

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II

LANDASAN TEORI

2.1 Bahasa Alay

Menurut Keraf (2004), bahasa merupakan alat untuk berkomunikasi antara anggota masyarakat yang berupa simbol bunyi yang dihasilkan dari alat ucap manusia. Bahasa digunakan sebagai salah satu cara untuk menyampaikan ide, perasaan, dan keinginan-keinginan yang sungguh manusiawi dengan menggunakan simbol-simbol dengan sengaja. Bahasa adalah suatu alat yang digunakan untuk berkomunikasi dalam sehari-hari baik itu lisan maupun tulisan.

Pada tahun 1970-an, bahasa “Prokem” muncul dan digunakan oleh para pencopet dan bandit. Bahasa ini digunakan sebagai bahasa kode dalam kalangan kelompok dengan tujuan agar percakapan mereka tidak diketahui oleh orang lain. Lambat laun bahasa prokem mengalami pergeseran fungsi dari bahasa rahasia menjadi bahasa pergaulan “bahasa gaul” anak-anak remaja.

Bahasa “Gaul” muncul pada akhir tahun 2000-an. Bahasa ini dikenal sebagai bahasanya para preman atau anak jalanan. Bahasa ini digunakan untuk menyamarkan apa yang mereka bicarakan agar tidak terlacak oleh orang lain, terutama polisi. Tetapi, semakin hari bahasa gaul menjadi bahasa keseharian yang digunakan oleh para remaja. Hal ini di karenakan remaja memasuki tahapan psikososial yaitu tahap pencarian dan pembentukan jati diri. Remaja ingin di akui sebagai individu yang unik yang memiliki identitas sendiri terlepas dari dunia anak-anak dan dewasa.

Seiring perkembangan teknologi, bahasa gaul tidak hanya sekedar istilah, namun juga sudah meluas ke penulisannya. Tulisan gaul ini disebabkan karena maraknya penggunaan jejaring sosial seperti *Facebook*, *Yahoo Messenger*, *BBM* dan *Twitter*. Banyak bermunculan istilah-istilah baru yang semakin sulit untuk dipahami, dalam hal ini munculah komunitas baru yaitu “Alay”. Alay muncul pada tahun 2010-an yang penggunaannya pertama kali di kenalkan oleh para remaja

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Jakarta. Bahasa Alay dapat di katakan sama dengan bahasa prokem atau bahasa gaul (Kelana, 2011).

Menurut Kelana (2011), Alay merupakan singkatan dari “anak layangan” atau “anak lebay”. Alay menggambarkan gaya hidup yang norak, kampungan dan berlebihan untuk menarik perhatian orang disekitarnya. Gaya hidup tersebut di alami oleh remaja Indonesia yang ingin selalu terlihat eksis atau narsis mereka dalam segala hal misalnya berpakaian, perilaku, dan berbahasa baik lisan maupun tulisan. Bahasa Alay cenderung seperti bahasa anak kecil yang kesulitan dalam menyebutkan kata dan menunjukkan sikap tutur yang manja dengan tujuan agar diperhatikan oleh orang lain. Hal ini biasanya dilakukan oleh para remaja. karena menurut Kelana (2011) Alay merupakan ciri khas remaja.

Hermawan (2016), menyatakan Masa remaja antara usia 17 – 24 tahun untuk wanita dan usia 17 – 25 tahun untuk pria. Sedangkan rentang usia menurut Depkes RI yaitu :

1. Masa balita : 0 – 5 tahun
2. Masa kanak- kanak : 5 – 11 tahun
3. Masa remaja awal : 12 – 16 tahun
4. Masa remaja akhir : 17 – 25 tahun
5. Masa dewasa awal : 26 – 35 tahun
6. Masa dewasa akhir : 36 – 45 tahun
7. Masa lansia awal : 46 – 45 tahun
8. Masa lansia akhir : 56 – 65 tahun
9. Masa manula : 65 ke atas.

2.1.1 Fungsi Bahasa Alay

Berikut ini fungsi menggunakan variasi bahasa Alay (Kelana, 2011) :

1. Fungsi gaul. seseorang yang mengikuti segala sesuatu yang sedang trend dikalangan masyarakat, dengan mengikuti trend tersebut seseorang baru dapat dikatakan “gaul” seperti *music, fashion, style*, rambut, film, *actor* dan lain sebagainya.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2. Fungsi identitas diri. Variasi bahasa Alay juga menunjukkan keberadaan anak-anak Alay.
 - a. Identitas diri, Pengguna variasi bahasa Alay menggunakan bentuk penulisan untuk menunjukkan identitas dirinya.
 - b. Identitas kelompok, Agar orang awam mengenali mereka sebagai kelompok Alay mereka juga membuat identitas mereka dengan menggunakan tulisan dan *style* serta dandanan tersendiri yang menandakan mereka komunitas Alay.
 - c. *Filter password*, dikarenakan penulisan Alay yang unik, banyak orang yang menggunakan bahasa Alay sebagai password agar tidak mudah dibobol.

Penambahan kesan lucu dan unik, Pada mulanya penulisan bahasa Alay digunakan hanya untuk menambah kesan lucu dan unik pada tulisan mereka, tetapi lama kelamaan mereka nyaman dengan gaya penulisan yang seperti itu.

2.1.2 Ciri-Ciri Alay

Ciri-ciri anak Alay dapat dilihat dari gaya bahasa, cara berpenampilan, cara berfoto, sering berganti-ganti warna rambut, memiliki tempat tongkrongan tersendiri, cara memakai aksesoris yang berlebihan, terlalu sering meng-*update* status di media sosial, suka meniru kebiasaan idolanya dan dari gaya penulisan.

Menurut Sumarni (2015), ciri-ciri anak alay di media sosial yaitu sebagai berikut :

1. Menulis *tweet* dengan ciri khas penulisan anak Alay
2. Terlalu sering meng-*update* status di media sosial
3. Sering meng-*upload* foto dengan gaya dua jari, atau dengan gaya dari berbagai sisi yang tidak jelas misalnya gaya yang kesamping, ke atas dan lain-lain.
4. Foto yang di-*upload* juga telah di-*edit* agar terlihat lebih cantik
5. Nama akun pengguna Twitter biasanya menggunakan nama samaran tidak nama sendiri (ahmad slalu bersyukur, putrii chie imuet). Hal tersebut menunjukkan bahwa mereka ingin terlihat lebih baik lebih hebat dan lebih keren di bandingkan dengan yang lain.

Beberapa istilah yang terdapat pada Twitter yaitu *Direct Message*, *Favorite*, *Follow*, *Unfollow*, *Follower*, *Following*, *Hashtag*, *Retweet*, *Timeline* dan *Tweet*. *Tweet* adalah teks tulisan yang hanya bisa memuat 140 karakter. *Tweet* dapat dilihat secara publik, namun pengirim dapat membatasi pengiriman pesan (*direct message*) ke daftar *follower* mereka saja. Pengguna bisa melihat *Tweets* pengguna lain (*follower*).

Muhammad Moeen Uddin dkk (2014) terdapat 17 fitur yang bisa digunakan untuk mengklasifikasikan Akun Twitter yaitu: *Favorite*, *Verified*, *Plain status*, *Replies received*, *Replies given*, *Retweet*, *Mention*, total *URLs*, total *hashtag*, *Promotion score*, *Life time*, *Tweet spread*, *Std URLs*, *Std Hashtags*, *User collective activeness*, *Degree of influence* dan *collective influence*.

2.3.1 Text Preprocessing

Text preprocessing berfungsi untuk menseleksi data tidak terstruktur menjadi data yang terstruktur. Proses *preprocessing* dibagi menjadi lima tahapan yaitu: *cleaning*, *case folding*, *tokenizing*, *filtering*, *stemming*.

1. *Cleaning* berfungsi untuk membuang atribut-atribut yang tidak penting seperti *hashtag*, *URL*, *mention* dan *emoticon*.
2. *Case Folding* berfungsi untuk mengubah seluruh teks atau dokumen menjadi huruf kecil.
3. *Tokenizing* berfungsi untuk memotong suatu kalimat menjadi per-kata.
4. *Filtering* berfungsi untuk mengambil kata-kata penting dari hasil tokenisasi. *Filtering* dibagi menjadi dua yaitu: *stopword* (membuang kata kurang penting dari hasil token) contohnya “yang”, “di”, “dari”, dan *wordlist* (menyimpan kata-kata yang penting dari hasil token).
5. *Stemming* berfungsi untuk mengubah hasil *filtering* menjadi kata dasar. Contohnya, “membela” menjadi “bela”.

2.4 Metode Naïve Bayes Classifier

Naïve Bayes Classifier merupakan salah satu metode yang digunakan untuk mengklasifikasikan sebuah data dengan menggunakan probabilitas. Probabilitas

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

untuk pohon keputusan dapat dihitung dengan syarat, bahwa kelas keputusan sudah benar mengingat vector informasi proyek.

Dalam segi mengolah data metode ini mudah dan sangat sederhana, selain itu metode ini juga memberikan tingkat akurasi yang baik dan akurat dan pengklasifikasian data. Cara kerja *Naive Bayes* lebih baik jika dibandingkan dengan model classifier yang lain (Xhemali, Hinde dan Stone, 2009) dalam jurnalnya " *Naive bayes vs Decission tree vs Neural Networks in the Classification of training web pages* " mengatakan bahwasanya *Naive Bayes* memiliki tingkat akurasi yang lebih baik dibanding dengan *classifier* yang lain. Kelebihan metode *Naive Bayes* yaitu metode hanya membutuhkan jumlah data pelatihan yang kecil untuk menentukan estimasi parameter yang dibutuhkan pada saat proses pengklasifikasi. *Naive Bayes* juga mudah dilakukan pada saat implementasi, memberikan hasil yang baik untuk banyak kasus. Tapi metode ini juga memiliki kekurangan yaitu bahwa antar fitur tidak terkait dalam realita, karena keterkaitan tersebut tidak dapat di modelkan oleh metode *Naive Bayes*. Berikut ini rumus *Naive Bayes Classifier* :

$$P(H/X) = \frac{P\left(\frac{X}{H}\right)P(H)}{P(X)} \quad (2.1)$$

Keterangan:

- X : Data dengan kelas yang belum diketahui
- H : Hipotesa data yg merupakan spesifik
- P(H/X) : Probabilitas hipotesis berdasarkan posterior
- P(H) : Probabilitas hipotesis
- $P\left(\frac{X}{H}\right)$: Probabilitas berdasarkan kondisi hipotesis
- P(X) : Probabilitas H

Konsep dasar yang digunakan *Naive Bayes* adalah Teorema Bayes (teorema dalam statiska untuk menghitung peluang), Bayes Optimal Classifier (menghitung peluang satu kelas menjadi beberapa kelompok atribut yang ada, lalu menentukan kelas mana yang paling optimal). Terdapat dua jenis peluang yaitu: peluang posterior dan peluang prior. Contohnya seperti: "Jika X melambangkan set atribut

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

data dan Y melambangkan kelas variabel. Peluang bersyarat $P(Y/X)$ dikenal dengan posterior peluang untuk Y, dan peluang untuk prior Y adalah $P(Y)$ ”.

Perhitungan *Naïve Bayes* dapat berupa data *Kontinu* dan data *diskrit*, tetapi cara penyelesaiannya menggunakan cara yang berbeda, jika datanya berupa data diskrit maka cara penyelesaiannya dengan menggunakan cara *Naïve Bayes* yang biasa, tetapi jika datanya berupa data *kontinu* maka proses penyelesaiannya menggunakan pendekatan *Densitas Gauss*. Mencari nilai *mean* dan nilai *standar deviasinya*. Setelah itu, baru melakukan tahap Perhitungan probabilitasnya dilakukan untuk setiap variabel *diskrit* dan *kontinu*.

Rumus nilai *mean* :

$$\mu = \frac{x_1+x_2+x_3+\dots+x_n}{n} \quad (2.2)$$

Keterangan :

- X_i : Nilai x ke- i
 μ : Rata-rata hitung
 n : Jumlah Sampel

Rumus *Standart Deviasi* :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \quad (2.3)$$

Keterangan :

- σ : *Standar Deviasi*
 x_i : Nilai x ke -i
 μ : Nilai Mean
 n : Jumlah sampel

Rumus *Densitas Gauss*:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.4)$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Ket:

- π : Nilai *Phi* (3.14)
- σ : Nilai *Standart Deviasi*
- e : Nilau Eksponden
- X : Data x
- μ : Nilai *Mean*

Berikut ini rumus nilai *LikeLihood* :

$$p(d|cj) = p(A1|cj)p(A2|cj) \dots p(An|cj) \tag{2.5}$$

2.5 Pengujian

Tahap terakhir yaitu melakukan tahap pengujian terhadap sistem yang telah dibuat. Pengujian sistem dilakukan dengan menggunakan *confusion matrix* :

Rumus *Confusion Matrix* :

$$\text{Akurasi} = \frac{(TP+TN)}{TP+TN+FP+FN} \tag{2.6}$$

Keterangan *Confusion Matrix* :

1. Jika hasil prediksi negatif dan data sebenarnya negatif
2. Jika hasil prediksi positif dan data sebenarnya negatif
3. Jika hasil prediksi negatif dan data sebenarnya positif
4. Jika hasil prediksi positif dan data sebenarnya positif

2.6 Penelitian Terkait

Berikut ini adalah penelitian-penelitian terkait klasifikasi akun Twitter, diantaranya yaitu:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 2.1 Penelitian Terkait

Nama	Judul	Tahun	Kesimpulan
Susilo dan Hastho	Pilihan Bahasa dalam Iklan Televisi	2007	Pemilihan ragam bahasa yang sesuai untuk bahasa iklan, dan variasi bahasa Alay termasuk bahasa yang bisa digunakan untuk bahasa iklan.
Pardosi	<i>The Description Of Alay Language Found In Twitter</i>	2014	Mengetahui karakteristik bahasa alay dalam status Twitter, hasil yang ditemukan yaitu (9,88%) menggunakan akronim, menggunakan huruf tambahan (22,08%). percampuran huruf dan angka (9,88%). menggunakan kombinasi antara huruf, angka (16 %). Menggunakan istilah Alay secara dominan.
Hermawan	<i>Gangguan Psikogenik Orang "Alay"</i>	2016	Orang yang menggunakan bahasa Alay dikatakan sebagai "Psikogenik", yaitu gangguan dibidang mental, yang mana penuturnya ingin mencari perhatian orang lain.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Nama	Judul	Tahun	Kesimpulan
Uddin dkk	<i>Understanding Types of Users on Twitter</i>	2014	Mengklasifikasikan Akun Twitter kedalam 6 kelas yaitu, personal, professional, spam, Business, Viral, and news. dengan menggunakan 17 fitur.
Ugheoke dkk	<i>Detecting the Gender of a Tweet Gender</i>	2014	Mendeteksi jenis kelamin pengirim <i>tweet</i> dengan tingkat akurasi 86,8% dan 95,4%.
Liantoni,	Klasifikasi Daun Herbal Menggunakan Metode Nive Bayes Classiffier dan K-nearest Neighbor	2015	Hasil klasifikasi untuk daun herbal, menunjukkan bahwa metode Naïve Bayes lebih baik dengan tingkat akurasi 75%. Knearest Neighbor nilai akurasinya 70,83%
Nagpal	<i>Application of Machine Learning Techniques for Twitter User Classification</i>	2014	Membandingkan tiga metode Decision tree, SVM dan Naïve Bayes untuk mengklasifikasikan pengguna twitter, dan Naïve bayes merupakan metode yang paling tinggi tingkat akurasinya dibanding SVM dan decision tree, yaitu 80. 9%.