

## BAB II

### LANDASAN TEORI

#### 2.1 Data Analytic

*Data analytic* adalah proses pemeriksaan set data untuk menarik kesimpulan tentang informasi yang dikandungnya. Teknologi dan teknik *data analytic* banyak digunakan dalam industri komersial untuk memungkinkan organisasi membuat keputusan bisnis berdasarkan informasi yang diperoleh, serta oleh para ilmuwan dan peneliti untuk memverifikasi atau menyangkal model, teori, dan hipotesis ilmiah (Stedman, 2020).

*Data analytic* merupakan cara menganalisis data mentah untuk membuat kesimpulan tentang suatu informasi. Teknik *data analytic* dapat mengungkapkan trend dan matrik yang hilang dalam massa informasi. Informasi ini kemudian dapat digunakan untuk mengoptimalkan proses untuk meningkatkan efisiensi keseluruhan bisnis atau sistem (Frankenfield, 2021).

*Data analytic* merupakan proses inspeksi serangkaian data yang berguna untuk mendapatkan kesimpulan yang berguna untuk mendapatkan kesimpulan dari informasi yang ada dan meningkatkan sistem pada software. Teknologi *data analytics* dan teknik digunakan di industri komersial yang memudahkan perusahaan mendapatkan hasil yang lebih baik dan akurat. *Data analytic* mencakup berbagai jenis teknik menganalisa data. Semua jenis informasi dapat memanfaatkan teknik *data analytic* untuk mendapatkan pengetahuan baru yang dapat digunakan untuk meningkatkan berbagai hal. *Data analytic* membantu perusahaan, bisnis, dapat membantu mengurangi biaya dengan mengidentifikasi cara yang lebih efisien dalam melakukan bisnis yang lebih baik dan membantu menganalisis kepuasan pelanggan, yang dapat mengarah pada produk dan layanan yang baru dan lebih baik.

Langkah – langkah yang dilakukan dalam *data analytic* sebagai berikut:

- a. Menentukan persyaratan data, atau bagaimana data dikelompokkan. Data yang dipilih dapat dipisahkan berdasarkan usia, pendapatan, atau jenis kelamin. Nilai data dapat berupa sebuah nilai atau angka ataupun bisa dibagi berdasarkan kategori.
- b. Proses pengumpulan data melalui berbagai sumber, seperti komputer, sumber online, kamera, sumber lingkungan, atau melalui personil.
- c. Setelah data dikumpulkan, kemudian diatur untuk dapat dianalisa. Pengaturan dapat

dilakukan dalam bentuk *spreadsheet* atau perangkat lunak lain yang dapat mengolah data statistik.

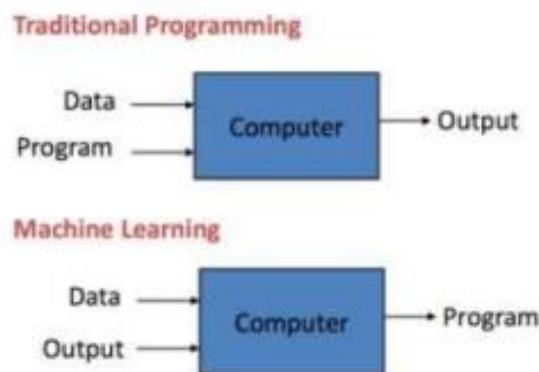
- d. Data kemudian dibersihkan, yakni diperiksa dan dihapus untuk memastikan tidak ada duplikasi, kesalahan, atau tidak lengkap. Langkah ini membantu memperbaiki kesalahan apapun sebelum data tersebut dianalisa.

## 2.2 Machine Learning

Secara umum, machine learning mengacu kepada teknologi yang memungkinkan sebuah komputer melakukan proses pembelajaran. Belajar adalah proses mengubah pengalaman menjadi keahlian atau pengetahuan. Pada algoritma pembelajaran, input merupakan data pelatihan yang mewakili pengalaman dan output itu sendiri berupa keahlian yang biasanya mengambil kerja program lain yang dapat melakukan beberapa tugas.

*Machine learning* adalah hasil ketika ilmu komputer dan statistik bergabung. Ilmu komputer berfokus pada membangun mesin yang memecahkan masalah tertentu, dan mencoba mengidentifikasi apakah masalah dapat dipecahkan. Sedangkan pendekatan yang dikerjakan oleh statistik secara mendasar adalah inferensi data, pemodelan hipotesis, dan pengukuran reliabilitas kesimpulan

Dalam *machine learning*, program menganalisis kumpulan data yang besar untuk menemukan pola.



**Gambar 2.1** Perbedaan program tradisional dan *Machine learning*

Melalui algoritma *machine learning*, user perlu melatih komputer sedemikian rupa sehingga bisa memahami model objek yang dikenali manusia. Pada pemrograman tradisional, data dan program dijalankan komputer untuk menghasilkan *output*, sedangkan pada pemrograman

menggunakan *machine learning*, data dan *output* dijalankan pada komputer untuk membuat program, dan kemudian program tersebut bisa digunakan dalam pemrograman tradisional [2].

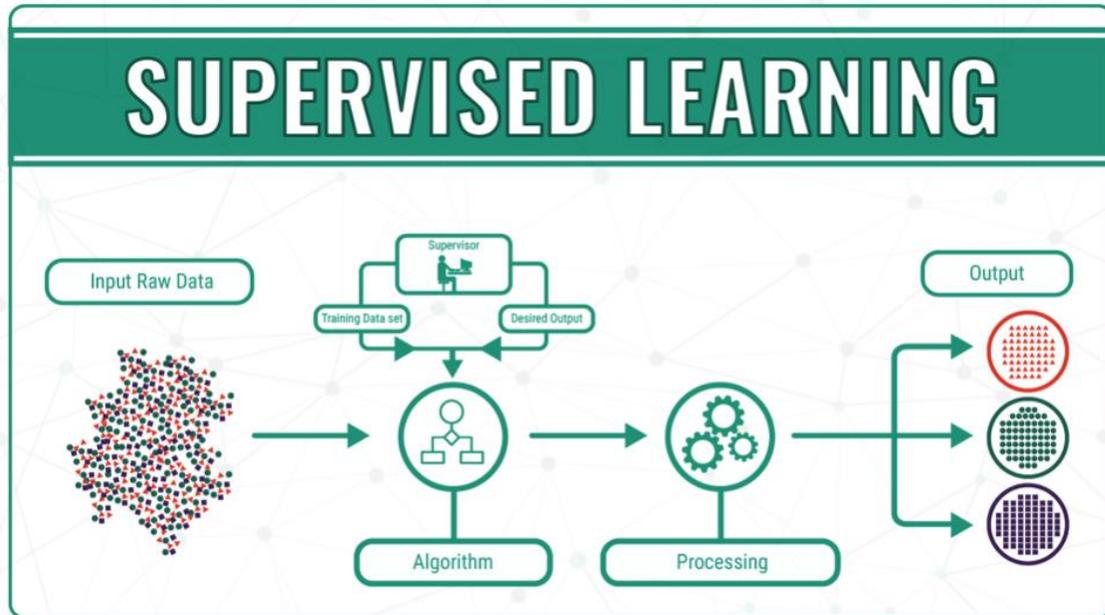
*Machine learning* mempunyai dua tipe teknik yaitu *supervised learning* dan *unsupervised learning*. *Supervised learning* adalah salah satu tipe algoritma *machine learning* yang menggunakan *dataset* yang dikenal (*training dataset*) untuk membuat prediksi. *Unsupervised learning* adalah salah satu tipe algoritma *machine learning* yang digunakan untuk menarik kesimpulan dari *datasets* yang terdiri dari input data *labeled response* [3].

Pada era *digital*, *machine learning* telah menjadi faktor penting untuk setiap bidang teknik. *Machine learning* biasa digunakan untuk keperluan analisis data, klasifikasi dan prediksi. Saat ini, dalam aplikasi web atau ponsel umum, berbagai teori *machine learning* berlaku sehingga aplikasi yang digunakan menjadi lebih cerdas dan dapat memperoleh kemampuan untuk memahami pikiran penggunanya [4].

Dari sisi metode pembelajaran, algoritma *Machine Learning* dapat dikategorikan sebagai *Supervised Learning*

### **2.2.1 Supervised Learning**

Secara literal, *supervised learning* berarti pembelajaran yang terarah/terawasi. Tujuan pembelajaran ini secara umum adalah untuk memperkirakan fungsi pemetaannya, sehingga ketika terdapat input baru, sistem dapat memprediksi output untuk input tersebut. Pendekatan *supervised learning* mempunyai input dan *desired* output yang dapat dibuat menjadi suatu model hubungan matematis sehingga mampu melakukan prediksi dan klasifikasi berdasarkan data yang telah ada sebelumnya.



**Gambar 2.1. Contoh ilustrasi *supervised learning***

(sumber:<https://www.techfor.id/konsep-supervised-learning-dalam-membangun-model-machine-learning/>)

*Supervised Learning* menggunakan satu set pelatihan untuk mengajarkan model untuk menghasilkan output yang diinginkan. Dataset pelatihan ini mencakup input dan output yang benar, yang memungkinkan model untuk belajar dari waktu ke waktu. Algoritma nantinya mengukur akurasi melalui fungsi kerugian (*Loss Function*), lalu menyesuaikan hingga kesalahan cukup diminimalkan.

Beberapa algoritma yang termasuk dalam *supervised learning* antara lain:

- a. Regresi Linier;
- b. Analisis Deret Waktu;
- c. *Decision Tree* dan *Random Forest*;
- d. *Naive Bayes Classifier*;
- e. *Nearest Neighbour Classifier*;
- f. *Artificial Neural Network*; dan
- g. *Support Vector Machine*.

Permasalahan pada *supervised learning* dapat dikelompokkan menjadi masalah klasifikasi (*classification problems*) dan masalah regresi (*regression problem*).

## 1. Klasifikasi

Klasifikasi merupakan salah satu metode statistika dalam pengelompokan suatu data yang disusun secara sistematis. Masalah klasifikasi muncul ketika terdapat sejumlah ukuran yang terdiri dari satu atau beberapa kategori yang tidak dapat diidentifikasi secara langsung tetapi harus menggunakan suatu ukuran (K. P. Suniantara. 2018)

Teknik dari klasifikasi adalah dengan melihat variabel dari kelompok data yang sudah ada. Klasifikasi bertujuan untuk memprediksi kelas dari suatu objek yang tidak diketahui sebelumnya. Klasifikasi terdiri dari tiga tahap, yaitu pembangunan model, penerapan model, dan evaluasi. Pembangunan model adalah membangun model menggunakan data latih yang telah memiliki atribut dan kelas. Kemudian, data-data tersebut diterapkan untuk menentukan kelas dari data atau objek yang baru. Setelah itu, data dievaluasi untuk melihat tingkat akurasi dari pembangunan dan penerapan model terhadap data baru. Proses klasifikasi terdiri dari dua fase, yaitu fase *training* dan fase *testing*. Fase *training* adalah fase di mana data digunakan untuk membangun sebuah model sedangkan fase *testing* adalah pengujian model yang telah dibuat dengan data lainnya untuk mengetahui akurasi dari model tersebut (D. A. Nasution, H. 2019)

Pada bidang *machine learning*, teknik klasifikasi sering dimanfaatkan untuk berbagai hal antara lain untuk prediksi kinerja siswa, klasifikasi jenis penyakit, memprediksi kecurangan pada transaksi kartu kredit dan masih banyak hal lagi yang dapat dibantu dengan menggunakan teknik klasifikasi (Y. Pristyanto. 2019).

## 2. Regresi

Regresi merupakan salah satu teknik untuk meramalkan data di masa yang akan datang menggunakan variabel *independent* dan variabel penjelas/bebas). Berbeda dengan klasifikasi yang memprediksi nilai variabel yang bersifat diskret, regresi melakukan fungsi pembelajaran yang memetakan sebuah unsur data ke sebuah variabel prediksi bernilai nyata, yang digunakan untuk memprediksi nilai variabel yang bersifat kontinyu.

Terdapat beberapa jenis regresi antara lain:

- a. Regresi Linier, merupakan sebuah pendekatan untuk memodelkan hubungan antara variabel terikat Y dan satu atau lebih variabel X yang merupakan variabel bebas, dimana seluruh variabelnya adalah data kuantitatif. Disebut linier karena setiap estimasi atas nilai diharapkan mengalami peningkatan atau penurunan mengikuti garis lurus.

Metode ini digunakan untuk mengetahui bagaimana variabel dependent dapat diprediksikan melalui variabel independent atau variabel prediktor. Dampak dari penggunaan regresi dapat digunakan untuk memutuskan apakah naik dan menurunnya variabel dependent dapat dilakukan melalui menaikkan dan menurunkan keadaan variabel independent, atau meningkatkan keadaan variabel *dependent* dapat dilakukan dengan meningkatkan variabel *independent* dan atau sebaliknya.

Berdasarkan jumlah variabel bebas X, regresi linier dibagi menjadi dua jenis yakni regresi linier sederhana dan regresi linier berganda (*multiple linear regression*).

- b. Regresi Non-Linier, merupakan hubungan antara variabel Y dan X yang tidak linier. Tidak linier artinya, laju perubahan Y akibat laju perubahan X tidak konstan untuk nilai-nilai X tertentu. Seperti regresi kuadratik, kubik. Misal: produksi padi akan meningkat saat diberi pupuk taraf rendah ke sedang. Namun apabila diberi pupuk dengan taraf tinggi, maka tingkat produksinya malah semakin menurun.
- c. Regresi *Dummy*, merupakan hubungan antara variabel y (data kuantitatif) dan variabel x (data kualitatif). Misal: melihat pengaruh kemasan terhadap harga jual makanan. Dengan kode '1' mewakili kemasan menarik dan '0' jika kemasan tidak menarik. Kode '1' dan '0' adalah variabel *dummy*.
- d. Regresi Logistik, merupakan hubungan antara variabel y (data kualitatif) dan variabel x (data kuantitatif). Misal: Apabila ingin diketahui apakah konsumen akan membeli makanan di rumah makan berdasarkan penilaian konsumen terhadap lokasi, pelayanan, pendapatan. Dalam kasus ini hanya ada 2 kemungkinan respon konsumen, yaitu konsumen membeli (1) dan tidak membeli (0).

Beberapa algoritma yang dapat digunakan pada metode regresi antara lain:

- a. *Simple linear regression*
- b. *Multiple linear regression*
- c. *Polynomial regression*
- d. *Support vector regression*
- e. *Decision tree regression*
- f. *Random forest regression*

## 2.2.2

### 2.2.3

## 2.3 Cacat Perangkat Lunak (*Software Defect*)

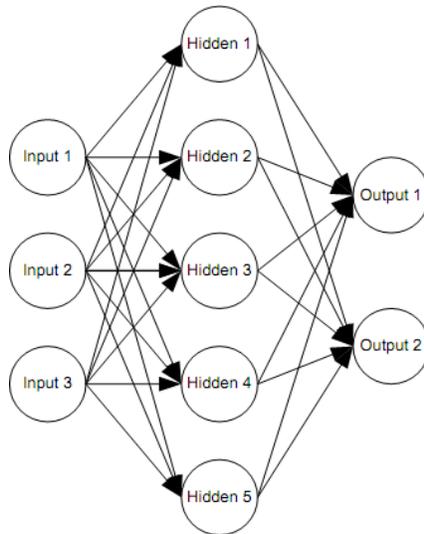
Cacat perangkat lunak (*software defect*) didefinisikan sebagai cacat pada perangkat lunak seperti cacat pada dokumentasi, pada kode program, pada desain dan hal – hal lain yang menyebabkan kegagalan perangkat lunak. Cacat perangkat lunak dapat muncul pada berbagai tahap proses pengembangan perangkat lunak (Roger S. Pressman. 2005). Cacat perangkat lunak merupakan faktor penting yang mempengaruhi kualitas perangkat lunak. Kualitas perangkat lunak dapat ditingkatkan dengan mencegah munculnya cacat perangkat lunak melalui perbaikan aksi yang mungkin menghasilkan cacat perangkat lunak pada proses pengembangan perangkat lunak (B. Boehm and S. California. 2001)

kegagalan dalam sistem perangkat keras sulit dihindari yang dimana kejadiannya disebabkan kecatatan bahan sedangkan setiap kegagalan dalam sistem perangkat lunak disebabkan oleh kesalahan manusia (Dick 2005),. laporan *Software defect* dapat diartikan sebagai gambaran dari ketidakpuasan pelanggan terhadap kinerja sistem software, biasanya kegagalan software atau kesalahan (error) (Gustafson. 2005),.

## 2.4 *Neural Network*

*Neural Network* adalah processor yang terdistribusi paralel, terbuat dari unit-unit yang sederhana, dan memiliki kemampuan untuk menyimpan pengetahuan yang diperoleh secara eksperimental dan siap pakai untuk berbagai tujuan [9]. *Neural Network* ini meniru otak manusia dari sudut:

1. Pengetahuan diperoleh oleh *network* dari lingkungan, melalui suatu proses pembelajaran.
2. Kekuatan koneksi antar unit yang disebut *synaptic weights*, berfungsi untuk menyimpan pengetahuan yang telah diperoleh oleh *network* tersebut.



Gambar 2. 1 Model *Neural Network*

*Neural Network* (NN) atau Jaringan syaraf tiruan (JST) juga diartikan sebagai suatu usaha untuk meniru fungsi otak manusia. Otak manusia diyakini terdiri dari jutaan unit pengolahan kecil, yang disebut neuron, yang bekerja secara paralel. Neuron saling terhubung satu sama lain melalui koneksi neuron. setiap individu neuron mengambil *input* dari satu set neuron. Ini kemudian memproses input tersebut dan melewati output untuk satu set neuron. Keluaran dikumpulkan oleh neuron lain untuk diproses lebih lanjut. Para Otak manusia adalah jaringan kompleks neuron di mana koneksi tetap melanggar dan membentuk

## 2.5 KFold