SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika

https://ejurnal.itats.ac.id/snestik dan https://snestik.itats.ac.id

Informasi Pelaksanaan :
SNESTIK I - Surabaya, 26 Juni 2021
Ruang Seminar Gedung A, Kampus Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:
DOI : 10.31284/p.snestik.2021.1463

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Pengaruh Penggunaan Information Gain dalam Peningkatan Akurasi Deteksi Berita Palsu

M.Teguh Septiyan[1], Weny Mistarika Rahmawati[2], dan Septiyawan Rosetya Wardhana[3]
Teknik Informatika, Institut Teknologi Adhi Tama Surabaya[1,2,3]
Email: teguh.septiyan111@gmail.com

ABSTRACT

The rise of news in the public, especially on social media, requires readers to be very clever in choosing whether the news is true or not before finally deciding to share the news. This selection of news became even more important when a law was issued for the spreaders of fake news. Several algorithms or methods can be used to classify whether a news is real or fake. One of the hoax detection is the Naïve Bayes method which is widely used by researchers. In this research, it is proposed to add the feature selection process with information gain before classification using the Naïve Bayes method. The result turns out that the addition of this information gain can significantly increase the accuracy value from 43,14% to 72,71%. It shows that this method is very good to be used as a detection of whether a news is real or fake.

Keywords: *Fake news, hoax detection, information gain, Naïve Bayes*

ABSTRAK

Maraknya berita yang ada di khalayak masyarakat terutama pada social media mengharuskan pembaca untuk pandai-pandai memilih apakah berita tersebut benar atau tidak sebelum akhirnya memutuskan untuk membagikan berita tersebut. Pemilihan berita ini menjadi semakin penting ketika munculnya undang-undang bagi para penyebar berita palsu atau hoax. Beberapa algoritma atau metode dapat dilakukan untuk melakukan deteksi hoax. Salah satunya adalah metode Naïve Bayes yang banyak dilakukan oleh para peneliti. Pada penelitian ini diajukan penambahan proses pemilihan fitur dengan information gain sebelum dilakukan pengklasifikasian dengan metode Naïve Bayes. Hasilnya ternyata penambahan information gain ini dapat meningkatkan nilai akurasi secara signifikan dari 43,14% meningkat menjadi 72,7%. Peningkatan

tersebut menunjukkan bahwa metode ini sangat baik untuk digunakan sebagai pendeteksi apakah sebuah berita asli atau palsu.

Kata kunci: berita palsu, deteksi hoax, Information Gain, *Naïve Bayes.*

PENDAHULUAN

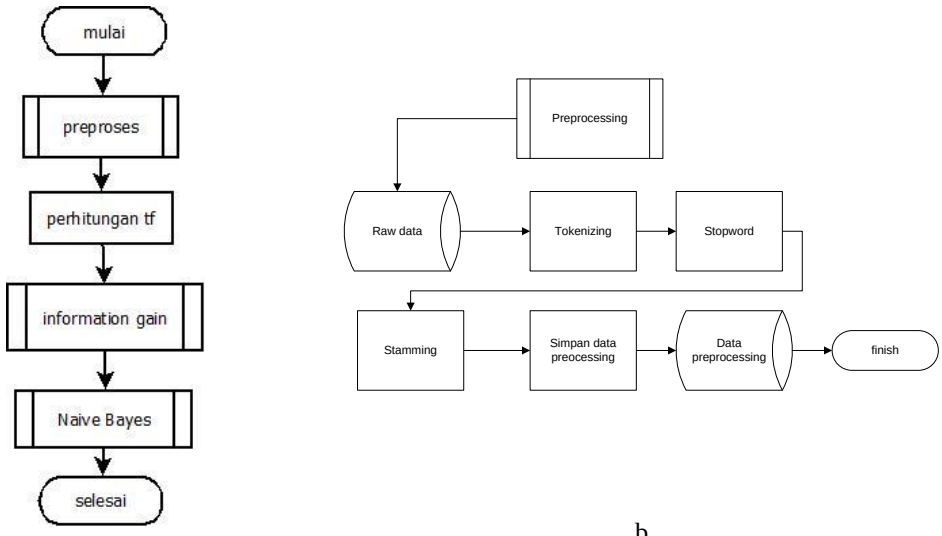
Luasnya pengguna internet saat ini memungkinkan siapa saja untuk membuat, menyebarkan atau mendapatkan banyak berita dan kabar. Tidak menutup memungkinan bagi seseorang untuk menyebar kebohongan dengan berbagai alasan dan kepentingan. Berita yang belum tentu benar adanya atau berita palsu ini dikenal oleh masyarakat sebagai berita hoax. Berita palsu ini kebanyakan muncul pada social media. Karena hampir semua orang menggunakan social media orang saat ini. Dengan banyaknya berita yang muncul, masyarakat harus pandai-pandai memilih, apakah sebuah berita dapat dijamin keterpercayaannya. Apakah berita tersebut memang benar adanya atau hanya sekedar isu yang tidak dapat dipertanggungjawabkan kebenarannya atau palsu.

Banyak peneliti yang mencoba melakukan deteksi berita palsu dengan beberapa algoritma klasifikasi. Diantaranya adalah algoritma naïve Bayes. Metode Naïve Bayes sering digunakan dalam penelitian tentang klasifikasi teks karena kesederhanaan dan efektivitasnya yang menggunakan ide dasar probabilitas gabungan dari kata-kata dan kategori untuk memperkirakan probabilitas kategori pada suatu dokumen [1]. Banyak penelitian yang dilakukan dalam hal deteksi berita hoax menggunkan algoritma naive bayes [2], [3] .

Pada kenyataannya, berita merupakan sebuah teks yang memiliki beragam kata. Ada yang pendek hanya beberapa kata, sampai yang panjang terdiri ribuan sampai jutaan kata. Banyaknya kata ini akan mempengaruhi hasil deteksi. Peneliti harus pandai-pandai memilih kata mana yang relevan untuk melakukan deteksi. Kata-kata ini dalam proses deteksi akan menjadi fitur. Semakin banyak fitur, maka kemungkinan kesalahan deteksi akan semakin besar sehingga diperlukan cara bagaimana untuk mengurangi jumlah fitur tanpa menghilangkan informasi penting dari sebuah berita. Metode Information Gain merupakan salah satu algoritma seleksi fitur yang digunakan untuk memilih fitur terbaik [4]. Information Gain dimanfaatkan untuk merangking kata-kata penting dari hasil reduksi fitur. Hasil dari proses Information Gain adalah kata penting yang bersifat informatif. Metode Information Gain dapat melihat setiap fitur untuk memprediksi label kelas yang benar, karena memilih nilai yang tertinggi dan lebih efektif untuk mengoptimalkan hasil klasifikasi [5]. Pada penelitian ini, pemilihan fitur dilakukan dengan menggunakan information gain yang dilakukan pada deteksi berita palsu menggunakan metode Naïve Bayes dengan harapan akan mendapatkan akurasi yang lebih baik.

METODE

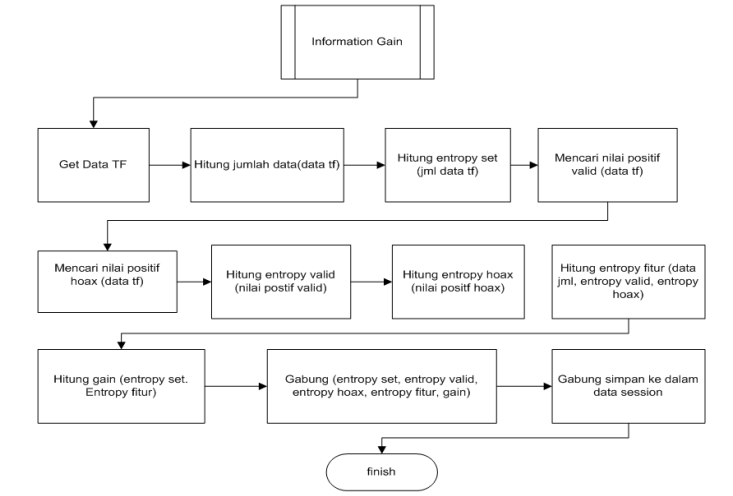
Pada penelitian ini, yang dilakukan pertama kali adalah pengumpulan data. Data yang dimaksud disini tentu saja merupakan data berita yang akan dideteksi apakah berita tersebut asli atau palsu. Data didapat dari peneliti [3] yang mengambil berita dari portal berita secara manual dan menjadikannya dalam format csv. Proses pengumpulan dataset yang telah dilakukan oleh peneliti tersebut terdapat 600 dataset berita yang terdiri dari 12 kata kunci berita dengan masing-masing kata kunci terdapat 50 data. Dari dataset yang diperoleh akan diolah per kata kunci. Dalam setiap kata kunci, data yang tersedia akan dibagi menjadi data training dan data testing. Data yang didapat secara garis besar akan diolah sebagaimana dapat dilihat pada Gambar 1a.



Gambar 1a Flowchart keseluruhan sistem b flowchart tahap preprocessing

Yang dilakukan pertama kali terhadap data yang diperoleh adalah dilakukan preprosesing yang prosesnya dapat dilihat pada gambar 1b.dari gambar tersebut dapat dilihat bahwa pada langkah preprosesing, data pertama kali akan dilakukan *tokenizing*. *Tokenizing* disini bertujuan untuk memilah-milah kata bredasarkan spasi sehingga lebih mudah diolah. Karakter tertentu seperti simbol dan angka akan dihilangkan pada tahap ini. Langkah selanjutnya adalah proses stopwords. Pada langkah ini menghapus kata *grammatical* yang tidak sesuai dengan konteks seperti preposisi dan konjungsi. Langkah berikutnya adalah stemming, yaitu mengubah kata yang telah diidapat menjadi bentuk dasarnya dengan menghilangkan imbuhan yang melekat pada kata tersebut.

Setelah melakukan tahap preprocessing, langkah berikutnya adalah menghitung term frekuensi atau kemunculan kata yang ada pada setiap berita yang akan diolah. Hasil dari pencarian preprocessing akan dilakukan pemilihan fitur menggunakan information gain dimana prosesnya dapat dilihat pada gambar 2.



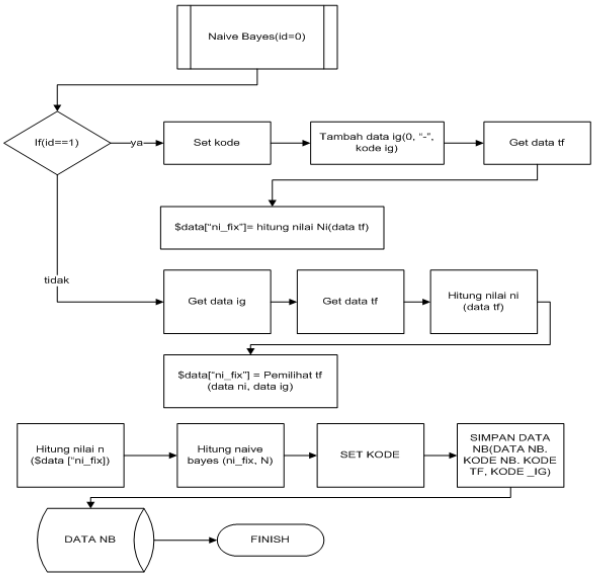
Gambar 2 Flowchart pencarian fitur dengan information gain

Pencarian entropi menggunakan persamaan berikut:

$$Entropi(S) = \sum_{i=1}^{K} (P_i) \log_2 (P_i)$$

Dimana k merupakan jumlah kelas yang akan digunakan. Dalam penelitian ini k=2 yaitu kelas hoax dan bukan hoax..

Setelah fitur-fitur yang dianggap penting didapatkan, selanjutnya melakukan klasifikasi menggunakan Naïve Bayes yang flowchartnya dapat dilihat pada gambar 3.



Gambar 3 Flowchart Naive Bayes

HASIL DAN PEMBAHASAN

Pembahasan Data I

Dari penelitian yang telah dilakukan, penulis menyampaikan hasilnya pada table 1. Dari table 1 dapat dilihat bagaimana perbandingan akurasi hasil deteksi menggunakan Naïve Bayes saja tanpa information gain dan dengan information gain.

Tabel 1 Perbandingan hasil akurasi tanpa dan dengan information gain

Kelompok berita	Akurasi dengan Naïve Bayes saja (%)	Akurasi dengan information gain dan Naïve Bayes(%)
1	34	67
2	60	67
3	27	80
4	27	54
5	60	94
6	60	67
7	34	80
Rata-rata	43,14	72,7

Dari table 1 dapat dilihat bahawa akurasi dari Naïve Bayes saja sangat rendah. Terutama pada kelompok berita 3 dan 4 yang hanya menghasilkan akurasi 27% saja. Namun dengan penambahan information gain dapat meningkatkan hasil akurasi dengan minimal 54% pada kelompok berita 4 dan bahkan sampai 94% pada kelompok berita 5. Dari perbandingan ini deteksi dengan penambahan information gain meningkatkan akurasi dengan sangat signifikan, bahkan ada yang hampir 3xnya yaitu pada kelompok berita 3.

Peningkatan ini disebabkan karena jumlah fitur yang digunakan pada hassil information gain lebih tepat untuk dijadikan klassifikasi. Dengan jumlah yang lebih sedikit, namun sangat relevan untuk proses pendeteksian.

KESIMPULAN

Pada kasus deteksi berita palsu dengan metode Naïve Bayes saja pada penelitian ini menghasilkan akurasi yang sangat rendah. Dengan penambahan proses pemilihan fitur menggunakan information gain terbukti dapat meningkatakan hasil akurasi yang sangat signifikan bahkan ada yang yang hampir 3x dari akurasi tanpa penambahan information gain. Untuk rata-rata akurasipun rata-rata akurasi tanpa information gain 43,14% sedangkan dengan information gain 72,7%, hampir dua kalinya. Hal ini membuktikan bahwa pengurangan jumlah fitur yang dihasilkan oleh information gain sangat baik untuk meningkatkan hasil akurasi suatu algoritma. Sehingga kedepannya information gain ini juga bisa diterapkan untuk metode klasifikasi yang lain sehingga diharapkan akan menambah keakurasian dalam mendeteksi sebuah berita itu asli atau palsi.

UCAPAN TERIMA KASIH

Pada kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada kampus Institut Teknologi Adhi Tama Surabaya yang telah memfasilitasi penulis untuk melakukan penelitian ini.

DAFTAR PUSTAKA

[1] A. N. Chy, M. H. Seddiqui, and S. Das, “Bangla news classification using naive Bayes classifier,” in *16th Int’l Conf. Computer and Information Technology, ICCIT 2013*, 2014, pp. 366–371.

[2] I. Y. R. Pratiwi, R. A. Asmara, and F. Rahutomo, “Study of hoax news detection using naïve bayes classifier in Indonesian language,” in *Proceedings of the 11th International Conference on Information and Communication Technology and System, ICTS 2017*, 2018, vol. 2018-January, pp. 73–78.

[3] F. Rahutomo, I. Y. R. Pratiwi, and D. M. Ramadhani, “Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia,” *J. Penelit. Komun. DAN OPINI PUBLIK*, vol. 23, no. 1, Jul. 2019.

[4] E. Y. Sari, Yuita Arum.;Puspaningrum, “(PDF) Pencarian Semantik Dokumen Berita Menggunakan Essential Dimension of Latent Semantic Indexing dengan Memakai Reduksi Fitur Document Frequency dan Information Gain Thresholding,” *Seminar Nasional Teknologi Informasi dan Multimedia*, 2013. [Online]. Available: https://www.researchgate.net/publication/259571815_Pencarian_Semantik_Dokumen_Berita_Menggunakan_Essential_Dimension_of_Latent_Semantic_Indexing_dengan_Memakai_Reduksi_Fitur_Document_Frequency_dan_Information_Gain_Thresholding. [Accessed: 22-Jan-2021].

[5] A. Hatta, H.R.; Arifin, ZA;Yuniarti, “(PDF) METODE HIBRIDASI ANT COLONY

OPTIMIZATION DAN INFORMATION GAIN UNTUK SELEKSI FITUR PADA DOKUMEN TEKS ARAB," *SCAN vol.III*, 2013. [Online]. Available: https://www.researchgate.net/publication/272165724_METODE_HIBRIDASI_ANT_COLONY_OPTIMIZATION_DAN_INFORMATION_GAIN_UNTUK_SELEKSI_FITUR_PADA_DOKUMEN_TEKS_ARAB. [Accessed: 22-Jan-2021].