

## **BAB II**

### **LANDASAN TEORI**

Landasan teori merupakan sekumpulan teori atau bahan pustaka yang berhubungan dengan permasalahan yang diteliti sehingga dapat memperdalam pengetahuan mengenai permasalahan yang bersangkutan. Dengan adanya landasan teori diharapkan dapat digunakan untuk memecahkan permasalahan yang diangkat dalam penelitian ini yaitu : Perbandingan Metode *Naïve Bayes* dan *Certainty Factor* Untuk Menagnosa Penyakit Epilepsi dengan studi kasus pada RSUD Selasih Kab. Pelalawan. Adapun landasan teori yang digunakan adalah :

#### **2.1 Data Mining**

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Data mining merupakan proses iterative dan interaktif untuk menemukan pola atau model baru yang sempurna, bermanfaat dan dapat dimengerti dalam suatu *database* yang sangat besar. Data mining berisi pencarian *trend* atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya.(Gorunnescu, Florin ,2011)

## 2.2 *Naïve Bayes*

Menurut Oslan dan Delen dalam (Kurniawan, Afif F, 2013) *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes yaitu dengan memprediksi masa depan berdasarkan data dari masa lalu. *Naïve Bayes* untuk setiap kelas keputusan menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vector informasi objek. Algoritma ini mengasumsikan bahwa atribut objek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari “master” tabel keputusan.

Rumus dari perhitungan *Naïve Bayes* dalam buku berjudul *Data Mining Concept and Techniques*, Nirmala dalam (Via, Y. V., Nugroho, B., & Syafrizal, A. 2015) yaitu :

$$P(C_i|X) = \frac{P_i}{P(X)} \quad (1)$$

atau dapat ditulis dengan :

$$Posterior = \frac{\text{likelihood} \times \text{prior probability}}{\text{evidence}} \quad (2)$$

Keterangan :

$P(C_i|X)$  = Peluang kategori  $C_i$  jika diberikan fakta atau bukti  $X$  (*posterior*)

$P(X|C_i)$  = Peluang pada kategori  $C_i$ , dimana fakta atau bukti  $X$  muncul pada kategori tersebut (*likelihood*)

$P(C_i)$  = Peluang dari kategori yang diberikan, dibandingkan dengan kategori lainnya yang dianalisa (*prior probability*)

$P(X)$  = Jumlah peluang dari fakta atau bukti  $X$  (*evidence*)

Alur kerja dari Naïve Bayes (Bustami , 2013) adalah sebagai berikut:

1. Baca *data training*
2. Hitung jumlah peluang untuk setiap variabel , cari nilai *likelihood* dan nilai probabilitas.
3. Mendapatkan nilai dalam tabel *mean*,standar deviasi,dan *likelihood* serta probabilitas.

Pada penelitian sebelumnya metode naïve bayes telah digunakan untuk melakukan diagnosa penyakit mata pada manusia (Sam'ani, 2016). Pada penelitian ini penyakit mata pada manusia di diagnosa dengan menggunakan data – data histories dari pasien – pasien yang telah dilakukan diagnosa oleh pakar atau dokter mata di rumah sakit yang telah ditentukan. Berdasarkan gejala – gejala yang sama dengan pasien yang telah terdiagnosa maka didapat penyakit mata yang mungkin dialami oleh seorang pasien. Dari hasil percobaan pada penelitian maka di dapat penyakit yang mungkin dialami oleh pasien dengan gejala yang telah di diagnosa adalah penyakit Pterigium, dengan contoh perhitungan manualnya adalah sebagai berikut :

Penyakit Pterigium mempunyai probabilitas 0,6

Tabel 2. 1 Contoh Kasus Aturan Penyakit Pterigium

<b>Gejala</b>	<b>Probabilitas Gejala</b>
Mata Kering	0,4
Terdapat sesuatu Yang mengganjal pada mata	0,7
Terdapat selaput berbentuk segitiga yang menutupi kornea atau bola mata	0,9

Maka probabilitas penyakit Pterigium berdasarkan gejala yang dipilih adalah sebagai berikut :

$$P(\text{Pterigium}|G1|G3) = \frac{(0,7 \times 0,6) + (0,9 \times 0,6)}{(0,4 \times 0,6) + (0,7 \times 0,6) + (0,9 \times 0,6)}$$

Dari hasil perhitungan maka penyakit Pterigium berdasarkan gejala yang dialami adalah  $0,8 \times 100 \% = 80 \%$

Tabel 2. 2 Kajian Pustaka Sistem Diagnosa Penyakit Mata pada Manusia

Penulis	M. Haris Qoamarruzaman & Sam'ani
Judul	Sistem Pakar Untuk Mendiagnosa Penyakit Mata Pada Manusia Menggunakan Teorema Bayes
Tahun Jurnal	2016
Objek penelitian	Penyakit Mata Pada Manusia
Metode yang ada	<i>Naive Bayes</i>
Masalah penelitian	Banyak penyakit mata pada manusia yang lambat di ketahui, sehingga meyebabkan penyakit yang dialami menjadi semakin parah.
Solusi	Dibuat sebuah aplikasi sistem pakar diagnosa penyaki mata pada manusia dengan memanfaatkan metode <i>Naive Bayes</i> .
Tujuan	Mengimplementasikan metode <i>Naive Bayes</i> untuk mendiagnosa penyakit mata pada manusia. Sehingga kelainan mata pada seseorang dapat di diagnosa dan ditangani secara dini.

### 2.3 Metode *Certainty Factor*

Faktor kepastian (*Certainty Factor*) diperkenalkan oleh Shortliffe Buchanan dalam pembuatan MYCIN. *Certainty Factor* (CF) merupakan nilai parameter klinis yang diberikan MYCIN untuk menunjukkan besarnya kepercayaan. *Certainty Factor* menunjukan kepastian terhadap suatu fakta atau aturan (Rosnelly, 2011).

Teori Certainty Factor (CF) diusulkan oleh Shortliffe dan Buchanan pada 1975 untuk mengakomodasi ketidakpastian pemikiran (*inexact reasoning*) seorang pakar. Seorang pakar, (misalnya dokter) sering kali menganalisis informasi yang ada dengan ungkapan seperti “mungkin”, “kemungkinan besar”, “hamper pasti”. Untuk mengakomodasi hal ini kita menggunakan *Certainty Factor* (CF) guna menggambarkan tingkat keyakinan pakar terhadap masalah yang sedang dihadapi. Ada dua cara dalam mendapatkan tingkat keyakinan (CF) dari sebuah rule Sujoto, T dalam (Tutur Larasati. 2016):

Dalam mengekspresikan derajat keyakinan, *certainty theory* menggunakan suatu nilai yang disebut *certainty factor* (CF) untuk mengasumsikan derajat keyakinan seorang pakar terhadap suatu data. *Certainty factor* memperkenalkan konsep *belief* atau keyakinan dan *disbelief* atau ketidakpercayaan. Konsep ini kemudian diformulasikan dalam rumus dasar sebagai berikut (Nila Rohmika F.H. 2016)

$$CF(H, E) = MB(H, E) - MD(H, E) \quad (3)$$

$$MB[h, e \wedge e2] = MB[h, e1] + MB[h, e2] \cdot (1 - MB[h, e1]) \quad (4)$$

$$MD[h, e \wedge e2] = MD[h, e1] + MD[h, e2] \cdot (1 - MD[h, e1]) \quad (5)$$

$$CF = \frac{MB - MD}{1 - \min(MB, MD)} \quad (6)$$

CF (H, E) : *Certainty Factor* dari hipotesis H yang dipengaruhi oleh gejala (*evidence*) E. Besarnya CF berkisar antara -1 sampai 1. Nilai -1 menunjukkan ketidakpercayaan mutlak sedangkan nilai 1 menunjukkan kepercayaan mutlak.

MB (H, E) : Ukuran kenaikan kepercayaan terhadap hipotesis H yang dipengaruhi oleh gejala E.

MD (H, E) : Ukuran kenaikan ketidakpercayaan terhadap hipotesis H yang dipengaruhi oleh gejala E.

Pada penelitian sebelumnya algoritma *Certainty Factor* dalam mendiagnosa penyakit. Dalam kasus sebelumnya algoritma *Certainty Factor* digunakan untuk mendiagnosa jenis penyakit Dyslexia (Sasmitha Ajeng D, 2016). Dengan mengetahui gejala – gejala pada penyakit dyslexia dan factor keyakinan yang telah ditentukan oleh pakar, maka dapat hasil diagnosa jenis penyakit dyslexia.

Tabel 2. 3 Kajian Pustaka Sistem Diagnosa Penyakit Dyslexia

Penulis	Dwi Ajeng Sasmitha
Judul	Sistem Pakar Untuk Mendiagnosa Jenis Dyslexia Menggunakan <i>Certainty Factor</i>
Tahun Jurnal	2016
Objek penelitian	Penyakit Dyslexia
Metode yang ada	<i>Certainty Fctor</i>
Masalah penelitian	Banyaknya penyakit Dyslexia yang tidak diketahui secara dini, membuat penderita penyakit ini lambat mendapat penanganan, sehingga tidak menutup kemungkinan dapat mengakibatkan stres pada penderita penyakit ini.
Solusi	Dibuat sebuah aplikasi sistem pakar diagnosa penyaki Dylexsia dengan memanfaatkan metode <i>Certainty Factor</i> .
Tujuan	Mengimplementasikan metode <i>Certainty Factor</i>

	untuk mendiagnosa penyakit Dyslexia. Sehingga kelainan pada penderita penyakit ini dapat di diagnosa dan ditangani secara dini.
--	---

#### 2.4 *Black Box Testing*

*Blackbox Testing* merupakan tehnik pengujian yang terfokus pada spesifikasi fungsional dari perangkat lunak. Pengujian fungsional bertujuan untuk mengevaluasi pemenuhan sistem atau komponen dengan kebutuhan fungsional tertentu. *Balckbox testing* hanya mengamati hasil eksekusi atau keluaran melalui data uji (Sunarya .2016).

#### 2.5 Penelitian yang sudah dilakukan

Menurut Jianguo Sun, dkk. Tahun 2009. Tingkat kinerja watermark adalah CF (W); CF (W, T) menunjukkan keefektifan target; CF (T, st) menunjukkan proporsi efektivitas kontras dengan target induknya; CF (st) adalah efektivitas sub target, itu bisa dihitung secara langsung.

Menurut Zhongjun Den, dkk. Tahun 2009. Dalam Jurnal ini, aturan asosiasi Untuk klasifikasi hanya terdiri dari aturan yang dihasilkan oleh Itemset tertutup. *Certainty Factor*, yang merupakan salah satu Model yang paling populer untuk representasi dan manipulasi pengetahuan yang tidak pasti pada awal aturan berbasis Sistem pakar, digunakan untuk merinci aturan klasifikasi.

*Certainty factor* teori itu sederhana untuk mengatasi ketidakpastian, berdasarkan pemilihan, hasilnya dapat diterima di sebagian besar aplikasi. Tapi faktor kepastian selalu berdasarkan pengalaman manusia dan kuat subyektivitas. Dalam tulisan ini, kita menggunakan pelengkap Kelebihan

dari set kasar dan teori *Certainty factor* untuk dipecahkan masalah di bidang teknik. Menurut Yi Ren, dkk. Tahun 2010.

Menurut Gautam Bhattacharya, dkk. Tahun 2015. Kontribusi pokok dari karya ini adalah proposisi Algoritma estimasi  $k$  dinamis baru berdasarkan kepadatan tetangga dari titik uji, varians kelas dan informasi faktor kepastian untuk poin pelatihan Masalah optimasi dengan menggunakan tetangga perkiraan kerapatan diformulasikan dimana  $k$  secara dinamis dipilih untuk berbagai titik uji.

Menurut Fitrah Rumaisa dan Dananng Junaedi. Tahun 2016. Diagnosis sistem pakar meningitis bisa menjadi media ketrampilan informasi, Pengetahuan dan alat deteksi (berdasarkan Usia, gejala atau keluhan) untuk awan mendeteksi kondisi awal meningitis mandiri dengan bantuan teknologi.

Metode menghasilkan vektor output dari Wikipedia yang terkait entitas dari teks tertentu dengan cara yang sama seperti ESA. Disini, Jurnal mendefinisikan skor probabilistik dan memperkenalkan Naïve Bayes untuk agregat entitas terkait. Menurut Masumi Shirakawa, dkk. Tahun 2015.

Menurut Kalla Orphanou, dkk. Tahun 2016. Classifier yang diusulkan dibandingkan dengan yang merupakan klasifikasi Naive Bayes yang fiturnya sederhana TAR. Melalui perbandingan ini, pemanfaatannya dari urutan yang lebih tinggi, yaitu abstraksi temporal yang lebih kompleks catatan pasien, menghasilkan kinerja yang lebih baik untuk yang diberikan pemecahan masalah medis.



Menurut Tahira Mahboob, dkk. Tahun 2016. Teknik efisien yang digunakan untuk menggabungkan Berbagai algoritma pembelajaran untuk meningkatkan akurasi prediksi keseluruhan. Dalam penelitian ini, memilih tiga teknik ensemble, yaitu J48, Naive Bayes dan Random Forest yang ada telah diterapkan untuk memprediksi hasil akhir siswa sesuai model yang diusulkan.

Menurut Harshit Sinha, dkk. Tahun 2016. Pengelompokan Naive Bayes adalah klasifikasi probabilistik yang Berdasarkan penerapan teorema Bayes dengan kuat yang mana menunjukkan eksklusivitas mutual antara kata-kata. (independent feature model) itu juga disebut sebagai pengelompokan dari kata classifier.

Menurut Rif'at Ahdi, dkk. Tahun 2016. Metode smoothing bagi Naïve Bayes memberikan kinerja yang berbeda untuk teks klasifikasi. Domain data sampel yang digunakan dalam penelitian ini terbatas pada kata kunci terkait fasilitas umum di Indonesia. Tambahan domain bisa ditambahkan dengan melibatkan kata kunci tertentu sehingga Pola karakteristik analisis sentimen dan pengaruhnya.

Tabel 2. 6 Persamaan dan Perbedaan Jurnal dengan Penelitian yang diangkat.

No	Studi Pustaka	Persamaan	Perbedaan
1	<i>Wikipedia-Based Semantic Similarity Measurements for Noisy Short Texts Using Extended Naïve Bayes</i>	Menggunakan <i>Naïve Bayes</i>	Makalah ini mengusulkan metode pengukuran kesamaan semantik berbasis Wikipedia
2	<i>Combining Naïve Bayes Classifiers with Temporal Association Rules for Coronary Heart Disease Diagnosis</i>	Menggunakan <i>Naïve Bayes</i>	Makalah ini klasifikasi Naive Bayes dimana Fituranya adalah aturan asosiasi temporal (TARs)
3	<i>A machine learning approach for Student Assessment in E-Learning Using Quinlan's C4.5, Naïve Bayes and Random Forest Algorithms</i>	Menggunakan <i>Naïve Bayes</i>	Penekanan fokus dari penelitian ini Adalah untuk memprediksi evaluasi siswa yang adil
4	<i>An Analysis of ICON Aircraft Log Through Sentiment Analysis Using SVM</i>	Menggunakan <i>Naïve Bayes</i>	Makalah ini Menggunakan Algoritma SVM dan <i>Naïve</i>

	<i>and Naïve Bayes Classification</i>		<i>Bayes</i>
5	<i>Comparison of Naive Bayes Smoothing Methods for Twitter Sentiment Analysis</i>	Menggunakan <i>Naïve Bayes</i>	bandingkan kinerja Naive Bayes Metode smoothing dalam meningkatkan kinerja sentiment analisis tweets.
6	<i>Digital Watermarking Appraisement based on Certainty Factor</i>	Menggunakan <i>Certainty Factor</i>	Penilaian watermarking memainkan peran penting dalam menguji kinerja sistem watermark
7	<i>Building Accurate Associative Classifier Based on Closed Itemsets and Certainty Factor</i>	Menggunakan <i>Certainty Factor</i>	Makalah ini menggunakan klasifikasi asosiatif
8	<i>Product Reliability Design Knowledge Reasoning Method Based on Rough Sets and Certainty Factors Theory</i>	Menggunakan <i>Certainty Factor</i>	Makalah ini menguraikan metode berdasarkan perangkat kasar dan teori faktor kepastian untuk memecahkan dua masalah
9	<i>A Probabilistic Framework for Dynamic k Estimation in kNN Classifiers with Certainty Factor</i>	Menggunakan <i>Certainty Factor</i>	Makalah ini menggunakan Algoritma kNN klasik dengan estimasi kestabilan titik uji
10	<i>Expert System For Early Diagnosis Of Meningitis Disease Using Certainty Factor Method</i>	Menggunakan <i>Certainty Factor</i>	Makalah ini membahas tentang diagnose penyakit meningitis