

BAB II

LANDASAN TEORI

2.1 Penelitian Terkait

2.1.1 *Native Bayes*

Kinerja *naïve bayes* yang kompetitif dalam proses pengklasifikasi, bahkan dengan asumsi ketergantungan Atribut (tidak ada hubungan antara atribut [2] *Naïve Bayes* merupakan pengklasifikasian dengan 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. Teorema tersebut dikombinasikan dengan ”*naive*” dimana diasumsikan kondisi antar atribut saling bebas. [3]

Hasil Penelitian ini untuk melakukan prediksi tingkat kelulusan tepat waktu dengan melakukan analisis menggunakan teknik klasifikasi data mining. Algoritma *Naïve Bayes* yang digunakan untuk penelitian ini akan dibahas juga sebagai acuan dalam melakukan penelitian. Penulis melakukan serangkaian skenario percobaan yang berbeda / cross validation untuk melakukan perbandingan yang dapat memberikan perbedaan dalam tingkat akurasi yang dihasilkan dari penelitian ini. Hasil penelitian ini menunjukkan dengan penambahan skenario pengujian *Cross Validation* terdapat peningkatan akurasi sebesar 2% dari pengujian yang dilakukan. [4]

Dari beberapa pengertian di atas penulis dapat menyimpulkan bahwa Salah satu algoritma pembelajaran induktif yang paling sukses dan efisien untuk machine learning serta data mining adalah Nave Bayes.

Berikut ini adalah rumus untuk teorema *naïve bayes*:

$$1. P(H_j|X_i) = \frac{P(H_j)P(X_i|H_j)}{P(X_i)}$$

Keterangan:

- a. $P(H_j|X_i)$: peluang pada kelas j saat ada munculnya kata
- b. (H_j) : peluang munculnya sebuah kelas j
- c. $P(X_i|H_j)$: peluang kata i masuk kedalam kelas j
- d. (X_i) : peluang munculnya suatu kelas

2.1.2 *K-Nearest Neighbor*

K-Nearest Neighbor merupakan salah satu algoritma yang bekerja melakukan prediksi dengan pembelajaran terawasi dimasa hasil query instance berdasarkan mayoritas kategoriya.[2]

Algoritma ini merupakan pendekatan mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama. Disebut juga dengan pembelajar yang malas (*lazy learners*) karena hanya melihat kedekatan dengan tetangga (*neighbor*). [1]

Algoritma *K-Nearest Neighbor* Merupakan metode klasifikasi yang mengelompokkan data baru berdasarkan jarak data baru itu kebeberapa data/tetangga (*neighbor*) terdekat. [5]

Dari beberapa pengertian di atas penulis dapat menyimpulkan bahwa metode untuk melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Adapun rumus dari kedekatan jarak menggunakan persamaan sebagai berikut:

$$1. d = \sqrt{(X1 - Y1)^2 + (X2 - Y2)^2 + \dots}$$

Keterangan:

- a. $d = \textit{Euclidean Distance}$
- b. $X1 = \text{Data Pertama}$
- c. $Y1 = \text{Data Baru Pertama}$

2.2 Penelitian Terdahulu

Tahun/Penulis	Judul	Metode	Hasil
R. Situmorang, Woro Isti Rahayu, dkk (2023)	MODEL ALGORITMA K-NEAREST NEIGHBOR (K-NN) DAN NAÏVE BAYES	Metode KNN dan Naïve Bayes	hasil akurasi yang didapat dari metode k-nearest neighbor lebih besar di banding hasil prediksi dari metode naïve bayes. [2]
S. Syarli, A. muin (2016)	Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)	Metode Naïve Bayes	Nilai Presentase keakuratan menunjukkan keefektifan dataset Penerimaan Mahasiswa Baru yang diterapkan ke dalam metode Naïve Bayes Clasification. Impelementasi [6]
S. Widaningsih (2019)	PERBANDINGAN METODE DATA MINING UNTUK PREDIKSI NILAI DAN WAKTU KELULUSAN MAHASISWA PRODI TEKNIK INFORMATIKA DENGAN ALGORITMA C4.5, NAÏVE BAYES, KNN, DAN SVM	Metode C4,5, Naïve Bayes, KNN dan SVM	Naïve Bayes memiliki nilai yang paling baik untuk semua kategori performansi dibandingkan dengan algoritma lainnya. [1]
yohakim benedictus samponu, kusrini (2017)	View of Optimasi Algoritma Naive Bayes Menggunakan Metode Cross Validation Untuk Meningkatkan Akurasi Prediksi Tingkat Kelulusan Tepat Waktu.pdf	Metode Naïve Bayes	Dari hasil penelitian maka dapat diambil kesimpulan terjadi peningkatan akurasi dengan menambahkan merode pengujian Cross Validation sebesar 2% dari data pengujian yang dilakukan berdasarkan data yang ada. [4]
Asrory, Faris Irawan, Joseph Dedy Wahid, Abdul (2020)	Aplikasi Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Knn	Metode K-Nearest Nighbor (KNN)	Hasil penelitian menunjukkan bahwa prediksi menggunakan metode k-NN dengan parameter parameter di atas sebagai tolak ukur perhitungan dapat di lakukan