

BAB 2 LANDASAN TEORI

Fokus utama dalam penelitian ini adalah penyaringan kalimat SARA serta penggunaan algoritma Naive Bayes. Algoritma Naive Bayes berkaitan dengan pengambilan keputusan pada peristiwa yang berurutan dimana informasi yang diperoleh digunakan untuk merevisi peluang kejadian awal [12].

2.1 Pembelajaran Mesin

Pembelajaran mesin telah banyak diterapkan pada bidang sains karena kemampuannya yang dapat beradaptasi dengan mempelajari dan menganalisis data yang diperoleh sehingga dapat melaksanakan tugas tertentu. Pembelajaran mesin banyak digunakan dalam bidang matematika, statistika, dan data mining. Penggunaan algoritma pembelajaran mesin pada bidang-bidang tersebut bertujuan untuk membuat klasifikasi atau prediksi dalam pengembangan data.

Alur kerja dari pembelajaran mesin, dimulai dari tahap pengumpulan data, eksplorasi data, pemilihan model, proses pelatihan terhadap model yang dipilih, evaluasi model, dan prediksi. Pembelajaran mesin membutuhkan data untuk dipelajari. Data yang dibutuhkan yaitu berupa *data training* yang digunakan untuk melatih algoritma untuk mencari model yang cocok dan *data testing* yang digunakan untuk mengetes dan mengetahui performa dari model yang telah didapat pada saat tahap *testing* [13].

Berikut beberapa contoh algoritma pembelajaran mesin:

1. Naive Bayes

Naive Bayes Classifier merupakan algoritma yang digunakan untuk melakukan klasifikasi suatu data dengan metode probabilitas dan statistik. *Naive Bayes* ditemukan oleh Thomas Bayes yang mengemukakan prediksi peluang di masa depan berdasarkan peluang dari masa lalu. *Naive Bayes* sangat populer baik pada dunia komersial maupun open-source dalam memfilter data [14].

2. Support Vector Machine (SVM)

Algoritma Support Vector Machine (SVM) digunakan dalam melakukan pengelompokan pada algoritma multidimensi.

3. K-Nearest Neighbors (KNN)

Prinsip kerja KNN adalah mengklasifikasikan data baru ke kelas tertentu berdasarkan jarak data baru tersebut dengan k data terdekat.

2.2 Naïve Bayes Classifier

Naive Bayes Classifier merupakan algoritma yang digunakan untuk melakukan klasifikasi suatu data dengan metode probabilitas dan statistik. Naive Bayes Classifier ditemukan oleh ilmuwan asal Inggris bernama Thomas Bayes, yang mengemukakan bahwa prediksi peluang masa depan dapat diukur berdasarkan pengalaman masa sebelumnya. Algoritma Naive Bayes disebut sebagai "naif" dikarenakan asumsi bahwa kemunculan suatu fitur tidak bergantung dengan kemunculan fitur lain.

Proses pada algoritma Naive Bayes Classifier terbagi dalam dua tahap, yaitu tahap *training* dan *classification*. Berikut persamaan algoritma Naive Bayes [15]:

$$P(C_j | W_i) = \frac{P(C_j) \times P(W_i | C_j)}{P(W_i)} \quad (2.1)$$

Keterangan:

- i : indeks kategori yang berawal dari 1 sampai dengan kata ke-k
- j : indeks kategori yang berawal dari 1 sampai dengan kategori ke-n
- $P(C_j/W_i)$: Probabilitas kategori j berdasarkan kemunculan kata i (*posterior*)
- $P(C_j)$: Probabilitas kemunculan suatu kategori j (*prior*)
- $P(W_i/C_j)$: Probabilitas sebuah kata i masuk ke dalam kategori j (*likelihood*)
- $P(W_i)$: Probabilitas kemunculan sebuah kata (*evidence*)

Berikut tiga jenis model Naïve Bayes dalam *library* Python [16]:

1. Gaussian Naïve Bayes

Gaussian Naïve Bayes merupakan pengklasifikasian Naïve Bayes dengan asumsi data masing-masing label diambil dari distribusi Gaussian. Distribusi Gaussian merupakan asumsi pendistribusian nilai kontinu dengan fitur yang saling terkait dan berisi nilai numerik.

2. **Multinomial Naïve Bayes**

Multinomial Naive Bayes mampu memprediksi multi-kelas. Multinomial Naive bayes digunakan untuk klasifikasi fitur diskrit sehingga cocok digunakan dalam pengklasifikasian data berupa teks. Multinomial Naïve Bayes merupakan pengklasifikasian Naïve Bayes dimana fiturnya diambil dari distribusi Multinomial sederhana.

3. **Bernoulli Naïve Bayes**

Bernoulli Naïve Bayes merupakan pengklasifikasian naive bayes dimana fiturnya diasumsikan biner (0s dan 1s). Pengklasifikasiannya lebih berfokus pada hasil ya atau tidak.

2.3 **SARA**

SARA adalah singkatan dari Suku, Agama, Ras, dan Antargolongan. Indonesia terdiri dari beragam Suku, Agama, Ras, dan Antargolongan dan telah dirumuskan dalam "Bhinneka Tunggal Ika" yang memiliki makna meskipun berbeda Suku, Agama, Ras, dan Antargolongan, namun pada hakikatnya bangsa Indonesia merupakan satu kesatuan. Keragaman menuntun setiap masyarakat untuk dapat menghargai dan menerima setiap perbedaan antar sesama karena perbedaan itu merupakan karunia Tuhan.

Perbedaan yang ada juga dapat menjadi sumber permasalahan dalam masyarakat karena kurangnya toleransi terhadap setiap perbedaan. Tindakan dalam masyarakat yang berkaitan dengan SARA merupakan tindakan yang melibatkan adanya kekerasan, pelecehan, dan diskriminasi yang berkaitan dengan keturunan, agama, kebangsaan, dan golongan. Kemunculan SARA diikuti dengan kemunculan budaya negatif, seperti stereotip dan etnosentrisme yang bertentangan dengan budaya Indonesia. Seiring dengan perkembangan zaman, sosial media saat ini menjadi sarana utama bagi masyarakat dalam menyebarkan isu SARA sehingga tindakan SARA dapat mudah tersebar melalui media sosial [17].

2.4 **Confusion Matrix**

Confusion Matrix digunakan untuk mengukur performa klasifikasi. *Confusion Matrix* terdiri dari empat tabel dengan kombinasi berbeda dari nilai prediksi dan nilai aktual [18]. Penggunaan *Confusion Matrix* dalam penelitian adalah untuk

menguji sistem dan melakukan evaluasi terhadap nilai *accuracy*, *precision*, *recall*, dan *f-1 score*. Berikut tabel *confusion matrix*:

Tabel 2.1. Tabel Confusion Matrix

	Negative	Positive
Negative	TN	FP
Positive	FN	TP

Komponen *confusion matrix* [19]:

- **TP**
TP merupakan *True Positive*. Kondisi dimana jumlah data positif dan berhasil diklasifikasikan sebagai kelas positif.
- **FP**
FP merupakan *False Positive*. Kondisi dimana jumlah data negatif dan diklasifikasikan sebagai kelas positif.
- **FN**
FN merupakan *False Negative*. Kondisi dimana jumlah data positif dan diklasifikasikan sebagai kelas negatif.
- **TN**
TN merupakan *True Negative*. Kondisi dimana jumlah data negatif dan berhasil diklasifikasikan sebagai kelas negatif.

Berikut perhitungan dari Confusion Matrix untuk mendapatkan tingkat *accuracy*, *precision*, dan *recall* [20]:

Accuracy merupakan pengukuran seberapa banyak prediksi yang benar dari model yang dibuat untuk seluruh set data uji. Rumus *accuracy* dirumuskan dalam formula (2.3)

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.2)$$

Precision merupakan perbandingan antara *True Positive* (TP) dengan banyaknya data yang diprediksi positif. Rumus *precision* dirumuskan dalam formula (2.4).

$$precision = \frac{TP}{TP + FP} \times 100\% \quad (2.3)$$

Recall merupakan perbandingan antara *True Positive* (TP) dengan keseluruhan data yang diprediksi benar positif. Rumus *recall* dirumuskan dalam formula (2.5)

$$recall = \frac{TP}{TP + FN} \times 100\% \quad (2.4)$$

F1 Score merupakan perbandingan rata-rata presisi dan recall yang dibobotkan. Rumus *F1 score* dirumuskan dalam formula (2.6)

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (2.5)$$

2.5 Discord

Discord merupakan aplikasi media sosial berbasis iOS, Windows, dan Android. Discord merupakan *platform* yang didesain khusus untuk mendukung para pemain *game* untuk dapat berkomunikasi pada saat bermain *game*. Dilansir dari Liputan 6, Discord dirilis pada tahun 2015 oleh Jason Citron dan pada tahun 2020 Discord telah memperoleh rekor pengguna aktif mencapai 100 juta. Salah satu hal yang menyebabkan Discord dapat digemari oleh penggunanya adalah fiturnya yang dapat mendukung para pemain *game* dalam bekerjasama dalam tim pada saat permainan berlangsung. Fitur-fitur yang terdapat dalam Discord di antaranya sebagai berikut [21]:

1. Server

Server merupakan ruang pada Discord yang memungkinkan pengguna untuk membuat suatu grup atau komunitas yang bersifat privat maupun publik.

2. Channel (Kanal)

Channel (kanal) merupakan fitur di dalam server yang dapat digunakan oleh pengguna untuk berdiskusi dengan pengguna lain dalam satu server. Channel dibagi menjadi dua, yaitu *text channel* dan *voice channel*. Penggunaan *text channel* yaitu untuk pengiriman pesan, pengunggahan file, gambar, video, serta gif. Sedangkan, penggunaan *voice channel* yaitu untuk melakukan panggilan berupa suara atau video.

3. Direct Messages (DMs) dan Group Direct Messages (GDMs)

Direct Messages (DMs) dan Group Direct Messages (GDMs) merupakan fitur yang berfungsi agar pengguna dapat mengirimkan pesan, melakukan panggilan

lan suara, serta melakukan panggilan video yang bersifat privat kepada pengguna lain.

4. **Streaming**

Streaming berfungsi agar pengguna dapat melakukan *sharing screen* kepada pengguna lain yang berada dalam satu server serta dalam obrolan yang bersifat privat.

5. **Server Boosts**

Server Boosts berfungsi untuk meningkatkan kualitas pada server, yaitu memberikan peningkatan pada kualitas streaming, audio, video, serta penambahan slot emoji.

Berikut kelebihan yang dimiliki Discord sehingga banyak digemari oleh para pengguna:

1. **Lintas Platform**

Aplikasi Discord tersedia dalam bentuk *mobile* dan PC. Aplikasi ini dapat diunduh di Google Play store, App Store, dan website resmi Discord, sehingga pengguna dapat terhubung secara *multi-platform*.

2. **Penggunaan Fitur Voice Chat Tidak Mengganggu Performa dalam Permainan**

Ketika pengguna bermain game dengan melakukan *voice chat* pada aplikasi Discord, pengguna tidak perlu khawatir akan penurunan *frame rate* dikarenakan Discord memiliki sistem tersendiri sehingga tidak mengganggu sistem yang berjalan di CPU.

3. **Komunitas yang Luas**

Terdapat bermacam-macam komunitas dalam Discord yang memungkinkan pengguna untuk bergabung dengan suatu komunitas. Komunitas yang terdapat dalam Discord memiliki beberapa keuntungan, contohnya adalah komunitas para *gamer* dimana pengguna Discord dapat saling berbagi informasi mengenai tips dan trik dalam suatu permainan. Selain itu, para pengguna yang tergabung dalam komunitas juga dapat menambah relasi pertemanan sehingga lingkup pertemanan menjadi semakin luas.