BAB II

TINJAUAN PUSTAKA

II.1 Landasan Teori

Landasan teori berisi tentang teori-teori yang mendukung penelitian yang sedang dilakukan. Adapun teori-teori tersebut adalah sebagai berikut.

II.1.1 Klasifikasi

Proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *Decision/classification trees, Bayesian classifiers/ Naïve Bayes classifiers, Neural networks, Analisa Statistik, Algoritma Genetika, Rough sets, k-nearest neighbor, Metode Rule Based, Memory based reasoning, dan Support vector machines (SVM)* (Annur, 2018).

II.1.2 Algoritma Naïve Bayes

1. Pengertian Naïve Bayes

Naïve Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari Naïve

Bayes Classifier ini adalah asumsi yang sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian (Ardianto, dkk, 2020). Naïve Bayes juga merupakan sebuah pengelompokan statistik yang bisa di dipakai untuk memprediksi probabilitas anggota suatu *class* (Pratiwi dan Nugroho, 2016).

Menurut Olson Delen (2008) menjelaskan Naïve Bayes untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari ”master” tabel keputusan.

Naïve Bayes berasumsi bahwa efek dari suatu pada kelas yang diberikan adalah independen terhadap nilai atirbut yang lainnya. Asumsi ini biasa disebut dengan class conditional inedependence. Itu dibuat untuk menyederhanakan komputasi yang tekait dan dalam hal ini disebut sebagai “naïve”.(Rezdy Anugrah Perdana hal.7)

Naïve Bayes Classifier merupakan sebuah pengklasifikasian probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidak tergantungan (independent) yang tinggi (Vangara, Vangara, 2020). Keuntungan penggunaan Naïve Bayes Classifier adalah metode ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. (Rezdy Anugrah Perdana hal.8)

2. Kegunaan Naïve Bayes

- Mengklasifikasikan dokumen teks seperti teks berita ataupun teks akademis
- Sebagai metode machine learning yang menggunakan probabilitas
- Untuk membuat diagnosis medis secara otomatis
- Mendeteksi atau menyaring spam

3. Kelebihan Naïve Bayes

- Bisa dipakai untuk data kuantitatif maupun kualitatif
- Tidak memerlukan jumlah data yang banyak
- Tidak perlu melakukan data training yang banyak
- Jika ada nilai yang hilang, maka bisa diabaikan dalam perhitungan.
- Perhitungannya cepat dan efisien
- Mudah dipahami
- Mudah dibuat
- Pengklasifikasian dokumen bisa dipersonalisasi, disesuaikan dengan kebutuhan setiap orang
- Jika digunakan dalaam bahasa pemrograman, *code*-nya sederhana
- Bisa digunakan untuk klasifikasi masalah biner ataupun *multiclass*

4. Kekurangan Naïve Bayes

- Apabila probabilitas kondisionalnya bernilai nol, maka probabilitas prediksi juga akan bernilai nol

- Asumsi bahwa masing-masing variabel independen membuat berkurangnya akurasi, karena biasanya ada korelasi antara variabel yang satu dengan variabel yang lain
- Keakuratannya tidak bisa diukur menggunakan satu probabilitas saja. Butuh bukti-bukti lain untuk membuktikannya
- Untuk membuat keputusan, diperlukan pengetahuan awal atau pengetahuan mengenai masa sebelumnya. Keberhasilannya sangat bergantung pada pengetahuan awal tersebut
- Banyak celah yang bisa mengurangi efektivitasnya
- Dirancang untuk mendeteksi kata-kata saja, tidak bisa berupa gambar

5. Perhitungan Naïve Bayes

Naive bayes merupakan salah satu metode pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya (Nofriansyah, 2016). Dalam metode ini perhitungan ditunjukkan melalui Persamaan 1 (Sutoyo, Almaarif, 2020):

$$P(H|X) = \frac{P(X|H)}{P(X)} . P(H) \quad (1)$$

Keterangan

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

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

P(X) : Probabilitas X

II.1.3 K-Nearest Neighbor

1. Defenisi K-Nearest Neighbor

Algoritma k-NN adalah suatu metode yang menggunakan algoritma *supervised* Perbedaan antara *supervised learning* dengan *unsupervised learning* adalah pada *supervised learning* bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru (Giovani, dkk, 2020). Sedangkan pada *unsupervised learning*, data belum memiliki pola apapun, dan tujuan *unsupervised learning* untuk menemukan pola dalam sebuah data. Tujuan dari algoritma *k*-NN adalah untuk mengklasifikasi objek baru berdasarkan atribut dan *training samples* (Saadatfar, dkk, 2020). Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada *k*-NN (Farokhah, 2020). Pada proses pengklasifikasian, algoritma ini tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori (Muhathir, dkk, 2020). Algoritma *k*-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru (Krisandi, dkk, 2013).

2. Langkah-Langkah K-Nearest Neighbor

Teknik *K-Nearest Neighbor* dengan melakukan langkah-langkah yaitu mulai input (Rohman, 2016): Data training, label data traning, *k*, data testing (Wahyono, 2020).

a. Untuk semua data testing, hitung jaraknya ke setiap data training
b. Tentukan k data training yang jaraknya paling dekat dengan data testing
c. Periksa label dari k data ini
d. Tentukan label yang frekuensinya paling banyak
e. Masukan data testing ke kelas dengan frekuensi paling banyak
f. Berhenti.

3. Rumus K-Nearest Neigbor

K-Nearest Neighbor menggunakan rumus perhitungan jarak. Jarak yang digunakan adalah jarak *Euclidean Distance*. Jarak *Euclidean* adalah jarak yang paling umum digunakan pada data numerik. *Euclidean distance* didefinisikan sebagai berikut (Krisandi, dkk, 2013):

$$d(x_{i,}x_j) = \sqrt{\sum_{r=1}^{n}(a_r(x_i) - a_r(x_j))^2} \quad (2)$$

Keterangan :

$d(x_i,x_j)$: Jarak *Euclidean (Eucledean Distance).*

(x_i) : *record* ke- *i*

(x_j) : *record* ke- *j*

(a_r) : data ke-*r*

i,j :1,2,3,…n

II.1.4 Precision, Recall dan Akurasi

Dalam "dunia" pengenalan pola (pattern recognition) dan temu kembali informasi (information retrieval), precision dan recall adalah dua perhitungan yang banyak digunakan untuk mengukur kinerja dari sistem / metode yang digunakan. Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Accuracy didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual. Tahap berikutnya melakukan pengujian terdapat tiga bagian :

1. Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.
2. Presisi tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.
3. Recall tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

Sedangkan di "dunia lain" seperti dunia statistika dikenal juga istilah accuray. Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.

Secara umum presisi, recall dan akurasi dapat dirumuskan sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

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

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

II.1.5 Tool Weka Dalam Klasifikasi

Weka adalah aplikasi data mining open source berbasis Java. Aplikasi ini dikembangkan pertama kali oleh Universitas Waikato di Selandia Baru sebelum menjadi bagian dari Pentaho. Weka terdiri dari koleksi algoritma machine learning yang dapat digunakan untuk melakukan generalisasi /formulasi dari sekumpulan data sampling. Walaupun kekuatan Weka terletak pada algoritma yang makin lengkap dan canggih, kesuksesan data mining tetap terletak pada faktor pengetahuan manusia implementornya. Tugas pengumpulan data yang berkualitas tinggi dan pengetahuan pemodelan dan penggunaan algoritma yang tepat diperlukan untuk menjamin keakuratan formulasi yang diharapkan.

II.1.6 Komparasi Akurasi Prediksi Kelulusan Mahasiswa Tepat Waktu

Menggunakan Algoritma Naïve Bayes dan Decision Tree Kerangka pemikiran dari penelitian ini bersumber dari fenomena akademik pada Fakultas Kedokteran Universitas Muhammadiyah Sumatera Utara yaitu ada ketidaksesuaian jumlah mahasiswa yang masuk dan yang lulus tepat waktu pada perguruan tinggi sehingga perlunya sebuah solusi yang tepat agar terdapat kesinambungan proses akademik yang relevan dengan mengevaluasi transaksi aktifitas kuliah mahasiswa 5 (lima) semester pertama.

Untuk mendapatkan nilai akurasi dari masa studi mahasiswa berdasarkan atribut tersebut menggunakan Algoritma Naïve Bayes kemudian mengkombinasikannya dengan algoritma Decision Tree yang bertujuan menemukan pola yang terdapat pada data mahasiswa berdasarkan data Indeks Pretasi Sementara (IPS) selama 5 (lima) semester pertama untuk memprediksi kelulusan mahasiswa tepat waktu.

Menghitung tingkat akurasi metode naive bayes dengan cara membandingkan data yang sudah ada dan juga prediksi yang telah dihitung. Sedangkan Menghitung tingkat akurasi metode decision tree dengan cara membandingkan data yang sudah ada dan juga prediksi yang telah dihitung.

II.2 Penelitian Terkait yang pernah dilakukan

Dalam hal ini peneliti melakukan penelitian berdasarkan beberapa kasus yang telah dilakukan terdahulu oleh peneliti-peneliti diantaranya :

1. Penelitian Indera Cahyo Wibowo, Abd Charis Fauzan, Marshella Dwi Putri Yustiana, Fiqih Ainul Qhabib dengan penelitian “Komparasi Algoritma Naive Bayes dan Decision Tree Untuk Memprediksi Lama Studi Mahasiswa” tahun 2019.
2. Penelitian Agus Romadhona, Suprapedi, H. Himawan dengan judul penelitian Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree.

3. penelitian yang dilakukan oleh Claudia Clarentina Ciptohartono yang berjudul “Algoritma Klasifikasi Naive Bayes untuk Menilai Kelayakan Kredit”.
4. Penelitian yang dilakukan oleh Arief Jananto. Penelitian yang berjudul “Algoritma Naive Bayes untuk Mencari Perkiraan Waktu 7 Studi Mahasiswa” Metode yang digunakan pada peneliti ini adalah Algoritma Klasifikasi Naive Bayes.
5. Penelitian yang dilakukan bersama Agus Budiyantara; Irwansyah; Egi Prengki; Pandi Ahmad Pratama; Ninuk Wiliani dengan judul “Komparasi Algortima Decision Tree, Naïve Bayes dan K-Nearest untuk Memprediksi Mahasiswa Lulus Tepat Waktu”.

II.3 Fakultas Kedokteran UMSU

Prodi Pendidikan Dokter merupakan pendidikan pendidikan kedokteran tahap akademik yang terdiri dari 144 sks dalam 8 semester. Kurikulum prodi pendidikan dokter menggunakan metode pendekatan hybrid problem based leaning (hybrid PBL) yang disusun secara spiral yaitu secara bertahap dari yang sederhana ke yang kompleks.

Strategi pendekatan pembelajaran berdasarkan metode SPICES (Student Centered, Problem-based, Integrated, Community Oriented, Early Clinical Exposure dan Systematic).

Student centered merupakan strategi pembelajaran dimana mahasiswa berperan aktif dan mandiri, serta sebagai pribadi dewasa dan bertanggung jawab

sepenuhnya atas pembelajarannya. Pembelajaran ini dicapai melalui Small Group Discussion (SGD), mahasiswa belajar secara aktif dari suatu pemicu skenario dan difasilitasi oleh tutor. Dosen sebagai tutor akan memfasilitasi diskusi kelompok mahasiswa dan menstimulus mahasiswa dalam mencapai tujuan pembelajaran.

Problem Based merupakan konsep pembelajaran dimana mahasiswa belajar berdasarkan masalah sebagai pemicu atau ilustrasi kasus yang digunakan untuk mencari, menggali, dan mengumpulkan informasi. Pembelajaran ini diterapkan pada SGD dan Keterampilan Klinik Dasar (KKD).

Integrated merupakan suatu pendekatan dalam proses pembelajaran yang menggunakan pendekatan antar bidang studi, menggabungkan bidang studi dengan cara menetapkan kurikulum, keterampilan dan konsep yang terintegrasi baik secara vertikal maupun horizontal, sehingga mmahasiswa diharapkan tidak berfikir secara terkotak-kotak dalam masing-masing disiplin ilmu tetapi dapat menghubungkan dan mengintegrasikan pengetahuan dan keterampilan yang dimiliki secara utuh lintas disilpin ilmu.

Community Oriented merupakan pendekatan yang mengarahkan kepada sistem pembelajaran yang berorientasi pada kebutuhan dan kepentingan masyarakat. Mahasiswa tidak dibatasi oleh ruang kelas tetapi juga mempelajari kehidupan masyarakat yang ada di lingkungan nyata.

Early Clinical exposure salah satu metode pembelajaran dimana mahasiswa dikenalkan lebih awal pada kondisi klinis sesuai dengan blok dan modul yang sedang dijalankan. Konsep ini diimplementasikan pada kegiatan

pembelajaran berbasis komunitas seperti kegiatan keluarga binaan kesehatan (KBK), Project Based Learning, orientasi ke RS dan puskesmas.

Systematic merupakan proses pembelajaran dengan tujuan, materi dan tahapan-tahapan yang jelas, logis dan tertib sehingga mahasiswa dapat memperoleh pemahaman yang lebih baik dan dapat mencapai kompetensi yang diharapkan.