



Klasifikasi Berita Hoaks Menggunakan Algoritma Vector Space Model *Classification of Hoax News Using the Vector Space Model Algorithm*

Ayuni Asistiyasari*, Bibit Sudarsono, Umi Faddilah

* Program Studi Sistem Informasi, Universitas Bina Sarana Informatika-Indonesia

* Jl. Ciledug Raya No.168, RW.4, Ulujami, Kec. Pesanggrahan, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta-Indonesia 12250

Informasi Artikel

Article History:

Submission: 30-11-2021

Revised: 24-12-2021

Accepted: 27-12-2021

Kata Kunci:

Hoaks; VSM; Text Mining;
CRISP-DM

Keywords:

Hoax; VSM; Text Mining;
CRISP-DM

*** Korespondensi:**

Ayuni Asistiyasari
ayuni.yas@bsi.ac.id

Abstrak

Sebuah berita terkait suatu informasi yang beredar di media cetak atau mainstream akan menjadikan opini publik tentang suatu masalah baik yang bersifat informasi positif atau negatif, perkembangan teknologi informasi sekarang ini menyebabkan penyebaran informasi bisa uptodate setiap harinya. Dengan semakin mudahnya sebuah informasi menyebar maka akan semakin mudah pula mempengaruhi kehidupan dalam sosial masyarakat sekarang ini. Namun pada kenyataannya informasi yang beredar di media itu tidak semuanya benar atau bisa dikatakan adanya suatu berita hoax atau tidak benar. Dalam penelitian ini bertujuan untuk mengklasifikasi sistem temu kembali informasi berita hoaks menggunakan metode vektor space model untuk memastikan kebenaran suatu berita apakah berita hoax atau tidak. Dalam penelitian tersebut menghasilkan klasifikasi kebenaran berita dengan akurasi terbaik pada K-6 sebesar 83%, artinya dengan akurasi tersebut bisa memvalidasi klasifikasi terkait informasi berita benar ataupun hoax sebesar 83%.

Abstract

A news related to an information circulating in print or mainstream media will make public opinion about a problem that is either positive or negative information, the development of information technology today causes the dissemination of information to be updated every day. The easier it is for information to spread, the easier it will be to influence life in today's society. But in reality the information circulating in the media is not all true or it can be said that there is a hoax news or not true. This study aims to classify the hoax news information retrieval system using the vector space model method to ensure the truth of a news whether it is hoax news or not. This study resulted in a classification of the truth of news with the best accuracy at K-6 of 83%, meaning that with this accuracy it can validate the classification related to true news or hoax information by 83%.

1. PENDAHULUAN.

Saat ini informasi dapat dengan mudah kita terima melalui media sosial maupun aplikasi mesenger yang terkadang kita tidak tau itu berita benar ataukah hoax. Lebih parahnya lagi terkadang kita turut berperan dalam penyebaran berita tersebut. Kata hoaks memiliki arti "tidak



benar bohong (tentang berita, pesan, dan sebagainya)”[1]. “Kita harus jauhi. Bagi bangsa Indonesia yang kita butuhkan adalah membangun kepercayaan, saling percaya. Karena itu hoax-hoaxan jadi berbahaya”[2].

Beda penelitian sebelumnya menyatakan bahwa *Vector Space Model* pada sistem pencarian lagu berdasarkan judul dapat mempercepat proses pencarian lagu yang relevan sesuai dengan lagu yang ingin dicari pengunjung[3].

Penelitian sebelumnya mengenai penilaian *thematic apperception test*[4] menghasilkan test kepribadian dengan nilai precision 78.2%, recall 86,7 dan f-measure 78.3.

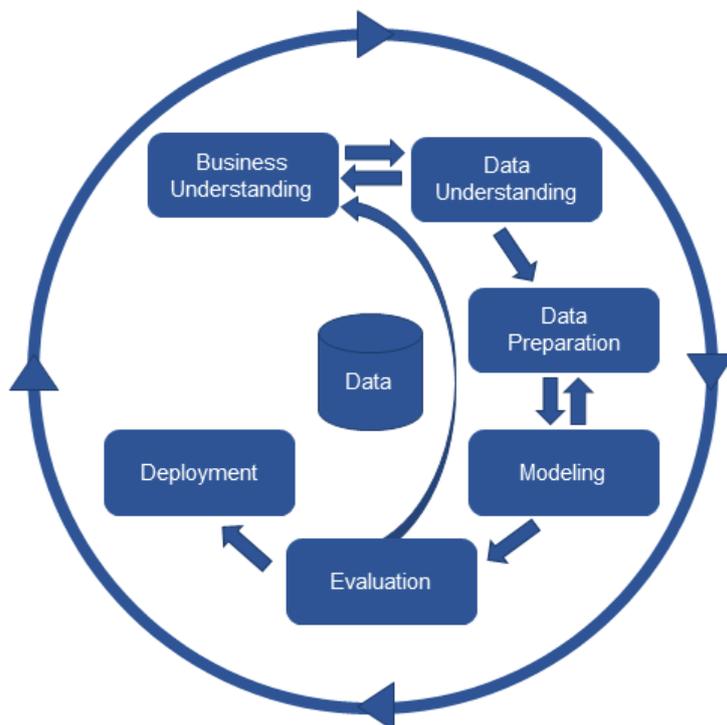
Kemudian pada penelitian lainnya juga menyimpulkan bahwa sistem temu Kembali informasi dengan metode *Vector Space Model* efektif untuk digunakan sebagai system pencarian bahan Pustaka dipergustakaan[5].

Penelitian lainnya yang berkaitan yaitu pencarian dokumen skripsi dengan menggunakan metode TF.IDF.ICF dapat meningkatkan relevansi dibandingkan hanya menggunakan TF.IDF dengan hasil *Mean Average Precision* TF.IDF sebanyak 71,12% dan *Mean Average Precision* TF.IDF.ICF 72,39%[6],[7],[8].

Atas dasar tersebut, penulis mencoba mengklasifikasikan berita-berita hoax menggunakan *vector space model* yang diharapkan dapat menghasilkan rekomendasi untuk membuat aplikasi.

2. METODE

Dalam penelitian ini menggunakan model CRISP-DM (*Cross-Industry Standard Proses for Data Mining*), yang terdiri dari 6 tahap [9][10][11][12] sesuai dengan gambar 1.



Gambar 1. Complete CRISP-DM Approach [13]

A. *Business/research understanding phase*

Dengan adanya aplikasi messenger, sering kita menerima berita-berita dari lingkungan sekitar. Berita-berita tersebut kadang ada yang hoax maupun tidak. Untuk mempermudah dalam klasifikasi berita hoax, kami mencoba melakukan pengklasifikasian berita hoax menggunakan metode *vector space model*.

B. *Data understanding phase*

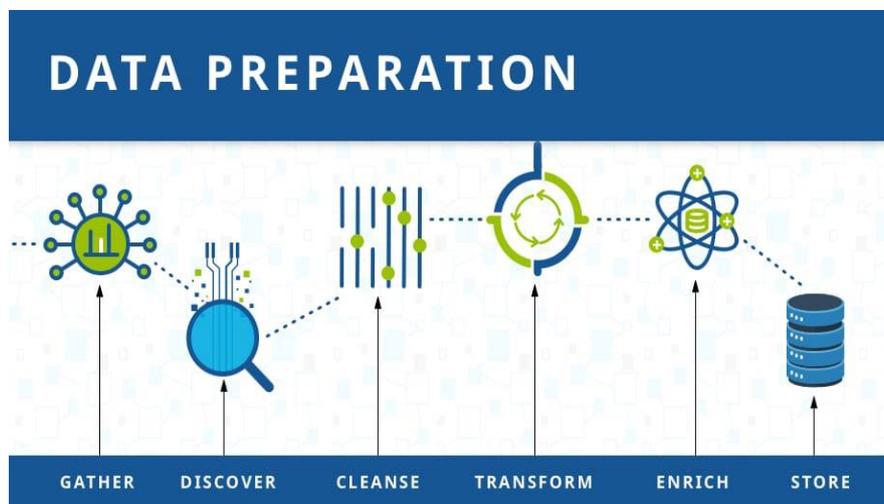
Data yang digunakan yaitu berasal dari berita-berita yang kita dapat melalui aplikasi messenger berjumlah 20 berita dengan rincian seperti pada table 1.

Tabel 1. Jumlah data

Jenis Berita	Jumlah
Hoaks	10
Bukan Hoaks	10
Jumlah	20

C. Data preparation phase

Pada persiapan data untuk lebih jelasnya seperti terlihat pada gambar 2, setelah data-data sudah didapat, maka data tersebut kita persiapkan kedalam sebuah tabel dengan membentuk ID berita, berita dan kelas seperti yang tampak pada tabel 2.



Gambar 2. Preparation Phase[14]

Tabel 2. Berita

ID BERITA	Data Berita	Kelas
Berita 1	Yang melintas dijalanan KOTA Jakarta Selatan khususnya Radio Dalam: Besok selasa ada uji coba tilang e-CCTV. Berikut daftar Jl yang akan diuji coba antara lain: - Jl. Mampang prapatan raya - Jl. Pasar minggu raya - Jl. Kalibata makam pahlawan - Jl.Lenteng agung	Hoaks
Berita 2	Yang melintas dijalanan KOTA DEPOK: Besok senin depan ada uji coba tilang e-CCTV. Berikut daftar Jl yang akan diuji coba antara lain: - Jl. Margonda - Jl. A. Rahman Hakim - Jl. Nusantara - Jl. Akses UI - Jl. Raya Bogor	Hoaks

	- Jl. Raya Sawangan - Jl. Raya Cimanggis	
Berita 3	AP (sengaja disebut inisial-red) 13thn Tinggal di dusun cibuyaya IV, rt/rw 001/007 desa cibuyaya, kecamatan.cibuyaya, kabupaten karawang. Dia yatim, Yg sll bekerja keras, wlu seharusnya anak seusianya msh sklh Dan bermain. Namun dia Kini Tak sekolah lagi krn tdk ada biaya.	Hoaks
Berita 4	Telah berpindah tangan satu lagi BUMN strategis yang dibangun oleh Alm Soeharto dan Prof DR Ing BJ Habibie... Dengan ditandatanganinya pelunasan pembayaran dari pemerintah Cina kepada pemerintah Indonesia pada awal April 2017 oleh Presiden Joko Widodo (disetujui dan ditandatangani oleh Ketua KPK, ketua DPR RI dan ketua dewan pertimbangan presiden) maka secara resmi kepemilikan PT Dirgantara Indonesia/PTDI dh. IPTN NURTANIO berpindah tangan menjadi milik pemerintah Tjina.	Hoaks
Berita 5	Untuk Menghindari Agar Kaca Tidak Pecah Waktu Dilemparkan Pecahan Busi Trik Cegah Mobil Dibobol dengan Busi Pemecah Kaca Pencurian dengan bermodus pecah kaca kembali mengintai. Namun ada cara untuk menghindarinya.	Hoaks
Berita 20	"Di Ukraina Jika Anggota DPR Ga Becus Langsung Dibuang Ke Tempat Sampah di Pinggir jalan, Kalau Di Indonesia....?"	Bukan Hoaks

Dan berikut tahapan-tahapan proses didalam text mining:

- 1) *Casefolding*: mengubah semua huruf dalam dokumen menjadi huruf kecil (*lowercase*).
- 2) *Tokenizing*: memotong tiap kata dalam kalimat atau parsing dengan menggunakan spasi sebagai delimiter yang akan menghasilkan token berupa kata.
- 3) *Filtering*: menyaring kata yang didapat dari *tokenizing* yang dianggap tidak penting atau tidak memiliki makna dalam proses *text mining* yang disebut *stopword*. *Stopword* berisi kata kata umum yang sering muncul dalam sebuah dokumen dalam jumlah banyak namun tidak memiliki kaitan dengan tema tertentu.
- 4) *Stemming*: mengembalikan kata-kata yang diperoleh dari hasil *filtering* ke bentuk dasarnya, menghilangkan imbuhan awal (*prefix*) dan imbuhan akhir (*suffix*) sehingga didapat kata dasar.

D. Modeling phase

Modeling yang digunakan adalah menggunakan algoritma *Vector Space Model* (VSM). VSM adalah metode untuk melihat tingkat kedekatan atau kesamaan (*similarity*) term dengan cara pembobotan term. Dokumen dipandang sebagai sebuah vektor yang memiliki magnitude (jarak) dan direction (arah). Pada *Vector Space Model*, sebuah istilah direpresentasikan dengan sebuah dimensi dari ruang vektor. Relevansi sebuah dokumen ke sebuah query didasarkan pada similaritas diantara vektor dokumen dan vektor query" [15],[16].

E. Evaluation Phase

Evaluasi dilakukan untuk mengetahui keakurasian model yang digunakan dalam hal ini yaitu mencari akurasi yang dihasilkan oleh *algoritma Vector Space Model*

F. Deployment Phase

Dan tahap terakhir adalah penerapan model yang sudah dibentuk kedalam sistem klasifikasi berita hoak yang akan dibuat.

3. HASIL DAN PEMBAHASAN

Dari data tabel 2 lakukan proses *casefolding*, *tokenisazing*, *filtering* dan *stemming* yang hasilnya terlihat pada tabel 3.

Tabel 3. Proses tokenisasi

Term	Doc ID
13	3
almarhum	4
anggota	6
Ap	3
bangun	4
becus	6
besok	1 2
buang	6
bumn	4
busi	5
cibuaya	3 3
Coba	1 2
dalam	1
depan	2
depok	2

Sebelum menghitung bobot dari tiap-tiap dokumen, kita harus menentukan query yang digunakan sebagai tolak ukur pencarian. Query yang dimaksud pada penelitian ini adalah “yang melintas di Jakarta selatan “. Dan hasil pembobotan tersebut seperti yang tampak pada tabel 4.

Tabel 4. Hasil pembobotan setiap term.

Term	Doc Id						Term Frekuensi	Doc/ Tf	TF.IDF	
	Q	1	2	3	4	5				6
13			1				1	6	0.77	
Alma				1			1	6	0.77	
Angg						1	1	6	0.77	
Ap			1				1	6	0.77	
Bang				1			1	6	0.77	
Becu						1	1	6	0.77	
Jakar	1	1					1	6	0.77	
jalan	1	1	1	1		1	1	5	1.2	0.07

Kemudian hitung bobot tiap term pada masing-masing dokumen yang hasilnya tampak seperti tabel 5.

Tabel 5. Hasil pembobotan pada setiap dokumen.

Term	Bobot						
	Q	1	2	3	4	5	6
13	0	0	0	0.77	0	0	0
Alma	0	0	0	0	0.778	0	0
Angg	0	0	0	0	0	0	0.7782
Ap	0	0	0	0.77	0	0	0

Bang	0	0	0	0	0.7782	0	0
Becu	0	0	0	0	0	0	0.7782
Jakar	0.77	0.77	0	0	0	0	0
jalan	0.07	0.07	0.07	0.07	0	0.07	0.079

Selanjutnya hitung jarak vector masing masing dokumen. Hasilnya tampak pada tabel 6.

Tabel 6. Hasil *dot of product*

Term	Bobot					
	1	2	3	4	5	6
hendar	0	0	0	0	0	0
jakarta	0.606	0	0	0	0	0
jalan	0.006	0.006	0.006	0	0.006	0.006
kaca	0	0	0	0	0	0
lansung	0	0	0	0	0	0
lempar	0	0	0	0	0	0
lintas	0.228	0.228	0	0	0	0
selasa	0	0	0	0	0	0
selatan	0.228	0	0.228	0	0	0
waktu	0	0	0	0	0	0
	2.067	2.234	3.234	4	5.006	6.006

Kemudian hitung *vector space* sehingga menghasilkan tabel 7.

Tabel 7. Perhitungan jarak vektor

Term	Jarak Vektor						
	Q	1	2	3	4	5	6
13	0	0	0	0.606	0	0	0
Almarhum	0	0	0	0	0.606	0	0
Anggota	0	0	0	0	0	0	0.606
Tilang	0	0.228	0.228	0	0	0	0
Tinggal	0	0	0	0.606	0	0	0
Uji	0	0.228	0.228	0	0	0	0
Ukraina	0	0	0	0	0	0	0.606
Waktu	0	0	0	0	0	0.606	0
	1.067	5.855	5.416	6.867	8.844	9.85	11.46
	1.033	2.42	2.327	2.621	2.974	3.139	3.385

Selanjutnya hitung *consine similiarity* untuk mengetahui kemiripan query dengan masing-masing dokumen. Hasil *consine similiarity* dengan nilai terbesar merupakan dokumen termirip dan hasil tersebut seperti tabel 8.

Tabel 8. Hasil *consine similiarity*

Consine Similarity	Nilai
Con (Q, D1)	0.469
Con (Q, D2)	0.123
Con (Q, D3)	0.115
Con (Q, D4)	0

Con (Q, D5)	0.003
Con (Q, D6)	0.003

Dari berikut hasil yang sudah diurutkan:

Tabel 9. Hasil *consine similiarity*

Consine Similarity	Nilai
Con (Q, D1)	0.469
Con (Q, D2)	0.123
Con (Q, D3)	0.115
Con (Q, D5)	0.003
Con (Q, D6)	0.003
Con (Q, D4)	0

Apabila hasil pencarian menggunakan k1 hingga k6 hasil akurasi adalah 83%.

4. SIMPULAN

Sesuai hasil penelitian dan pengujian akurasi sistem temu kembali informasi berita hoaks menggunakan metode algoritma vektor space model untuk memastikan kebenaran suatu berita. Hasil penelitian menunjukkan klasifikasi kebenaran berita dengan akurasi terbaik pada K-6 sebesar 83%, sehingga dengan akurasi tersebut bisa memvalidasi klasifikasi terkait kebenaran berita suatu berita.

DAFTAR PUSTAKA

- [1] A. Budiwiyanto, "Penyerapan Kosakata Bahasa Daerah ke dalam Bahasa Indonesia pada Kamus Besar Bahasa Indonesia Edisi IV," *Mabasan*, vol. 3, no. 1, pp. 1–14, 2019.
- [2] A. Baswedan, "https://www.cnnindonesia.com/nasional/20170109013304-20-184857/anies-baswedan-berita-hoax-itu-berbahaya," *Diakses : 20 Oktober 2021*, 2021. .
- [3] Anna and A. Hendini, "Implementasi Vector Space Model Pada Sistem Pencarian Mesin Karaoke," *Evolusi J. Sains dan Manaj.*, vol. 6, no. 1, pp. 1–6, 2018.
- [4] U. Nurhasan, R. Arianto, A. G. Kurnia, J. Soekarno, and H. No, "Penerapan Vector Space Model Dalam Klasifikasi Penilaian Thematic Apperception Test," vol. 5, no. September, pp. 551–564, 2021.
- [5] B. S, "Aplikasi Pencarian Bahan Pustaka Di Perpustakaan Menggunakan Metode Vector Space Model," *J I M P - J. Inf.*, vol. 5, no. 2, 2021.
- [6] I. R. Priandono, M. Hakimah, and N. F. Rozi, "Implementasi Vector Space Model Dengan Pembobotan Berbasis Kelas Pada Mesin Pencari Dokumen Skripsi," vol. 5, no. 2, pp. 54–58, 2020.
- [7] A. D. P, Lutfiyatul, A. Z, Maryamah, R. W, and Rarasmaya, "Metode Pembobotan Kata Berbasis Cluster Untuk Perangkingan Dokumen Berbahasa Arab," *Techno.COM*, vol. 20, no. 2, pp. 259–267, 2021.
- [8] H. A. Gozali, M. A. Rosid, and Sumarno, "Klasifikasi Keluhan Mahasiswa dengan Metode Naive Bayes dan Sastrawi," *JOINCS (Journal Informatics, Network, Comput. Sci.*, vol. 3, no. 1, pp. 22–26, 2020.
- [9] Aprilia et al., *Belajar Data Mining dengan Rapid Miner*. 2013.
- [10] C. Schröer, F. Kruse, J. Marx, F. Kruse, and J. Marx, "A Systematic Literature Review on Applying Process Model on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021.

- [11] j. Martin dan S. Oxman, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "DMME : Data mining methodology for engineering applications – a holistic extension the CRISP-DM model," *Procedia CIRP*, vol. 79, pp. 403–408, 2018.
- [12] W. Y. Ayele, "Adapting CRISP-DM for Idea Mining A Data Mining Process for Generating Ideas Using a Textual Dataset," *IJACSA (nternational J. Adv. Comput. Sci. Appl.)*, vol. 11, no. 6, pp. 20–32, 2020.
- [13] J. Taylor, "Complete CRISP-DM Approach," *Diakses : Sabtu 25 Desember 2021, 6:10*, 2017. .
- [14] <https://devopedia.org/>, "Preparation Phase," *Diakses : Sabtu 25 Desember 2021, 6:10*, 2021. .
- [15] N. R. Baeza R.Y., "Modern Information Retrieval," Addison We., Iinternational edition, Ed. Boston, USA, 1999.
- [16] S. Vijayarani, M. J. Ilamathi, and M. Nithya, "Preprocessing Techniques for Text Mining - An Overview," *Int. J. Comput. Sci. Commun. Networks*, vol. 5, no. February 2015, pp. 7–16, 2020.