Terakreditasi SINTA Peringkat 2

Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 10/E/KPT/2019

masa berlaku mulai Vol. 1 No. 1 tahun 2017 s.d Vol. 5 No. 3 tahun 2021

Terbit online pada laman web jurnal: **http://jurnal.iaii.or.id**

|  |  |  |
| --- | --- | --- |
| logok.png | JURNAL RESTI  (Rekayasa Sistem dan Teknologi Informasi) | |
| Vol. x No. x (20xx) xx - xx | ISSN Media Elektronik: 2580-0760 |

**KLASIFIKASI BERITA HOAX MENGGUNAKAN METODE**

Alam Asrorul Haq1, I Made Wirawan2, Agusta Rakhmat Taufani3

1,2,3 Teknik Elektro, Fakultas Teknik, Universitas Negeri Malang

1kurniaanis73@gmail.com, 2muladi@um.ac.id, 3utomo.pujianto.ft@um.ac.id

# Abstract

News is a collection of some information which is summarized and distributed. News has 2 categories, false news and facts news. Fake news gives very bad to all who receive it. This can cause divisions, dissension and changes in attitude to someone.Example is the news that is spread on social media in order to pit two class sheep. Therefore, this research will be collected to find news that is presented in the false news category or not by using machine learning. The data obtained was collected from data contained on the turnbackhoax.id site. This study uses the Tf-Idf and Apriori methods for weighting words in the news which will then look for word values ​​indicated by the hoax news offered to create a library of hoax words. Hoax word library is searched by using 100 training data. For the testing process, it is done by comparing the word and word libraries in the news that are obtained value to the authenticity of the news. The news testing process uses 14 pieces of correct news and 16 pieces of false news or hoaxes. The results of this apriori method are news values ​​/ weights indicated hoax or not. From the test data that has been used, obtained a test value of 83.3% for testing both true and false news. With this research, it is expected to minimize the spread of hoax news and spread hoax news early on.

*Keywords: Tf-Idf, Apriori, Fake News, Fact News*

# Abstrak

Berita adalah kumpulan dari beberapa informasi yang dirangkum dan disebarkan. Berita memiliki 2 buah kategori ,berita bohong/ *fake* dan berita benar /*fact*. Berita *fake* memberikan dampak yang sangat buruk bagi semua kalangan yang menerimanya. Hal tersebut dapat menimbulkan perpecahan,pertikaian dan perubahan sikap pada seseorang. Salah satu contohnya yakni berita yang disebar di sosial media guna mengadu domba 2 buah kalangan. Oleh karena itu, pada penelitian ini akan dilakukan klasifikasi guna mengetahui apakah berita yang disajikan termasuk kategori berita *fake*  atau tidak dengan menggunakan *machine learning.*Data yang didapat yakni dikumpulkan dari data yang terdapat di situs *turnbackhoax.id*. Penelitian ini menggunakan metode *Tf-Idf* dan Apriori untuk pembobotan kata dalam berita yang selanjutnya akan dicari nilai kata yang terindikasi berita *hoax* yang bertujuan untuk membuat library kumpulan kata *hoax*. Library kumpulan kata hoax di cari dengan menggunakan 100 buah data latih. Untuk proses pengujian, dilakukan dengan membandingkan *library* kata dan kata pada sebuah berita sehingga didapat nilai ke aslian dari berita.Proses pengujian berita menggunakan 14 buah berita benar dan 16 buah berita salah atau *hoax.* Hasil penelitian dengan metode apriori ini yakni sebuah nilai / bobot berita terindikasi *hoax* atau tidak*.* Dari data uji yang telah digunakan, didapat nilai akurasi sebesar 83,3% untuk pengujian baik berita benar maupun berita salah. Dengan adanya penelitian ini diharapkan dapat meminimalisir penyebaran berita *hoax* dan mendeteksi berita-berita *hoax* sejak dini.

© 20xx Jurnal RESTI

# Pendahuluan

Informasi adalah sebuah pesan yang disampaikan dari satu orang kepada orang lain baik melalui media maupun dari mulut ke mulut.Selain itu, informasi dapat diartikan dengan hasil dari pengolahan sebuah atau beberapa data (Darmawan, 2012). Tanpa adanya sebuah informasi seseorang tidak dapat saling mengetahui satu sama lain dan tidak dapat belajar dari pengalaman orang lain. Informasi juga sangat bermanfaat untuk kehidupan bersosial maupun personal. Kumpulan dari beberapa informasi baru yang tersusun dan disajikan dalam bentuk media cetak, siaran ataupun internet disebut berita. Berita juga dapat menjadi sebuah acuan informasi untuk di konsumsi khalayak umum sehingga apapun informasi yang sedang atau masih baru terjadi akan ditampilkan sehingga orang dapat mengetahui informasi tersebut dengan mudah.

Dalam berkomunikasi, informasi menjadi suatu hal yang penting untuk disampaikan. Dengan perkembangan teknologi yang begitu pesat informasi menjadi salah satu konsumsi masyarakat.Akibatnya, batas- batas ruang dan waktu menjadi semakin kabur (Ardoni, 2015). Seiring berkembangnya teknologi penyebaran informasi menjadi sangat mudah. Salah satunya yakni penyebaran informasi melalui berita baik melalui media social, situs berita *online* maupun situs-situs penyedia layanan informasi lain.Jejaring social seakan memperkuat kedudukan internet sebagai *new* *media communication*,dimana jarak seakan tidak lagi terlihat(Kesuma, 2016).Sangat tingginya pengguna internet juga mempengaruhi penyebaran berita melalui media online.Pengguna internet di Indonesia mencapai 38 juta lebih dengan penetrasi mencapai 15 persen pada tahun 2014 (Lestari, 2017). Adanya berita dapat berdampak positif maupun berdampak negative bagi pembacanya. Sehingga pembaca dituntut untuk bijak menyikapi berita yang disebar oleh penyedia layanan berita. Banyak sekali persoalan yang timbul akibat penyalahgunaan berita sebagai sarana kampanye, politik dan lain sebagainya.

Pertukaran sebuah informasi sangatlah penting bagi kehidupan bersosial. Akan tetapi ada beberapa oknum yang menyalahgunakan informasi atau menyelewengkan informasi guna memenuhi kepentingan pribadi sehingga dapat merugikan beberapa pihak terkait khususnya dalam sebuah berita. Pembuatan berita-berita palsu oleh oknum-oknum terkait biasa disebut dengan berita *hoax* / berita yang masih diragukan fakta dan status aslinya. Berita hoax biasa didasari dari dendam, dengki atau pencarian sensasi oleh si pembuat kepada yang dituju. Seperti berita pada 5 februari 2019 tentang es krim magnum mengandung babi yang disebarkan melalui jejaring social facebook. Berita ini tidak terbukti kebenaranya sehingga disimpulkan sebagai berita hoax. Tujuan dari pembuatan berita palsu ini yakni dikarenakan adanya dendam kepada seseorang ataupun hanya untuk mencari sensasi agar mendapat view dan follower yang cukup banyak. Selain itu media sosial juga memiliki fungsi positif yakni memberikan informasi terkini tentang peristiwa yang terjadi pada masyarakat (Aulia., 2018).

Media social merupakan sarana yang efektif untuk menyampaikan suatu informasi. Dengan semakin cepatnya pertukaran informasi, semakin cepat pula penyebaran berita-beriat *hoax* yang dinilai sangat merugikan dikalangan masyarakat melalui media social (Rahadi dkk, 2017). Berita hoax juga dapat memicu perpecahan sehingga sangat berbahaya dikonsumsi oleh masyarakat umum. Dengan begitu masyarakat dituntut untuk memilah dan memilih berita serta memastikan keaslian berita terlebih dahulu sebelum mencerna mentah-mentah isi dari berita.Hoax berkembang karenda didasari beberapa faktor meliputi : Motif politik kekuasaan, Kurangnya kesadaran social masyarakat , penyalahgunaan pengaruh tokoh-tokoh guna mempengaruhi pengikutnya (Septanto, 2017). Dalam kasus berita hoax yang ada di Indonesia ini sangat beragam. Semakin banyak pula penyebaran berita hoax dan segala ujaran kebencian guna kepentingan politik yang kian masif jelang Pilpres 2019. Kementerian Sekretariat Negara menyebut penyebaran berita hoaks menjelang Pemilu merupakan satu fenomena yang timbul di tengah masyarakat. Ini berpotensi menciptakan disintegrasi dan memecah belah bangsa Indonesia.(Syafirdi, 2019). Dalam kasus ini, terbukti bahwa berita hoax sangat signifikan berperan dalam pemecah belah masyarakat. Sehingga perlu adanya solusi dan penindak lanjutan oknum-oknum pembuatan berita hoax. Dengan adanya penelitian *Klasifikasi Berita Hoax Menggunakan Metode Apriori* diharapkan dapat meminimalisir perpecahan masyarakat akibat berita hoax.

# Metode Penelitian

1. Rancangan Penelitian

Metode penelitian yang digunakan adalah metode eksperimen. Metode eksperimen adalah metode cara penyajian sebuah data melalui proses percobaan-percobaan untuk membuktikan sendiri suatu pernyataan atau hipotesis tertentu dalam sebuah penelitian.Penelitian eksperimen merupakan penelitian yang paling valid dikarenakan metode eksperimen dilakukan dengan pengontrolan variable yang sangat ketat sehingga didapat hasil yang maksimal(Jaedun, 2011). Keuntungan utama penelitian dengan penggunaan metode eksperimen adalah adanya kendali ditangan peneliti dan ketepatan sebuah logika yang terkandung di dalam metode tersebut. Metode ini juga sering digunakan karena dapat dibuat perencanaan yang sistematis sehingga metode eksperimen adalah metode yang paling kuat dalam sebuah penelitian. Berikut adalah tahapan metode eksperimen :

Pengumpulan Data

Pre Prosses NLP

Proses TF-IDF

Proses Apriori

Experimen dan Pengujian

Evaluasi dan Validasi Hasil

Gambar 1 Metode yang Digunakan

2.2. Sumber Data

Berita sendiri memiliki arti yakni laporan tercepat dari suatu peristiwa atau kejadian yang faktual. Pada penelitian ini terdapat kumpulan dokumen berita tervalidasi berita *hoax* yangnantinya digunakan untuk pembuatan *library* kata. Berita *hoax* ini didapat dari website *turnbackhoax.id* yang memilki beberapa berita yang terbukti bukan berita benar atau tervalidasi berita *hoax.* Penggunaan klasifikasi dinilai tepat untuk identifikasi berita hoax dikarenan dalam metode klasifikasi terdapat teknik untuk mengidentifikasi pola dari kata pada berita yang terindikasi *hoax.* Data dokumen yang sudah dikumpulkan dibagi menjadi 2 buah bagian yakni data uji dan data latih. Data latih yakni berita tervalidasi hoax yang nantinya akan digunakan untuk pembuatan *library* kata. Data uji yakni data baik berita *hoax* maupun non *hoax* yang digunakan untuk mengukur performa sistem yang digunkan untuk klasifikasi berita *hoax*.

## Data Preprocessing

Data berita yang didapat dari situs turnbackhoax.id ataupun dari detik.com adalah sebuah data mentah yang belum dapat diolah. Agar data tersebut dapat diolah, maka dilakukan pre prosessing dengan menggunakan NLP (*Neural Language Processing*). Pada tahap pre prosessing ini memiliki beberapa *step* yakni :

1. Tokenizing

Proses pemisahan text menjadi per kata sehingga tiap kata dapat diolah dengan mudah. Proses pemisahan kata dilakukan menggunakan fungsi dalam python.

*2. Stoward Removal*

Stopword merupakan kata umum yang digunakan / kata yang tidak penting untuk dianalisi / diproses. Seperti contoh kata : dan, atau, tetapi. Tujuan dari penghapusan kata ini yakni mengurangi pembengkakan jumlah index yang digunakan.

3. *Stemming*

Sebuah penghapusan kata yang memilki awalan / akhiran sehingga didapat kata dasar. Guna dari *stemming* ini yakni meminimalisir persamaan kata yang memiliki perbedaan awalan / akhiran sehingga dapat memperkecil jumlah indeks yang akan diproses. Proses stemming ini menggunakan fungsi *stemmer factory* yang terdapat pada python dengan *import* *package* sastrawi

2.3 Proses Tf-Idf

. Setelah dilakukan *pre processing* dengan menggunakan NLP, maka langkah selanjutnya yakni menacari bobot dari sebuah kata dengan menggunakan *Term Frequent dan Inverse Document Frequent.* Berikut adalah alur kerja system :

1. Data yang sudah di proses dengan NLP dimasukkan kedalam perhitungan Tf
2. Selanjutnya akan dihitung nilai Idf dari data

c. Pencarian bobot dengan menggabungkkan Tf dan Idf

d. Kata dan bobot di simpan dalam bentuk .csv

1. Proses Apriori

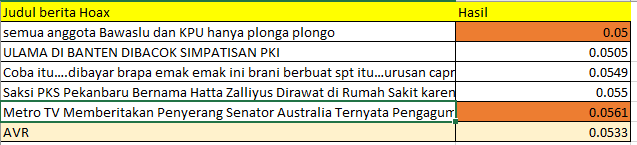
Tahap selanjutnya yakni prosesing dengan menggunakan metode apriori. Dalam metode ini diambil sebuah nilai support dari perbandingan antara jumlah berita yang mengandung kata dengan jumlah total berita. Yang selanjutnya akan dimasukkan kedalam data *index library* kata yang tersimpan beserta bobot. Selanjunya, data akan dicari rataan dari nilai bobot dan nilai support sehingga didapat nilai akhir. Berikut proses alur kerja sistem:

1. Masukan data yang sudah diproses NLP
2. *System* menghitung nilai *support* dengan menggunakan rumus apriori
3. Proses selanjutnya yakni pencarian nilai akhir dengan cara menari rataan dari nilai *support*  dan nilai bobot
4. Data kata dan nilai akhir di simpan dalam bentuk .csv

**3. Hasil dan Pembahasan**

Pada tahap ini adalah tahap pengujian *system* dengan menggunakan data uji yang telah di siapkan berupa 16 data terindikasi berita *hoax* dan 14 buah data berita berupa berita real. Untuk menentukan nilai batas dilakukan uji coba menggunakan 5 buah data terindikasi *hoax*  yang nantinya dicari rataan nilai sehingga didapat hasil nilai batas. Selanjutnya, *system* akandiuji menggunakan batas yang telah ditentukan dan dicari hasil dari *precision and recall*  serta tingkat akurasi.

3.1 Pencarian Nilai Batas

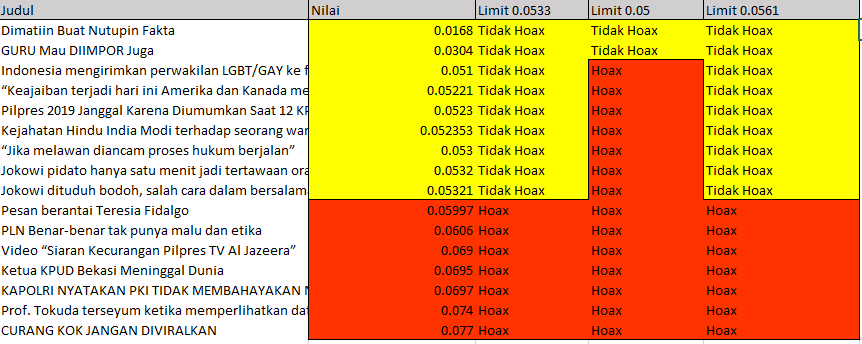


Gambar 2 Data Uji Nilai Batas

Untuk mencari nilai batas dilakukan uji coba dengan menggunakan 5 buah data uji berita terindikasi *hoax*.Hasil rata-rata, batas atas dan batas bawah dari nilai data uji akan digunakan untuk nilai batas dari pengujian data uji selanjutnya.

3.2  Proses Uji Hasil Data Uji

Pada tahap ini dilakukan 2 buah pengujian yakni pengujian data berupa berita hoax dan pengujian data berupa data real.Proses pengujian dilakukan dengan menggunakan 3 buah nilai batas yang telah ditentukan yakni 0.0533, 0.0561 dan 0.05.



Gambar 3 Data Uji Berita Hoax



Gambar 4 Data Uji Berita Real

3.3 Skenario 1

Skenario 1 proses uji dengan menggunakan nilai batas rata-rata dari pencarian nilai limit yakni 0,0533.

Tabel 1 Nilai dari Komponen Perhitungan dengan Batas 0.0533

|  |
| --- |
| Berita Jumlah |
| *True Positive* (tp) 7  *False Positive* (fp) 3  *False Negative* (fn) 9  *True Negative* (tn) 11 |

Berdasarkan nilai dari setiap komponen, maka dapat ditentukan nilai *Precision, Recall* dan *Accuracy* sebagai berikut :

*Precision :* tp/(tp+fp) *Recall :* tp/(tp+fn)

= 7/10 = 7/16

= 0,7 = 0,437

*Accuracy = x100*

*=* (7 + 11) / 30 x100

= 60 %

3.3 Skenario 2

Skenario 2 proses uji dengan menggunakan nilai batas bawah dari pencarian nilai limit yakni 0,05.

Tabel 2 Nilai dari Komponen Perhitungan dengan Batas 0.05

|  |
| --- |
| Berita Jumlah |
| *True Positive* (tp) 14  *FalsePositive*(fp) 3  *FalseNegative*(fn) 2  *TrueNegative*(tn) 11 |

Berdasarkan nilai dari setiap komponen, maka dapat ditentukan nilai *Precision, Recall* dan *Accuracy* sebagai berikut :

*Precision :* tp/(tp+fp) *Recall :* tp/(tp+fn)

=14/17 = 14/16

=0,82 = 0,875

*Accuracy = x100*

*=* (14 + 11 ) / 30 x100

= 83,33 %

3.4 Skenario 3

Skenario 3 proses uji dengan menggunakan nilai batas atas dari pencarian nilai limit yakni 0,0561.

Tabel 3 Nilai dari Komponen Perhitungan dengan Batas 0.0561

|  |
| --- |
| Berita Jumlah |
| *TruePositive*(tp) 7  *FalsePositive*(fp) 2  *FalseNegative*(fn) 9  *TrueNegative*(tn) 12 |

Berdasarkan nilai dari setiap komponen, maka dapat ditentukan nilai *Precision, Recall* dan *Accuracy* sebagai berikut :

*Precision :* tp/(tp+fp) *Recall :* tp/(tp+fn)

=7/9 = 7/16

=0,77 = 0,437

*Accuracy = x100*

*=* (7 + 12) / 30 x100

= 63,33 %

## 3.5 Pembahasan

Berdasarkan hasil pengujian *system* nilai batas 0.0533 menghasilkan akurasi sebesar 60% , untuk nilai batas 0,561 mendapatkan akurasi sebesar 63,33 % dan untuk nilai batas 0.5 mendapatkan nilai akurasi sebesar 83,33%.Dengan data dari ke 3 buah skenario dapat disimpulkan bahwa nilai batas yang paling efektif digunakan untuk proses klasifikasi berita *hoax* pada system yakni 0.5 dengan tingkat akurasi 83,33 % dengan nilai *precission and recall* 82% dan 87% . Dapat dijelaskan bahwa pengubahan nilai limit sangatlah berpengaruh pada tingkat akurasi system.Dikarenakan nilai dari data uji yang berbeda-beda sehingga untuk dapat mencari kefektifan sebuah system digunakan ujicoba beberapa skenario.

**4 Kesimpulan**

Berdasarkan dari hasil penelitian yang telah dilakukan, maka dapat disimpulkan bahwa :

1. Metode Apriori dapat digunakan untuk mengklasifikasi berita terindikasi hoax.
2. Pada penelitian yang dilakukan metode apriori dapat menghasilkan tingkat akurasi terbesar 83,33 %
3. Perpaduan antara *apriori* dan Tf-Idf terbukti mampu membedakan antara berita *hoax* dan *real.*
4. Kelebihan dari algoritma ini yakni dapat digunakan untuk mendeteksi berita hoax murni (bukan berita yang diubah).
5. Kekurangan dari algoritma ini yakni sangat bergantung pada seberapa banyak library yang dipunyai oleh system dan proses pembuatan library yang memakan cukup waktu.
6. Dalam skenario 2 digunakan nilai batas 0.5 untuk mendeteksi 16 buah berita *hoax* dan 14 buah berita *real* memiliki nilai *precission* 82% dan nilai *recall* 87% sehingga semakin banyak data latih yang dimasukkan semakin akurat pula *system* melakukan proses klasifikasi

# Daftar Rujukan

[1] Ardoni. (2015). Teknologi Informasi: Kesiapan Pustakawan Memanfaatkannya Ardoni. *Jurnal Aplikasi Iptek Untuk Masyarakat*, *4*(1), 59–62. Retrieved from http://jurnal.unpad.ac.id/dharmakarya/article/download/9042/4064

[2] Bim, I. M. (2012). *Menerapkan Metode Probabilistik Binary Metode Probabilistik Binary Independence Model ( Bim ) Periode Wisuda : 28 Februari 2013 Fakultas Sains Dan Teknologi*.

[3] Bird, S., & Loper, E. (2016). The natural language toolkit NLTK: The Natural Language Toolkit. *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, (March), 63–70. https://doi.org/10.3115/1225403.1225421

[4] Cardiologia, S. B. de, Francisco, P. M. S. B., Segri, N. J., Borim, F. S. A., Malta, D. C., Fontbonne, A., … Brasil, N. (2018). https://doi.org/10.1590/s1809-98232013000400007

[5] Darmawan, D. (2012). *Mengenal Teknologi Informasi*. 92.

[6] Hendri. (2003). *Cepat Mahir python*. 1–92.

[7] Jaedun, A. (2011). Oleh : Amat Jaedun. *Metodologi Penelitian Eksperimen*, 0–12.

[8] Kesuma, H. W. A. (2016). *Penerapan Metode TF-IDF dan Cosine Similarity dalam Aplikasi Kitab Undang-Undang Hukum Dagang*.

[9] *2014-Python 3 Text Processing with NLTK 3 Cookbook*. (n.d.). https://doi.org/10.1017/CBO9781107415324.004\

[10] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2017). *Natural Language Processing: State of The Art, Current Trends and Challenges*. (Figure 1). Retrieved from http://arxiv.org/abs/1708.05148

[11] KPM. (2013). *Buku Panduan Pemrograman Python. Buku*, *84*, 487–492. Retrieved from http://ir.obihiro.ac.jp/dspace/handle/10322/3933

[12] Lestari, R. D. (2017). *Quality News dan Popular News Sebagai Trend Pemberitaan Media Online (Studi Deskriptif Kualitatif Trend Pemberitaan Quality News dan Popular News pada Media Online Nasional di Indonesia Periode 2016)*. *Channel*, *5*(1), 83–94.

[13] Listriani, D., Setyaningrum, A. H., & M.A, F. E. (2016). *Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Pola Belanja Konsumen ( Studi Kasus Toko Buku Gramedia Bintaro )*. *Jurnal Teknik Informatika Vol 9 No. 2, Universitas Islam Negeri Jakarta*, *9*(2), 120–127. Retrieved from http://journal.uinjkt.ac.id/index.php/ti/article/view/5602/3619

[14] Pascapraharastyan, R. A., Supriyanto, A., & Sudarmaningtyas, P. (2014). *Rancang Bangun Sistem Informasi Manajemen Arsip Rumah Sakit Bedah Surabaya Berbasis Web*. *Sistem Informasi*, *3*(1), 72–77.

[15] Saputro, G. A. (2017). *Perapan Algoritma Apriori Untuk Mencari Pola Penjualan di Cafe*.

[16] Saputro, H. (2016). Pembelajaran Basis Data (Mysql). *Modul Pembelajaran Praktek Basis Data (MySQL)*, 1–34. Retrieved from http://dinus.ac.id/repository/docs/ajar/materi\_1.pdf

[17] Septanto, H. (2017). *Jurnal Sosbud 3*. 157–162.

[18] Vibriza, J., Rahadi, D. R., Marwan, M. R., & Ahyad. (2017). Perilaku pengguna dan informasi. *Jurnal Gunadarma*, *4*(1), 192–208. https://doi.org/https://doi.org/10.22146/jps.v4i2.28586

[19] Weddiningrum, F. G., Studi, P., Informasi, S., Sains, F., Teknologi, D. A. N., Islam, U., & Ampel, N. (2018). *Deteksi Konten Hoax Berbahasa Indonesia Pada Media Sosial Menggunakan Metode Levenshtein Distance*.