

Perbandingan Algoritma Winnowing dan Algoritma Manber dalam Mendeteksi Berita Hoax di Media Sosial

A. M. A. K. Parewe^{1*}, A. Aman², D. P. M. Dewang³

¹STMIK AKBA

Email: *amaulidinnawati@akba.ac.id

Abstrak

Saat ini berbagai Informasi dapat beredar dengan cepat dan luas dari berbagai belahan dunia melalui Internet. Kemudahan untuk mengakses internet yang sudah bisa dinikmati di mana saja merupakan salah satu faktor utama yang menyebabkan informasi dapat dengan mudah diperoleh. Namun, dikarenakan kemudahan informasi yang didapat tersebut, tidak sedikit tindak kejahatan yang terjadi di dunia maya contohnya adalah Berita Hoax. salah satu kategori hoaks yang berbahaya dan banyak terjadi di kalangan pengguna media sosial yaitu kategori konten false, konten ini sangat menipu masyarakat. Sistem deteksi hoaks juga mampu memberikan informasi Penggunaan Web Sistem deteksi berita hoaks sebagai sarana dalam membandingkan algoritma Winnowing dan Manber sangat praktis dan sebagai solusi, serta mampu menangkal isu-isu berita hoaks yang beredar di kalangan pengguna media sosial dan meningkatkan intelektual kepada masyarakat agar tidak mudah percaya dengan pemberitaan yang menyesatkan. Dari algoritma winnowing dan Manber informasi menemukan perbandingan keefektifan langkah dan pengaruh jumlah parameter dalam menghitung kesamaan teks berita hoaks yang di uji. Hasil persentase 65% dari segi kecepatan maka Manber lebih baik dari pada Winnowing karena Manber memiliki langkah lebih sedikit dalam mencari nilai Persentase kesamaan teks berita hoaks yang di uji. Perbandingan dari segi ketepatan yang di dapat adalah winnowing lebih baik dari pada manber di karenakan winnowing lebih akurat pada langkah pemilihan nilai windows terkecil sehingga mendapatkan nilai persentase yang lebih akurat.

Kata Kunci: Winnowing, Manber, Hoax

1. Pendahuluan

Saat ini berbagai Informasi dapat beredar dengan cepat dan luas dari berbagai belahan dunia melalui Internet. Kemudahan untuk mengakses internet yang sudah bisa dinikmati di mana saja merupakan salah satu faktor utama yang menyebabkan informasi dapat dengan mudah diperoleh. Namun, dikarenakan kemudahan informasi yang didapat tersebut, tidak sedikit tindak kejahatan yang terjadi di dunia maya contohnya adalah Berita Hoax.

Salah satu berita hoax yang terjadi di kalangan masyarakat misalnya pada berita yang di posting oleh sebuah akun media sosial facebook bernama Eventorganizerjakarta Kocekoce mengunggah seorang dokter Rumah Sakit Angkatan Laut (RSAL) dr Ramelan Surabaya yang diklaim meninggal akibat Covid-19 viral di media sosial sejak Minggu (21/6/2020) malam. Dalam foto disebutkan bahwa dokter tersebut bernama Dokter Tirka Nandadan Humas RSAL dr Ramelan Surabaya, drg Aldiah, membantah jika ada tenaga kesehatan di RSAL dr Ramelan Surabaya meninggal akibat Covid-19, dia menyebut Tidak ada tenaga kesehatan yang meninggal dunia kemarin. Dia bahkan menyebut tidak ada nama dokter di RSAL dr Ramelan Surabaya bernama Dr Tirka

Nandanan seperti yang dicantumkan dalam unggahan foto tersebut. Yang dia tahu, di kesatuan Angkatan Laut, ada nama dr I Ketut Tirka Nandaka. Ini merupakan kategori Informasi salah. Hoax menurut KBBI adalah berita bohong yang tidak bersumber yang biasanya disebar dari media sosial. Hoax dapat menimbulkan perpecahan dalam masyarakat [1]

Berita bohong disusun tidak berdasarkan fakta. (Arsitu Hidayatullah, 2017) Secara umum, hoaks tersebut bisa dikategorikan menjadi dua golongan besar, yakni misinformasi dan disinformasi. Secara sederhana bisa dikatakan bahwa misinformasi adalah bentuk informasi yang salah, sementara disinformasi adalah informasi yang (justru) sengaja dibuat salah. Persebaran misinformasi dan disinformasi menjadi lebih mudah dengan adanya teknologi. Tehnologi yang mampu menciptakan media baru pula, sehingga hoaks akan menjadi lebih mudah dipercaya jika didukung dengan adanya foto dan video.

Di Indonesia, Pemerintah telah mengatur dalam Pasal 28 ayat 1 Undang Undang ITE yang berbunyi “Setiap orang yang dengan sengaja dan atau tanpa hak menyebarkan berita bohong dan menyesatkan, ancamannya adalah bisa terkena pidana maksimal 6 tahun “penjara dan denda maksimal Rp. 1 miliar. Beberapa algoritma yang termasuk dalam metode fingerprinting yaitu Algoritma Rabin Karp, Algoritma Winnowing dan Algoritma Manber. Dalam penelitian ini akan dilakukan analisis performansi dari dua buah algoritma yang termasuk dalam metode.

Pada penelitian sebelumnya dengan judul “Deteksi Konten Hoax Berbahasa Indonesia Pada Media Sosial Menggunakan Metode Levenshtein Distance”[2] dilakukan pra-pemrosesan kata berupa stemming bahasa dengan merubah suatu kata menjadi kata dasarnya yang dapat merubah makna dari sebuah kata tersebut, sehingga bisa mengurangi ke akuratan dalam pendeteksian berita. Pada judul “Perbandingan Algoritma Winnowing dengan Algoritma Rabin karp untuk mendeteksi plagiarisme pada kemiripan teks judul skripsi” yang melakukan proses analisis input jumlah nilai gram di mulai dari yang terkecil. dengan judul “Rancang Bangun Aplikasi Deteksi Kemiripan Dokumen Teks Menggunakan Algoritma Ratcliff/Obershelp [3]. Dalam analisis data dilakukan penganalisaan data dengan melakukan text preprocessing dan processing. Dimana dalam text preprocessing akan dilakukan pembersihan data teks dan pada tahap processing.

[4] Untuk menghindari berita hoax dikalangan masyarakat sangat penting agar tidak terjadi kesalahpahaman dari berita tersebut. Penelitian ini menghitung tingkat kesamaannya (similarity) dengan berita hoax pada database tanpa melakukan pra-pemrosesan kata yang dapat merubah makna dari kata tersebut. Maka ditentukannya nilai gram [5] dan window pada perhitungan Algoritma Winnowing[7] dan Algoritma Manber [6], diharapkan bisa memudahkan alur penelitian untuk mengukur banyaknya kesamaan yang dimiliki setiap berita saat diproses, sehingga menghasilkan persentase yang akurat dari kemungkinan berita tersebut mengandung hoax atau tidak mengandung hoax serta peneliti dapat menentukan perbandingan algoritma yang di gunakan dari segi kecepatan dan efektifitas .

2. Metode

2.1 Algoritma Winnowing

Algoritma *winnowing* adalah salah satu algoritma yang digunakan untuk mendeteksi bentuk kesamaan (*similarity*) [8] pada dokumen teks. Setiap kata dalam dokumen teks diubah terlebih dahulu menjadi sebuah kumpulan nilai hash. teknik yang digunakan untuk mencari nilai hash dalam *winnowing* adalah *rolling* hash. Nilai hash merupakan nilai numerik yang terbentuk dari perhitungan ASCII tiap karakter [9].

Ada beberapa tahapan dalam penerapan algoritma *winnowing*:

- a. Tahap Pertama: Membuang karakter yang tidak relevan seperti tanda baca, spasi, dan simbol-simbol lainnya.
- b. Tahap Kedua: Membentuk rangkaian gram.
- c. Tahap Ketiga: Melakukan proses *rolling* hash untuk mencari nilai hash dari setiap gram.
- d. Tahap Keempat: Membentuk *window* yang terdiri dari nilai hash yang dihasilkan.
- e. Tahap Kelima: Membentuk nilai *fingerprint* yang unik, dengan memilih nilai terendah dari setiap baris di dalam *window*.

2.2 Algoritma Manber

Algoritma Manber yakni metode Pendekatan Manber dengan cara memilih hasil dari proses hashing dengan cara memilih semua hasil hashing dengan yang memenuhi kriteria $0 \text{ mod } p$. Dengan cara ini fingerprints yang terpilih tidak tergantung dari posisinya [6].

secara singkat, konsep dasar algoritma Manber dimulai dari tahap awal baik penghapusan *noise* dan *whitespace* hingga hasil akhirnya berupa persentase adalah:

1. Penghapusan *noise* dan *whitespace*.
2. Pembentukan rangkaian gram dengan panjang N karakter.
3. Penghitungan nilai hash dari setiap gram menggunakan fungsi hash.
4. Pemilihan beberapa nilai hash menjadi *fingerprint* dokumen.

Menentukan persentase kemiripan antar dokumen menggunakan persamaan Jaccard *Coefficient*.i [5].

$$= \frac{\text{hasil pengujian}}{\text{target pengujian}} \times 100 \%$$

2.3. Penghapusan Noise & Whitespace Berdasarkan Algoritma Manber

[10], artinya dalam melakukan pendeteksian terhadap dokumen teks, algoritma tidak boleh dipengaruhi oleh spasi, jenis huruf (kapital atau normal), tanda baca dan sebagainya. Oleh sebab itu, dilakukan penghapusan terhadap karakter yang tidak relevan tersebut sehingga nantinya hanya akan tersisa satu rangkaian yang terdiri dari huruf dan angka

[10] Noise Suppression, artinya dalam melakukan pendeteksian, algoritma harus dapat menghindari adanya kata yang tidak penting, misal: “di”, “ke”, dan sebagainya. Panjang kata yang ditengarai harus cukup untuk membuktikan bahwa kata-kata tersebut telah di jiplak dan bukan merupakan kata yang umum digunakan.

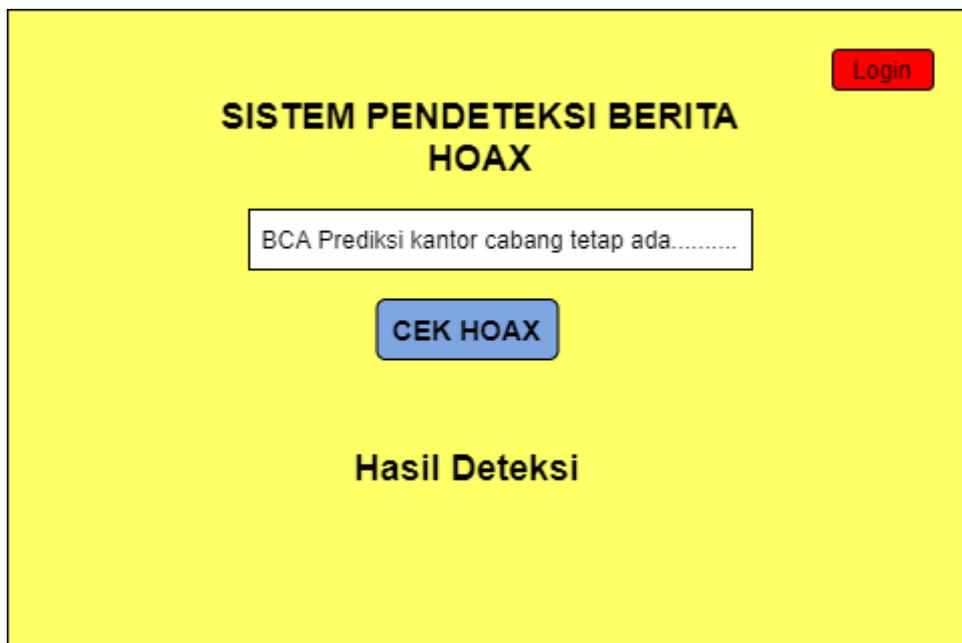
Position Independence, artinya pendeteksian tidak boleh bergantung pada posisi kata sehingga apabila posisi kata berbeda maka pendeteksian tetap dapat dilakukan.

Metode N-Gram Algoritma Manber memiliki metode utama yaitu metode N-Gram. Metode N-Gram merupakan metode yang berfungsi untuk memecah kata atau kalimat menjadi sebuah rangkaian dengan panjang N karakter. Sebagai contoh : "KEMEJA"

Dengan menggunakan nilai N = 2, maka akan dihasilkan :
"KE", "EM", "ME", "EJ", "JA"

[10] berperan penting karena merupakan langkah awal dalam proses pembentukan fingerprint. Dengan kata lain, metode

3. Hasil dan Pembahasan



Gambar 1. Tampilan Halaman Utama Aplikasi Sistem Deteksi *Hoax*

Hasil Pengujian = Total jumlah pengujian yang menjawab kuisisioner

Target pengujian = Didapat dari tota Jumlah pertanyaa dikalikan dengan responden
yang pada table 4.13 Hasil pengujian Kuisisioner

Dari hasil pengujian yang telah dilakukan dapat di lihat bahwa 88% pengujian merasa Sistem deteksi hoaks Layak di gunakan dalam mendeteksi berita di sosial media.

Hasil 88% di dapat dari :

$(\text{hasil pengujian}) / (\text{target pengujian}) \times 100\%$

$$\frac{\text{hasil pengujian}}{\text{target pengujian}} \times 100\%$$

Dimana hasil pengujian 173

Target pengujian 200

$$88/200 \times 100\% = 86,5\%$$

4. Kesimpulan

Berdasarkan hasil implementasi dan pengujian dapat disimpulkan Perbandingan algoritma Winnowing dan Manber dalam mendeteksi berita hoaks di media sosial yaitu:

1. Perbandingan dari segi kecepatan, maka Algoritma manber lebih baik dari pada Winnowing. Di karenakan Manber memiliki langkah lebih sedikit dalam mencari nilai Persentase kesamaan teks berita hoaks yang di uji.
2. Perbandingan dari segi ketepatan yang di dapat adalah winnowing lebih baik dari pada manber di karenakan winnowing lebih akurat pada langkah pemilihan nilai windows terkecil sehingga mendapatkan nilai persentase yang lebih akurat.
3. Pada Algoritma Manber pemilihan nilai P dapat mempengaruhi kelonggaran data akurasi kemiripan sehingga melewatkan data yang sebaiknya lebih baik pada saat setelah proses rolling hash

Referensi

- [1] F. Nurseha, Aji Teguh, Restu Alamsyah, 'Scripta: jurnal ilmiah mahasiswa', *J. Ilm.*, vol. 1, no. 2, pp. 106–115, 2019.
- [2] E. Wahyono, R. Amalia, and I. C. Ranteallo, 'Menelaah Lebih Dekat "Post Factual/Post Truth Politics, Studi Kasus Brexit" (Analsis Resensi Media)', *Nyimak (Journal Commun.*, vol. 1, no. 1, pp. 23–31, 2017, doi: 10.31000/nyimak.v1i1.272.
- [3] Y. L. Joane, A. Sinsuw, and A. Jacobus, 'Rancang Bangun Aplikasi Deteksi Kemiripan Dokumen Teks Menggunakan Algoritma Ratcliff / Obershelp', vol. 11, no. 1, 2017.
- [4] A. Mayssara A. Abo Hassanin Supervised, '濟無No Title No Title No Title', *Pap. Knowl. . Towar. a Media Hist. Doc.*, 2014.
- [5] M. Muryanto and D. Setyawan, 'Penentuan Kemiripan Manfaat Tanaman Obat dengan Gejala Sakit Menggunakan Algoritma Winnowing', *Just TI (Jurnal Sains Terap. Teknol. Informasi)*, vol. 13, no. 1, p. 12, 2021, doi: 10.46964/justti.v13i1.605.
- [6] M. Faisal, F. Nugroho, M. M. El Sulthan, F. Amini, M. A. Hariyadi, and A. Sedayu, 'Plagiarism detection using manber and winnowing algorithm', *Int. J. Adv. Sci. Technol.*, vol. 29, no. 6 Special Issue, pp. 2130–2136, 2020.
- [7] J. Jarwati, A. C. Prihandoko, and W. E. Yulia R, 'Penerapan Algoritma Winnowing Pada Sistem Rekomendasi Penentuan Dosen Pembimbing Skripsi (Studi Kasus Program Studi Sistem Informasi)', *Berk. Sainstek*, vol. 5, no. 1, p. 11, 2017, doi: 10.19184/bst.v5i1.5369.

- [8] W. Susanti, 'Penerapan Algoritma Cosine Similarity Pada Sistem Pengajuan Judul Tugas Akhir Berbasis Web', 2021.
- [9] R. Purnamasari, M. Fairuzabadi, and A. Riyadi, 'Sistem Pengecekan Plagiasi Judul Tugas Akhir Menggunakan Algoritma Winnowing di Fakultas Sains dan Teknologi Universitas PGRI Yogyakarta', pp. 16–20.
- [10] S. Agustian and A. Sucipto, 'Source Retrieval pada Deteksi Plagiarisme Berdasarkan Biword Fingerprint dengan Model Ruang Vektor', pp. 79–86, 2020.