

BAB II

LANDASAN TEORI

2.1 Hama Tanaman Padi

Budidaya tanaman padi banyak mengalami kendala terutama tanaman padi yang ditanam di daerah pasang surut. Kendala tersebut antara lain serangan hama dan penyakit tanaman. Berdasarkan pengamatan penulis penelitian di Desa Turi Kecamatan Turi Kabupaten Lamongan bulan juni 2015 (belum dipublikasikan), ada beberapa hama yang menyerang tanaman padi diantaranya wereng coklat, penggerak batang, wereng hijau, kepinding tanah, walang sangit dan belalang.

Serangan hama dan penyakit yang timbul dilapangan dapat dipengaruhi oleh lingkungan misalnya keadaan air, kemasaman tanah, suhu, kelembaban udara, penggunaan bibit unggul dan cara budidaya. Penggunaan varietas tanaman dan pemupukan yang tidak tepat dapat memicu timbulnya serangan hama. Iklim atau musim yang tidak menentu dapat mempengaruhi tingkat serangan hama (Hasibuan, 2008).

Berikut ini adalah beberapa hama tanaman yang menyerang tanaman padi:

1. Wereng Coklat

Wereng coklat (WCK) menjadi salah satu hama utama tanaman padi di Indonesia sejak pertengahan tahun 1970-an. Penggunaan pestisida yang melanggar kaidah-kaidah PHT (tepat jenis, tepat dosis, dan tepat waktu aplikasi) turut memicu ledakan wereng coklat. Tergantung pada tingkat kerusakan, serangan wereng coklat dapat meningkatkan kerugian hasil padi dari hanya beberapa kuintal gabah. Selain itu, juga merupakan penyebab penyakit virus kerdil rumput dan kerdil hampa. Dengan menghisap cairan dari dalam jaringan pengangkutan tanaman padi, WCK dapat menimbulkan kerusakan ringan sampai berat pada hampir semua fase tumbuh, sejak fase bibit, anakan, sampai fase masak susu (pengisian). Gejala WCK pada individu rumpun dapat terlihat dari daun- daun yang menguning, kemudian tanaman mengering dengan cepat (seperti terbakar).

Dalam suatu hamparan, gejala terlihat sebagai bentuk lingkaran, yang menunjukkan pola penyebaran WCK yang dimulai dari satu titik, kemudian meyebar ke segala arah dalam bentuk lingkaran. Dalam keadaan demikian, populasi WCK biasanya sudah sangat tinggi. Wereng ini menyerang tanaman padi pada bagian batangnya. Hama wereng cokelat terdiri dari 2 jenis *Nilaparvata lugens*, yang berciri panjang badan berkisar 3-4 mm. Pada bagian punggung terdapat 3 buah garis samar-samar. *Sogatela furcifera* yang panjang badanya kurang lebih 3-4 mm dan pada punggungnya terdapat 3 buah baris berwarna cokelat hitam dengan warna putih disebelah tengahnya. (Hudi, 1989: 17)

2. Penggerak Batang

Penggerak batang termasuk hama paling sering menimbulkan kerusakan berat dan kehilangan hasil yang tinggi. Di lapangan, keberadaan hama ini ditandai oleh kehadiran ngengat (kupu-kupu), kematian tunas-tunas padi (sundep, dead heart), kematian malai (beluk, white head). Hama ini dapat merusak tanaman pada semua fase tumbuh, baik pada saat di pembibitan, fase anakan, maupun fase berbunga. Bila serangan terjadi pada pembibitan sampai fase anakan, hama ini disebut sundep dan jika terjadi pada saat berbunga, disebut beluk. Sampai saat ini belum ada varietas yang tahan penggerak batang. Oleh karena itu gejala serangan hama ini perlu diwaspadai, terutama pada pertanaman musim hujan. Waktu tanam yang tepat, merupakan cara yang efektif untuk menghindari serangan penggerak batang. Penggerak batang padi di bedakan menjadi beberapa macam di antaranya:

- a. Penggerak batang putih (*Tryporiza innotata*)
 - b. Penggerak batang kuning (*Tryporiza intertulas*)
 - c. Penggerak batang bergaris (*Chillo suppressalis*)
 - d. Penggerak batang merah (*Sesamia inferens*) (Sartono dan Indriati, 2007)
- Keempat jenis penggerak tersebut bekerja dengan cara yang sama. Kerusakan ditimbulkan pada stadium vegetatif dan generatif. Serangan pada stadium vegetatif menimbulkan gejala yang disebut sundep karena

pucuk tanaman mati karena dimakan larva. Sedangkan pada stadium generatif menimbulkan gejala beluk yaitu malai menjadi hampa berwarna putih dan berdiri tegak karena tangkai malai putus di gerak. (Y.T. Prasetio, 2002: 31).

3. Wereng Hijau

Peran wereng hijau (WH) dalam sistem pertanaman padi menjadi penting oleh karena WH merupakan penyebab penyakit tungro, yang merupakan salah satu penyakit virus terpenting di Indonesia. Kemampuan WH sebagai penghambat dalam sistem pertanian padi sangat tergantung pada penyakit virus tungro. Sebagai hama, WH banyak ditemukan pada sistem sawah irigasi teknis, ekosistem tadah hujan. WH menghisap cairan dari dalam daun bagian pinggir, tidak menyukai pelepah, ataupun daun-daun bagian tengah. WH menyebabkan daun-daun padi berwarna kuning sampai kuning oranye, penurunan jumlah anakan, dan pertumbuhan tanaman yang terhambat (memendek). Pemupukan unsur nitrogen yang tinggi sangat memicu perkembangan WH. Disebut wereng padi hijau karena warnanya memang hijau. serangga ini masih muda berwarna hijau muda, sedangkan yang dewasa mempunyai bintik-bintik hitam pada ujung dan tengah sayap. Pada serangga jantan bintik-bintik ini sangat jelas. Wereng ini menghisap daun dan juga menularkan virus, dibanding dengan wereng coklat kerusakan yang ditimbulkan tidak begitu berarti. (Pracaya, 2007: 76)

4. Kepinding Tanah

Jenis kepinding tanah ini sering mencapai jumlah berlimpah dan karena pengendalian dengan pestisida sulit dilakukan, hama ini sering menimbulkan kerugian besar. Pada siang hari, kepinding tua yang hitam coklat mengkilat bergerombol di pangkal batang padi, persis di batas genangan air pada siang hari. Pada malam hari mereka naik batang padi dan mengisap cairan dari dalam jaringan tanaman. Selama musim kemarau, kepinding tanah menghabiskan waktunya di belahan tanah-tanah yang ditumbuhi rumput. Kepinding tanah dapat terbang ke pertanaman

padi dan berkembang biak dalam beberapa generasi. Kepinding dewasa dapat berpindah menempuh jarak yang jauh. Kepinding dewasa tertarik pada sinar dengan intensitas yang kuat dan penangkapan tertinggi diperoleh pada saat bulan purnama. Pengisapan cairan oleh kepinding tanah menyebabkan warna tanaman berubah menjadi coklat kemerahan atau kuning. Buku pada batang merupakan tempat isapan yang disukai karena menyimpan banyak cairan. Pengisapan oleh kepinding tanah pada fase anakan, menyebabkan jumlah anakan berkurang dan pertumbuhan terhambat (kerdil).

5. Walang Sangit

Walang sangit merupakan hama yang umum merusak bulir padi pada fase pemasakan. Mekanisme merusaknya yaitu menghisap butiran gabah. Apabila diganggu, serangga akan mempertahankan diri dengan mengeluarkan bau. Selain sebagai mekanisme pertahanan diri, bau yang dikeluarkan juga digunakan untuk menarik walang sangit lain dari spesies yang sama. Kerusakan yang ditimbulkannya menyebabkan beras berubah warna dan mengapur, serta gabah menjadi hampa. Binatang ini berbau hidup bersembunyi di rerumputan, tuton, paspalum, alang alang, sehingga berinvansi pada padi muda ketika bunting, berbunga atau berbuah. (Amelia, 2007 : 4). Hama ini aktif menyerang pada pagi dan sore hari. Walang sangit merusak tanaman padi dengan cara menghisap buah padi saat masih masak susu sehingga buah menjadi kopong dan perkembangannya kurang baik. (Y.T. Prasetio, 2002: 31).

6. Belalang

Belalang adalah serangga herbivora yang terkenal sebagai hama dengan kemampuan melompat mumpuni dapat mencapai jarak hingga 20 kali panjang tubuhnya. Klasifikasi menurut Kalshoven, L.G.E, (1981)

Kelas : Insekta,

Ordo : Orthoptera,

Famili : Acridida,

Genus : Locusta,

Species : *Locusta migratoria*

Gejala serangan yang ditimbulkan adalah terdapat robekan pada daun, dan pada serangan yang hebat dapat terlihat tinggal tulang-tulang daun saja. Gejala serangan belalang tidak spesifik, bergantung pada tipe tanaman yang diserang dan tingkat populasi. Serangan pada daun biasanya bagian daun pertama. Hampir keseluruhan daun habis termasuk tulang daun.

2.2 Sistem Pakar

Sistem pakar adalah sistem yang menggunakan pengetahuan manusia yang dimasukkan kedalam komputer. Sehingga setiap orang dapat menggunakannya untuk dapat memecahkan berbagai masalah yang bersifat spesifik (Turban dan Aronson, 2001). Sistem pakar disusun oleh dua bagian utama, yaitu bagian lingkungan pengembangan (*development environment*) dan lingkungan konsultasi (*consulation environment*) (Turban, 1995). Seorang pakar yang dimaksud disini adalah orang yang mempunyai keahlian dalam bidang tertentu, yaitu pakar yang mempunyai *knowledge* atau kemampuan khusus yang tidak dimiliki orang lain.

Menurut (Turban, 1995). Sistem pakar (*expert system*) adalah sistem berbasis komputer yang berusaha mengadopsi pengetahuan manusia ke komputer, agar komputer dapat menyelesaikan masalah seperti yang biasa dilakukan oleh para ahli melalui antarmuka (*Interface*). Pada dasarnya sistem pakar diterapkan untuk mendukung aktivitas pemecahan masalah. Beberapa aktivitas pemecahan masalah yang dimaksud antara lain: pembuatan keputusan (*decision making*). Pemaduan pengetahuan (*knowledge fusing*). Pembuatan desain (*designing*). Perencanaan (*planning*). Prakiraan (*forecasting*). Pengaturan (*regulating*). Pengendalian (*controlling*). Diagnosis (*diagnosing*). Perumusan (*prescribing*). Penjelasan (*explaining*). Pemberian nasihat (*advising*) dan pelatihan (*toturing*). Sistem pakar dibuat pada wilayah pengetahuan tertentu untuk suatu kepakaran tertentu yang mendekati kemampuan manusia di salah satu bidang. Sistem pakar mencoba mencari solusi yang tepat sebagaimana yang dilakukan oleh seorang

pakar. Selain itu, sistem pakar juga dapat memberikan penjelasan terhadap langkah yang diambil dan memberikan alasan atas saran atau kesimpulan.

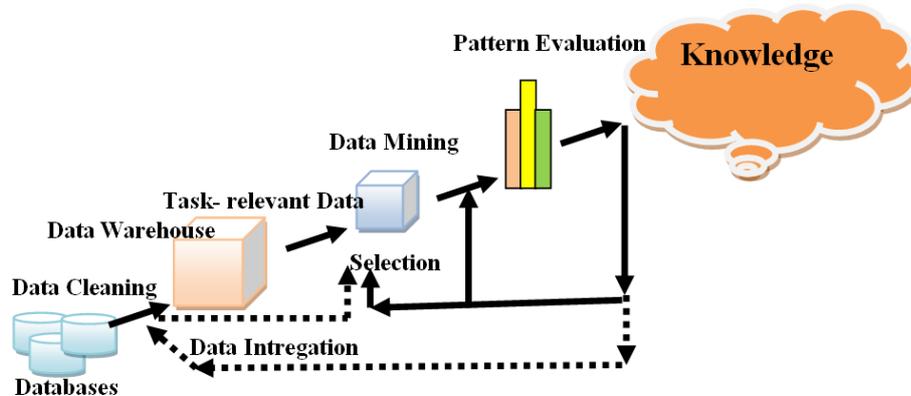
2.3 Pengertian Data Mining

Menurut Han dan Kamber (2011:6) menjelaskan bahwa data mining merupakan pemilihan atau menggali pengetahuan dari jumlah data yang banyak. Berbeda dengan segall, Guha dan Noris (2008:127) menjelaskan data mining disebut penemuan pengetahuan atau menemukan pola tersembunyi dalam data. Data mining adalah proses menganalisis data dari perspektif yang berbeda dan meringkas menjadi informasi yang berguna. Bisa disimpulkan data mining adalah proses menganalisis data yang banyak dan membuat suatu pola untuk menjadi informasi yang berguna.

Data mining adalah sebuah proses untuk menemukan pola atau pengetahuan yang bermanfaat secara otomatis atau semi otomatis dari sekumpulan data dalam jumlah besar. *Data mining* hadir dianggap sebagai bagian dari Knowledge Discovery in Database (*KDD*) yaitu sebuah proses mencari pengetahuan yang bermanfaat dari data. *KDD* terdiri dari beberapa langkah yaitu :

1. Pembersihan data (membuang nois dan data yang tidak konsisten)
2. Intregasi data (penggabungan data dari beberapa sumber)
3. Seleksi data (memilih data yang relevan digunakan untuk analisa)
4. *Data mining*
5. Evaluasi pola
6. Presentasi pengetahuan dengan teknik visualisasi

Berikut adalah gambar tahapan – tahapan yang ada didalam *KDD*.



Gambar 2.1. Proses di dalam *Knowledge Discovery in Database*

Menurut Gartner Group data mining adalah suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Data mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah bahwa kenyataan data mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, data mining bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani:

- Jumlah data yang sangat besar.
- Dimensi data yang tinggi.
- Data yang heterogen berbeda sifat.

Menurut para ahli, data mining merupakan sebuah analisa dari observasi data dalam jumlah besar untuk menemukan hubungan yang tidak diketahui sebelumnya dan metode baru untuk meringkas data agar mudah dipahami serta kegunaannya untuk pemilik data.

Data-data yang ada tidak dapat langsung diolah dengan menggunakan sistem data mining. Data-data tersebut harus dipersiapkan terlebih dahulu agar hasil yang diperoleh dapat lebih maksimal. Proses persiapan data ini sendiri dapat

mencapai 60 % dari keseluruhan proses dalam data mining. Karena itu sebenarnya data mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelegent*), *machine learning*, statistik dan database. Beberapa metode yang sering disebut dalam literatur data mining antara lain *clustering*, *clasification*, *assosiation rules mining*, *neural network genetic algorithm* dan lain-lain (Pramudiono 2007).

2.4 Tahap-tahap Data Mining

Sebagai suatu rangkaian proses, *data mining* dapat dibagi menjadi beberapa tahap yang diilustrasikan di Gambar, tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung dengan perantara *knowlege based*.

Tahap-tahap data mining ada 6 yaitu :

1. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidan konsisten atau data yang tidan relevan. Pada umumnya data diperoleh, baik dari database suatu erusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan lebih baik dibuang. Pembersihan data juga akan mempengaruhi performasi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. integrasi data (*data integraton*)

integrasi data merupakan penggabungan data dari berbagai database kedalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomer pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa

menghasilkan hasil yang menyimpang bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

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. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

4. Transformasi data (*data transformation*)

Data diubah atau digabung kedalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standart seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal, karenanya data berupa angka numbering yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data

5. proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi Pola (*Pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik kedalam knowlege based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikan umpan balik untuk

memprediksi proses data mining, mencoba metode data mining yang sesuai, atau menerima hasil ini sebagai suatu hal yang diluar dugaan yang mungkin bermanfaat.

2.5 Pengelompokan Data Mining

Pada umumnya *data mining* dapat di kelompokkan ke dalam dua kategori yaitu: deskriptif dan prediktif. Deskriptif bertujuan untuk mencari pola yang dapat dimengerti oleh manusia yang menjelaskan karakteristik dari data. Prediktif menggunakan ciri-ciri tertentu dari data yang melakukan prediksi.

pengelompokan yang ada dalam *data mining* adalah sebagai berikut:

1. *Classification*

Klasifikasi (*Classification*) merupakan proses untuk menemukan sekumpulan model yang menjelaskan dan membedakan kelas-kelas data, sehingga model tersebut dapat digunakan untuk memprediksi nilai suatu kelas yang belum diketahui pada sebuah objek. Untuk mendapatkan model, kita harus melakukan analisis terhadap data latih (*training set*). Sedangkan data uji (*test set*) digunakan untuk mengetahui tingkat akurasi dari model yang dihasilkan. Klasifikasi dapat digunakan untuk memprediksi nama atau nilai kelas dari suatu obyek data.

2. *Clustering*

Pengelompokan (*Clustering*) merupakan proses untuk melakukan segmentasi. Digunakan untuk melakukan pengelompokan secara alami terhadap atribut suatu set data, termasuk kedalam supervised task. Contoh *clustering* seperti mengelompokkan dokumen berdasarkan topiknya.

3. *Assosiation*

Untuk menghasilkan sejumlah rule yang menjelaskan sejumlah data yang berhubung kuat satu dengan yang lainnya. Sebagai contoh *assosiation analysis* dapat digunakan untuk menentukan produk yang datang secara bersamaan oleh banyak pelanggan, atau bisa juga disebut dengan basket analysis.

4. *Regression*

Regression mirip dengan klasifikasi. Perbedaan utamanya adalah terletak pada atribut yang diprediksi berupa nilai yang kontinyu.

5. *Forecasting*

Prediksi (*Forecasting*) berfungsi untuk melakukan kejadian yang akan datang berdasarkan data sejarah yang ada.

6. *Sequence Analysis*

Tujuan dari metode ini adalah untuk mengenali pola dari data diskrit. Sebagai contoh adalah menemukan kelompok gen dengan tingkat ekspresi yang mirip.

7. *Deviation Analysis*

Untuk menemukan penyebab perbedaan antara data yang satu dengan data yang lain dan biasa disebut dengan *outlier detection*. Sebagai contoh adalah apakah sudah terjadi peniapaun terhadap pengguna kartu kredit dengan melihat catatan transaksi yang tersimpan dalam penggunaan basis data perusahaan kartu kredit.

2.6 Naive Bayes

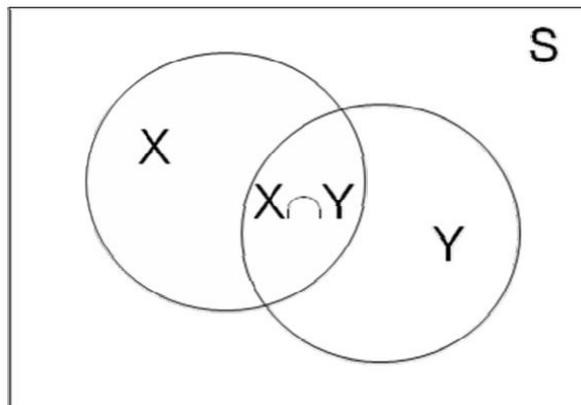
Naive Bayes adalah metode probabilitas dan statistic (memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya). sebuah metode *information retrieval* yang menggunakan pendekatan probabilistik dalam menginferensi, yakni berbasis pada teorema Bayes secara umum. Menurut Olson dan Delen (2008:102) menjelaskan naive bayes untuk setiap keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi objek. Aplikasinya yang paling banyak digunakan yaitu untuk klasifikasi teks. Arti kata "*naive*" di sini adalah metode *Naive Bayes* mengasumsikan bahwa probabilitas kemunculan sebuah kata independen terhadap posisinya di dalam teks. Tujuan dari asumsi independensi posisi kata adalah untuk mengurangi kompleksitas dalam perhitungan nilai-nilai probabilitas bersyarat pada teorema Bayes secara umum.

Naive Bayes juga merupakan salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk machine learning dan data mining. Performa naive bayes yang kompetitif dalam proses klasifikasi walaupun menggunakan asumsi keidependenan atribut (tidak ada kaitan antar atribut). Asumsi keidependenan atribut ini pada data sebenarnya jarang terjadi, namun walaupun

asumsi keindependenan atribut tersebut dilanggar performa pengklasifikasian naïve bayes cukup tinggi, hal ini dibuktikan pada berbagai penelitian empiris.

Metode bayes merupakan pendekatan statistic untuk melakukan inferensi induksi pada persoalan klasifikasi. Pertama kali dibahas terlebih dahulu tentang konsep dasar dan definisi pada teorema bayes, kemudian menggunakan teorema ini untuk melakukan klasifikasi dalam Data Mining. Metode Bayes menggunakan probabilitas bersyarat sebagai dasarnya. Dalam ilmu probabilitas bersyarat dinyatakan sebagai:

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$$



Probabilitas X di dalam Y adalah probabilitas inteseksi X dan Y dari probabilitas Y, atau dengan bahasa lain $P(X|Y)$ adalah prosentase banyaknya X di dalam Y.

Bayes' adalah teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema bayes (atau aturan bayes) dengan asumsi independensi (ketidak tergantungan) yang kuat (naif). Dengan kata lain, dalam *Naive Bayes*, model yang digunakan adalah “model fitur independen”. Dalam Bayes (terutama Naive Bayes) maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama. Teorema yang berlaku umum di semua interpretasi soal memainkan peran penting dalam perdebatan sekitar dasar statistik: frequentist dan Bayesian interpretasi tidak setuju tentang cara dalam kendala yang harus diberikan dalam aplikasi. Frequentists kendala untuk menetapkan peristiwa acak sesuai dengan frekuensi kejadian atau untuk subset

dari populasi sebagai proporsi keseluruhan, sementara Bayesians menjelaskan kendala dalam hal kepercayaan dan derajat ketidakpastian. Artikel tentang kemungkinan Bayesian dan frequentist kemungkinan mendiskusikan perdebatan ini lebih lanjut.

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} \dots\dots\dots \text{Persamaan (2.1)}$$

teorema ini menjadi dasar keputusan modern

Dimana:

- a. $P(H|E)$ = Probabilitas posterior bersyarat (*Conditional Probability*) suatu hipotesis H terjadi jika diberikan evidence/bukti E terjadi.
- b. $P(E|H)$ = Probabilitas sebuah evidence E terjadi akan mempengaruhi hipotesis H.
- c. $P(H)$ = Probabilitas awal (priori) hipotesis H terjadi tanpa memandang evidence apapun.
- d. $P(E)$ = Probabilitas awal (priori) evidence E terjadi tanpa memandang hipotesis/evidence yang lain.

Ide dasar dari aturan Bayes adalah bahwa hasil dari hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan pada beberapa bukti (E) yang diamati. Ada beberapa hal penting dari aturan Bayes tersebut yaitu:

- 1. Sebuah probabilitas awal/prior H atau $P(H)$ adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
- 2. Sebuah probabilitas akhir H atau $P(H|E)$ adalah probabilitas dari suatu hipotesis setelah bukti diamati.

Naive Bayes dinyatakan sebagai sebuah hipotesa yang disebut dengan HMAP (*Hypothesis Maximum Appriori Probability*). Secara matematis HMAP dirumuskan seperti persamaan 2.2:

$$\begin{aligned} H_{\text{Map}} &= \arg \max P(h|e) \\ &= \arg \max \frac{P(e|h) \cdot e(h)}{p(e)} \dots\dots\dots (2.2) \\ &= \arg \max p(e|h) * p(h) \end{aligned}$$

Dalam konteks data mining atau *machine learning*, data e adalah set training, dan h adalah ruang dimana fungsi yang akan ditemukan tersebut terletak. HMAP juga seringkali dituliskan seperti persamaan 2.3 [5]:

$$H_{Map} = \arg \max_j P(a_1, a_2, a_3, \dots, a_n | h_j) * P(h_j) \dots\dots\dots (2.3)$$

Dimana:

- a. H_{Map} = Nilai *output* hasil klasifikasi *Naive Bayes*.
- b. $P(a_1, a_2, a_3, \dots, a_n | h_j)$ = Peluang a .
- c. $P(h_j)$ = Keadaan atau kategori j .

HMAP menyatakan hipotesa yang diambil berdasarkan nilai probabilitas berdasarkan kondisi prior yang diketahui.

2.6.1 *Naive Bayes* Untuk Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya kedalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu: pertama, pembangunan model sebagai *prototype* untuk disimpan sebagai memori dan kedua, penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada suatu objek data lain agar diketahui dikelas mana objek data tersebut dalam model yang mudah disimpan.

Klasifikasi dengan *Naive Bayes* bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik *Naive Bayes* sebagai berikut:

1. Metode *Naive Bayes* bekerja teguh (*robust*) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outliner*). *Naive Bayes* juga bisa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi.
2. Tangguh menghadapi atribut yang tidak relevan.

- Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi *Naive Bayes* karena asumsi independensi atribut tersebut sudah tidak ada.

2.6.2 Contoh Perhitungan

Berikut ini akan dijelaskan ilustrasi dari alur proses perhitungan algoritma *Naive Bayes* data set yang digunakan pada contoh ini adalah data untuk menentukan Ya atau Tidak dengan beberapa atribut yaitu atribut *Day*, *Cuaca*, *Temperatur* dan *Kecepatan Angin*. Dimana atribut tersebut bertipe kategorikal. Sedangkan kolom Ya Tidak adalah kelas tujuannya atau label kelasnya.

Tabel 2.1 Contoh Data Set

Day	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
D1	Cerah	Normal	Pelan	Ya
D2	Cerah	Normal	Pelan	Ya
D3	Hujan	Tinggi	Pelan	Tidak
D4	Cerah	Normal	Kencang	Ya
D5	Hujan	Tinggi	Kencang	Tidak
D6	Cerah	Normal	Pelan	Ya

Dari data diatas dapat dinyatakan pengertian tentang data konsisten dan tidak konsisten.

- Data konsisten
Suatu data disebut konsisten, jika setiap atributnya memiliki nilai target yang sama.

Day	Cuaca	Temperatur	Kecepatan Angin	Berolah raga
#	Cerah	Normal	Pelan	Ya
#	Cerah	Normal	Pelan	Ya
#	Cerah	Normal	Kencang	Ya

Atribut Cuaca, Temperatur mempunyai nilai target yang sama (Berolah-raga), maka data ini adalah data yang konsisten.

- Data tidak konsisten

Suatu data disebut tidak konsisten, jika setiap atributnya memiliki nilai target yang sama, tapi nilai yang berbeda untuk atributnya.

Day	Cuaca	Temperatur	Kecepatan Angin	Berolah raga
#	Cerah	Normal	Pelan	Ya
#	Cerah	Tinggi	Pelan	Ya
#	Hujan	Normal	Kencang	Ya

Tidak satupun atribut yang mempunyai nilai yang sama dalam satu keputusan (berolah-raga).

- Data bias

Suatu data disebut data bias jika memiliki target atau keputusan yang berbeda sedangkan instance pada semua atributnya sama

Day	Cuaca	Temperatur	Kecepatan Angin	Berolah raga
#	Cerah	Normal	Pelan	Ya
#	Cerah	Normal	Pelan	Ya
#	Cerah	Normal	Pelan	Tidak

Dataset yang digunakan sebagai data training bias bersifat konsisten, tidak konsisten atau bias. Data set tersebut digunakan untuk memprediksi suatu kejadian dari fakta atau kenyataan yang diketahui sebelumnya. Prediksi dari suatu kejadian disebut Hipotesa.

Hipotesa dituliskan dengan:

$$H(\text{atribut}_1, \text{atribut}_2, \dots, \text{atribut}_n) = \text{keputusan}$$

Menghitung Hipotesa:

Day	Cuaca	Temperatur	Kecepatan Angin	Berolah-raga
D1	Cerah	Normal	Pelan	Ya
D2	Cerah	Normal	Pelan	Ya
D3	Hujan	Tinggi	Pelan	Tidak
D4	Cerah	Normal	Kencang	Ya
D5	Hujan	Tinggi	Kencang	Tidak
D6	Cerah	Normal	Pelan	Ya

1. $H(\text{cuaca}=\text{cerah}, \text{temperature}=\text{normal}, \text{kec.angin}=\text{pelan})=\text{ya}$

Hipotesa ini menunjukkan bahawa keputusan untuk berolah raga bila cuaca=cerah, temperature=normal, kec.angin=pelan, untuk singkatnya dituliskan hanya instance pada setiap atribut dengan **H(cerah, normal, pelan)=ya**

2. $H(\text{cuaca}=\text{cerah}, \text{kec.angin}=\text{pelan})=\text{ya}$

Hipotesa ini menunjukkan bahawa keputusan untuk berolah raga bila cuaca=cerah, dan kec.angin=pelan, untuk singkatnya dituliskan hanya instance pada setiap atribut dengan **H(cerah, *, pelan)=ya**

3. $H(\text{cuaca}=\text{cerah})=\text{ya}$

Hipotesa ini menunjukkan bahawa keputusan untuk berolah raga bila cuaca=cerah, untuk singkatnya dituliskan hanya instance pada setiap atribut dengan **H(cerah, *, *)=ya**

Perhitungan

Tabel 2.2 Menghitung nilai probabilitas kelas

Ya	Tidak
Ya = 4	Tidak = 2
Ya = $4/6 = 0.67$	Tidak = $2/6 = 0.34$

Tabel 2.3 Menghitung Tiap Fitur Ya

Cuaca	Temperatur	Kecepatan Angin
Cerah = 4	Normal = 4	Pelan = 3
Hujan = 0	Tinggi = 0	Kencang = 1
P(cuaca = cerah ya) = 4/4 = 1	P(temperatur = normal ya) = 4/4 = 1	P(kecepatan angin = pelan ya) = 3/4 = 0.75
P(cuaca = hujan ya) = 0/4 = 0	P(temperatur = tinggi ya) = 0/4 = 0	P(kecepatan angin = kencang ya) = 1/4 = 0.25

Tabel 2.4 Menghitung tiap fitur Tidak

Cuaca	Temperatur	Kecepatan Angin
Cerah = 0	Normal = 0	Pelan = 1
Hujan = 2	Tinggi = 2	Kencang = 1
P(cuaca = cerah tidak) = 0/2 = 0	P(temperatur = normal tidak) = 0/2 = 0	P(kecepatan angin = pelan tidak) = 1/2 = 0.5
P(cuaca = hujan tidak) = 2/2 = 1	P(temperatur = tinggi tidak) = 2/2 = 1	P(kecepatan angin = kencang tidak) = 1/2 = 0.5

Menghitung probabilitas akhir kelas kategorikal.

P (Ya)	= 0.67
P (Cuaca cerah)	= 1
P (Temperatur normal)	= 1
P (Kecepatan Angin pelan)	= 0.75

$$\begin{aligned}
 P(X | Ya) &= P(\text{cuaca} = \text{cerah} | \text{ya}) \times P(\text{temperatur} = \text{normal} | \text{ya}) \times P(\text{kecepatan} \\
 &\text{angin} = \text{pelan} | \text{ya}). \\
 &= 0.67 \times 1 \times 1 \times 0.75 \\
 &= 0.5025
 \end{aligned}$$

P (Tidak)	= 0.34
P (Cuaca hujan)	= 1
P (Temperatur tinggi)	= 1
P (Kecepatan Angin kencang)	= 0.5

$$\begin{aligned}
P (X | \text{Tidak}) &= P(\text{cuaca} = \text{hujan} | \text{tidak}) \times P(\text{temperatur} = \text{tinggi} | \text{tidak}) \times \\
&P(\text{kecepatan angin} = \text{kencang} | \text{tidak}). \\
&= 0.34 \times 1 \times 1 \times 0.5 \\
&= 0.17
\end{aligned}$$

Keterangan nilai probabilitas akhir terbesar ada di kelas YA dengan hasil akhir 0.5025.

2.7 Penelitian Sebelumnya

Penelitian sebelumnya dilakukan Abdul Sani Sembiring (2013) dengan judul Sistem pakar diagnosa penyakit dan hama tanaman padi menggunakan metode *Forward chaining dan Backward chaining*. Dengan menggunakan sistem pakar diharapkan para petani dapat memberikan penalaran untuk data-data yang tidak pasti, dapat mengemukakan rangkaian alasan-alasan yang diberikannya dengan cara yang dapat dipahami dan mendapatkan Pengetahuan tentang hama-hama padi dan penanggulangannya. Dalam penelitiannya dapat membantu dalam hal pengembangan media pengetahuan tentang perkembangan dan solusi hama pada tanaman padi tersebut dengan menggunakan bantuan komputer beserta dengan aplikasi pendukungnya tersebut, guna menciptakan data yang akurat dan *up to date* (berkembang) seiring dengan bertambahnya waktu atas perkembangan atau penemuan penyakit dan hama baru pada tanaman padi, berikut dengan bagaimana tahap atau proses penanggulangan panyerangan penyakit dan hama pada tanaman padi tersebut.

Penelitian selanjutnya dilakukan Rika Sofa (2012) dengan judul Pembangunan aplikasi sistem pakar untuk diagnosis penyakit tanaman padi menggunakan metode *Forward chaining*. Untuk mendiagnosis hama tanaman peneliti menggunakan seorang pakar/ahli dibidang pertanian khususnya tanaman

padi. Dalam hal ini peran seorang *expert* atau pakar sangat diandalkan untuk mendiagnosis dan menentukan jenis penyakit serta memberikan cara mengendalikan guna mendapatkan solusinya. Sistem pakar ini diharapkan dapat membantu memecahkan permasalahan dalam mendiagnosis penyakit tanaman padi, yang didasarkan pada seorang ahli pertanian yang bertugas sebagai penyuluh khususnya dibidang produksi tanaman pangan khususnya seksi hama dan penyakit padi.

Penelitian selanjutnya dilakukan Endang Trigiyanti (2010) dengan judul Pembuatan program aplikasi untuk mengidentifikasi hama dan penyakit padi menggunakan metode *Forward dan Backward chaining*. Diharapkan dapat mempermudah masyarakat atau petani dalam menyelesaikan permasalahan tanaman padi yang terserang hama atau penyakit, maka dibuatlah suatu program aplikasi yang dapat mengidentifikasi hama dan penyakit padi. Program aplikasi ini meniru cara berpikir seorang ahli spesialis hama dan penyakit padi dalam melakukan identifikasi suatu penyakit. Program aplikasi ini dibuat untuk membantu dalam mencari kesimpulan tentang penyakit yang menyerang beserta pencegahan atau solusi yang sesuai untuk mengatasinya. Program aplikasi ini menganalisa gejala-gejala dari suatu penyakit. Pengembangan Program aplikasi ini menggunakan metode inferensi *forward chaining* dan *backward chaining*.