

BAB II

LANDASAN TEORI

2.1 Katarak

Katarak adalah setiap keadaan kekeruhan pada lensa yang dapat terjadi akibat hidrasi lensa, denaturasi protein lensa, atau akibat kedua-duanya. Biasanya mengenai kedua mata dan berjalan progresif (Mansjoer. 2008). Katarak berasal dari bahasa Yunani *Katarrhakies*, Inggris *Cataract*, dan Latin *Cataracta* yang berarti air terjun, karena orang yang menderita katarak mempunyai penglihatan yang kabur seolah-olah tertutup oleh air terjun. Pada mata sehat, lensa yang jernih berfungsi meneruskan cahaya ke dalam mata agar mata dapat memfokuskan benda dari jarak yang berbeda-beda. Sebaliknya pada penderita katarak, lensa mata yang keruh menyebabkan cahaya yang masuk ke dalam mata dapat terpecah dan mengakibatkan penglihatan kabur. Penyebab utama katarak adalah penuaan usia, tetapi banyak hal lain yang dapat terlibat seperti genetik, trauma mata, penyakit diabetes melitus, tekanan darah tinggi (hipertensi), merokok, mengkonsumsi alkohol dan radiasi UV (Ilyas. 2014). Secara umum stadium katarak dapat di klasifikasikan menjadi 4 kategori :

1. Katarak Insipien

Ini merupakan stadium satu (awal), dimana kekeruhan lensa mata masih sangat minim, bahkan terkadang sulit terlihat tanpa menggunakan alat bantu periksa. Pada tahap ini seringkali penderitanya tidak merasakan keluhan atau gangguan pada penglihatannya, sehingga cenderung diabaikan. Kekeruhan mulai dari tepi ekuator berbentuk jeruji menuju korteks anterior dan posterior. Pada katarak insipient kekeruhan ini dapat menimbulkan poliopia oleh karena indeks refraksi yang tidak sama pada semua bagian lensa. Bentuk ini kadang-kadang menetap untuk waktu yang lama.

2. Katarak Immatur

Merupakan stadium dua, dimana kekeruhan lensa mata yang lebih tebal tetapi belum mengenai seluruh lensa sehingga masih terdapat bagian-

bagian yang jernih pada lensa. Pada stadium ini terjadi hidrasi kortek yang mengakibatkan lensa menjadi bertambah cembung. Pencembungan lensa akan memberikan perubahan indeks refraksi dimana mata akan menjadi mioptik. Kecembungan ini akan mengakibatkan pendorongan iris kedepan sehingga bilik mata depan akan lebih sempit.

3. Katarak Matur

Merupakan stadium tiga (tahap lanjut), dimana kekeruhan telah mengenai seluruh massa lensa. Kekeruhan ini terjadi akibat deposit *ion Ca* yang menyeluruh. Pada tahap ini bila dibiarkan saja akan menjadi katarak dengan kandungan air berlebih, yang dapat memblok pupil dan menyebabkan tekanan pada bola mata meningkat.

4. Katarak Hiper matur

Merupakan stadium empat (tahap sangat lanjut), dimana kekeruhan telah mengenai seluruh massa lensa dan telah mengalami proses degenerasi lanjut, dapat menjadi lembek dan mencair. Masa lensa telah mencair dan sudah merembes melalui kapsul lensa dan bisa menyebabkan perdarahan pada struktur mata yang lainnya. Pada tahap ini penglihatan sudah sangat menghilang dan sudah masuk ke fase glaukoma atau kebutaan.

2.2 Etiologi

Etiologi berasal dari bahasa Yunani (*aitia*), yang berarti sebab. Etiologi telah digunakan dalam bahasa Inggris sejak 1555, yang diambil dari bahasa Yunani. Secara teori etiologi merupakan studi yang mempelajari tentang asal muasal atau faktor penyebab. Dalam kasus seorang pasien, memahami etiologi dapat menjadi sangat penting, karena dapat memberikan bagian penting dari teka-teki ketika datang untuk mengobati kondisi. Ketika penyebab tidak dapat ditentukan, penyakit ini dikatakan “idiopatik,” yang berarti bahwa ia tidak memiliki etiologi yang diketahui. Kondisi idiopatik bisa jadi sangat frustrasi untuk dokter dan pasien, karena kurangnya etiologi dapat membuat sulit untuk dijabarkan sifat yang tepat dari penyakit. Semakin cepat etiologi dapat ditentukan, semakin cepat dapat dikendalikan, karena faktor yang menciptakan penyakit ini biasanya dapat

dihilangkan, dikendalikan, atau bahkan dihindari. Berikut beberapa faktor penyebab (etiologi) pada katarak (Ilyas. 2014) :

1. Umur

Katarak pada umumnya terjadi karena proses penuaan. Besarnya jumlah penderita katarak berbanding lurus dengan jumlah penduduk umur lanjut. Proses penuaan menyebabkan lensa mata menjadi keras dan keruh, umumnya terjadi pada umur diatas 50 tahun.

2. Trauma Mata

Trauma mata menyumbang sebagian besar bertambahnya jumlah penderita katarak. Katarak terjadi akibat trauma mata dapat terjadi pada semua umur. Trauma atau cedera pada mata mengakibatkan terjadinya erosi epitel pada lensa. Pada keadaan ini dapat terjadi hidrasi korteks hingga lensa mencembung dan mengeruh.

3. Diabetes Melitus

Penyakit diabetes melitus pun ikut menyumbang terhadap tingginya jumlah penderita katarak, sejalan dengan bertambahnya jumlah kasus penderita diabetes melitus. Pembentukan katarak yang terkait dengan diabetes sering terjadi karena kelebihan kadar sorbitol (gula yang terbentuk dari glukosa), yang membentuk penumpukan dalam lensa dan akhirnya membentuk kekeruhan lensa.

4. Hipertensi.

Hipertensi memainkan peranan penting terhadap perkembangan katarak. Hipertensi bisa menyebabkan konformasi struktur perubahan protein dalam kapsul lensa, sehingga memperburuk pembentukan katarak, sehingga dapat memicu katarak.

5. Genetika

Faktor genetik atau keturunan merupakan salah satu faktor yang menyebabkan terjadinya katarak. Sebab beberapa kelainan genetik yang diturunkan dapat menyebabkan gangguan kesehatan lainnya yang dapat meningkatkan resiko katarak, seperti kelainan kromosom mampu mempengaruhi kualitas lensa mata sehingga dapat memicu katarak.

6. Merokok

Merokok secara signifikan meningkatkan resiko katarak dibandingkan non-perokok. Sebab merokok dapat mengubah sel-sel lensa melalui oksidasi, merokok dapat juga menyebabkan akumulasi logam berat seperti cadmium dalam lensa sehingga dapat memicu katarak.

7. Alkohol

Meminum minuman beralkohol secara berlebihan juga dapat memicu terkena penyakit katarak. Alkohol dapat mengganggu homeostasis kalsium dalam lensa dan meningkatkan proses seperti kerusakan membran sehingga dapat memicu katarak.

8. Radiasi Ultraviolet

Radiasi sinar ultraviolet pada siang hari cukup tinggi dan paparannya untuk jangka waktu yang lama dapat menjadi pemicu katarak. Sebab sinar ultraviolet mampu merusak jaringan mata, dapat merusak saraf pusat penglihatan serta makula dan dapat merusak bagian kornea dan lensa.

2.3 Definisi Sistem

Sistem secara fisik adalah kumpulan dari elemen-elemen yang beroperasi bersama-sama untuk menyelesaikan suatu sasaran (Gordon, 1991).

Jogianto (2005:2) mengemukakan bahwa sistem adalah kumpulan dari elemen-elemen yang berinteraksi untuk mencapai suatu tujuan tertentu.

2.3.1 Karakteristik Sistem

Jogianto (2005: 3) mengemukakan sistem mempunyai karakteristik atau sifat-sifat tertentu, yakni :

1. Komponen

Suatu sistem terdiri dari sejumlah komponen yang saling berinteraksi, yang artinya saling bekerja sama membentuk satu kesatuan. komponen-komponen sistem atau elemen-elemen sistem dapat berupa suatu subsistem atau bagian-bagian dari sistem. setiap subsistem mempunyai sifat-sifat dari

sistem untuk menjalankan suatu fungsi tertentu mempengaruhi proses sistem secara keseluruhan.

2. Batasan sistem

Batasan sistem (boundary) merupakan daerah yang membatasi antara suatu sistem dengan sistem yang lainnya atau dengan lingkungan luarnya. batasan suatu sistem menunjukkan ruang lingkup dari sistem tersebut.

3. Lingkungan Luar Sistem.

Lingkungan luar (*evinronment*) dari suatu sistem adalah apapun diluar batas sistem yang mempengaruhi operasi. Lingkungan luar sistem dapat bersifat menguntungkan dan dapat juga bersifat merugikan sistem tersebut.

4. Penghubung Sistem

Penghubung (interfance) merupakan media penghubung antara satu subsistem dengan subsistem yang lainnya. melalui penghubung ini memungkinkan sumber-sumber daya mengalir dari satu subsistem ke subsistem yang lainnya. Dengan penghubung satu subsistem dapat berintegrasi dengan subsistem yang lainnya membentuk satu kesatuan.

2.4 Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai database besar (Turban, dkk., 2005).

Menurut (Goela., et al, 2012), Data Mining (DM) adalah proses menganalisis data dengan berbagai perspektif dan meringkasnya kedalam informasi yang berguna, dimana informasi tersebut bisa digunakan untuk meningkatkan pendapatan, memotong biaya atau keduanya.

Istilah *data mining* kadang disebut juga *knowledge discovery*. Selanjutnya perbedaan *data mining* dan *data warehouse* adalah *data mining* merupakan bidang yang sepenuhnya menggunakan apa yang sepenuhnya digunakan oleh *data warehouse*, bersama dengan bidang yang menangani masalah pelaporan dan manajemen data. Sementara *data warehose* sendiri bertugas untuk menarik/meng-

query data dari basis data mentah untuk memberikan hasil data yang nantinya digunakan oleh bidang yang menangani manajemen, pelaporan, dan *data mining*. Dengan *data mining* inilah, penggalian informasi baru dapat dilakukan dengan bekal data mentah yang diberikan oleh *data warehouse*. Hasil yang diberikan oleh ketiga bidang tersebut berguna untuk mendukung aktivitas bisnis cerdas (*business intellegent*). Secara umum *data mining* mempunyai tugas utama, yaitu :

1. Klasifikasi (*Classification*)

Klasifikasi bertujuan untuk mengklasifikasikan *item* data menjadi satu dari beberapa kelas standart. Sebagai contoh, suatu program email dapat mengklasifikasikan email yang sah dengan email spam. Beberapa algoritma klasifikasi antara lain *pohon keputusan*, *nearest neighbor*, *naïve bayes*, *neural networks*, dan *support vector machines*.

2. Regresi (*Regression*)

Regresi merupakan permodelan dan investigasi hubungan dua atau lebih variabel. Dalam analisis regresi ada satu lebih variabel independent / prediktor yang biasa diwakili dengan notasi dengan notasi x dan satu variabel respon yang biasa diwakili dengan notasi y (Santosa, 2007).

3. Pengelompokan (*Clustering*)

Clustering merupakan metode pengelompokan sejumlah data ke dalam kluster(*group*) sehingga dalam setiap kluster berisi data yang semirip mungkin.

4. Pembelajaran Aturan Asosiasi (*Association Rule Learning*)

Pembelajaran aturan asosiasi mencari hubungan antara variabel. Sebagai contoh suatu toko mengumpulkan data kebiasaan pelanggan dalam berbelanja. Dengan menggunakan pembelajaran atur asosiasi, toko tersebut dapat menentukan produk yang sering dibeli bersamaan dan menggunakan informasi ini untuk tujuan pemasaran.

2.5 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. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Model itu sendiri bisa berupa aturan "Jika-maka", berupa pohon keputusan, atau formula matematis (Bustomi, 2011)

Tahapan dari klasifikasi dalam data mining menurut (Han dan Kamber, 2006) terdiri dari :

1. Pembangunan Model

Pada tahapan ini dibuat sebuah model untuk menyelesaikan masalah klasifikasi class atau atribut dalam data. Tahap ini merupakan fase pelatihan, dimana data latih dianalisis menggunakan algoritma klasifikasi, sehingga model pembelajaran direpresentasikan dalam bentuk aturan klasifikasi.

2. Penerapan Model

Pada tahapan ini model yang sudah dibangun sebelumnya digunakan untuk menentukan atribut/class dari sebuah data baru yang atribut/classnya belum diketahui sebelumnya. Tahap ini digunakan untuk memperkirakan keakuratan aturan klasifikasi terhadap data uji. Jika model dapat diterima, maka aturan dapat diterapkan terhadap klasifikasi data baru.

2.6 Naive Bayes Classifier

2.6.1 Teori Bayes

Bayes merupakan teknik prediksi probalistik 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 di gunakan adalah model fitur independen.

Dalam Bayes (terutama Naïve Bayes), maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidak adanya fitur lain dalam data yang sama. Contohnya, pada kasus klasifikasi hewan dengan fitur penutup kulit, melahirkan, berat dan menyusui. Disini ada ketergantungan pada fitur menyusui karena hewan yang menyusui biasanya melahirkan, atau hewan bertelur tidak menyusui. Dalam Bayes, hal tersebut tidak dipandang sehingga masing-masing fitur seolah tidak memiliki hubungan apa pun. *Naive Bayes* memiliki beberapa keuntungan dan kekurangan yaitu sebagai berikut :

1. Keuntungan *Naive Bayes*
 - a. Cepat dan efisiensi ruang.
 - b. Kokoh terhadap atribut yang tidak relevan.
 - c. Hanya memerlukan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata dan variansi dari variabel) yang dibutuhkan untuk klasifikasi.
 - d. Menangani Kuantitatif dan data diskrit.
2. Kekurangan *Naive Bayes*
 - a. Tidak berlaku jika *Probabilitas* kondisionalnya adalah nol, apabila nol maka *Probabilitas* prediksi akan bernilai nol juga.
 - b. Mengasumsikan Variabel bebas.

Prediksi Bayes didasarkan pada teori Bayes dengan formula umum sebagai berikut :

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

Keterangan :

- $P(H | E)$ = *Probabilitas* akhir bersyarat (conditional probability) suatu hipotesis H terjadi jika diberikan bukti (evidence) E terjadi.
- $P(E | H)$ = *Probabilitas* sebuah bukti E terjadi akan mempengaruhi hipotesis H .
- $P(H)$ = *Probabilitas* awal hipotesis H terjadi tanpa memandang bukti apapun.

$P(E)$ = *Probabilitas* awal bukti E terjadi tanpa memandang hipotesis atau bukti yang lain.

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

1. Sebuah probabilitas awal/priori H atau PH adalah *probabilitas* dari suatu hipotesis sebelum bukti diamati.
2. Sebuah *probabilitas* akhir H atau (PH) adalah *probabilitas* dari suatu hipotesis setelah bukti diamati.

Teorema Bayes juga bisa menangani beberapa bukti, misalnya ada $E_1, E_2,$ dan E_3 sehingga akhir untuk hipotesis (H) dapat dihitung dengan cara berikut.

$$P(H | E_1, E_2, E_3) = \frac{P(E_1, E_2, E_3 | H) \cdot P(H)}{P(E_1, E_2, E_3)} \dots\dots\dots(2.2)$$

Karena yang digunakan untuk bukti adalah independen, bentuk di atas dapat diubah menjadi.

$$P(H | E_1, E_2, E_3) = \frac{P(E_1 | H) \cdot P(E_2 | H) \cdot P(E_3 | H) \cdot P(H)}{P(E_1) \cdot P(E_2) \cdot P(E_3)} \dots\dots\dots(2.3)$$

2.6.2 Naive Bayes untuk Klasifikasi

Klasifikasi *Naïve Bayes* adalah metode yang berdasarkan *probabilitas* dan Teori Bayes dengan asumsi bahwa setiap variabel bersifat bebas (*independence*) dan mengasumsikan bahwa keberadaan sebuah fitur tidak ada kaitannya dengan keberadaan fitur yang lain. Asumsi keidependenan atribut akan menghilangkan kebutuhan banyaknya jumlah data latih dari perkalian kartesius seluruh atribut yang dibutuhkan untuk mengklasifikasikan suatu data. Formulasi Naïve Bayes untuk klasifikasi adalah :

$$P(Y | X) = \frac{P(Y) \prod_{i=1}^n P(X_i | Y)}{P(X)} \dots\dots\dots(2.4)$$

Keterangan :

$P(Y | X)$ = *probabilitas* data dengan vektor X pada kelas Y

$P(Y)$ = *probabilitas* awal kelas Y

$\prod_{i=1}^q P(X_i|Y)$ = *probabilitas* independen kelas Y dari semua fitur dalam vector X

Karena $P(X)$ selalu tetap, sehingga dalam perhitungan prediksi nantinya cukup hanya dengan menghitung $P(Y) \prod_{i=1}^q P(X_i|Y)$ dengan memilih yang terbesar sebagai kelas yang di pilih sebagai hasil prediksi. Sementara *probabilitas* independen $\prod_{i=1}^q P(X_i|Y)$ tersebut merupakan pengaruh semua fitur dari data terhadap setiap kelas Y , yang dinotasikan dengan :

$$P(X|Y=y) = \prod_{i=1}^q P(X_i|Y=y) \dots \dots \dots (2.5)$$

setiap fitur $X = \{X_1, X_2, X_3, \dots, X_q\}$ terdiri atas q atribut (q dimensi).

Umumnya, metode Naïve Bayes ini mudah dihitung untuk fitur bertipe kategoris. Namun untuk tipe numerik (kontinu), ada perlakuan khusus sebelum dimasukkan dalam Naïve Bayes, yaitu :

1. Melakukan diskretisasi pada setiap fitur kontinu dan mengganti nilai fitur kontinu tersebut dengan nilai interval diskret. Pendekatan ini dilakukan dengan mentransformasi fitur kontinu kedalam fitur ordinal.
2. Dari distribusi *probabilitas* diasumsikan bentuk tertentu untuk fitur kontinu dan memperkirakan parameter distribusi dengan data peralihan. Distribusi Gaussian biasanya dipilih untuk mempresentasikan *probabilitas* bersyarat dari fitur kontinu pada sebuah kelas $P(X_i|Y)$, sedangkan distribusi Gausaian dikarakteristikan dengan dua parameter. mean μ , dan varian, σ^2 . Untuk setiap kelas y_j , *probabilitas* bersyarat kelas y_j untuk fitur X_i adalah

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \exp \left[-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \right] \dots \dots \dots (2.6)$$

Keterangan :

μ_{ij} : mean sampel X_i (\bar{X}) dari semua data latih yang menjadi milik kelas y_j

σ_{ij}^2 : varian sampel (S^2) dari data latih.

2.7 Penelitian Sebelumnya

Naive Bayes merupakan metode yang cukup populer dan banyak di gunakan untuk klasifikasi. Penelitian atau riset sebelumnya yang berkaitan dengan kasus klasifikasi atau prediksi yang menggunakan metode Naïve Bayes, antara lain :

Penelitian pertama dilakukan oleh Rudi Hariyanto, lulusan Universitas Brawijaya (2016). Penelitiannya berjudul “Klasifikasi Penyakit Mata Katarak berdasarkan Kelainan Patologis dengan menggunakan Algoritma *Learning Vector Quantization (LVQ)*”. Data yang digunakan sebanyak 85 *record* dan data uji sebanyak 50 *record*. Atribut yang digunakan adalah gejala umum katarak. Hasil penelitian menunjukkan bahwa penggunaan metode LVQ memberikan tingkat akurasi penentuan sebesar 99%.

Penelitian kedua dilakukan oleh Mohamad Baharuddin Rabani, lulusan Universitas Muhammadiyah Gresik Tahun 2014. Penelitiannya berjudul : “Sistem klasifikasi keluarga siswa mampu dan tidak mampu untuk mendapatkan beasiswa di MA Muhammadiyah 1 Sumberrejo Kab. Bojonegoro dengan menggunakan metode *Naive Bayes*”. Adapun data yang diambil dalam penelitian ini adalah sampel dari 105 *record* dengan kelas “Mampu” dan “Tidak Mampu” masing-masing berjumlah 68 dan 37 yang akan dibagi menjadi data latih data uji. Dan menggunakan 8 variable, adapun variabel yang dipakai : dari saudara kandung, dari saudara tiri, saudara yang berkerja, rata-rata penghasilan saudara perbulan pekerjaan ayah, pekerjaan ibu, rata-rata penghasilan orang tua perbulan, kelas. Dan keakurasian hasil penelitian menggunakan metode Naïve Bayes ini menunjukkan akurasi tertinggi pada pengujian pertama adalah 86.67 %.

Penelitian ketiga di lakukan oleh Meinggian Vilian Sari, lulusan Universitas Muhammadiyah Gresik Tahun 2014. Penelitiannya berjudul : “Sistem prediksi prestasi (IPK) mahasiswa berdasarkan latar belakang sekolah asal dan atribut mahasiswa ketika awal masuk kuliah menggunakan *Naive Baye*”. Adapun

data yang diambil dalam penelitian ini adalah sampel dari 103 *record* dengan kelas “Tinggi” dan “Rendah” masing-masing berjumlah 69 dan 34 yang akan dibagi menjadi data latih data uji. Dan menggunakan 6 variable, adapun variabel yang dipakai : Instansi Sekolah, Status Sekolah, Jurusan Sekolah, Motivasi Pilihan Kuliah, Status Kerja, Nilai Danem. Dan keakurasian hasil penelitian menggunakan metode Naïve Bayes ini adalah 84.62%.