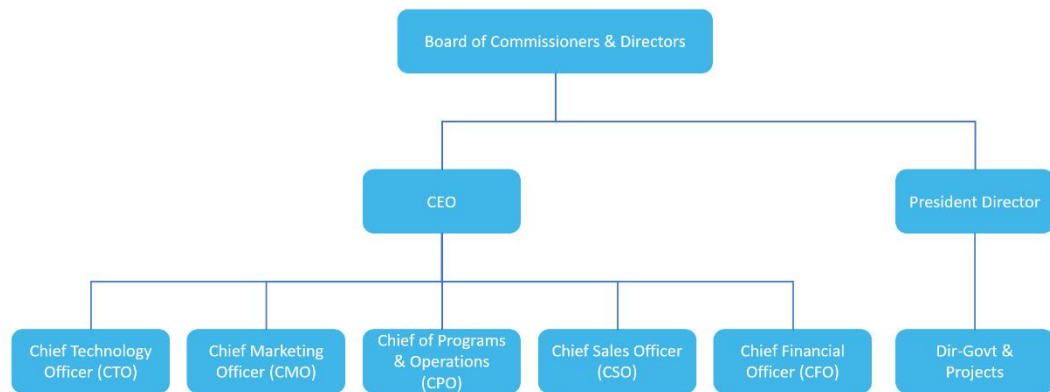
Memberikan pembelajaran berbasis keterampilan transformatif terbaik untuk para pencari kerja & pencipta lapangan kerja.

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 Normalisasi Data selama pengerjaan PA, dengan deskripsi pekerjaan sebagai berikut:

- a. Membuat Project Cycle (menentukan tema, mencari sumber data dan merapikan data).
- b. Melakukan model (Menggunakan algoritma KNN dengan n-neighbor 5 dan SVM dengan jenis kernel Polynomial).
- c. Normalisasi Data (Melakukan normalisasi data agar data yang digunakan tidak memiliki penyimpangan yang besar).
- d. Menyusun Paper.

II.4 Jadwal Kerja

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

Tabel 2.1 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

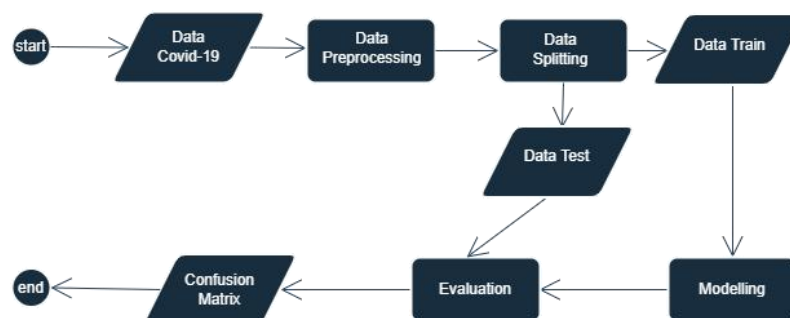
III.1 Latar Belakang Proyek Akhir

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. 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[11]. 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

3.2.1 Alur Penelitian

Gambar 3.1 Alur Penelitian



3.2.2 Data Acquisition

Dalam penelitian ini data yang digunakan didapat melalui situs informasi dan koordinasi COVID-19 milik web <https://www.andrafarm.com/>. Data yang digunakan merupakan data dengan format CSV dan dibuat manual.

No.	Negara, Provinsi, Kota, Kabupaten, Kecamatan, Kelurahan, Desa > (update 13 Juni, jam 22:41 WIB)	Kasus Positif (Konfirmasi) >	Meninggal >	Sembuh >	Masih Sakit, Positif Aktif, Dirawat >
1	Sumatera Pulau	811.667	24.661	786.783	223
2	Sumatera Barat (Sumbar) Provinsi	103.840	2.350	101.460	30
3	Sumatera Selatan (Sumsel) Provinsi	80.501	3.343	77.133	25
4	Sumatera Utara (Sumut) Provinsi	155.104	3.257	151.793	54

Gambar 3.2 Data Keseluruhan Pulau Sumatera

Berdasarkan data yang diambil dari *website* resmi dapat dilihat pada gambar di atas yang merupakan hasil dari kasus serta konfirmasi berupa positif, meninggal, sembuh dan juga sakit. Selain itu, pada penelitian ini juga terdapat penambahan variabel dengan tingkat risiko seperti yang tertera pada gambar 3 di bawah ini.

	kota_kabupaten	kasus_positif_total	total_meninggal	masih_sakit	total_sembuh	tingkat_risiko
0	Kota Padang	25.421	424.0	1.0	24.996	Resiko Tinggi
1	Kab. Agam	4.475	115.0	104.0	4.256	Resiko Tinggi
2	Kab. Tanah Datar	2.818	87.0	64.0	2.667	Resiko Rendah
3	Kota Bukittinggi	2.566	52.0	26.0	2.488	Resiko Rendah
4	Kab. Padang Pariaman	2.148	93.0	78.0	1.977	Resiko Tinggi
5	Kab. Pesisir Selatan	2.127	68.0	53.0	2.006	Resiko Rendah
6	Kab. Solok	2.052	70.0	52.0	1.930	Resiko Rendah
7	Kab. Limapuluh Kota	1.768	49.0	33.0	1.686	Resiko Rendah
8	Kota Payakumbuh	1.672	28.0	13.0	1.631	Resiko Rendah
9	Kota Padang Panjang	1.666	31.0	11.0	1.624	Resiko Rendah

Gambar 3.3 Dataset Sumatera.csv

3.2.3 Data Exploration

Data Exploration merupakan pendekatan yang mirip dengan analisis data awal, di mana seorang analis data menggunakan eksplorasi visual untuk memahami apa yang ada dalam dataset dan karakteristik data, daripada melalui sistem manajemen data tradisional.

Data Selection merupakan proses pemilihan data dari sekumpulan data operasional yang ada sebelum masuk ke tahap mining data maupun informasi. Pada tahap ini akan dilakukan langkah-langkah sebagai berikut:

A. Pivot Data

Data yang sudah dipivot sebanyak 150 data berdasarkan kota dan kabupaten serta mengakumulasi angka kasus dari masing-masing kabupaten dan kota terhadap penyebaran COVID-19 di Sumatera ini nantinya yang akan digunakan dalam proses data mining menggunakan algoritma *K-Nearest Neighbor*.

B. 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, dan 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.

3.2.4 Modeling

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.

3.2.5 Evaluation

Pada Tahapan evaluasi ini KNN dan SVM digunakan untuk membentuk suatu prediksi kelompok suatu masalah dengan akurasi yang tepat.

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

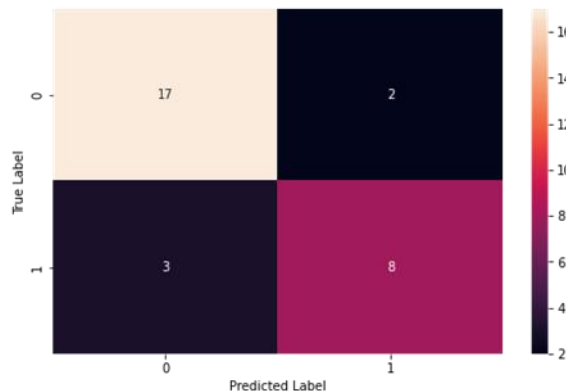
berbeda untuk algoritma K-NN dan pengujian terhadap 4 jenis kernel dari algoritma SVM. Maka dari itu hasil dan pembahasan dalam penelitian ini terbagi menjadi ke beberapa skema, yaitu:

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 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 3.4 Nilai Performansi Matriks n Neighbor 3



Gambar 3.5 Visualisasi confusion matrix n Neighbor 3

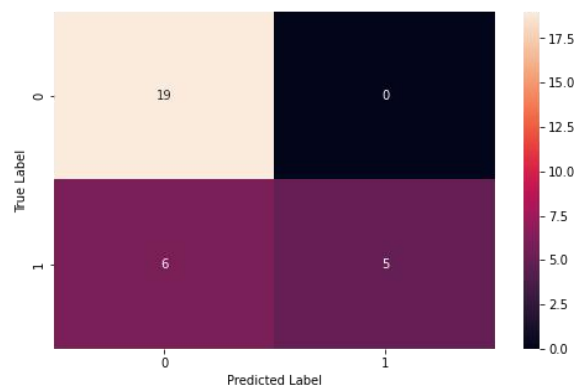
Berdasarkan visualisasi confusion matrix dari gambar 5, dari 20 kota tingkat resiko rendah atau 0, model memprediksi terdapat 3 kota yang diprediksi tingkat resiko tinggi (FN). Sedangkan dari 10 kota tingkat resiko tinggi atau 1, model memprediksi terdapat 2 kota yang tingkat resiko rendah.

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 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 3.6 Nilai Performansi Matriks n Neighbor 4



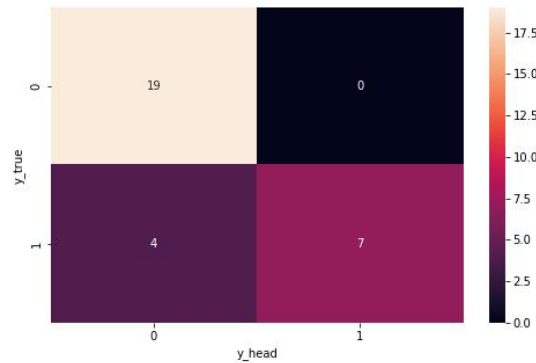
Gambar 3.7 Visualisasi Confusion Matrix n Neighbor 4

Berdasarkan visualisasi confusion matrix dari **Gambar 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 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 3.8 Visualisasi Confusion Matrix n Neighbor 5

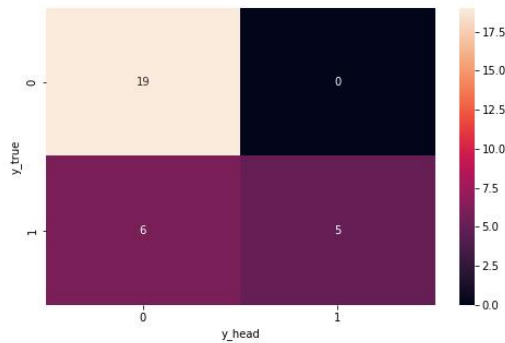
	precision	recall	f1-score	support
0	0.83	1.00	0.90	19
1	1.00	0.64	0.78	11
accuracy			0.87	30
macro avg	0.91	0.82	0.84	30
weighted avg	0.89	0.87	0.86	30

Gambar 3.9 Nilai Performansi Matriks n Neighbor 5

Berdasarkan hasil dari gambar 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 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 3.10 Visualisasi Confusion Matrix n Neighbor 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 3.11 Nilai Performansi Matriks n Neighbor 6

Berdasarkan nilai ketepatan prediksi gambar 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

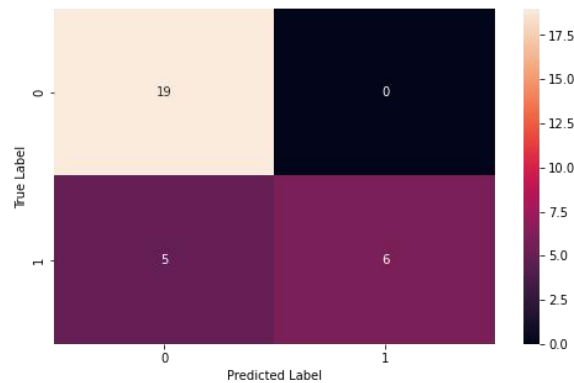
3.3.2.1 Hasil Pengujian Model SVM dengan Jenis Kernel Linear

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 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 3.12 Nilai Performansi Matriks Kernel Linear



Gambar 3.13 Visualisasi Confusion Matrix Kernel Linear

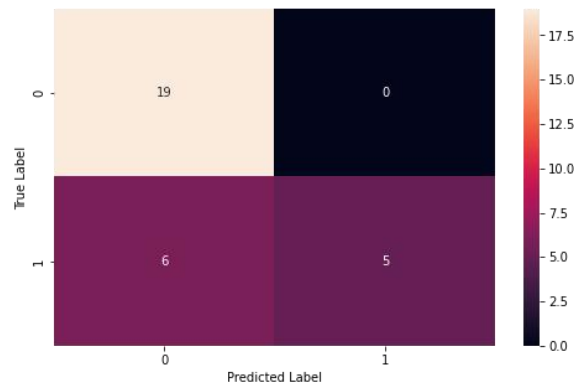
Berdasarkan **Gambar 13** Confusion matriks di atas , menunjukkan bahwa dari kota 24 kota dengan tingkat resiko rendah, model svm memprediksi terdapat 5 kota yang diprediksi masuk ke tingkat resiko tinggi (FN). Sedangkan dari 6 kota dengan tingkat resiko tinggi, model memprediksi dengan sempurna bahwa 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 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 3.14 Nilai Performansi Matriks Kernel Sigmoid



Gambar 3.15 Visualisasi Confusion Matrix Kernel Sigmoid

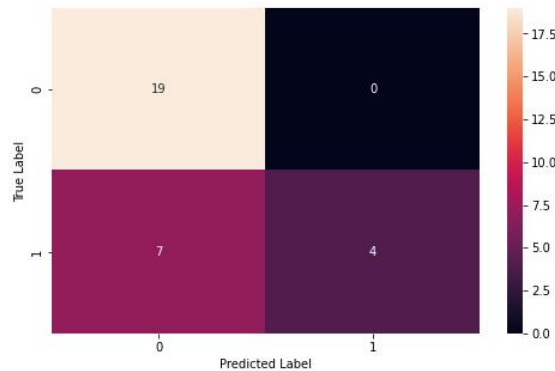
Berdasarkan **Gambar 15** confusion matriks di atas, menunjukkan bahwa dari 25 kota dengan tingkat resiko rendah, model svm memprediksi terdapat 6 kota yang diprediksi masuk ke tingkat resiko tinggi (FN). Sedangkan dari 5 kota dengan tingkat resiko tinggi, model memprediksi dengan sempurna bahwa ke 5 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 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

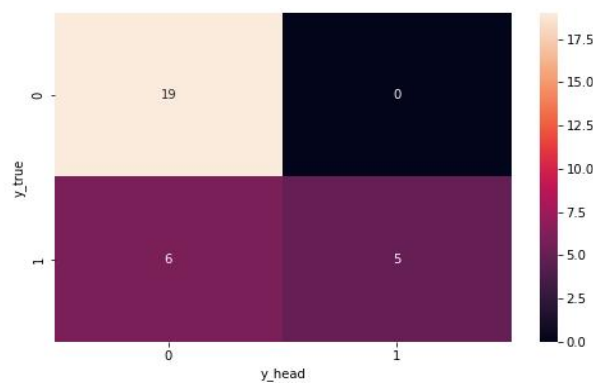
Gambar 3.16 Nilai Performansi Matriks Kernel Polynomial



Gambar 3.17 Visualisasi Confusion Matrix Kernel Polynomial

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 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 3.18 Visualisasi Confusion Matrix Kernel RBF

	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 19 Nilai Performansi Matriks Kernel RBF

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 3.1 Hasil Algoritma dan Parameter

Algo	Parameter	Nilai	Akurasi	Press	Rec	F1
SVM	Kernel	Linear	83%	87%	83%	82%
		Sig	80%	85%	80%	78%
		Poly	77%	83%	77%	73%
		RBF	80%	85%	80%	78%
KNN	n	3	83%	83%	83%	83%
		4	80%	85%	80%	78%
		5	87%	89%	87%	86%
		6	80%	85%	80%	78%

Bab IV Penutup

IV.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan oleh penulis, 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

Perlu pembahasan yang lebih dalam dan detail dalam proyek.

Disarankan untuk mempelajari dan mendalami teori atau studi literatur terlebih dahulu.

Bab V Referensi

- [1] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S. D. Langhans, M. Tegmark and F. F. Nerini, “The role of artificial intelligence in achieving the Sustainable Development Goals,” *NATURE COMMUNICATIONS*, pp. 1-10, 2020.

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,

Depok, 21 Februari 2022

Peserta Program,



Angel Metanosa Afinda S.Kom
2201043



Ghifary Catur Achmad Dhany
12119597

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	Tidak ada hasil
M-10/26 Apr 2022	Tidak ada kegiatan	Tidak ada hasil
M-11/02 Mei 2022	Libur Nasional	Tidak ada hasil
M-11/03 Mei 2022	Libur Nasional	Tidak ada hasil
M-11/04 Mei 2022	Libur Nasional	Tidak ada hasil
M-11/05 Mei 2022	Libur Nasional	Tidak ada hasil
M-11/06 Mei	Libur Nasional	Tidak ada hasil

2022		
M-12/ 09 Mei 2022	Libur Nasional	Tidak ada hasil
M-12/ 10 Mei 2022	Bimbingan dengan Coach membahas tentang data	Memutuskan data yang akan digunakan
M-12/ 11 Mei 2022	Tidak ada kegiatan	Tidak ada hasil
M-12/ 12 Mei 2022	Berdiskusi dengan kelompok membahas dataset	Fiksasi data, yaitu dataset pulau sumatera barat
M-12/ 13 Mei 2022	Diskusi di group membahas dan mencoba codingan	Error dalam codingan dan masih kendala dengan dataset
M-13/ 16 Mei 2022	Membahas lanjutan dataset	Masih bermasalah dengan dataset pulau Sumatera Barat
M-13/ 17 Mei 2022	Pertemuan kelompok membahas lanjutan dataset	Dataset masih belum menentukan data yang cocok
M-13/ 18 Mei 2022	Tidak ada kegiatan	Tidak ada hasil
M-13/ 19 Mei 2022	Mengeksplere dataset	Menentukan dataset yang cocok
M-13/ 20 Mei 2022	Pertemuan kelompok dan merapihkan dataset	Data siap digunakan
M-14/ 23 Mei 2022	Diskusi Bersama kelompok melalui google meet	
M-14/ 24 Mei 2022	Diskusi Bersama kelompok lewat google meet	
M-14/ 25 Mei 2022	Diskusi bersama kelompok untuk mencari hitungan pada data yang akan	

	kelompok ambil	
M-14/ 26 Mei 2022	Mencari referensi lain untuk menjadi acuan dan mulai menggarap paper sedikit-sedikit	
M-14/ 27 Mei 2022	Mencari referensi lain untuk menjadi acuan pada paper	Referensi paperbased semakin bervariasi
M-15/ 30 Mei 2022	Sudah mulai data acquisition dari dataset	Data acquisition
M-15/ 31 Mei 2022	Disukusi bersama kelompok melalui google meet membahas permodelan	Mencari Model
M-15/ 01 Jun 2022	Mencari referensi lain untuk menjadi acuan	Mencari referensi
M-15/ 02 Jun 2022	Diskusi bersama kelompok melalui google meet membahas permodelan	Fiksasi tentang model
M-15/ 03 Jun 2022	Konsultasi bersama coach mengenai isi kodingan, yang tadinya banyak permodelan KNN	Penambahan model algoritma KKN
M-16/ 06 Jun 2022	Membahas paper	Mendapatkan hasil paper selanjutnya
M-16/ 07 Jun 2022	Memulai mengcoding dengan dataset baru	Mengisi codingan dengan dataset baru
M-16/ 08 Jun 2022	Memulai menyicil paper	Menambah isi paper
M-16/ 09 Jun 2022	Bimbingan bersama coach angel membahas codingan dan isi paper	Mengubah dan menambahkan isi paper
M-16/ 10 Jun 2022	Melanjutkan mengerjakan paper	Menambah isi paper
M-17/ 14 Juni 2022	Melakukan presentasi dengan coach	Mengetahui progres sudah sampai mana
M-17/ 15	Coach merevisi paper	revisi

Juni 2022		
M-17/ 16 Juni 2022	Menyelesaikan laporan individu	Laporan individu selesai
M-17/ 17 Juni 2022	Menyelesaikan paper dan presentasi untuk laporan progres proyek coach	Paper selesai

*Log activity **dimulai dari pengerjaan Proyek Akhir (PA)**. Isi log activity bisa berbeda tergantung peran/tupoksi Anda dalam tim. Isi log activity bisa disamakan dengan isi laporan harian/mingguan Anda di web Kampus Merdeka.*

Bab VIII Lampiran C. Jurnal

1. Judul.

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

2. Data penulis.

Fadhilah Nur

Amaliah#1, Ghifary Catur Achmad Dhany#2, Mohammad Luthfan

Faohan#3, Naila Ameera Larasati*4, Nova Nurul Putri#5

mluthfanfaohan@upi.edu, novanurul@upi.edu,

nailaameera@student.gunadarma.ac.id,

ghifarycaturachmaddhany@student.gunadarma.ac.id ,

fadhilahamaliah@student.gunadarma.ac.id.

3. 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, dibutuhkannya 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 f1-score masing-

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

Keywords— 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[11]. 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.

A. 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 [2]. Konsep dasar dari K-NN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga terdekatnya dalam data pelatihan [3]. 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 [4].

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

- 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.

$$d_{ij} = q + r$$

- Minkowski Distance

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

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

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 [5].

B. Support Vector Machine (SVM)

Support Vector Machine (SVM) dikenalkan pertama kali oleh Vapnik tahun 1992 [6]. 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 [7].

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.

$$K(\vec{X}_i, \vec{X}_j) = (\vec{X}_i, \vec{X}_j + 1)^q$$

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

$$K(\vec{X}_i, \vec{X}_j) = \exp\left(\frac{\|\vec{X} - \vec{X}_i\|^2}{2\sigma^2}\right)$$

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

C. 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

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

Confusion Matrix atau Error Matrix, membandingkan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya [8]. 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.

$$\text{Rumus : Recall} = \text{TP} / (\text{TP} + \text{FN})$$

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

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

6. Hasil dan Analisa.

Algo	Parameter	Nilai	Akurasi	Press	Rec	F1
SVM	Kernel	Linear	83%	87%	83%	82%
		Sig	80%	85%	80%	78%
		Poly	77%	83%	77%	73%
		RBF	80%	85%	80%	78%
KNN	n	3	83%	83%	83%	83%
		4	80%	85%	80%	78%
		5	87%	89%	87%	86%
		6	80%	85%	80%	78%

7. 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.

8. Ucapan terima kasih. Mencantumkan sumber dana (lembaga, tahun kontrak) dan kepada siapa hibah dana tersebut diberikan beserta nama orang atau pihak yang membantu penelitian ini (jika ada).

9. *Author Contributions Statement** Spesifikasi kontribusi (tanggung jawab) dari penulis yang berbeda dalam *paper*. Pernyataan tersebut menjelaskan siapa dan sejauh mana tanggung jawab, misalnya, perumusan hipotesis, pemikiran tentang penelitian, pengumpulan data, analisis data, interpretasi data, dan penulisan artikel.

10. 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] 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.
- [3] 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.
- [4] D. M. U. Atmaja, A. R. Hakim, D. Haryadi, dan N. Suwaryo, “PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK PREDIKSI PENGELOMPOKAN TINGKAT RISIKO PENYEBARAN COVID-19 JAWA BARAT,” vol. 1, hlm. 9, 2021.
- [5] A. Tangu 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.
- [6] 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.
- [7] 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.
- [8] Nugroho. K, “Confusion Matrix untuk Evaluasi Model pada Supervised Learning,” 2022. <https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-pada-unsupervised-machine-learning-bc4b1ae9ae3f>.

- [9] 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-
- [10] 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>
- [11] algoritma. (2022). “DATA MINING: PROSES PENTING DALAM DATA SCIENCE”[Online]. Tersedia: <https://algoritma.blog/data-science/data-mining/>

Catatan:

Rekomendasi format Jurnal:

1.Format IEEE (English)([Link](#))

Link Template : <https://ieeaccess.ieee.org/wp-content/uploads/2020/04/Access-Template.doc>

- *free submission fee*

- format jurnal internasional yang sering digunakan

2. Format CST (English) ([Link](#))

Link Template : https://cst.kipmi.or.id/public/journals/1/CST_template.doc

- dikelola oleh orang Indonesia,

- no article processing charge (FREE),

- terindeks Scopus

3. Format JRSI dari Univ Telkom (Bahasa) ([Link](#))

Link Template : https://docs.google.com/document/d/1RV-96WQ_VMCv75fkkPzntRWhU438iOIF/edit

- tidak memungut biaya untuk pemrosesan artikel, pengiriman, review, hingga publikasi artikel.

- Sinta 4