

BAB 2

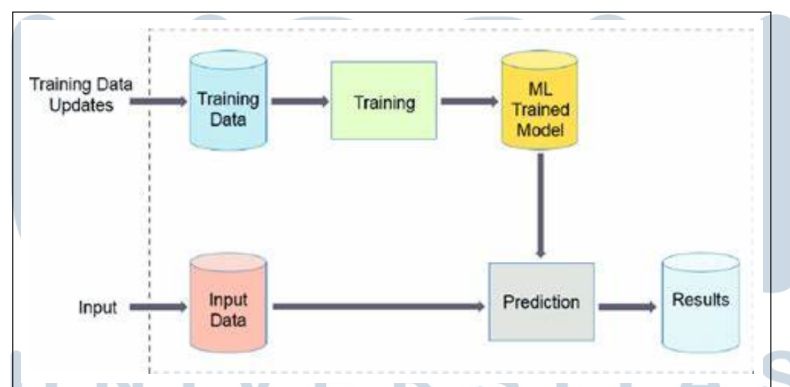
LANDASAN TEORI

2.1 Merchant

Merchant adalah penjual produk barang atau jasa yang mempunyai jenis bisnis offline maupun online. Selain itu, biasanya merchant juga bekerjasama dengan pihak lembaga keuangan atau lembaga lainnya untuk bisa memudahkan mereka dalam melakukan transaksi bisnisnya. Sebagai contoh bentuk kerjasama dengan pihak bank adalah dengan menyediakan layanan penerimaan pembayaran produk atau jasa secara debit, kredit ataupun dengan uang elektronik lainnya.

2.2 Pembelajaran Mesin

Pembelajaran mesin adalah kecerdasan buatan (Artificial Intelligence) pada bidang ilmu komputer yang memberikan kemampuan pembelajaran kepada komputer untuk mengetahui sesuatu tanpa pemrograman yang jelas. Pembelajaran Mesin merupakan serangkaian teknik yang dapat membantu dalam menangani dan memprediksi data yang sangat besar dengan cara mempresentasikan data-data tersebut dengan algoritma pembelajaran.

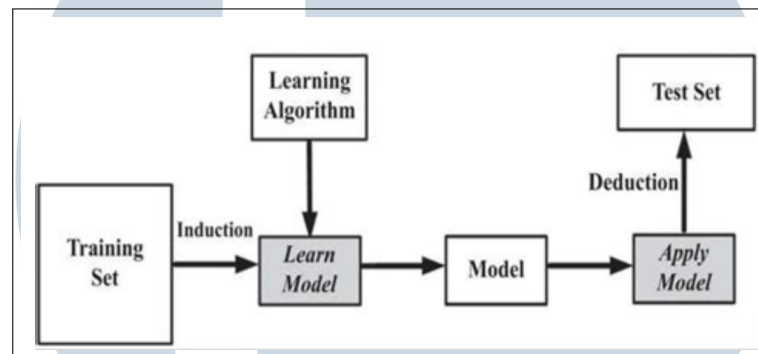


Gambar 2.1. Tahapan pembelajaran mesin [7]

Pembelajaran mesin memiliki fokus pada pengembangan sistem yang mampu belajar sendiri untuk memutuskan sesuatu tanpa harus berulang kali diprogram oleh manusia. Hal ini menjadikan mesin tidak hanya mampu berperilaku mengambil keputusan, namun juga dapat beradaptasi dengan perubahan yang terjadi.

2.3 Pembelajaran Terbimbing

Pembelajaran terbimbing merupakan pembelajaran menggunakan masukan data pembelajaran yang telah diberi label. Setelah itu membuat prediksi dari data yang telah diberi label. Algoritma ini belajar dengan membandingkan output sebenarnya dengan output yang benar untuk menemukan error atau kesalahan [8].



Gambar 2.2. Tahapan pembelajaran terbimbing [7]

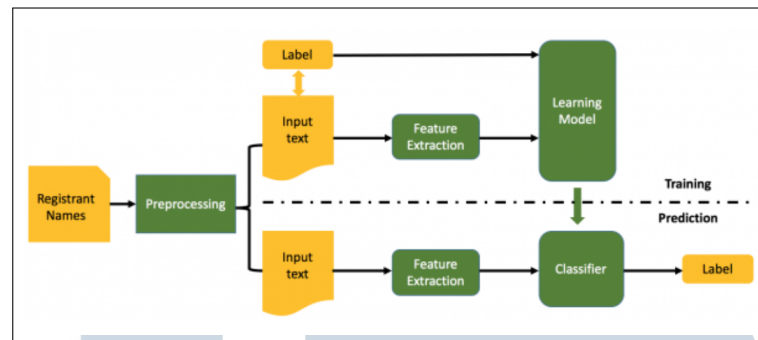
Pada Gambar 2.2. Merupakan contoh ilustrasi dari tahapan pembelajaran terbimbing, email yang masuk kedalam inbox akan disaring dan diklasifikasi sesuai dengan label pada pesan. Jika email tidak memiliki label maka akan masuk kedalam spam.

Dalam pembelajaran terbimbing, algoritma dapat memodifikasi model agar sesuai dengan hasil yang diinginkan. Pembelajaran terbimbing bertujuan untuk memetakan suatu input ke dalam sebuah output yang nilainya sudah disediakan.

2.4 Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui [9]. Metode-metode dalam klasifikasi digunakan untuk pembelajaran fungsi-fungsi berbeda yang memetakan data terpilih ke dalam suatu kelompok kelas yang telah ditentukan sebelumnya.

Dengan menggunakan atribut, dan data latihan (data latih), metode-metode klasifikasi tersebut dapat memprediksi kelas dari data lain yang belum diklasifikasikan [9].



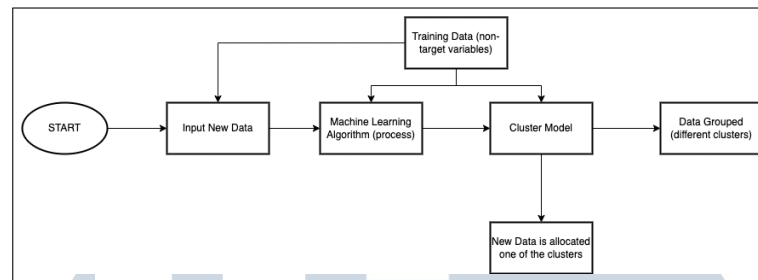
Gambar 2.3. Tahapan klasifikasi [10]

2.5 Regresi

Regresi adalah metode statistik yang dipakai untuk memperkirakan hubungan antara sebuah variabel terikat dan satu variabel independen atau lebih. Metode ini juga bisa digunakan untuk menilai kekuatan hubungan antara variabel dengan perkiraan masa depan. Analisis regresi termasuk beberapa variasi, yakni linear, linear majemuk, dan nonlinear. Model yang paling umum ialah linear dan linear majemuk. Sementara itu, nonlinear biasa dipakai untuk kelompok data yang lebih kompleks—karena hubungan antarvariabel tidak sejalan. Analisis regresi dapat diaplikasikan ke berbagai disiplin, termasuk finansial, investasi, dan pemasaran.

2.6 Pembelajaran Tidak Terbimbing

Pembelajaran tidak terbimbing menggunakan algoritma *machine learning* untuk menganalisis dan mengelompokkan data tidak berlabel. Algoritma ini menemukan pola tersembunyi dalam data tanpa perlu campur tangan manusia, itulah kenapa disebut sebagai “unsupervised” atau “tanpa pengawasan”[11]. Model *unsupervised learning* bekerja sendiri untuk menemukan struktur yang saling berhubungan dari data tidak berlabel. Model ini cenderung tidak membutuhkan intervensi manusia. Berbeda dengan *supervised learning* yang “belajar” dari kumpulan data lalu membuat prediksi dari data dan menyesuaikan jawaban yang benar.



Gambar 2.4. Pembelajaran tidak terbimbing

2.7 Algoritma K-Nearest Neighbor

Algoritma K-Nearest Neighbor merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (data latih), diantaranya euclidean distance dan manhattan distance atau city block distance [6]. Untuk pembobotan pada KNN, klasifikasi akan dilakukan dengan cara pendekatan mencari kasus dengan menghitung kedekatan jarak antara kasus baru dan kasus lama berdasarkan pada pencocokan bobot dari sejumlah atribut atau variabel yang ada. Adapun langkah-langkah yang dilakukan dalam pendekatan Nearest Neighbor adalah sebagai berikut

1. Menentukan bobot atribut/variabel. Bobot antara satu atribut dengan atribut yang lain dapat didefinisikan dengan nilai yang berbeda, berdasarkan tingkat pengaruhnya.
2. Mencari nilai kedekatan antar kriteria pada kasus baru dan lama.
3. Kedekatan biasanya berada pada nilai antara 0 s/d 1. Nilai 0 berarti kedua kasus mutlak mirip.
4. Menghitung jarak kedekatan antara kasus baru dengan kasus lama. Rumus untuk melakukan penghitungan kedekatan antara dua kasus yaitu: Euclidean

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

5. Data yang digunakan diambil pada situs Kaggle pada tahun 2022 [6].

2.8 Akurasi

Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasikan secara benar dengan total sample data testing yang diuji [12]. Setelah menghitung *euclidean*, selanjutnya akan menghitung uji ketepatan nilai yang digunakan untuk menghitung tingkat akurasi pada dataset yang digunakan, berikut persamaannya:

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

2.9 Presisi

Perhitungan Menghitung nilai precision dengan cara membagi jumlah data benar yang bernilai positif (True Positive) dibagi dengan jumlah data benar yang bernilai positif (True Positive) dan data salah yang bernilai positif (False Negative) [11].

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

2.10 Recall

Recall dihitung dengan cara menbagi data benar yang bernilai positif (TP) dengan hasil penjumlahan dari data benar yang bernilai positif dan data salah yang bernilai negatif (FN) [12].

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

2.11 F1 Score

Nilai F1-Score didapat dari perhitungan pembagian hasil dari perkalian precision dan recall dengan hasil penjumlahan precision dan recall kemudian dikalikan dua [12].

$$F_1 = 2 \times \frac{P \times R}{P + R}$$