

## **PENDEKATAN *TEXT MINING* SEBAGAI SISTEM PENDETEKSI PEMBERITAAN PALSU YANG TERSEBAR DALAM TWITTER**

Budi Setiawan Tanjung  
Fakultas Teknik, Program Studi Teknik Informatika, Universitas Widya Kartika  
Budisetiawan760@gmail.com

### **ABSTRAK**

Berita dari *tweet* menjadi informasi yang dibaca oleh ribuan pengguna *twitter*. Pengguna *twitter* memiliki kebebasan menambahkan berita menurut opininya sendiri dan dapat dibaca oleh hampir seluruh pengguna *twitter*. Sering kali ada berita yang tidak benar tersebar sangat luas tanpa diketahui sumber beritanya. Inilah yang menjadi penyebab penyebaran berita palsu menjadi sangat luas. Untuk mengurangi penyebaran berita palsu maka perlu ada pengelolaan *tweet* agar menjadi berita yang bermanfaat dengan mengecek apakah berita tersebut asli ataupun palsu. Sistem pendeteksi berita palsu ini menjadi suatu alat yang membantu dalam melakukan klasifikasi data dari *twitter* ke dalam kategori berita asli dan berita palsu. Klasifikasi yang dilakukan terhadap *tweet* menggunakan metode *naïve bayes classifier*. Dengan metode *Naïve bayes classifier* sebuah kalimat dapat dikelompokkan berdasarkan bobot probabilitas kata yang dihitung dalam kategori yang ditentukan. Dari hasil penjumlahan bobot per kategori tersebut maka dapat diambil nilai tertinggi yang menentukan suatu kalimat *tweet* merupakan kalimat palsu ataupun tidak.

**Kata Kunci:** *Text mining, naïve bayes, klasifikasi, tweet.*

### **1. PENDAHULUAN**

Sosial media merupakan sumber informasi yang banyak digunakan saat ini. Digunakan oleh dua orang atau lebih. Biasanya digunakan untuk berinteraksi dengan teman dan tempat berbagi tujuan bersama. Bentuk informasi yang dibagikan pun sangat beragam, mulai dari berita, gambar, maupun video. Platform sosial media seperti *Facebook, Whatsapp, Instagram* selama ini belum bisa menyaring secara penuh informasi masuk di dunia maya. Konten yang tersebar ke pengguna terkadang belum pernah disaring secara signifikan, tidak ada pengecekan fakta, dan juga tidak melewati pengecekan editor. Dalam berberapa kasus sebuah berita dapat menjadi viral dan mendapatkan sorotan utama dalam berita nasional.

*Twitter* adalah salah satu media sosial yang populer digunakan orang saat ini. *Twitter* ini bisa dibilang sangat mudah untuk menyebarkan informasi dan opini seseorang dengan menggunakan fitur *hashtaging*, yang berfungsi sebagai kata kunci untuk pencarian suatu informasi. Dalam *twitter* terdapat 3 bentuk informasi yang dapat dibagikan yaitu adalah teks, tautan link gambar dan juga video.

Kebenaran bukan lagi menjadi suatu hal yang utama karena berita palsu atau yang biasa disebut

*hoax* yang membuat suatu fakta alternatif yang dapat membelokkan sudut pandang setiap individual. Isi pesan dari suatu berita yang dibaca dapat berubah-ubah sesuai dengan sudut pandang dan pengertian orang yang membaca. Banyak orang yang percaya dengan berita palsu yang akan menjadi pertimbangan baru dalam masyarakat. Sosial media memungkinkan seseorang atau grup untuk membuat suatu 'fakta alternatif / palsu' dan menerbitkan informasi tersebut, sehingga ada kecenderungan berita tersebut dapat tersebar luas.

Hal ini menjadi isu serius yang dihadapi oleh pemerintah Indonesia dalam komando kementerian komunikasi dan informatika. Presiden Joko Widodo menegaskan bahwa pemerintah akan bertarung melawan penyebaran berita palsu, atau *hoax* atau berita burung yang semakin merajalela (Kompas, 2017). Hal ini juga sejalan dengan menteri agama, Lukman Hakim Saifuddin mengingatkan mahasiswa untuk bersikap kritis terhadap dinamika informasi. Mencermati suatu berita, siapa yang menulis, tujuannya, lalu apa kepentingan pembuat tulisan dan dampak pada masyarakat yang membaca berita tersebut (Kompas, 2017). Menurut Menkominfo pemerintah dan seluruh segmen masyarakat harus bekerja sama dalam memerangi berita palsu yang menyebar di sosial media. Menkominfo juga

membuat sistem *Trust+* sebagai alat penyaring konten berita palsu, saat ini berisi sudah berisi 800 ribu konten *blacklist* (Industri, 2017). Meskipun begitu bentuk dari konten yang tersebar cepat berubah dan berkembang seiring perubahan jaman. Sehingga masih banyak konten yang masih belum tersaring.

Salah satu topik berita yang berbahaya jika salah penyebaran informasinya adalah topik kesehatan. Menurut dr. Astrawinata.G ada banyak tipe hoax tentang kesehatan mulai dari informasi yang salah tentang sebuah penyakit sampai cara pengobatan yang salah. Tak jarang juga pelaku penyebar memiliki tujuan untuk mencari keuntungan dari efek penyebaran berita tersebut. Berita yang tersebar ini jelas akan merugikan pembaca sebab seringkali ada biaya yang dikeluarkan pasien dan jumlahnya tidak rasional (analisedaily, 2017). Jangka panjang dari berita dalam topik ini adalah dapat merubah kebiasaan hidup seseorang dari yang sebelumnya benar menjadi salah dan dapat menghilangkan nyawa manusia.

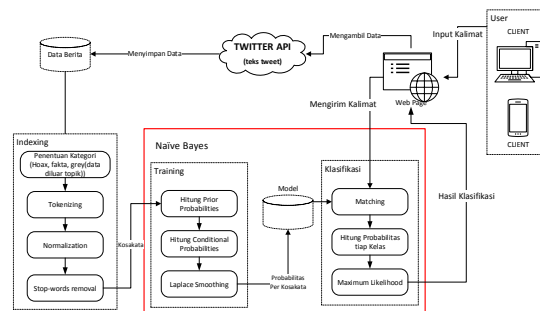
Berdasarkan latar belakang permasalahan diatas, maka penulis merasa perlu untuk membuat sebuah sistem untuk dapat melakukan pengecekan apakah berita tersebut adalah hoax atau fakta. Dirancanglah penelitian yang berjudul Pendekatan *Text Mining* sebagai Sistem Pendeteksi Pemberitaan Palsu yang Tersebar dalam Twitter. *Text mining* merupakan variasi dari *data mining* yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar. *Text mining* juga digunakan untuk menangani masalah klasifikasi, *clustering*, *information extraction*, dan *information retrieval* (Berry, M.W. & Kogan, J. 2010).

## 2. METODE PENELITIAN

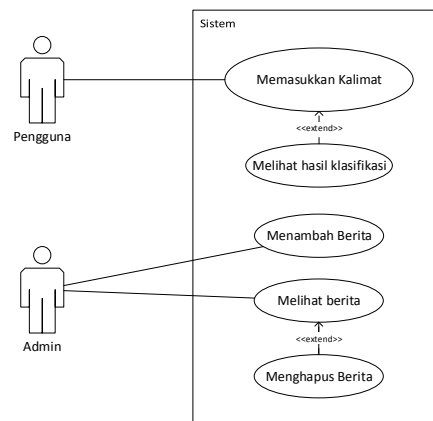
### 2.1. Rancangan Sistem

Gambar 1 merupakan rancangan sistem yang akan dibuat, data training didapat dari kalimat *tweet* yang diambil melalui *twitter api* yang kemudian disimpan dalam database. Kemudian kategori dari database dipilah menjadi 3 yaitu database kategori yaitu database fakta, database hoax dan database grey. Kalimat dari *tweet* dipecah menjadi kata kata individual (Kosakata). Bobot dari per kategori dihitung dan disimpan

sebagai *prior probability*. Dari kosakata yang ditemukan maka dihitung juga bobotnya dan disimpan sebagai *conditional probability*. Setelah semua bobot kosakata dalam kategori dihitung maka akan disimpan ke data model. Setelah itu data *input* yang dimasukkan user akan dicari bobotnya berdasarkan perbandingan dari input dan data model yang dibuat. Dari data input dibandingkan dalam 3 kategori yang ada. Nilai yang paling tinggi diambil sebagai hasil klasifikasi.



Gambar 1. Skema Sistem



Gambar 2. Use Case

### 2.2. Rancangan Database

Sistem pendeteksi berita palsu menggunakan database yang terdiri dari:

Tabel 1 Tabel Berita Hoax

No	Nama Kolom	Tipe Data
1	id	int(11) (Primary Key)
2	name	varchar(100)
3	screen_name	varchar(100)
4	text	varchar(1000)
5	created_at	varchar(100)
6	id_str	varchar(100)
7	pembenaran	varchar(1000)

Tabel 2 Tabel Berita Fakta

No	Nama Kolom	Tipe Data
1	id	int(11) (Primary Key)
2	name	varchar(100)
3	screen_name	varchar(100)
4	text	varchar(1000)
5	created_at	varchar(100)
6	id_str	varchar(100)

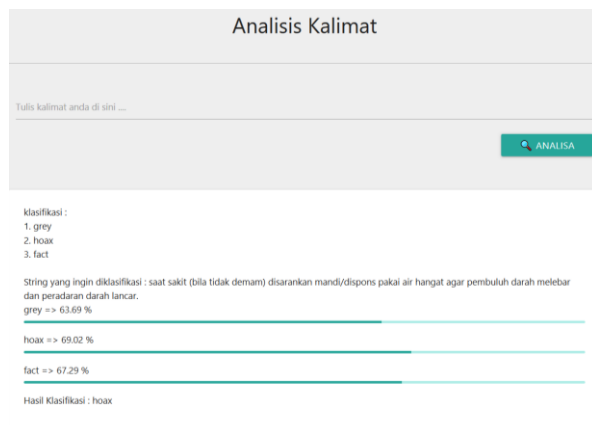
Tabel 3 Tabel Berita Grey

No	Nama Kolom	Tipe Data
1	id	int(11) (Primary Key)
2	name	varchar(100)
3	screen_name	varchar(100)
4	text	varchar(1000)
5	created_at	varchar(100)
6	id_str	varchar(100)

### 3. HASIL DAN PEMBAHASAN

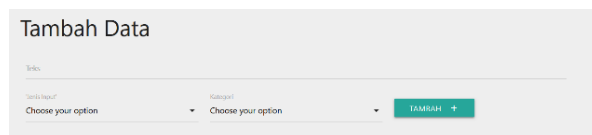
#### 3.1. Tampilan Program

Hasil implementasi sistem dibuat dalam bentuk website. Dibawah ini adalah tampilan dari sistem yang telah dibuat.



Gambar 3 Tampilan hasil klasifikasi

Gambar 3 merupakan tampilan hasil klasifikasi yang dimasukkan user. Akan ditampilkan dalam bentuk persentase per kategori.



Gambar 4 Form untuk membaca data

Gambar 4 merupakan form untuk menambah berita baru, data berita baru meliputi tweet dari

user yang dipilih, hasil search dari twitter, dan kalimat manual yang diisi sendiri.



Gambar 5 Tampilan untuk melihat data

Gambar 5 merupakan tampilan data yang tersimpan dalam database, digunakan untuk pengelolaan data berita.

#### 3.2. Pengujian

Dari training data ditemukan data model sebagai berikut

Tabel 4. Judul Tabel

No	Kategori	Jumlah kata	Jumlah kosakata	Prior probability
1	Grey	5108	1822	0.4571326 29
2	Hoax	1382	586	0.1236799 71
3	Fact	4684	1831	0.4191873 99

Dari data training diatas dilakukan pengujian sebanyak 30 data dengan hasil sebenarnya adalah 15 untuk berita hoax dan 15 untuk berita fakta. Dari hasil pengujian didapat *confusion matrix* sebagai berikut.

Tabel 5 Confusion Matrix

	Fakta	Hoax	
Actual Fakta	TP = 9	FP = 6	15
Actual Hoax	FN = 0	TN = 15	15
	9	21	

Dari tabel tersebut didapat data pengujian :

- **Akurasi**

Tingkat Akurasi pendeteksian berita yang benar pada sistem yang telah dibuat rumus (1).

$$\frac{TP+TN}{total} = \frac{15+9}{30} = \frac{24}{30} = 0,8 \quad (1)$$

- **Misclassification rate**

Tingkat akurasi pendeteksian berita yang salah pada sistem yang telah dibuat rumus (2).

$$\frac{FP+FN}{total} = \frac{6+0}{30} = \frac{6}{30} = 0,2 \quad (2)$$

- **True positive rate**

Saat *value* sebenarnya fakta, seberapa sering sistem mendeteksi berita dengan benar rumus (3).

$$\frac{TP}{Actual\ Fact} = \frac{9}{15} = 0,6 \quad (3)$$

- **False Positive Rate**

Saat *value* sebenarnya hoax, seberapa sering sistem mendeteksi berita menjadi fakta rumus (4).

$$\frac{FP}{Actual\ Hoax} = \frac{6}{15} = 0,4 \quad (4)$$

- **Specificity**

Saat *value* sebenarnya hoax, seberapa sering sistem mendeteksi berita menjadi hoax dengan rumus (5).

$$\frac{TN}{Actual\ hoax} = \frac{15}{15} = 1 \quad (5)$$

- **Precision**

Saat memprediksi fakta , seberapa sering prediksi benar dengan rumus (6).

$$\frac{TP}{Actual\ Fact} = \frac{9}{15} = 0,6 \quad (6)$$

- **Prevalence**

Total kondisi fakta dalam pengetesan dengan rumus (7).

$$\frac{Actual\ Fact}{total} = \frac{15}{30} = 0.5 \quad (7)$$

Dari hasil pengujian confusion matrix diatas sistem dapat mengkategorikan berita dengan baik dengan tingkat akurasi 80 % dari data sebanyak 30 berita, sehingga bisa dikatakan secara fungsional sistem sudah dapat mengklasifikasi berita secara tepat dan menghasilkan output yang diharapkan

#### 4. SIMPULAN

Kesimpulan yang dapat diambil dari Pendekatan *Text Mining* sebagai Sistem

Pendeteksi Pemberitaan Palsu yang Tersebar dalam Twitter adalah sistem ini telah berhasil mengklasifikasikan berita ke dalam 3 kategori yaitu berita palsu (*hoax*), berita asli (*fact*), dan berita yang tidak terindikasi (*grey*). Hasil klasifikasi ditentukan dari besarnya database yang dimiliki, dalam arti lain klasifikasi dengan metode *naïve bayes* menggunakan perhitungan bobot per kategori. Sehingga bobot terbesar dipilih sebagai hasil klasifikasi. Pola dari berita palsu belum dapat dikenali secara jelas karena banyak berita yang hampir sama persis, hanya berbeda 1-3 kata sehingga hasil klasifikasi yang dikeluarkan selisih kurang dari 10% antar kategori.

#### 5. DAFTAR PUSTAKA

- Andre . (2014, 15 Desember) . *Tutorial Belajar PHP Part 1: Pengertian dan Fungsi PHP dalam Pemrograman Web* . *DuniaIkom* . waktu akses : 28-10-2017 . <http://www.duniaikom.com/pengertian-dan-fungsi-php-dalam-pemrograman-web/>
- Andre . (2017, 17 Desember) . *Tutorial Belajar MySQL Part 2: Sejarah MySQL dan Masa Depan MySQL* . *DuniaIkom* . waktu akses : 26 - 12 -2017 . <http://www.duniaikom.com/tutorial-mysql-sejarah-dan-masa-depan-mysql/>
- Astrawinata G . (2017 , 6 Februari) . *Mengenal dan mencegah penyebaran hoax kesehatan* . *Analisa Daily* . Waktu akses: 27-10-2017 .<http://harian.analisadaily.com/kesehatan/news/mengenal-dan-mencegah-penyebaran-hoax-kesehatan/313430/2017/02/06>
- Berry, M.W. & Kogan, J. 2010. *Text Mining Application and theory*. WILEY : United Kingdom
- C. Lethbridge, Timothy dan Laganieri, Robert. (2002). *Object-Oriented Software Engineering*. McGraw-Hill, New York.
- Goker, A. and Davies, J. (2009) *Information Retrieval Searching in the 21st Century*. Wiley, West Sussex.
- Hamilton, Naomi (2008-06-31). "The A-Z of Programming Languages: JavaScript" . [computerworld.com.au](http://computerworld.com.au).
- Hancock, P. (2015). *Hoax Springs Eternal. In Hoax Springs Eternal: The Psychology of Cognitive Deception* (pp. 197-204).

- Cambridge: Cambridge University Press.  
doi:10.1017/CBO9781107785311.012
- Hariyanto . (2017, 28 Agustus) . *Menkominfo kerjasama antar elemen sangat penting untuk perangi konten negatif di media sosial* . *Kompas Cyber Media*. Waktu akses: 27-10-2017  
<http://www.industry.co.id/read/14981/menkominfo-kerjasama-antar-elemen-sangat-penting-untuk-perangi-konten-negatif-di-media-sosial>
- Kuwado, Fabian Januarius . (2017, 8 Juni) . *Jokowi stop penyebaran berita bohong*. *Kompas Cyber Media* . Waktu akses: 27-10-2017 . <http://nasional.kompas.com/read/2017/06/08/18261731/jokowi.stop.penyebaran.berita.bohong.fitnah.dan.kebencian.di.medsos>
- Manning, C.D., Raghavan .P., & Schutze, H. (2008) . *Introduction to Information Retrieval* . Cambridge University Press.
- Saraswati , Ni Wayan Sumartini . (2011) . *Text Mining Dengan Metode Naïve Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis* . Denpasar : Universitas Udayana
- Schmuller, Joseph. (1999) . *Sams Teach Yourself UML in 24 Hours* . Pearson Education, Inc.
- Tan, P.N., Steinbach, M., & Kumar .V .(2006). *Introduction to Data Mining*. Pearson Education.
- Winarno, Ahmad . (2017, 13 Januari). *Menteri agama minta perguruan tinggi ikut perangi hoax* . *Kompas Cyber Media*. Waktu akses: 27-10-2017  
<http://regional.kompas.com/read/2017/01/13/21470081/menteri.agama.minta.perguruan.tinggi.ikut.perangi.hoax>.
- Zallera,D .(2010) . *The Social Media Marketing Book* . Jakarta: PT. Serambi Ilmu Semesta Anggota IKAPI

Budi Setiawan Tanjung, Pendekatan *Text Mining* sebagai Sistem Pendeteksi Pemberitaan Palsu yang Tersebar dalam Twitter