PERANCANGAN SISTEM KLASIFIKASI UJARAN KEBENCIAN PADA BAHASA BALI DENGAN METODE NAIVE BAYES

Dewa Ayu Eka Ratna Dewi [1], Anak Agung Ngurah Mahendra Adhi Putra [2]
Program Studi Teknik Informatika K. Jembrana [1] , Program Studi Teknik Elektro K. Jembrana [2]
Fakultas Kesehatan Sains dan Teknologi, Universitas Triatma Mulya, Jembrana, Bali[1] [2]
ratna.dewi@triatmamulya.ac.id [1] , ngurah.mahendra@triatmamulya.ac.id [2]

ABSTRACT

One negative impact of social media is the freedom of everyone to express their opinions or utterances through writing, especially those containing hate content or often called hate speech. Hate speech through writing is difficult to detect because it takes time to read and determine an article that contains hate speech, especially in writing in Balinese. The hate speech is contrary to Balinese culture which upholds good manners in using everyday language. A system is needed that is able to classify Balinese writing containing hate speech. In this study a system design was made to classify hate speech in Balinese writing using the Naïve Bayes method. There are several stages in the design of this Balinese hate speech classification, including the identification of input data, preprocessing, classification using the naïve nayes classifier (NBC) method and results / outputs. Based on accuracy testing, the results obtained by 85% with the number of True Positive as much as 8 data, True Negative as much as 9 data, as much as 2 False Positive data, and False Negative as much as 1 data.
Keywords*: Hate Speech, Balinese, Naïve Bayes*

ABSTRAK

Salah satu dampak negatif dari media sosial adalah kebebasan setiap orang dalam mengemukakan pendapat atau ujaran melalui tulisan terlebih yang mengandung muatan kebencian atau sering disebut dengan hate speech. Ujaran kebencian melalui tulisan sulit untuk dideteksi karena membutuhkan waktu untuk membaca dan menentukan sebuah tulisan itu mengandung hate speech terlebih pada tulisan yang berbahasa Bali. Ujaran kebencian bertolak belakang dengan budaya masyarakat Bali yang sangat menjunjung tinggi sopan santun dalam menggunakan bahasa sehari-sehari. Diperlukan sebuah sistem yang mampu mengklasifikasi tulisan berbahasa Bali yang mengandung ujaran kebencian. Pada penelitian ini dibuat sebuah perancangan sistem untuk mengklasifikasi ujaran kebencian pada tulisan bahasa Bali menggunakan metode Naïve Bayes. Terdapat beberapa tahapan dalam perancangan klasifikasi ujaran kebencian bahasa Bali ini, diantaranya tahap identifikasi data masukan, pra-proses / preprocessing, klasifikasi dengan metode naïve nayes classifier (NBC) dan hasil/keluaran. Berdasarkan pengujian akurasi yang dilakukan diperoleh hasil sebesar 85% dengan jumlah True Positive sebanyak 8 data, True Negative sebanyak 9 data, False Positif sebanyak 2 data, dan False Negatif sebanyak 1 data.
Kata Kunci : *Ujaran Kebencian, Bahasa Bali, Naïve Bayes*

PENDAHULUAN

Perkembangan teknologi informasi membawa perubahan pada kehidupan manusia salah satunya adalah pergaulan melalui media sosial. Media sosial juga membawa dampak positif dan negatif terhadap kehidupan manusia. Salah satu dampak negatif dari media sosial adalah kebebasan setiap orang dalam mengemukakan pendapat atau ujaran melalui tulisan terlebih yang mengandung muatan kebencian atau sering disebut dengan *hate speech. Hate speech* (ujaran kebencian) adalah ujaran yang mempunyai unsur-unsur seperti segala tindakan dan usaha baik langsung maupun tidak langsung yang didasarkan pada kebencian atas dasar suku, agama, aliran keagamaan, keyakinan/kepercayaan, ras, atau

antar golongan yang dilakukan melalui berbagai sarana (HAM 2015).

Ujaran kebencian melalui tulisan sulit untuk dideteksi karena membutuhkan waktu untuk membaca dan menentukan sebuah tulisan itu mengandung *hate speech* terlebih pada tulisan yang berbahasa Bali. Ujaran kebencian bertolak belakang dengan budaya masyarakat Bali yang sangat menjunjung tinggi sopan santun dalam menggunakan bahasa sehari-sehari.

Dari permasalahan tersebut diperlukan sebuah sistem yang mampu mengklasifikasi tulisan berbahasa Bali yang mengandung ujaran kebencian. Dengan sistem maka secara otomatis mampu secara cepat menentukan sebuah tulisan mengandung ujaran kebencian atau tidak mengandung ujaran kebencian. Dalam perancangan sistem yang dapat mengklasifikasi *hate speech* tersebut dibutuhkan sebuah metode untuk klasifikasi, salah satunya adalah dengan metode Naïve Bayes.

Pada penelitian ini akan dibuat sebuah perancangan sistem untuk mengklasifikasi ujaran kebencian pada tulisan bahasa Bali menggunakan metode Naïve Bayes. Metode ini digunakan karena merupakan algoritma yang biasa digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Dengan adanya perancangan sistem ini, maka deteksi untuk ujaran kebencian yang menggunakan bahasa Bali dapat dilakukan secara mudah dan cepat. Penelitian ini juga penting dilakukan untuk menghindarkan masyarakat Bali yang penuh sopan santun dengan tulisan-tulisan yang mengandung ujaran kebencian.

TINJAUAN PUSTAKA

Ujaran Kebencian

Hate Speech (Ucapan Penghinaan atau kebencian) adalah tindakan komunikasi yang dilakukan oleh suatu individu atau kelompok dalam bentuk provokasi, hasutan, ataupun hinaan kepada individu atau kelompok yang lain dalam hal berbagai aspek seperti ras, warna kulit, etnis, gender, cacat, orientasi seksual, kewarganegaraan, agama, dan lain-lain. Dalam arti hukum, Hate speech adalah perkataan, perilaku, tulisan, ataupun pertunjukan yang dilarang karena dapat memicu terjadinya tindakan kekerasan dan sikap prasangka entah dari pihak pelaku Pernyataan tersebut ataupun korban dari tindakan tersebut (Mawarti 2018).

Penelitian lain menyebutkan bahwa ujaran kebencian (*hate speech*) merupakan tindakan komunikasi yang dilakukan oleh individu atau kelompok tertentu dalam bentuk provokasi, hasutan, hinaan, penistaan, pencemaran nama baik, serta penyebaran berita bohong dalam aspek seperti ras, warna kulit, gender, etnis, cacat fisik, orientasi seksual, kewarganeraaan, agama, dan lain-lain (Permatasari and Subyantoro2 2020).

Sedangkan penelitian lainya menyebutkan bahwa ujaran kebencian adalah tuturan yang diujarkan seseorang membawa dampak bagi pendengarnya baik itu tersurat maupun tersirat. Bahkan suatu ujaran juga akan membuat seseorang diseret ke meja hijau lantaran dianggap meresahkan. Misalnya ujaran kebencian atau sering disebut hate speech yang marak diperbincangkan di Indonesia saat ini terkait akan wacana penindakan secara hukum bagi pelaku karena dianggap menyulut kebencian bagi kelompok-kelompok tertentu (Dwi Puji Lestari 2016).

Bahasa Bali

Bahasa Bali adalah salah satu bahasa yang digunakan sebagai bahasa ibu oleh masyarakat Bali dan juga merupakan salah satu elemen budaya Bali. Bahasa Bali terkategori sebagai bahasa yang aman karena memiliki penutur di atas dua juta, memiliki tradisi tulis yang kuat dan memiliki peranan sebagai pendukung kebudayaan daerah (Seri Malini, Laksminy, and Sulibra 2018).

Bahasa Bali sebagai salah satu bahasa daerah yang ada di Indonesia, merupakan bahasa ibu dan bahasa pergaulan atau alat komunikasi bagi masyarakat suku Bali serta merupakan alat untuk mempelajari kebudayaan Bali yang berguna bagi pembinaan, pemeliharaan, dan perkembangan kebudayaan nasional. Dalam praktiknya, bahasa Bali sebagian besar dipakai oleh penutur suku Bali, bahkan bahasa daerah ini, pemakaiannya tidak terbatas di daerah Bali saja, tetapi di daerahdaerah lain pun bahasa Bali ini sering dipergunakan (I Nyoman Budi Yase 2019).

Text Preprocessing

Text preprocessing merupakan proses mengubah bentuk data yang belum memliki struktur menjadi data yang terstruktur sesuai dengan kebutuhan, untuk proses mining yang lebih lanjut (*sentiment analysis*, peringkasan, *clustering* dokumen, etc.). Sebuah teks yang ada harus dipisahkan, hal ini dapat dilakukan dalam beberapa tingkatan yang berbeda. Pengubahan bentuk dapat berupa memecah paragraf menjadi kalimat dan kalimat akhirnya menjadi kata serta dapat menghilangkan angka, simbol atau karaktek-karater lainya. Tahapan preprocessing berdasarkan, yaitu: *case folding, tokenizing/parsing, filtering, stemming* (Triawati 2009).

a. Case Folding

Case folding yaitu merubah semua karakter huruf pada sebuah kalimat menjadi huruf kecil dan menghilangkan karakter yang dianggap tidak valid seperti angka, tanda baca, dan *Uniform Resources Locator* (URL) (Indraloka and Santosa 2017). Contoh kata "KOMPUTER" akan menjadi "komputer".

b. Tokenizing

Tokenizing yaitu memotong sebuah kalimat berdasarkan tiap kata yang menyusunnya (Indraloka and Santosa 2017). Contohnya : kalimat "aku cinta bahasa" dipotong menjadi kata : aku | Cinta | bahasa.

c. Filtering

Filtering adalah tahap mengambil kata-kata penting dari hasil tokenizing. Proses filtering dapat menggunakan algoritma stoplist (membuang kata yang kurang penting) atau wordlist (menyimpan kata penting). Stoplist/stopword adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words. Contoh stopword adalah "yang", "dan", "di", "dari" dan lain – lain (Triawati 2009).

d. *Stemming* Bahasa Bali

Stemming adalah proses pemetaan dan penguraian berbagai bentuk (variants) dari suatu kata menjadi bentuk kata dasarnya (G. W. Putra and Made Sudarma 2016). Proses stemming untuk setiap Bahasa berbeda dengan Bahasa yang lain misal, proses stemming Bahasa Inggris dengan Bahasa Indonesia tentunya berbeda karena perbedaan pembentukan dan perubahan kata menjadi bentuk kata lain (Agusta 2009).

Morfologi Bahasa Bali hampir sama dengan Bahasa Indonesia. Dimana Bahasa Bali memiliki prefixes, suffixes, infexes dan confixes seperti Bahasa Indonesia (Denes et al. 1991). Hanya saja elemen pada setiap imbulan tidak semuanya sama.

Klasifikasi dengan *Naive Bayes Classifier*

Klasifikasi merupakan penentuan objek ke dalam suatu kategori atau kelas. Penentuan objek menggunakan beberapa model (Han 2006). Dalam memulai suatu klasifikasi data dengan membangun sebuah rule klasifikasi dengan algoritma tertentu yang digunakan pada data training dan data testing.

Naïve Bayes Classifier merupakan penyederhanaan dari teorema Bayes, penemu metode ini adalah seorang ilmuwan Inggris yang bernama Thomas Bayes. Algoritma dalam metode Naïve Bayes didasarkan dengan teknik klasifikasi (Kusumadewi 2009). Metode Naive Bayes dengan prinsip teorema Bayes mempunyai atribut yang saling berhubungan satu sama lain. Pendekatan yang digunakan teorema bayes yaitu menghitung probabilitas sebuah kejadian pada kondisi tertentu (Lukito and Chrismanto 2015). Dasar dari teorema Bayes dinyatakan dalam persamaan (Bustami 2014).

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

Keterangan:

X = Data kelas yang belum diketahui.
H = Hipotesis dari data X yaitu suatu kelas Spesifik.
P(H|X) = Probabilitas Hipotesis H berdasarkan kondisi X.
P(H) = Probabilitas Hipotesis H
P(H|X) = Probabilitas X berdasarkan kondisi H
P(X) = Probabilitas X

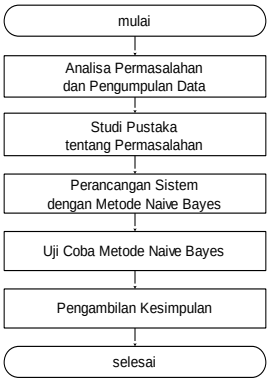
Pada rumus di atas dapat dijelaskan bahwa teorema naive bayes dibutuhkan sebuah petunjuk sebagai proses penentu kelas yang sesuai dengan sampel. Sehingga dibutuhkan kesesuaian terhadap teorema bayes sebagai berikut:

$$P\ (C \mid F_1 \dots F_n) = \frac{P\ (C)\ .\ P(F_1 \dots F_n \mid C)}{P(F_1 \dots F_n)}$$

Keterangan:
C = Sebagai kelas
F1…Fn = Petunjuk atau syarat kondisi

METODE PENELITIAN

Metode yang digunakan pada penelitian tentang klasifikasi ujaran kebencian ini adalah menggunakan metode *Naïve Bayes*. Klasifikasi *Naïve Bayes* ini biasa digunakan sebagai metode pembelajaran probabilistik untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (A. K. B. A. Putra et al. 2018). Alur penelitian dalam pembuatan perancangan sistem klasifikasi ujaran kebencian dengan metode *Naïve Bayes* dapat dilihat seperti bagan di bawah ini:



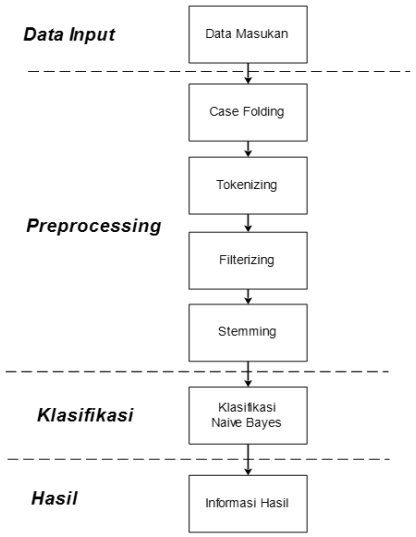
Gambar 1. Proses Pembuatan Perancangan Sistem

Pada bagan di atas terlihat beberapa proses atau tahapan dalam perancangan sistem klasifikasi ujaran kebencian dengan metode *Naïve Bayes* dapat dilihat sebagai berikut:

1. Analisa permasalahan dan pengumpulan data, pada tahap pertama adalah proses analisa permasalahan terkait dengan ujaran kebencian pada bahasa Bali. Dilanjutkan dengan pengumpulan data terkait dengan ujaran kebencian pada bahasa bali.
2. Studi pustaka tentang metode *Naïve Bayes*, tahap studi pustaka adalah pengumpulan referensi terkait permasalahan yaitu tentang ujaran kebencian, bahasa bali, dan metode klasifikasi *Naïve Bayes*.
3. Perancangan sistem dengan metode *Naïve Bayes*, pada tahap ini dilakukan perancangan untuk klasifikasi ujaran kebencian pada bahasa Bali dengan metode *Naïve Bayes.*
4. Uji coba metode Naïve Bayes, tahap selanjutnya adalah uji coba metode *Naïve Bayes* pada tulisan dengan bahasa Bali yang mengandung ujaran kebencian.
5. Pengambilan kesimpulan, pada tahap dilakukan penarikan kesimpulan terhadap perancangan yang dibuat sesuai dengan hasil yang diperoleh dari uji coba metode Naïve Bayes.

IMPLEMENTASI SISTEM

Dalam penelitian terdapat beberapa tahapan diantaranya tahap identifikasi data masukan, pra-proses/preprocessing, klasifikasi dengan metode naïve nayes classifier *(NBC)* dan keluaran. Gambaran proses pada tahapan perancangan klasifikasi ujaran kebencian ini adalah menggunakan metode *Naïve Bayes* terdapat pada gambar 2.



Gambar 2. Perancangan Klasifikasi Ujaran Kebencian Bahasa Bali

Data Masukan

Data masukan diambil dari *postingan* di sosial media seperti contoh di bawah ini :
*"**MEME CAINE MASIH NASKLENG SAJAN!!!!** 😢😢😢"*

Preprocessing

a. Case Folding

Pada tahap ini dilakukan pengubah huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan

dan dianggap sebagai delimiter. Berikut ini *case folding* yang dilakukan :

Tabel 1. Proses *Case Folding*

Data Uji		*Hasil Case Folding*	
Input	MEME CAINE MASIH NASKLENG SAJAN !!!! 😟😟😟	Output	meme caine masih naskleng sajan

b. *Tokenizing*

Tokenizing merupakan proses pemotongan string input berdasarkan tiap kata yang menyusunya serta membedakan karakter-karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan. Tahapan ini dilakukan setelah inputan data uji melewati tahap *Case Folding*. Berikut ini tokenizing yang dilakukan :

Tabel 2. Proses *Tokenizing*

Hasil *case folding*		Hasil *Tokenizing*	
Input	meme caine masih naskleng sajan	Output	| meme | | caine | | masih | | naskleng | | sajan |

c. *Filtering* / Penghapusan *Stopwords*

Kata-kata yang terkandung pada daftar stopwords yang terdapat pada daftar kata khusus stopword bahasa Bali seperti : masih, sajan, keto, lan, muah, utawi, dan yang lainnya. Adapun hasil *Filtering* adalah sebagai berikut :

Tabel 3. Proses *Filtering*

Hasil *Tokenizing*		Hasil *Filtering*	
Input	| meme | | caine | | masih | | naskleng | | sajan |	Output	| meme | | caine | | naskleng |

c. *Stemming* Bahasa Bali

Pada tahap ini dilakukan pembuangan imbuhan kata (*stemming*) bahasa Bali. Adapun stemming bahasa Bali pada penelitian ini menggunakan Kombinasi Metode *Rule-Based* dan N-Gram *Stemming* untuk Mengenali Stemmer Bahasa Bali hasil penelitian dari Made Agus Putra Subali dan Chastine Fatichah. Metode *rule-based* digunakan untuk membentuk aturan yang meluluhkan seluruh variasi imbuhan kata dalam bahasa Bali (Agus et al. 2019). Adapun hasil *stemming* menggunakan metode ini adalah :

Tabel 4. Proses *Stemming*

Hasil *Tokenizing*		Hasil *Stemming*	
Input	| meme | | caine | | naskleng |	Output	| meme | | cai | | naskleng |

Klasifikasi dengan Naïve Bayes

Setelah tahapan preprocessing text maka hasil kata yang didapatkan adalah sebagai berikut:

\| meme \| cai \| naskleng \|

Selanjutnya dilakukan klasifikasi data menggunakan algoritma Naïve Bayes. Pertama kali yang dilakukan adalah dengan membuat data training untuk kalimat yang mengandung ujaran kebencian dan tidak mengandung ujaran kebencian, yaitu seperti berikut ini:

Tabel 5. Proses *Stemming*

No	K1	K2	K3	Ujaran Kebencian
1	Cai	Jelema	Bansat	Ya
2	Cai	Jelema	Wanen	Tidak
3	Cai	Jelema	Naskleng	Ya
4	Bangka	Cai	Naskleng	Ya
5	Bangka	Celeng	Cai	Tidak
6	Bangka	Meme	Cai	Ya
7	Meme	Jelema	Wanen	Tidak
8	Meme	Cai	Bangsat	Ya
9	Meme	Manjus	Kejep	Tidak
10	Meme	Kiap	Jani	Tidak

Langkah berikutnya adalah menghitung data uji hasil preprocessing text yaitu : | meme | cai | naskleng | menggunakan rumus algoritma Naive Bayes sebagai berikut :

P (YA) = 5/10 = 0.5
P (TIDAK) = 5/10 = 0.5

P (meme | ya) = 1 / 5 = 0.2
P (cai | ya) = 4 / 5 = 0.8
P (naskleng | ya) = 2/5 = 0.4

P (meme | tidak) = 1/5 = 0.2
P (cai | tidak) = 2/5 = 0.4
P (naskleng | tidak) = 0/5 = 0

Ujaran Kebencian (Ya) :
= P(Ya) x P(meme | ya) x P(cai | ya) x P(naskleng | ya)
= 0.5 x 0.2 x 0.8 x 0.4
= 0,032

Ujaran Kebencian (Tidak) :
= P(Tidak) x P(meme | tidak) x P(cai | tidak) x P(naskleng | tidak)
= 0.5 x 0.2 x 0.4 x 0
= 0

Dari hasil perhitungan dengan menggunakan Naive Bayes didapatkan hasil : Ya > Tidak berarti tergolong Ujaran Kebencian.

Pengujian Sistem

Pada bagian ini dilakukan pengujian terhadap perancangan klasifikasi ujaran kebencian yang dibuat dengan menentukan akurasi dari sistem yang telah dibuat menggunakan persamaan sebagai berikut :

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

Keterangan:
TP = True Positive (Hasil pakar dan hasil klasifikasi menunjukkan data uji ujaran kebencian).
TN = True Negative (Hasil pakar dan hasil klasifikasi menunjukkan data uji bukan ujaran kebencian).
FP = False Positive (Hasil pakar menunjukkan data uji merupakan ujaran kebencian sedangkan hasil klasifikasi menunjukkan data uji bukan ujaran kebencian).
FN = False Negative (Hasil pakar menunjukkan data uji bukan ujaran kebencian sedangkan hasil klasifikasi menunjukkan data uji merupakan ujaran kebencian).

Pada hasil pengujian akurasi didapatkan data hasil pengujian klasifikasi ujaran kebencian bahasa bali dengan menguji data yang mengandung ujaran kebencian dan data yang bukan ujaran kebencian. Data yang digunakan diperoleh dari data 20 kalimat dalam bahasa Bali yang merupakan ujaran kebencian sebanyak 10 data dan yang bukan sebanyak 10 data. Diperoleh hasil pengujian terhadap TP, TN, FP, FN yang dijelaskan pada tabel berikut:

Tabel 1. Skenario pengujian

Data Uji	Hasil Pakar	Hasil Klasifikasi	TP	TN	FP	FN
K1	YA	YA	1	0	0	0
K 2	YA	YA	1	0	0	0
K3	YA	YA	1	0	0	0
K4	YA	YA	1	0	0	0
K5	YA	BUKAN	0	0	1	0
K6	YA	YA	1	0	0	0
K7	YA	YA	1	0	0	0
K8	YA	BUKAN	0	0	1	0
K9	YA	YA	1	0	0	0
K10	YA	YA	1	0	0	0
K11	BUKAN	BUKAN	0	1	0	0
K12	BUKAN	BUKAN	0	1	0	0
K13	BUKAN	BUKAN	0	1	0	0
K14	BUKAN	BUKAN	0	1	0	0
K15	BUKAN	BUKAN	0	1	0	0
K16	BUKAN	BUKAN	0	1	0	0
K17	BUKAN	YA	0	0	0	1

Maka diperoleh total TP = 8, TN = 9, FP =2, FN = 1, yang selanjutnya dimasukkan ke dalam perhitungan akurasi sehingga diperoleh hasil yaitu:

$$Akurasi = \frac{(8 + 9)}{(8 + 9 + 2 + 1)} = 85\ \%$$

Berdasarkan pengujian akurasi dari perancangan klasifikasi ujaran kebencian bahasa bali yang telah dibuat maka diperoleh akurasi sebesar 85% dengan jumlah *True Positive* sebanyak 8 data, *True Negative*

sebanyak 9 data, *False Positif* sebanyak 2 data, dan *False Negatif* sebanyak 1 data.

SIMPULAN

Dalam penelitian terdapat beberapa tahapan diantaranya tahap identifikasi data masukan, pra-proses / preprocessing, klasifikasi dengan metode naïve nayes classifier *(NBC)* dan keluaran Berdasarkan pengujian akurasi dari perancangan klasifikasi ujaran kebencian bahasa bali yang telah dibuat diperoleh akurasi sebesar 85% dengan jumlah *True Positive* sebanyak 8 data, *True Negative* sebanyak 9 data, *False Positif* sebanyak 2 data, dan *False Negatif* sebanyak 1 data.

DAFTAR PUSTAKA

[1] Agus, Made, Putra Subali, Chastine Fatichah, and Departemen Informatika. 2019. "Kombinasi Metode Rule-Based Dan N-Gram Stemming Untuk a Combination of Methods Rule-Based and N-Gram Stemming To Recognize Balinese Language Stemmer." *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK)* 6 (2). https://doi.org/10.25126/jtiik.201961105.

[2] Agusta, Ledy. 2009. "Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia." In *Konferensi Nasional Sistem Dan Informatika*.

[3] Bustami. 2014. "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi." *JURNAL INFORMATIKA* 8 (1): 884–98. https://doi.org/10.26555/jifo.v8i1.a2086.

[4] Denes, I Made, Ketut Reoni, Made Pasmidi, I wayan Jendra, and Bagus Nyoman Putra. 1991. *Morfologi Kata Benda Bahasa Bali*. Jakarta: Departemen Pendidikan dan Kebudayaan.

[5] Dwi Puji Lestari. 2016. "Ungkapan Kebencian Yang Muncul Pada Fenomena Islamophobia Di United Kingdom." Universitas Gajah Mada.

[6] HAM, Komisi Nasional. 2015. *Buku Saku Penanganan Ujaran Kebencian (Hate Speech)*. Komisi Nasional Hak Asasi Manusia Republik Indonesia.

[7] Han, J.K. 2006. *Concept and Techniques*. Morgan Kaufmann.

[8] I Nyoman Budi Yase. 2019. "PREFIKS PEMBENTUK VERBA BAHASA BALI DIALEK BULELENG DI KABUPATEN DONGGALA." *Jurnal Bahasa Dan Sastra* 4 (3).

[9] Indraloka, Dwi Smaradahana, and Budi Santosa. 2017. "Penerapan Text Mining Untuk Melakukan Clustering Data Tweet Shopee Indonesia." *Jurnal Sains Dan Seni ITS* 6 (2): 6–11. https://doi.org/10.12962/j23373520.v6i2.24419.

[10] Kusumadewi, Sri. 2009. "Klasifikasi Status Gizi Menggunakan Naive Bayesian Classification." *CommIT (Communication and Information Technology) Journal* 3 (1): 6. https://doi.org/10.21512/commit.v3i1.506.

[11] Lukito, Yuan, and Antonius R. Chrismanto. 2015. "Perbandingan Metode-Metode Klasifikasi Untuk Indoor Positioning System." *Jurnal Teknik Informatika Dan Sistem Informasi* 1 (2): 123–31. https://doi.org/10.28932/jutisi.v1i2.373.

[12] Mawarti, Sri. 2018. "FENOMENA HATE SPEECH Dampak Ujaran Kebencian." *TOLERANSI: Media Ilmiah Komunikasi Umat Beragama* 10 (1): 83. https://doi.org/10.24014/trs.v10i1.5722.

[13] Permatasari, Devita Indah, and Subyantoro Subyantoro2. 2020. "Ujaran Kebencian Facebook Tahun 2017-2019." *Jurnal Sastra Indonesia* 9 (1): 62–70. https://doi.org/10.15294/jsi.v9i1.33020.

[14] Putra, Aditya Kresna Bayu Arda, Mochammad Ali Fauzi, Budi Darma Setiawan, and Eti Setiawati. 2018. "Identifikasi Ujaran Kebencian Pada Facebook Dengan Metode Ensemble Feature Dan Support Vector Machine." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 2 (12).

[15] Putra, Gede Widnyana, and Satya Kumara Made Sudarma. 2016. "Klasifikasi Teks Bahasa Bali Dengan Metode Supervised Learning Naïve Bayes Classifier." *Teknologi Elektro* 15 (2).

[16] Seri Malini, Ni Luh Nyoman, Ni Luh Putu Laksminy, and I Nengah Ketut Sulibra. 2018. "Pilihan Bahasa Generasi Muda Di Destinasi Wisata Di Bali." *Jurnal Kajian Bali (Journal of Bali Studies)* 8 (1): 71. https://doi.org/10.24843/jkb.2018.v08.i01.p05.

[17] Triawati, C. 2009. "Metode Pembobotan Statistical Concept Based Untuk Klastering Dan Kategorisasi Dokumen Berbahasa Indonesia." Institut Teknologi Telkom Bandung.