

## BAB II

### LANDASAN TEORI

#### 2.1 Cuaca

(Puspita & Yulianti, 2016) Mengungkapkan cuaca adalah keadaan udara pada saat tertentu dan di wilayah tertentu yang relatif sempit dan dalam jangka waktu yang singkat. Unsur-unsur yang mempengaruhi cuaca dan iklim adalah :

1. Suhu udara perubahan suhu udara di satu tempat dengan tempat lainnya bergantung pada ketinggian tempat dan letak astronomisnya (lintang). Perubahan suhu karena perbedaan ketinggian jauh lebih cepat daripada perubahan suhu karena perbedaan letak lintang. Biasanya, perubahan suhu terjadi berkisar 0,6 derajat *celsius* tiap kenaikan 100 m, alat ukur adalah Termometer.
2. Tekanan udara tekanan udara adalah berat massa udara pada suatu wilayah. Tekanan udara menunjukkan tenaga yang bekerja untuk menggerakkan massa udara dalam setiap satuan luas tertentu. Tekanan udara semakin rendah jika semakin tinggi dari permukaan laut, alat ukur adalah Barometer.
3. Angin: angin adalah massa udara yang bergerak dari suatu tempat ke tempat lain. Tiupan angin terjadi jika di suatu daerah terdapat perbedaan tekanan udara, yaitu tekanan udara maksimum dan minimum. Angin bergerak dari daerah bertekanan udara maksimum ke minimum, alat ukur adalah Anemometer.
4. Kelembaban udara : kelembaban udara adalah kandungan uap air dalam udara. Uap air yang ada dalam udara berasal dari hasil penguapan air di permukaan bumi, air tanah, atau air yang berasal dari penguapan tumbuh-tumbuhan, alat ukur adalah Higrometer.

## 2.2 Data Mining

*Data Mining* merupakan proses pengekstraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data (Alfa Saleh, 2015). *Data mining* adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya. Secara teknis, *data mining* dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan *field* dari sebuah relasional *database* yang besar (Mabrur & Riani, 2012). Dapat ditarik kesimpulan bahwa *data mining* merupakan proses mengumpulkan data dalam jumlah yang *masive* untuk di ekstraksi menjadi informasi – informasi yang nantinya dapat digunakan. Ada enam kelompok fungsional *data mining*, yaitu:

1. Deskripsi (*description*), memberi gambaran secara ringkas terhadap sejumlah data yang berskala besar dan memiliki banyak jenis. Termasuk di dalamnya metode *Decision Tree*, *Exploratory Data Analysis* dan *Neural Network*.
2. Estimasi (*estimation*), menerka suatu nilai yang belum diketahui, misalnya menerka penghasilan seseorang ketika beberapa informasi mengenai orang tersebut sudah diketahui. Metode yang dapat digunakan adalah *Point Estimation*, *Confidence Interval Estimations*, *Simple Linear Regression*, *Correlation*, dan *Multiple Regression*.
3. Prediksi (*prediction*), memperkirakan suatu nilai di masa mendatang, misalnya memprediksi stok barang tiga tahun ke depan. Yang termasuk fungsi ini antara lain metode *Neural Network*, *Decision Tree*, dan *k-Nearest Neighbor*.
4. Klasifikasi (*Classification*), merupakan proses dalam menemukan suatu model atau fungsi yang dapat membedakan konsep atau kelas

data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Yang termasuk dalam fungsi ini antara lain *Neural Network*, *Decision Tree*, *k-Nearest Neighbor*, dan *Naive Bayes*.

5. Pengelompokan (*Clustering*), pengelompokan untuk mengidentifikasi data dengan karakteristik tertentu. Yang termasuk dalam fungsi ini diantaranya model *Hierarchical Clustering*, metode *K-Means*, dan *Self Organizing Map* (SOM)
6. Asosiasi (*Association*), biasa disebut juga analisis keranjang pasar dimana fungsi ini digunakan untuk mengidentifikasi item-item produk yang kemungkinan dibeli konsumen bersamaan dengan produk lain. Yang termasuk dalam metode atau algoritma dalam fungsi ini antara lain *Apriori*, *Generalized Sequential Pattern (GSP)*, *FP-Growth* dan *GRI Algorithm*

### 2.3 Algoritma *Naive Bayes*

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang ditemukan oleh ilmuwan Inggris Thomas *Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema *Bayes*. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya (Bustami, 2010). Persamaan dari teorema *Bayes* adalah :

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \dots \dots \dots (2.1)$$

Keterangan :

$X$  : Data dengan *class* yang belum diketahui.

$H$  : Hipotesis data  $X$  merupakan suatu *class* spesifik.

$P(H|X)$  : Probabilitas hipotesis  $H$  berdasar kondisi  $X$  (*Posteriori Probability*).

$P(H)$  : Probabilitas hipotesis  $H$  (*Prior Probability*).

$P(X|H)$  : Probabilitas  $X$  berdasarkan kondisi pada hipotesis  $H$ .

$P(X)$  : Probabilitas  $X$ .

Untuk menjelaskan teorema *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema *bayes* di atas disesuaikan sebagai berikut :

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(C|F_1 \dots F_n)} \dots \dots \dots (2.2)$$

Dimana Variabel  $C$  merepresentasikan kelas, sementara variabel  $F_1 \dots F_n$  merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas  $C$  (*Posterior*) adalah peluang munculnya kelas  $C$  (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas  $C$  (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana sebagai berikut :

$$Posterior = \frac{Prior \times likelihood}{evidence} \dots\dots\dots(2.3)$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus *Bayes* tersebut dilakukan dengan menjabarkan  $(C/F_1, \dots, F_n)$  menggunakan aturan perkalian sebagai berikut :

$$\begin{aligned} P(C|F_1 \dots F_n) &= P(C)P(C|F_1 \dots F_n|C) \\ &= P(C)P(C|F_1|C)P(C|F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\ &= P(C)P(C|F_1|C)P(C|F_2|C, F_1)P(F_3|C, F_1, F_2), P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\ &= P(C)P(C|F_1|C)P(C|F_2|C, F_1) \\ &P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \dots\dots\dots(2.4) \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya factor - faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan, disinilah digunakan asumsi independensi yang sangat tinggi (*naif*), bahwa masing-masing petunjuk  $(F_1, F_2 \dots F_n)$  saling bebas (*independen*) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \dots\dots\dots(2.5)$$

Untuk  $i \neq j$ , sehingga

$$P(F_i|C, F_j) = P(F_i|C) \dots\dots\dots(2.6)$$

Dari persamaan diatas dapat disimpulkan bahwa asumsi independensi naif tersebut membuat syarat peluang menjadi sederhana,

sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran  $P(C|F_1, \dots, F_n)$  dapat disederhanakan menjadi :

$$P(C|F_1 \dots F_n) = P(C)P(F_1|C)P(F_2|C)P(F_3|C) \dots$$

$$P(C) \prod_{i=1}^n P(F_i|C) \dots \dots \dots (2.7)$$

Persamaan diatas merupakan model dari teorema *Naive Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss* :

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \dots \dots \dots (2.8)$$

Keterangan :

$P$  : Peluang.

$X_i$  : Atribut ke  $i$ .

$x_i$  : Nilai atribut ke  $i$ .

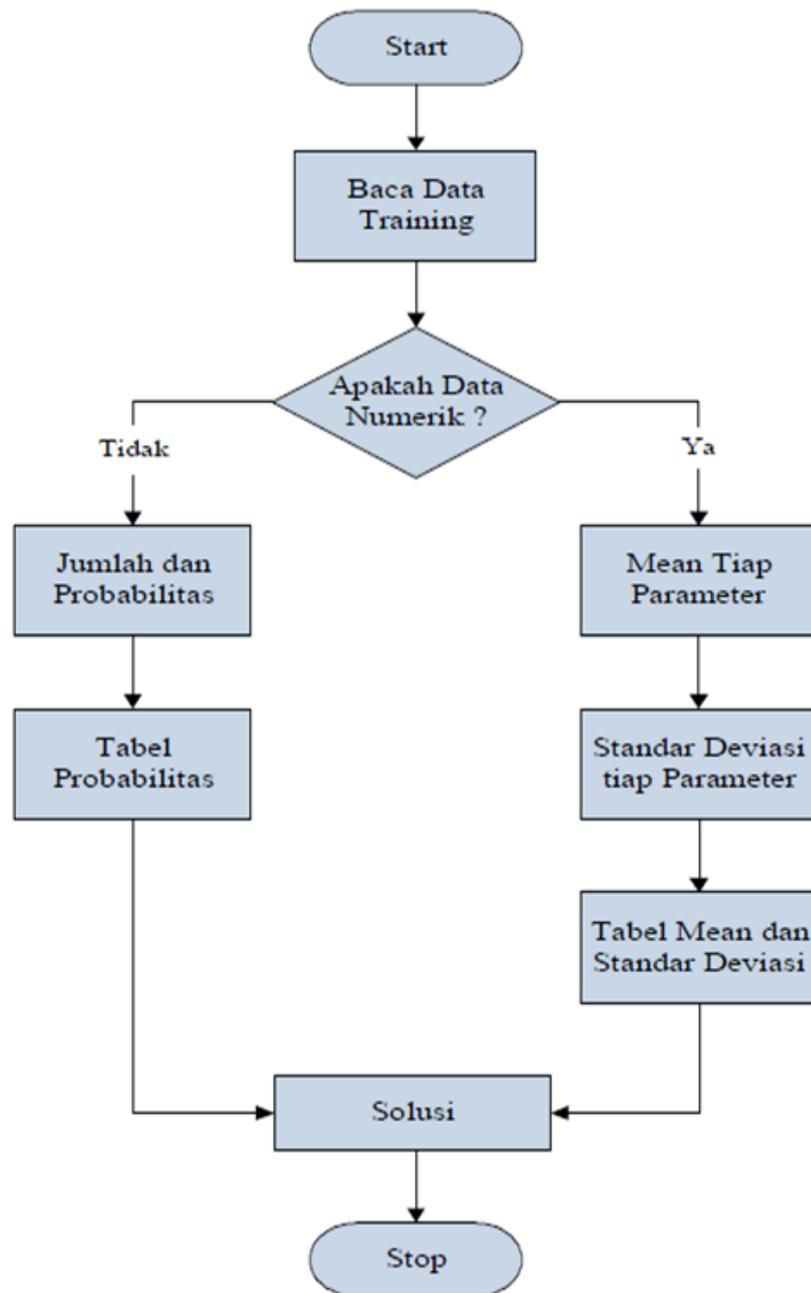
$Y$  : Kelas yang dicari.

$y_j$  : Sub kelas  $Y$  yang dicari

$\mu$  : *Mean*, menyatakan rata-rata dari seluruh atribut

$\alpha$  : *Deviasi standar*, menyatakan varian dari seluruh atribut

Adapun alur dari metode *Naive Bayes* adalah sebagai berikut :



Gambar 2.1. Alur Metode *Naive Bayes*.

1. Baca data training
2. Hitung Jumlah dan probabilitas, namun apabila data numerik maka:

- a. Cari nilai mean dan standar deviasi dari masing-masing parameter yang merupakan data numerik.

Adapun persamaan yang digunakan untuk menghitung nilai rata – rata hitung (*mean*) dapat dilihat sebagai berikut :

$$\mu = \frac{\sum_{i=1}^n x_i}{n}$$

atau

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \dots\dots\dots(2.9)$$

di mana :

$\mu$  : rata – rata hitung (*mean*)

$x_i$  : nilai sample ke -*i*

$n$  : jumlah sampel

Dan persamaan untuk menghitung nilai simpangan baku (standar deviasi) dapat dilihat sebagai berikut:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \dots\dots\dots(2.10)$$

di mana :

$\sigma$  : standar deviasi

$x_i$  : nilai x ke -*i*

$\mu$  : rata-rata hitung

$n$  : jumlah sampel

- b. Cari nilai probabilitas dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
3. Mendapatkan nilai dalam tabel *mean*, standart deviasi dan probabilitas.
  4. Solusi kemudian dihasilkan.

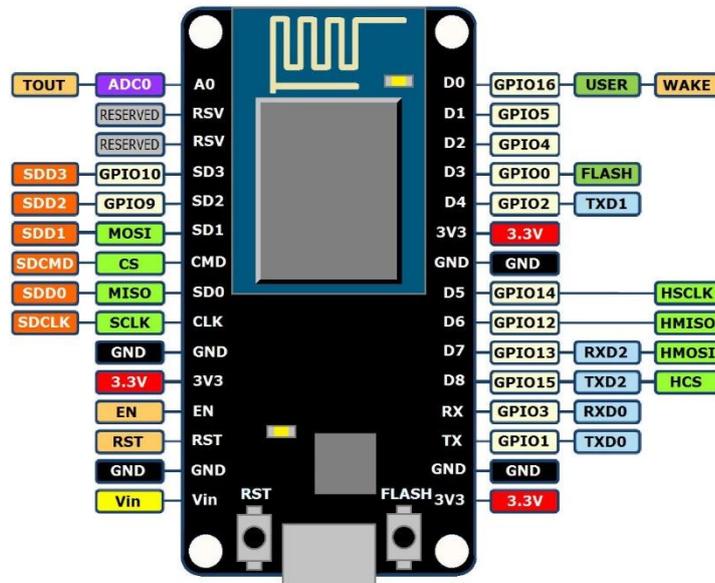
## 2.4 Prediksi

Prediksi adalah usaha menduga atau memperkirakan sesuatu yang akan terjadi di waktu mendatang dengan memanfaatkan berbagai informasi yang relevan pada waktu-waktu sebelumnya (historis) melalui suatu metode ilmiah. Prediksi (Forecasting) sangat membantu dalam perencanaan dan pengambilan keputusan dalam suatu aktivitas. Analisis sangat penting dalam pembelajaran, karena penelitian menjadi lebih tepat dan terarah (Purba & Wanto, 2018). Prediksi menjadi sangat penting karena proses analisis data yang membuat penelitian menjadi lebih tepat.

## 2.5 NodeMCU Esp8266

NodeMCU Esp8266 merupakan salah satu modul *board* yang dapat berfungsi dengan *arduino* khususnya untuk *project* yang mengusung konsep *IoT*. NodeMCU Esp8266 dapat *running stand-alone* karena sudah terdapat CPU yang dapat diprogram melalui serial *port* atau via OTA serta transfer program secara *wireless*

NodeMCU Esp8266 merupakan sebuah open source platform IoT dan pengembangan kit yang menggunakan bahasa pemrograman Lua untuk membantu makers dalam membuat prototype produk IoT atau bisa dengan memakai sketch dengan *arduino* IDE. Pengembangan Kit ini didasarkan pada modul ESP8266, yang mengintegrasikan GPIO, PWM (Pulse Width Modulation), IIC, 1-Wire dan ADC (Analog to Digital Converter) semua dalam satu board. Keunikan dari Nodemcu ini sendiri yaitu Board nya yang berukuran sangat kecil yaitu panjang 4.83cm, lebar 2.54cm, dan dengan berat 7 gram. Tapi walaupun ukurannya yang kecil, board ini sudah dilengkapi dengan fitur *wifi* dan *firmware*-nya yang bersifat opensource. (Marsda & Yogyakarta, 2016).



Gambar 2.2 Module NodeMCU Esp8266.

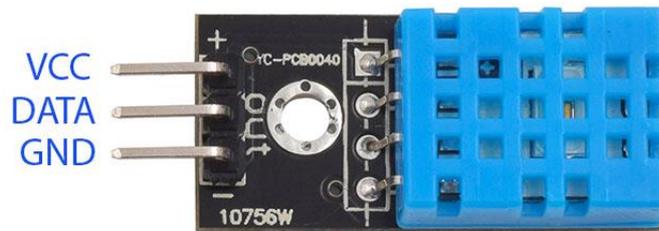
#### Spesifikasi dari NodeMCU Esp8266

1. Beroperasi pada tegangan operasional 3,3 – 5V
2. Ukuran board 57mm x 30mm
3. Memiliki 13 pin GPIO
4. *Memory flash* : 4Mbyte
5. *Clock speed* : 80/26/24 MHz
6. *Wifi IEEE* 802.11 b/g/n
7. Frekuensi 2.4 Ghz – 22.5 Ghz

## 2.5 Sensor DHT 11

Sedangkan untuk sensor suhu dan kelembaban digunakan modul sensor DHT11. Modul ini dapat digunakan sebagai alat pengindra suhu dan kelembaban dalam aplikasi pemantau suhu dan kelembaban relatif ruangan.

DHT11 memiliki output digital yang sudah terkalibrasi. Koefisien kalibrasinya telah diprogramkan ke dalam OTP memori. Koefisien tersebut akan digunakan untuk mengkalibrasi keluaran dari sensor selama proses pengukuran. DHT11 menggunakan *single write serial interface* yang cukup cepat dan mudah. Ukuran sensor yang kecil, kebutuhan konsumsi daya yang rendah dan mampu mentransmisikan *output*-nya dalam jarak 20 meter(D-robotics UK, 2010).



Gambar 2.3 *Module* sensor DHT 11.

DHT11 merupakan sensor yang cukup murah dan mudah didapatkan di pasaran. Sensor ini memiliki spesifikasi diantaranya (D-robotics UK, 2010):

1. *Humidity Range*: 20-90% RH
2. *Humidity Accuracy*:  $\pm 5\%$  RH
3. *Temperature Range*: 0-50 °C
4. *Temperature Accuracy*:  $\pm 2\%$  °C
5. *Operating Voltage*: 3V to 5.5

## 2.6 Sensor Anemometer

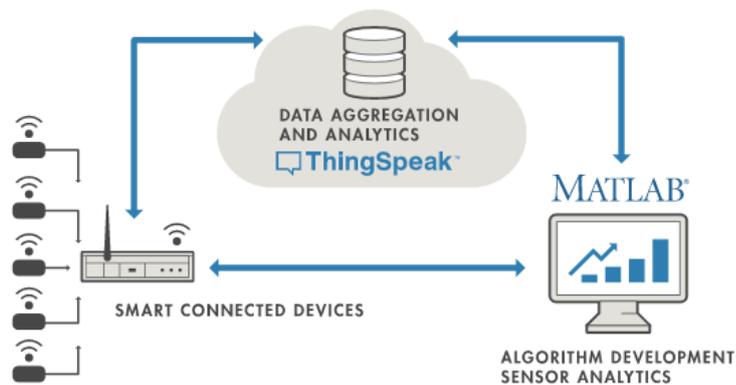
Anemometer merupakan sensor angin untuk mengukur kecepatan angin di sekitarnya dan juga banyak digunakan pada stasiun pengukuran cuaca. Pengukuran kecepatan/RPM angin yang bisa digunakan ada beberapa metode yang digunakan, salah satunya menghitung waktu yang terjadi tiap munculnya sinyal pulsa.



Gambar 2.4 *Module* Sensor Kecepatan Angin Anemometer

## 2.7 Cloud Server Thingspeak

*ThingSpeak* adalah aplikasi *open-source Internet of Things* dan API untuk menyimpan dan mengambil data dari berbagai hal menggunakan protokol HTTP melalui Internet atau melalui *Local Area Network*. *ThingSpeak* telah terintegrasi dengan *software MATLAB* dari *Mathworks*, menjadikan *user thingspeak* dapat melakukan analisa berupa visual dari data yang telah di upload ke web server tersebut, berikut adalah cara kerja dari *server thingspeak*.



Gambar 2.5 Cloud server thingspeak