



PENDETEKSIAN KALIMAT SINDIRAN MENGGUNAKAN ALGORITMA NAIVE BAYES

Muhammad Fikry¹, Safwandi²

^{1,2} Teknik Informatika Universitas Malikussaleh Lhokseumawe
Jl. Cot Tgk Nie-Reulet, Aceh Utara, 141 Indonesia
email: mfikry86@gmail.com

Abstrak – Dalam kehidupan masyarakat modern ini tidak terlepas dari dari media sosial, dimana penggunaan media sosial yang tidak bijak bisa mengakibatkan hal yang fatal. Dimasa sekarang ini banyak terjadi kasus pem-bully-an maupun kasus dimana masyarakat saling sindir menyindir sehingga terjadi saling permusuhan antar masyarakat. Pada penelitian kali ini penulis akan membuat sebuah sistem untuk mendeteksi kalimat-kalimat sindiran yang sering terjadi dalam kehidupan sehari-hari masyarakat indonesia. Penulis menggunakan metode Naive Bayes, Metode ini juga sangat sering digunakan oleh peneliti-peneliti lain karena tingkat akurasi yang tinggi.

Kata Kunci – Pendeteksian kalimat, Kalimat sindiran, Deteksi sindiran, Naive Bayes, Metode Naive Bayes.

Pendahuluan

Pada era globalisasi ini umumnya masyarakat sudah mengenal yang namanya gadget. Kehidupan masyarakat sehari-hari seperti kegiatan belanja, kegiatan saling bersosial sangat banyak dilakukan dengan melalui media sosial. Namun, tidak hanya itu karena bermedia sosial juga terdapat hal-hal yang negatif, salah satunya adalah saling sindir-menyindir.

Kegiatan saling sindir-menyindir ini sangat mempengaruhi kehidupan sosial seseorang, maka dari itu sudah sepatutnya media sosial melakukan pendeteksian terhadap kata-kata sindirian untuk mencegah terjadinya permusuhan maupun pem-bully-an yang sering terjadi akhir-akhir ini.

Naive Bayes Classifier bekerja sangat baik dibanding dengan model classifier lainnya. Hal ini dibuktikan pada jurnal Xhemali, Daniela, Chris J. Hinde, and Roger G. Stone. "Naive Bayes vs. decision trees vs. neural networks in the classification of training web pages." (2009), mengatakan bahwa "Naive Bayes Classifier memiliki tingkat akurasi yg lebih baik dibanding model classifier lainnya".

Keuntungan penggunaan adalah bahwa metoda ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi

parameter yg diperlukan dalam proses pengklasifikasian. Karena yg diasumsikan sebagai variabel independent, maka hanya varians dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians.

Landasan teori

Pengertian sindiran

Sindiran berasal dari kata dasar sindir. Sindiran memiliki arti dalam kelas nomina atau kata benda sehingga sindiran dapat menyatakan nama dari seseorang, tempat, atau semua benda dan segala yang dibendakan.

Nomina (kata benda)

Perkataan (gambar dan sebagainya) yang bermaksud menyindir orang celaan (ejekan dan sebagainya) yang tidak langsung.

Pengertian kalisifikasi

Klasifikasi adalah 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 [5]. 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).

Pengertian algoritma naive bayes

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Bayesian classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan dengan decision tree dan neural network. Bayesian classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. Metode Bayes merupakan pendekatan statistic untuk melakukan inferensi induksi pada persoalan klasifikasi. Pertama kali dibahas terlebih dahulu tentang konsep dasar dan definisi pada Teorema Bayes, kemudian menggunakan teorema ini untuk melakukan klasifikasi dalam Data Mining. Teorema Bayes memiliki bentuk umum sebagai berikut :

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

Keterangan :

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$ = Probabilitas hipotesis H (prior prob.)

















$P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut

$P(X)$ = Probabilitas dari X

Pengertian flowchart

Flowchart adalah adalah suatu bagan dengan simbol-simbol tertentu yang menggambarkan urutan proses secara mendetail dan hubungan antara suatu proses (instruksi) dengan proses lainnya dalam suatu program.

Dalam perancangan flowchart sebenarnya tidak ada rumus atau patokan yang bersifat mutlak (pasti). Hal ini didasari oleh flowchart (bagan alir) adalah sebuah gambaran dari hasil pemikiran dalam menganalisa suatu permasalahan dalam komputer. Karena setiap analisa akan menghasilkan hasil yang bervariasi antara satu dan lainnya. Kendati begitu secara garis besar setiap perancangan flowchart selalu terdiri dari tiga bagian, yaitu input, proses dan output.

	Flow Direction symbol Yaitu simbol yang digunakan untuk menghubungkan antara simbol yang satu dengan simbol yang lain. Simbol ini disebut juga connecting line.		Simbol Manual Input Simbol untuk pemasukan data secara manual on-line keyboard
	Terminator Symbol Yaitu simbol untuk permulaan (start) atau akhir (stop) dari suatu kegiatan		Simbol Preparation Simbol untuk mempersiapkan penyimpanan yang akan digunakan sebagai tempat pengolahan di dalam storage.
	Connector Symbol Yaitu simbol untuk keluar - masuk atau penyambungan proses dalam lembar / halaman yang sama.		Simbol Predefine Proses Simbol untuk pelaksanaan suatu bagian (sub-program)/prosedure
	Connector Symbol Yaitu simbol untuk keluar - masuk atau penyambungan proses pada lembar / halaman yang berbeda.		Simbol Display Simbol yang menyatakan peralatan output yang digunakan yaitu layar, plotter, printer dan sebagainya.
	Processing Symbol Simbol yang menunjukkan pengolahan yang dilakukan oleh komputer		Simbol disk and On-line Storage Simbol yang menyatakan input yang berasal dari disk atau disimpan ke disk.
	Simbol Manual Operation Simbol yang menunjukkan pengolahan yang tidak dilakukan oleh komputer		Simbol magnetik tape Unit Simbol yang menyatakan input berasal dari pita magnetik atau output disimpan ke pita magnetik.
	Simbol Decision Simbol pemilihan proses berdasarkan kondisi yang ada.		Simbol Punch Card Simbol yang menyatakan bahwa input berasal dari kartu atau output ditulis ke kartu
	Simbol input-Output Simbol yang menyatakan proses input dan output tanpa tergantung dengan jenis peralatannya		Simbol Dokumen Simbol yang menyatakan input berasal dari dokumen dalam bentuk kertas atau output dicetak ke kertas.

Gambar 1. Simbol Flowchart

Metodologi penelitian

Analisa

Analisa sistem adalah metode untuk menemukan kelemahan-kelemahan sistem guna memperoleh gambaran terhadap sistem yang akan dikembangkan sehingga dapat diusulkan perbaikan. Tahapan dalam menganalisa sistem diawali dengan mempelajari bagaimana mengidentifikasi masalah-masalah yang dihadapi, mengidentifikasi pengguna (user) sistem serta spesifikasi perangkat lunak yang akan dikembangkan. Analisis kebutuhan sistem dimaksudkan untuk mengetahui faktor -faktor apa saja yang mempengaruhi sistem. Faktor-faktor tersebut akan menjadi tolak ukur dalam proses pengembangan sistem selanjutnya.

Identifikasi Masalah

Berdasarkan hasil penelitian dan pengamatan, diperlukan sebuah sistem yang bisa mendeteksi kata-kata sindiran untuk meminimalisir permusuhan antara masyarakat. Maka dari itu penulis mencoba untuk mengidentifikasi masalah sebagai berikut.

- 1) Banyaknya kasus bullying
- 2) Banyaknya kasus pembunuhan yang terjadi akibat permusuhan di media sosial.
- 3) Banyaknya kasus pencemaran nama baik.
- 4) Tidak adanya sistem yang memadai untuk mendeteksi kata-kata sindiran.

Hasil dan pembahasan

Implementasi

Penjelasan yang meliputi cara, langkah-langkah serta jadwal pelaksanaan untuk mengimplementasikan rancangan perangkat lunak. Jadwal implementasi ini juga menjelaskan tentang aktifitas-aktifitas yang akan dilakukan. Berikut ini aktifitas yang dilakukan dalam merancang dan mengimplementasikan Pendeteksian Kalimat Sindiran Menggunakan Algoritma Naive Bayes, sebagai tahap awal yang dilakukan dalam rangka instalasi perangkat apa saja yang dibutuhkan untuk membuat Pendeteksian Kalimat Sindiran Menggunakan Algoritma Naive Bayes.

Kebutuhan Perangkat Keras (Hardware)

Spesifikasi perangkat keras minimum yang diperlukan dalam implementasi Sistem Pendeteksian Kalimat Sindiran Menggunakan Algoritma Naive Bayes adalah sebagai berikut:

1. Personal komputer/ Laptop

- a. Processor Intel core duo.
 - b. Hard disk 250 Gb.
 - c. RAM DDR3 2 Gb
2. Smartphone
- a. Processor 1,0 Ghz
 - b. Memory 16 Gb.
 - c. RAM 2 Mb.

Kebutuhan Perangkat Lunak(Software)

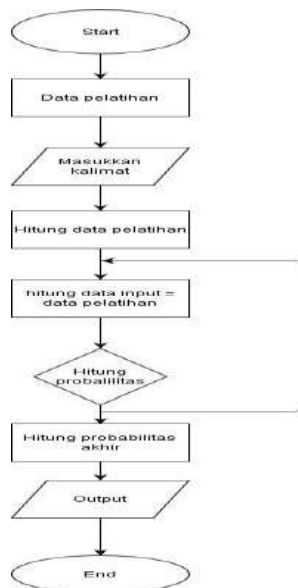
Untuk implementasi Sistem Pendeteksian Kalimat Sindiran Menggunakan Algoritma Naive Bayesini digunakan software adalah sebagai berikut:

Komputer/ Laptop

Operating System Microsoft Windows 7

PHP Server (XAMPP)

Visual Studio Code



Gambar 2. Flowchart Sistem

Pengujian

Pengujian dilakukan terhadap dua kalimat yang bernilai sindiran dan bukan sindiran :

Pengujian terhadap Kalimat Sindiran

Pengujian terhadap kalimat sindiran menggunakan kalimat "Teman datang

seiring banyaknya kebutuhan.



Gambar 3. Pengujian Terhadap Kalimat Sindiran

Pengujian terhadap kata bukan sindiran

Pengujian terhadap kata bukan sindiran menggunakan kata "Persahabatan adalah bayang-bayang malam, yang bertambah seiring terbenamnya matahari kehidupan."



Gambar 4. Pengujian Terhadap Kalimat Bukan Sindiran

Kesimpulan

Penerapan Sistem Pendeteksian Kalimat Sindiran Menggunakan Algoritma Naive Bayes adalah salah satu solusi untuk menciptakan media sosial yang lebih nyaman bagi masyarakat. Kesimpulan yang dapat diambil dari penelitian ini adalah:

1. Penerapan Sistem Pendeteksian Kalimat Sindiran Menggunakan Algoritma Naive Bayes memberikan hasil yang akurat
2. Penerapan Sistem Pendeteksian Kalimat Sindiran Menggunakan Algoritma Naive Bayes akan mampu mengurangi permusuhan masyarakat karena sindiran.
3. Penerapan Sistem Pendeteksian Kalimat Sindiran Menggunakan Algoritma Naive Bayes akan menjadi solusi media sosial yang nyaman.

4. Algoritma Naive Bayes mampu mendeteksi kalimat sindiran dan kalimat bukan sindiran dengan baik.

Daftar pustaka

- [1] B. Liu, —Sentiment Analysis and Subjectivity, *Handb. Nat. Lang. Process.*, no. 1, pp. 1–38, 2010. <http://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf> [http://people.sabanciuniv.edu/berrin/proj102/1-BLiu-Sentiment Analysis and Subjectivity- NLPHandbook-2010](http://people.sabanciuniv.edu/berrin/proj102/1-BLiu-Sentiment%20Analysis%20and%20Subjectivity-NLPHandbook-2010) [Online]. Available: [pdf](http://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf)
- [2] N. D. Putranti and E. Winarko, —Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine, *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 8, no. 1, pp. 91–100, 2014 [Online]. Available: <https://jurnal.ugm.ac.id/ijccs/article/view/3499>.
- [3] Turban, E., Aronson, J.E., and Liang, T.P., 2005, *Decision Support Systems and Intelligent Systems (Sistem Pendukung Keputusan dan Sistem Kecerdasan)*, Yogyakarta: Penerbit Andi