

BAB II

TINJAUAN PUSTAKA

2.1 Virus Covid-19

Virus Corona (Covid-19) adalah keluarga besar dari virus yang menyebabkan penyakit yaitu manusia dan hewan. Penyebab penyakit pada manusia biasanya yaitu infeksi saluran pernapasan, mulai dari flu biasa hingga penyakit serius seperti Sindrom Pernafasan Akut Berat / *Severe Acute Respiratory Syndrome* (SARS) dan *Middle East Respiratory Syndrome* (MERS). Pada manusia virus Corona ditemukan saat kejadian luar biasa muncul di Wuhan Cina, Desember 2019. Kemudian diberi nama *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-COV2), dan menyebabkan penyakit *Corona virus Disease-2019* (COVID-19) [7].

Virus Corona mulai menyebar disekitar wilayah Wuhan dan kini telah mewabah lebih dari 100 negara. Lebih dari 100.000 orang di dunia dinyatakan positif terinfeksi virus ini. Dilaporkan jumlah kasus baru di China memang menurun. Namun kasus justru meningkat di Korea Selatan, Italia dan Iran. Wabah Corona semakin meluas di berbagai belahan dunia dan menjadi ancaman serius bagi perekonomian global [8].

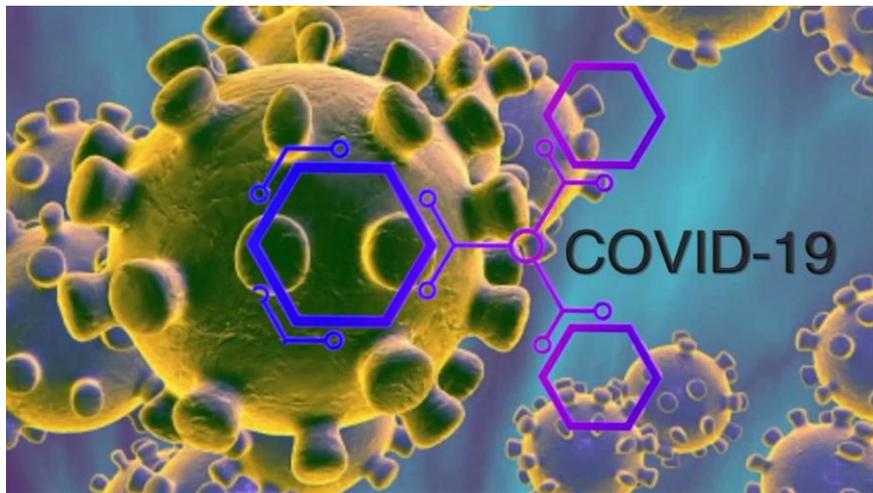
Saat ini Wabah *Corona Virus Disease* (Covid-19) menjadi isu kesehatan yang paling menghebohkan seluruh dunia. Penanggulangan disuatu negara pun dilakukan dalam upaya meminimalisir penyebaran penyakit seperti *lockdown* dan *social distancing*. Ditemukan bahwa diameter virus Corona diperkirakan mencapai 125 nanometer atau 0,125 mikrometer. Satu mikrometer sama dengan 1000 nanometer. Kecil sekali dan tak mungkin pandangan mata manusia mampu melihatnya. Sekalipun sangat kecil tetapi manusia yang tidak mau berpikir meremehkannya. Virus ini mampu bertahan lebih dari 10 menit di permukaan termasuk tangan. Bahkan WHO menyebut virus Corona baru (Covid-19) dapat bertahan selama beberapa jam, bahkan beberapa hari dan dapat bertahan hidup di suhu 26-27 derajat celcius. Ketika berita terkait virus Corona pertama kali terdengar, banyak negara-negara yang panik akan penyebaran virus tersebut, namun ada juga yang menanggapi dengan santai wabah virus ini [9].

Penyebaran virus Covid-19 sangat cepat, yaitu kontak fisik melalui hidung, mulut, dan mata, dan berkembang di paru. Adapun gejala atau tanda seseorang terkena Covid-19 adalah suhu tubuh naik, demam, mati rasa, batuk, sakit di tenggorokan, kepala pusing, susah bernafas apabila virus corona sudah mencapai paru-paru. Adapun gejala lain seperti nyeri otot, diare, kehilangan bau dan sakit perut. Pada tanggal 4 April 2020, lebih dari 1.100.000 kasus telah dilaporkan lebih dari 200 negara dan wilayah, mengakibatkan lebih dari 58.900 kematian. Lebih dari 226.000 orang telah pulih [10].

Berikut ada beberapa cara penularan virus Covid-19 yang sering terjadi sebagai berikut [11] :

1. Kontak dengan benda yang sering tersentuh. Benda merupakan media yang bisa menjadi cara penularan yang masif, sebab menurut penelitian virus corona Covid-19 dapat bertahan hidup hingga tiga hari dengan menempel pada permukaan benda. Dengan menempelnya virus tersebut di permukaan benda yang sering tersentuh, otomatis virus tersebut akan berpindah dan menemukan inang baru apabila orang lain menyentuhnya.
2. Tidak menjaga kebersihan tangan. Cara kedua yang efektif sebagai media penularan virus Covid-19 adalah tidak menjaga kebersihan tangan. Sebab, tangan adalah anggota tubuh yang paling banyak melakukan aktivitas dan melakukan interaksi dengan orang lain atau benda yang ada di sekitar. Apabila tangan tidak terjaga kebersihannya maka virus Covid-19 dapat mudah menyebar dengan cepat.
3. Tidak menjaga kebersihan setelah bepergian. Beraktivitas adalah hal yang wajar dilakukan oleh manusia namun penularan Covid-19 secara tidak langsung sering dilakukan oleh orang yang melakukan aktivitas di tempat tertentu apabila tempat tersebut terdapat droplet virus Covid-19 dan menempel pada pakaian dan benda yang digunakan sehingga dapat menular pada orang-orang terdekat dirumah.
4. Tidak menerapkan etika batuk dan bersin. Etika batuk dan bersin dapat dilakukan dengan menutup mulut dan hidung menggunakan siku bagian dalam atau tisu bersih bersih.

5. Terjadi interaksi dengan banyak orang. Berkumpul atau beraktivitas di tengah kerumunan menjadi salah satu cara penularan virus Covid-19 sebab virus ini dapat menempel secara tak kasat mata pada pakaian dan benda yang dibawa oleh orang lain yang juga dapat terjadi melalui droplet orang lain ketika batuk dan bersin.
6. Tidak isolasi diri setelah kembali dari wilayah pandemi. Cara lain yang dapat menularkan virus Covid-19 adalah tidak melakukan tindakan pencegahan setelah kembali dari wilayah atau negara pandemi. Cara ini diduga banyak terjadi di Indonesia. Mengingat saat ini banyaknya warga yang kembali ke kampung halaman namun tidak melakukan isolasi diri.



Gambar 2.1 Virus Covid-19

2.2 Implementasi

Implementasi merupakan proses umum tindakan yang dapat diteliti pada tingkat program tertentu. Implementasi menurut konsep dapat dibagi ke dalam dua bagian yaitu implementasi merupakan persamaan fungsi dari maksud, output dan outcome. Berdasarkan deskripsi tersebut, implementasi merupakan fungsi yang terdiri dari maksud dan tujuan, hasil sebagai produk, dan hasil dari akibat. Selanjutnya, implementasi merupakan persamaan fungsi dari kebijakan, formator, implementor, inisiator, dan waktu [12].

Implementasi adalah kegiatan yang dilakukan untuk menerapkan sistem yang diperoleh dari kegiatan seleksi dan menguji data. Jadi dapat disimpulkan

bahwa implementasi adalah proses untuk menguji antara konsep dengan konseptual atau konteks. Selanjutnya implementasi adalah suatu proses untuk melaksanakan suatu ide, program atau aktifitas baru dengan harapan orang lain dapat menerima dan melakukan perubahan. Berdasarkan pengertian diatas dapat disimpulkan juga bahwa implementasi adalah suatu proses untuk menilai, mengevaluasi dan mengukur apakah suatu peraturan atau kebijakan dapat berjalan dengan baik atau tidak, dengan begitu maka akan di nilai apakah harus ada evaluasi atau tidak terhadap program tersebut [13].

2.3 Dataset

Dataset adalah sebuah data yang diambil dari sumber data yang merepresentasikan data tabel dan relasinya di mana strukturnya mirip dengan data di *database* [14].

2.4 Data Mining

Data Mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Sehingga *data mining* menjadi alat yang semakin penting untuk mengubah data tersebut menjadi informasi. *Data mining* adalah teknik analisis data berbasis pada aplikasi statistik yang bertujuan untuk mengekstrak informasi. Dengan *data mining* kumpulan data dalam jumlah besar dapat dijadikan informasi lain yang bermanfaat. *Data mining* dapat melakukan pekerjaan seperti memperkirakan, mengklasifikasikan sampai mengelompokkan data [15].

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database (KDD)* secara keseluruhan [16].

Data mining adalah proses menemukan korelasi, pola, dan tren baru yang bermakna dengan memilah-milah sejumlah data besar yang disimpan dalam *repository*, menggunakan teknologi pengenalan pola serta teknik statistik dan

matematika [17].

Data mining memungkinkan pemakai menemukan pengetahuan dalam data database yang tidak mungkin diketahui keberadaanya oleh pemakai. *Data mining* adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan *knowledge base*. Tahap-tahap *data mining* adalah sebagai berikut :

1. Pembersihan data (*Data Cleaning*)
Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan.
2. Integrasi data (*Data Integration*)
Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru.
3. Seleksi data (*Data Selection*)
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.
4. Transformasi data (*Data Transformation*)
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam Data Mining.
5. Proses *Mining*
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan Data Mining.
6. Evaluasi pola (*Pattern Evaluation*)
Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.
7. Presentasi pengetahuan (*Knowledge Presentation*)
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2.5 Klasifikasi

Klasifikasi adalah proses menemukan kumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya, untuk dapat digunakan dalam memprediksi data yang belum memiliki kelas data tertentu [18].

Klasifikasi dokumen adalah pemberian kategori yang telah didefinisikan ke dokumen. Mengklasifikasikan dokumen merupakan salah satu cara untuk mengorganisasikan dokumen. Dokumen-dokumen yang memiliki isi yang sama akan dikelompokkan ke dalam kategori yang sama. Dengan demikian orang-orang yang melakukan pencarian informasi akan tertolong dalam hal navigasi dan dengan mudah kategori yang tidak relevan dengan informasi dapat segera dilewatkan [19].

Pada awalnya, kegiatan klasifikasi dokumen dilakukan secara manual oleh seorang ahli. Untuk melakukan klasifikasi, ahli harus menyediakan waktu untuk membaca isi setiap dokumen baru dan membandingkan isinya dengan skema kategori yang tersedia. Dengan demikian klasifikasi ini merupakan kegiatan yang menghabiskan waktu serta biaya. Selain alasan yang telah disebutkan sebelumnya ditambah dengan peningkatan jumlah dokumen telah mendorong berkembangnya metode pengklasifikasian secara otomatis yang dapat melakukan klasifikasi dengan cara belajar dari sekumpulan contoh dokumen yang telah diklasifikasikan sebelumnya. Dengan metode tersebut, keuntungan yang diperoleh adalah penghematan tenaga kerja dan efektivitas yang baik.

Klasifikasi adalah salah satu peran utama dari *data mining*. Klasifikasi termasuk dalam *supervised learning* karena dalam proses klasifikasi ada proses belajar dengan data masa lalu. Proses ini digunakan oleh algoritma untuk mengenali pola dari data yang nantinya dapat diterapkan ke data baru yang kelompoknya belum diketahui. Teknik klasifikasi diterapkan secara luas di dunia nyata maupun di dunia kedokteran, pendidikan, teknik bangunan, jaringan komputer dan banyak digunakan di bidang lain [20].

Label dalam klasifikasi atau bisa juga disebut dengan atribut tujuan adalah atribut yang akan dicari dalam perhitungan algoritma *data mining*. Untuk contoh dalam dunia medis jika ada pasien baru dengan gejala penyakit tertentu tetapi jenis

penyakitnya menderita belum diketahui. Jadi klasifikasi dapat menjadi alat untuk membuat keputusan. Adanya data masa lalu atau apa yang akan disebut *training data* akan banyak membantu dalam proses klasifikasi. Karena jumlah *training data* akan mempengaruhi akurasi klasifikasi algoritma *data mining*. Jumlah atribut juga akan mempengaruhi kinerja suatu algoritma [21], meskipun juga banyak atribut atau biasa disebut dengan *high-dimensional data* akan mempengaruhi kompleksitas waktu dari algoritma. Itu semakin banyak atribut yang digunakan, semakin mahal proses komputasi akan, atau semakin lama waktu komputasi. Untuk mengatasi ini, atribut data reduksi dapat dilakukan atau disebut juga dengan ekstraksi ciri dan pemilihan fitur [22].

Dalam melakukan suatu klasifikasi dibutuhkan data masa lalu yang akan nantinya akan diolah menjadi suatu aturan atau pengetahuan baru. Itu klasifikasi masalah pada dasarnya adalah sebagai berikut [23] :

1. Masalah klasifikasi berangkat dari yang tersedia data pelatihan.
2. Data latih akan diolah menggunakan klasifikasi algoritma.
3. Masalah klasifikasi berakhir dengan pembangkitan pengetahuan yang direpresentasikan dalam bentuk diagram, aturan atau pengetahuan.

Klasifikasi dimulai dengan data awal yang digunakan sebagai data pembelajaran algoritma atau disebut juga data latih. Tentu saja, data pelatihan yang dimaksud adalah data dengan tujuan atribut atau atribut label. Yang dimaksud dengan label adalah hasil akhir dari data yang nantinya akan dihitung menggunakan sebuah algoritma. Misalnya ada data pendaftaran siswa dengan label registrasi/tidak terdaftar. Data ini akan menjadi diproses oleh algoritma untuk mengetahui pola, aturan atau pengetahuan baru dari data tersebut. Nanti pola baru ini atau pengetahuan dapat digunakan sebagai alat untuk memprediksi jika ada yang baru rekaman dengan label yang tidak diketahui. Akurasi dari algoritma berbeda tergantung pada jenis data yang diprosesnya [24].

Dalam mengelompokkan data (klasifikasi) terdapat empat (4) komponen dasar yaitu kelas, prediktor, *training set* dan pengujian *dataset*. Komponen-komponen utama dari proses klasifikasi antara lain [25] :

1. Kelas, merupakan variable tidak bebas yang merupakan label dari hasil

klasifikasi. Sebagai contoh adalah kelas loyalitas pelanggan, kelas badai atau gempa bumi, dan lain-lain.

2. Prediktor, merupakan variable bebas suatu model berdasarkan dari karakteristik atribut data yang diklasifikasi, misalnya merokok, minuman beralkohol, tekanan darah, status perkawinan, dan sebagainya.
3. Set data pelatihan, merupakan sekumpulan data lengkap yang berisi kelas dan predictor untuk dilatih agar model dapat mengelompokkan ke dalam kelas yang tepat. Contohnya adalah grup pasien yang telah di-test terhadap serangan jantung, grup pelanggan di suatu supermarket, dan sebagainya.
4. Set data uji, berisi data-data baru yang akan dikelompokkan oleh model guna mengetahui akurasi dari model yang telah dibuat.

2.6 Machine Learning (ML)

2.6.1 Pengertian

Machine Learning adalah metode yang digunakan untuk membuat program yang bisa belajar dari data. Berbeda dengan program komputer biasa yang statis, program *machine learning* adalah program yang dirancang untuk mampu belajar sendiri. Cara belajar program *machine learning* mengikuti cara belajar manusia, yakni belajar dari contoh-contoh. *Machine learning* akan mempelajari pola dari contoh-contoh yang dianalisa, untuk menentukan jawaban dari pertanyaan-pertanyaan berikutnya. *Machine learning* adalah seperti membuat program yang bias menebak kotak hitam yang memiliki rumus fungsi yang belum diketahui. Kotak hitam itu diberikan sebuah input dan akan menghasilkan sebuah output tertentu. Dari data-data input dan output yang diperoleh, maka program akan menebak rumus fungsi yang paling mendekati keakuratan [26].

Machine learning adalah pembelajaran mesin yang sangat membantu dalam menyelesaikan masalah, membuat mudah dalam mengerjakan sesuatu. Dibidang rumah sakit atau bidang kesehatan, machine learning membuat mudah dalam mengerjakan sesuatu, contohnya dokter bisa mendiagnosa penyakit jantung dalam waktu cepat tanpa memakan waktu yang lama [27].

Machine learning (ML) adalah bidang studi yang fokus kepada desain dan

analisis algoritma sehingga memungkinkan komputer untuk dapat belajar. ML berisi sebuah algoritma yang bersifat generic (umum) dimana algoritma tersebut dapat menghasilkan sesuatu yang menarik atau bermanfaat dari sejumlah data tanpa harus menulis kode yang spesifik. Pada intinya, algoritma yang generik tersebut ketika diberikan sejumlah data maka ia dapat membangun sebuah aliran aturan atau model atau inferensi dari data tersebut.

Machine Learning (ML) digunakan untuk mengajari mesin cara menangani data dengan lebih efisien. Tujuan pembelajaran mesin adalah untuk belajar dari data. Banyak penelitian telah dilakukan tentang bagaimana membuat mesin belajar sendiri tanpa diprogram secara eksplisit [28].

Machine learning juga dapat diartikan sebuah komputer yang memiliki kemampuan belajar tanpa diprogram secara eksplisit. Program tersebut memanfaatkan data untuk membangun model dan mengambil keputusan berdasarkan model yang telah dibangun.

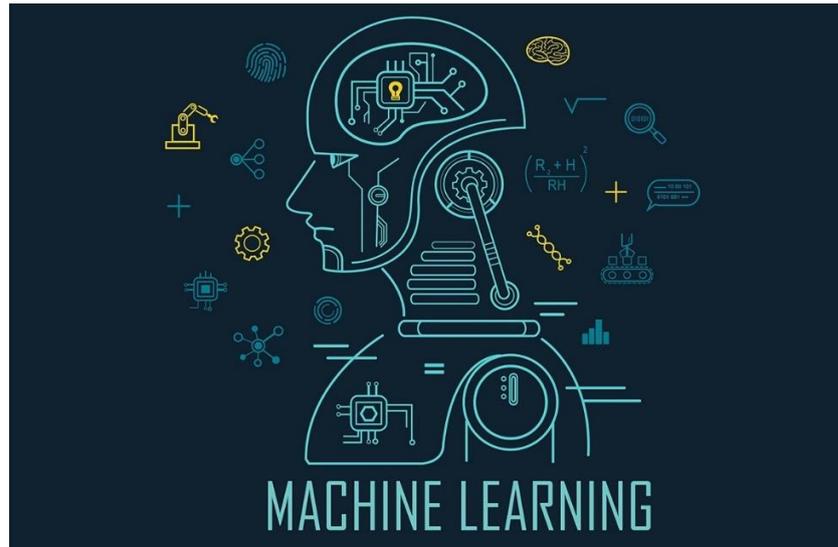
Machine learning adalah satu program komputer yang dikatakan telah melakukan pembelajaran dari pengamalan E (*Experience*) terhadap tugas T (*Task*) dan mengukur peningkatan kinerja P (*Performance Measure*), jika kinerja Tugas T diukur oleh kinerja P, maka meningkatkan pengalaman E. Dari definisi ini dapat dikatakan sebuah aplikasi *machine learning* memiliki 3 komponen yaitu *Task* T, *Performance Measure* P, dan *Experience* E. Oleh karena itu, untuk membangun sebuah aplikasi ML maka komponen T,P dan E harus dapat diidentifikasi [29]. Prinsip cara kerja *machine learning* masih sama, meliputi pengumpulan data, eksplorasi data, pemilihan model atau teknik, memberikan pelatihan terhadap model yang dipilih dan mengevaluasi hasil dari *machine learning*. *Machine Learning* mempunyai dua tipe yaitu:

1. *Supervised Learning*

Supervised Learning adalah metode yang digunakan saat semua data yang dimiliki sudah mempunyai label dan algoritmanya belajar memprediksi output dari input. *Supervised Learning* meliputi: Teknik *Regression* dan *Classification*.

2. *Unsupervised Learning*

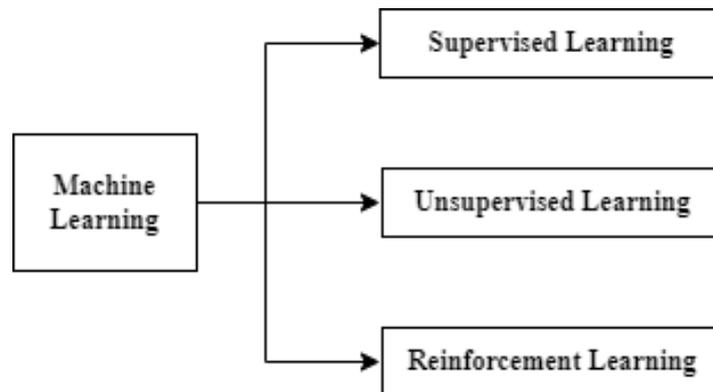
Unsupervised Learning adalah metode yang digunakan saat semua data yang dipunya tidak mempunyai label dan algoritmanya mempelajari struktur yang melakat dari data tersebut. *Unsupervised Learning* meliputi: Teknik *Clustering* dan *Association*



Gambar 2.2 *Machine Learning*

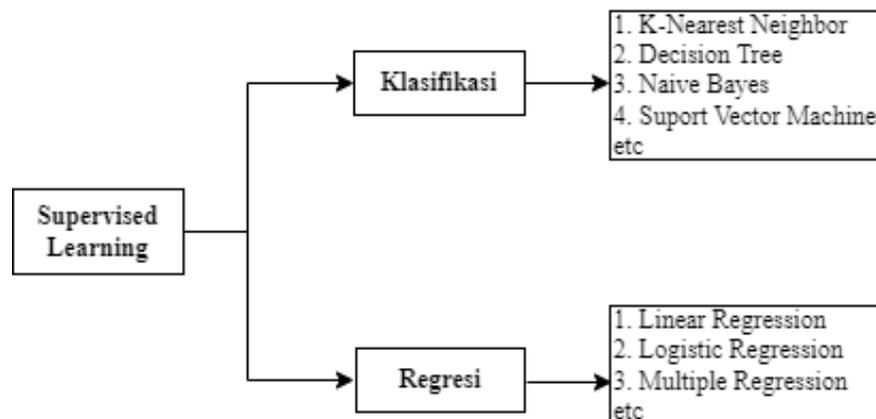
2.6.2 Penerapan Metode

Salah satu metode *machine learning* yang paling populer adalah *supervised learning*, yang menggunakan data latih untuk melakukan suatu pembelajaran. Syarat menerapkan metode *supervised learning* yaitu adanya data latih berupa dataset yang sudah berlabel. Data latih berlabel yang dimaksud adalah dataset penderita virus Covid-19 yang menghasilkan tingkat keparahan (*severity*) dengan 4 jenis yaitu tidak parah (*none*), parah (*severe*), sedang (*moderate*), ringan (*mild*). Tujuannya untuk mengidentifikasi label ketika ada input baru.



Gambar 2.3 Metode *Machine Learning*

Ada dua hal utama dalam menerapkan *supervised learning*, yaitu klasifikasi dan regresi. Namun yang dibahas adalah metode klasifikasi yang merupakan proses pengelompokkan data berdasarkan pelatihan dataset berlabel. Algoritma yang digunakan dalam tugas akhir ini adalah *decision tree* serta menggunakan metode *machine learning* dalam mengklasifikasikan penderita virus Covid-19. Penerapan metode *machine learning* sebagai metode pendukung untuk dilakukan sistem saat diprogram menggunakan Pycharm.



Gambar 2.4 Pembagian algoritma pada metode *supervised learning*

Adapun tahapan membangun model pada *Machine Learning* yaitu sebagai berikut :

- a. Memahami permasalahan, permasalahan yang dimaksud ialah bagaimana cara *machine learning* untuk mengetahui suatu masalah pada penderita virus Covid-19 ketika ada input data baru;
- b. Memilih metode yang digunakan, dalam hal ini, metode yang digunakan adalah klasifikasi menggunakan algoritma *decision tree*;

- c. Mempersiapkan dataset, dalam menerapkan metode klasifikasi *decision tree*, dataset berlabel yang disiapkan berupa data penderita virus Covid-19;
- d. Implementasi, melakukan training dataset menggunakan *website* dengan bahasa pemrograman python, *web framework* Flask dan *software* Pycharm.
- e. Integrasi dan evaluasi, yaitu mengukur kinerja, *error* dan akurasi.

2.7 *Decision Tree*

Decision Tree adalah sebuah struktur pohon, dimana setiap simpul (*node*) pohon merepresentasikan atribut yang telah diuji. Pohon keputusan (*Decision Tree*) adalah salah satu metode yang cukup mudah untuk diinterpretasikan oleh manusia. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. Manfaat utamanya adalah mem-*break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih mudah menemukan solusi dari permasalahan yang ada.

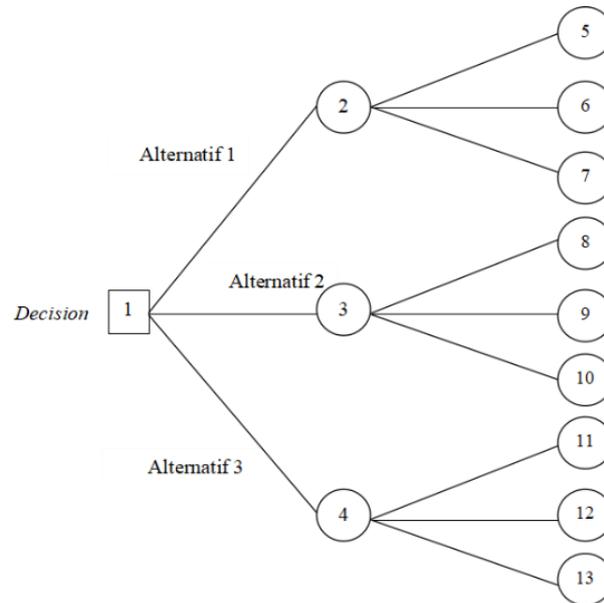
Decision Tree (Pohon Keputusan) adalah pohon dimana setiap cabangnya menunjukkan pilihan diantara sejumlah alternatif pilihan yang ada, dan setiap daunnya menunjukkan keputusan yang dipilih [30].

Decision Tree merupakan struktur *flowchart* yang berbentuk seperti pohon, dimana setiap simpul dalam menandakan suatu tes pada atribut, dengan cabang yang dihasilkan menunjukkan hasil tes, dan simpul daun merepresentasikan persebaran kelas [31]. Pohon (*tree*) adalah sebuah struktur data yang terdiri dari simpul (*node*) dan rusuk (*edge*). Simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (*root/node*), simpul percabangan/internal (*branch/internal node*) dan simpul daun (*leaf node*). Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas berhingga, dimana simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk-rusuknya diberi label nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda [32].

2.8 Analisis *Decision Tree*

Diagram pohon sering kali membantu dalam memahami dan menyelesaikan persoalan probabilitas. Diagram pohon biasanya digambarkan dengan lambang yang baku. Dimulai dengan suatu nokhta kemudian dibuat cabang-cabang sebanyak peristiwa yang mungkin dapat dihasilkan dari percobaan. Pada masing-masing cabang dituliskan probabilitas terjadinya peristiwa yang bersangkutan. Jika percobaan dilakukan lagi, maka langkah-langkah itu diulang. Setiap cabang berakhir pada nokhta yang kemudian diisi dengan probabilitas peristiwa bersama. Pada nokhta yang paling awal dituliskan angka 1 yang artinya jumlah probabilitas dari seluruh peristiwa yang mungkin. Kelebihan dari metode pohon keputusan [33] adalah :

1. Daerah pengambilan keputusan yang sebelumnya kompleks dan sangat global, dapat diubah menjadi lebih simpel dan spesifik.
2. Eliminasi perhitungan-perhitungan yang tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka *sample* diuji hanya berdasarkan kriteria atau kelas tertentu.
3. Fleksibel untuk memilih fitur dari internal node yang berbeda, fitur yang terpilih akan membedakan suatu kriteria dibandingkan kriteria yang lain dalam node yang sama. Kefleksibelan metode pohon keputusan ini meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode perhitungan satu tahap yang lebih konvensional.
4. Dalam analisis multivariat, dengan kriteria dan kelas yang jumlahnya sangat banyak. Metode pohon keputusan dapat menghindari munculnya permasalahan ini dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan.



Gambar 2.5 *Decision tree* (Pohon keputusan)

2.9 Confusion Matrix

Confusion matrix merupakan salah satu alat yang digunakan untuk mengukur kinerja suatu metode klasifikasi dalam mengestimasi objek yang bernilai true atau false. Pada dasarnya *confusion matrix* membandingkan kelas input awal yang memuat informasi nilai aktual dan prediksi pada hasil klasifikasi. *Confusion matrix* juga merupakan model pengukuran kinerja klasifikasi berdasarkan pengujian data dan setiap hasil prediksi data dengan proporsi yang tepat [34].

Rumus parameter berikut digunakan untuk mengevaluasi modelnya yaitu:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

(2.1)

Akurasi yang dihasilkan dihitung berdasarkan *confusion matrix*. Perhitungan pada *confusion matrix* dihitung sesuai dengan prediksi positif yang benar (*True Positif*), prediksi positif yang salah (*False Positif*), prediksi negatif yang benar (*True Negatif*) dan prediksi negatif yang salah (*False Negatif*). Semakin tinggi nilai akurasi yang didapat maka semakin baik pula metode yang dihasilkan.

$$Recall = \frac{TP}{TP+FN}$$

(2.2)

Recall adalah rasio prediksi benar positif (TP) yang dibandingkan dengan keseluruhan data yang benar positif.

$$Specificity = \frac{TN}{TN+FP}$$

(2.3)

Specificity adalah kebenaran memprediksi negatif dibandingkan dengan keseluruhan data negatif.

$$Precision = \frac{TP}{TP+FP}$$

(2.4)

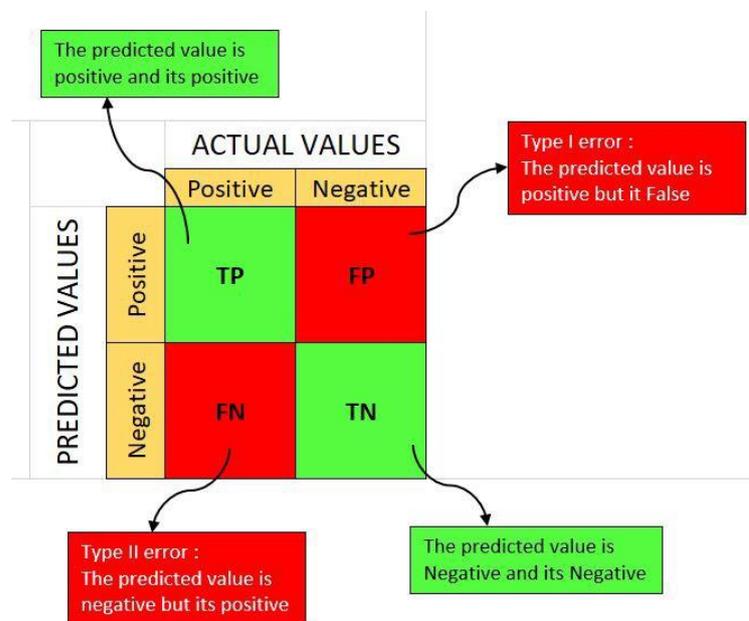
Precision adalah rasio prediksi benar positif (TP) yang dibandingkan dengan keseluruhan hasil yang diprediksi positif.

$$F1\ Score = \frac{2 \times (Recall \times Precision)}{Recall + Precision}$$

(2.5)

F1 Score adalah perbandingan rata-rata presisi dan recall yang dibobotkan.

Rumus-rumus diatas merupakan matrik utama untuk menentukan jumlah pasien Covid-19 yang teridentifikasi dengan benar.



Gambar 2.6 *Confusion matrix*

2.10 Python

Python adalah bahasa pemrograman interpretatif yang dianggap mudah dipelajari serta berfokus pada keterbacaan kode. Dengan kata lain, Python diklaim sebagai bahasa pemrograman yang memiliki kode-kode pemrograman yang sangat jelas, lengkap, dan mudah untuk dipahami.

Python dianggap memiliki kehebatan untuk menangani pembuatan aplikasi-aplikasi kekinian yang mengandung kata kunci *big data*, *data mining*, *deep learning*, *data science*, hingga *machine learning*. Dengan kata lain, Python adalah bahasa pemrograman simpel untuk pembuatan aplikasi berbasis kecerdasan buatan (*artificial intelligence*) [35].

Python secara umum berbentuk pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Python dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi. Pada prinsipnya, Python dapat diperoleh dan digunakan secara bebas oleh siapa pun, bahkan bagi para developer yang menggunakan bahasa pemrograman ini untuk kepentingan komersial. Namun demikian, penggunaan packages atau module-module dari pihak ketiga, mungkin saja membutuhkan lisensi yang berbeda, misalnya lisensi berbayar.



Gambar 2.7 Bahasa pemrograman python

2.10.1 Platform Pemrograman Python

Python dapat dijalankan di berbagai platform sistem operasi. Oleh karena itu, distribusi aplikasi yang dibuat menggunakan Python sangat luas dan multiplatform. Beberapa platform yang mendukung Python diantaranya adalah Linux/Unix, Windows, Mac OS X, Java Virtual Machine, OS/2, Amiga, Palm, Symbian (untuk produk-produk Nokia).

2.10.2 Fitur Bahasa Pemrograman Python

Sisi utama yang membedakan Python dengan bahasa lain adalah dalam hal aturan penulisan kode program. Bagi para programmer yang tidak terbiasa menggunakan python akan dibingungkan dengan aturan indentasi, tipe data, *tuple*, dan *dictionary*. Python memiliki kelebihan tersendiri dibandingkan dengan bahasa lain terutama dalam hal penanganan modul, ini yang membuat beberapa programmer menyukai python. Selain itu python merupakan salah satu produk yang *open source*, gratis, dan multiplatform. Beberapa fitur dan kelebihan yang dimiliki Python adalah [36] :

1. Memiliki koleksi kepustakaan yang banyak. Artinya, telah tersedia modul-modul siap pakai untuk berbagai keperluan, seperti pembuatan game hingga artificial intelligence (misal: Tensor Flow).
2. Memiliki struktur bahasa yang jelas, sederhana, dan mudah dipelajari.
3. Berorientasi prosedural dan objek sekaligus (multi-paradigma)
4. Memiliki sistem pengelolaan memori otomatis (*garbage collection*) seperti halnya Java.
5. Bersifat modular sehingga mudah dikembangkan dengan menciptakan moodul-modul baru, baik dibangun dengan bahasa Python maupun C/C++

2.10.3 Anaconda

Anaconda merupakan sebuah distribusi bahasa pemrograman Python dari *Continuum Analytics* yang berisi paket Python ditambah beberapa paket tambahan untuk keperluan komputasi ilmiah (*scientific computing*) seperti *data science*, *machine learning*, *data processing* skala-luas, analisis prediksi, dan lain

sebagainya dalam satu distribusi *platform* yang *user friendly*.



Gambar 2.8 Logo Anaconda

Anaconda diciptakan agar mempermudah pengguna manajemen paket python. Dengan menggunakan Anaconda, maka versi dari paket yang ada, di manajemen oleh package management system conda.

2.10.4 Jetbrains Pycharm

IDE yang cukup populer dikalangan developer Python adalah PyCharm. PyCharm sendiri memiliki dua versi yaitu *Professional Edition* dan *Community Edition*. PyCharm *Professional Edition* merupakan versi berbayar dari PyCharm dan *Community Edition* merupakan versi gratis yang tersedia bagi komunitas python dengan lisensi Apache 2. Pycharm digunakan sebagai tool dalam pensimulasian ini.

PyCharm merupakan text editor atau Integrated Development Environment (IDE). PyCharm Edu merupakan text editor dengan tampilan user interface yang mudah dipahami sehingga mudah digunakan dalam tujuan pembelajaran. File Python menggunakan format `.py`.



Gambar 2.9 Logo Pycharm

2.10.5 Flask

Dengan perkembangan teknologi saat ini, terdapat banyak alat untuk membantu kegiatan web *development*, salah satunya adalah web *framework* yang dibuat menggunakan bahasa Python, yaitu Flask. Flask adalah sebuah alat yang membantu membuat kerangka untuk sebuah web dan dengan menggunakan Flask, pengembang pemula pun dapat menciptakan sebuah web yang bagus.

Flask adalah sebuah web framework yang ditulis dengan bahasa Python dan tergolong sebagai jenis microframework. Flask berfungsi sebagai kerangka kerja aplikasi dan tampilan dari suatu web. Dengan menggunakan Flask dan bahasa Python, pengembang dapat membuat sebuah web yang terstruktur dan dapat mengatur behaviour suatu web dengan lebih mudah.

Flask termasuk pada jenis microframework karena tidak memerlukan suatu alat atau pustaka tertentu dalam penggunaannya [37]. Sebagian besar fungsi dan komponen umum seperti validasi form, database, dan sebagainya tidak terpasang secara default di Flask. Hal ini dikarenakan fungsi dan komponen-komponen tersebut sudah disediakan oleh pihak ketiga dan Flask dapat menggunakan ekstensi yang membuat fitur dan komponen-komponen tersebut seakan diimplementasikan oleh Flask sendiri. Selain itu, meskipun Flask disebut sebagai microframework, bukan berarti Flask mempunyai kekurangan dalam hal fungsionalitas. Microframework disini berarti bahwa Flask bermaksud untuk membuat core dari aplikasi ini sesederhana mungkin tapi tetap dapat dengan mudah ditambahkan. Dengan begitu, fleksibilitas serta skalabilitas dari Flask dapat dikatakan cukup tinggi dibandingkan dengan framework lainnya. Aplikasi yang dibuat dengan Flask disimpan dalam satu berkas “.py”.

Flask memiliki keunggulan serta kekurangan dibandingkan dengan web framework berbasis Python lainnya, seperti Django, CherryPy, dan sebagainya. Keunggulan-keunggulan dari Flask yaitu [38]:

1. Ringan untuk dijalankan karena mempunyai core yang sederhana dan desain modular.
2. Dapat menangani fungsi HTTP request dengan mudah.
3. API yang baik dan koheren.

4. Dokumentasi yang banyak dan terstruktur dengan baik, penuh dengan contoh yang dapat digunakan langsung.
5. Mudah untuk dipasang dan di-deploy untuk produksi.
6. Mudah untuk diperiksa secara menyeluruh. (Unit testability.)
7. Fleksibilitas tinggi, dengan konfigurasi yang sangat mudah diubah.

Selain memiliki keunggulan-keunggulan yang telah disebutkan, Flask juga memiliki kekurangan, diantaranya yaitu [38]:

1. Tidak memiliki ORM dan database layer bawaan, sehingga harus menggunakan aplikasi dari pihak ketiga.
2. Tidak async-friendly.

Karakteristik Flask yang sangat membutuhkan aplikasi pihak ketiga dalam menjalankan fitur-fiturnya menjadi pedang bermata dua yang bisa menjadi keunggulan maupun kekurangan dari Flask itu sendiri.



Gambar 2.10 Flask

2.11 Penelitian-Penelitian Sebelumnya

Pada tabel 2.1 menunjukkan penelitian-penelitian sebelumnya yang pernah dilakukan. Hal tersebut dibutuhkan sebagai bahan pertimbangan pembuatan tugas akhir.

Tabel 2.1 Penelitian-Penelitian Sebelumnya

Penulis (Tahun)	Judul	Algoritma/ Teknik Pemodelan	Kelebihan	Kekurangan	Tujuan	Hasil
Rian Rafiska, Sarjon Defit, Gunadi Widi Nurcahyo (2018) [3]	Analisis Rekam Medis untuk Menentukan Pola Kelompok Penyakit Menggunakan Algoritma C4.5	Metode klasifikasi dengan algoritma C4.5, rapid maner	Algoritma C4.5 telah berhasil diterapkan dalam menganalisis data rekam medis di RSUD Mayjen H.A. Thalib Kerinci	Hanya menggunakan data 1 bulan saja	Menemukan pola kelompok penyakit yang sering terjadi di masyarakat berdasarkan kode penyakit dalam <i>International Classification of Diseases (ICD)</i>	Pemilihan variabel umur, jenis kelamin, dan alamat adalah kriteria penilaian terhadap pelanggan aktif dan tidak aktif.
Akbar Mujahidin dan Denny Pribadi (2017) [4]	Penerapan Algoritma C4.5 untuk Diagnosa Penyakit Pneumonia pada Anak Balita Berbasis <i>Mobile</i>	Algoritma C4.5, bentuk aplikasi mobile berbasis android.	Sistem ini dirancang aplikasi mobile berbasis android, sehingga memudahkan para pengguna khususnya para orang tua dalam penggunaannya dan dapat langsung digunakan kapan pun dan dimanapun.	Hardware yang digunakan dalam aplikasi ini adalah semua perangkat mobile seperti handphone, smartphone, atau tablet yang memiliki sistem operasi dari mulai versi Android 2.2 Froyo sampai versi Android 5.1.1 Lollipop	Memenuhi nilai informasi yang lebih cepat dan fleksibel, Dapat membantu pengguna terutama orang tua dalam mendiagnosis penyakit pneumonia pada anak balita	10 aturan yang dihasilkan dari algoritma <i>decision tree</i> C4.5 dengan Jumlah kelas "Tidak" sebanyak 4 aturan dan jumlah kelas "iya" sebanyak 6 aturan.
Ivandari dan M. Adib Al Karomi (2021) [5]	Diagnosis of Coronavirus Disease 2019 (Covid-19) Surveillance Using C4.5	Algoritma C4.5, rapid miner dan 10 folds cross validation	Tingkat akurasi yang diperoleh dengan menggunakan algoritma	Rendahnya tingkat akurasi disebabkan oleh kurangnya record yang	Untuk mengklasifikasi dataset surveilans Covid-19	Pada dataset surveilans Covid-19 terdapat 3 varian

	Algorithm		decision tree adalah 65%. Walaupun pada penelitian ini nilai akurasi mengalami peningkatan sebesar 10%	ada, dan adanya varian dari atribut label yang dominan		label dengan rincian 8 record (PUS), 5 record (PIM), dan 1 record (PWS)
Yetti Fitriani, Sarjon Defit, Gunadi Widi Nurcahyo (2021) [6]	Prediksi Hasil Belajar Siswa Secara Daring pada Masa Pandemi Covid-19 Menggunakan Metode C4.5	Metode C4.5	Klasifikasi Metode C4.5 terhadap nilai hasil belajar siswa menghasilkan rule-rule dan knowledge yang dapat digunakan untuk memprediksi hasil belajar siswa secara daring pada masa pandemi Covid-19	Dapat dilakukan dengan data yang lebih banyak untuk mendapatkan rule-rule yang lebih bervariasi sehingga mendapatkan tingkat akurasi yang lebih baik.	Untuk memprediksi hasil belajar siswa pada masa pandemi Covid-19 sehingga hasil penelitian bisa digunakan untuk menjadi acuan dalam pengambilan kebijakan di sekolah	Hasil pengujian dengan akurasi sebesar 83,33% sudah dapat direkomendasikan untuk membantu pihak sekolah dalam membuat kebijakan
Ahmad Harun dan Dea Putri Ananda (2021) [16]	Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan <i>Naïve Bayes</i> dan <i>Decission Tree</i>	Algoritma <i>Naïve Bayes</i> dan <i>Decission Tree</i>	<i>Naïve Bayes Classifier</i> sebagai algoritma rekomendasi terkait dengan sentimen masyarakat di Indonesia pada Fanspages Kementerian Kesehatan Republik Indonesia	<i>Decision tree</i> memiliki nilai akurasi yang rendah	Untuk memberikan gambaran kepada masyarakat apakah vaksinasi Covid-19 cenderung ke opini positif atau negatif dan membandingkan tingkat akurasi algoritma <i>Naïve Bayes</i>	Analisis sentimen opini publik terhadap vaksinasi Covid-19 di Indonesia yang dilakukan dengan menggunakan <i>Naïve Bayes Classifier</i> dengan

					dan <i>Decision Tree</i>	nilai akurasi 100% dan <i>Decision Tree</i> dengan nilai akurasi 50.39%
--	--	--	--	--	---------------------------------	--