Klasifikasi Berita Clickbait Menggunakan K-Nearest Neighbor (KNN)

by Joins 3705

Submission date: 15-Aug-2020 10:30AM (UTC+0700)
Submission ID: 1369760923
File name: 3705-10562-1-BR.docx (214.82K)
Word count: 2838
Character count: 17064

Klasifikasi Berita *Clickbait* Menggunakan *K-Nearest Neighbor* (KNN)

Abstrak

Clickbait menjadi salah satu cara untuk mencari pendapatan dengan meningkatkan traffic pembaca dan pengunjung. Praktik clickbait pada saat ini sudah merambah pada dunia jurnalistik dimana sistem berita media online berbeda dengan media cetak. Sama hal nya dengan media online lainnya praktik clickbait ini memberikan pengaruh besar terhadap penyedia situs berita karena rasa keingintahuan dari para pembaca serta sulitnya para pembaca memilih berita clickbait atau bukan clickbait. Praktik clickbait ini sendiri sangat di andalkan oleh penyedia situs berita yang kerap menggunakan judul–judul yang menjebak untuk menarik para pembaca. Berdasarkan masalah tersebut dilakukan penelitian untuk mengklasifikasikan berita clickbait menggunakan metode K-Nearest Neighbors (K-NN). Dari hasil penelitian yang dilakukan, didapatkan hasil terbaik pada jumlah k=11 dengan menggunakan skenario 1 pada pembagian data dengan jumlah data sebanyak 800 data latih dan 200 data uji yang menghasilkan akurasi sebesar 71%, Precision 72%, dan Recall 71%. Hal ini menunjukan bahwa klasifikasi berita clickbait dapat di klasifikasikan menggunakan K-Nearest Neighbor.

Kata kunci: *klasifikasi, text mining, clickbait, k-nearest neighbor*

Abstract

Clickbait is one way to find revenue by increasing traffic of readers and visitors. The practice of clickbait has now penetrated the world of journalism where the online media news system is different from print media. Same thing with other online media, this clickbait practice has a big influence on the providers of news sites because of the curiosity of the readers and the difficulty of the readers to choose clickbait news or not clickbait. The practice of clickbait itself is highly relied on by news site providers who often use trapping titles to attract readers. Based on these problems, a study was conducted to classify clickbait news using the K-Nearest Neighbors (K-NN) method. From the results of the research conducted, obtained the best results at the number k = 11 using scenario 1 in the division of data with the amount of data as much as 800 training data and 200 test data that produces an accuracy of 71%, 72% Precision, and 71% Recall. this shows that the clickbait news classification can be classified using K-Nearest Neighbor.

Keywords*: classification, text mining, clickbait, k-nearest neighbor*

1. PENDAHULUAN

Internet adalah salah satu teknologi informasi yang sangat diminati oleh banyak kalangan, karena internet telah menyentuh berbagai aspek kebudayaan manusia mulai dari gaya hidup, pendidikan, penelitian hingga ke dunia bisnis. Salah satu manfaat dari internet adalah sebagai penyedia informasi yang cepat dan fleksibel, namun terkadang informasi yang diberikan tidak

sesuai dengan judul artikel di media elektonik. Judul yang menjebak (*Clickbait*) merupakan modus media *online* untuk meningkatkan *traffic* pengunjung [1]

Clickbait menjadi salah satu cara untuk mencari pendapatan dengan meningkatkan *traffic* pembaca dan pengunjung. Semakin banyak pengunjung situs, semakin banyak pula kemungkinan untuk mendapatkan pendapatan dari situs tersebut. Faktor pendorong maraknya *clickbait* di media *online* adalah persaingan yang semakin ketat antar media untuk mendapatkan pembaca [2] dan pendapatan yang melimpah pada *viewer,* iklan, sama dengan keuntungan materil walaupun pada dasarnya antara judul dan isi tidak memiliki maksud yang sama atau judul yang bombastis tersebut hanya dijelaskan secara singkat atau tidak utuh [3]

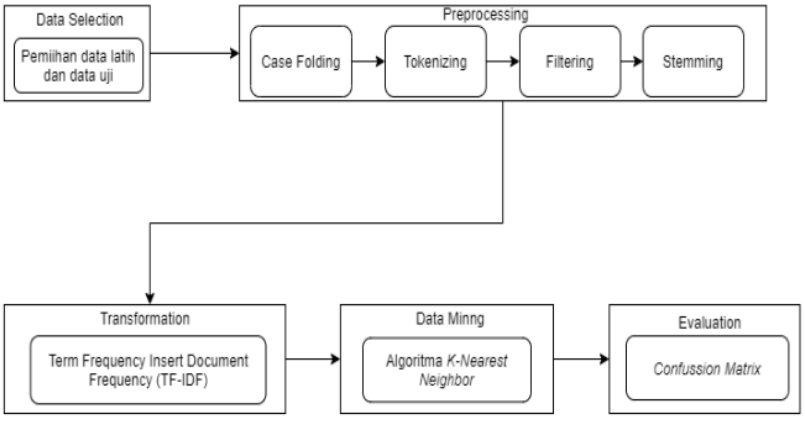
Praktik *clickbait* pada saat ini sudah merambah pada dunia jurnalistik dimana sistem berita media *online* berbeda dengan media cetak, dimulai dari pembaca melihat judul lalu melakukan klik barulah masuk pada isi berita. Sama hal nya dengan media online lainnya praktik *clickbait* ini memberikan pengaruh besar terhadap penyedia situs berita karena rasa keingintahuan dari para pembaca serta sulitnya para pembaca memilih berita *clickbait* atau bukan *clickbait*.

Berdasarkan praktik *clickbait* yang marak terjadi saat ini untuk pengklasifikasian *clickbait* atau bukan *clickbait* pada berita pengelompokan berita dilakukan dengan teknik *text mining*. *Text Mining* dapat didefinisikan penemuan serta ekstrasi pengetahuan yang menarik dari sebuah teks yang bebas maupun tidak struktur dan mencakup segala sesuatu mulai dari pengambilan informasi yaitu pengambilan dokumen atau pengambilan web untuk klasifikasi. Dalam penelitian [4] di dapatkan hasil bahwa KNN mendapatkan nilai akurasi yang lebih tinggi dibandingkan dengan SVM. Pada penelitian [5] didapatkan bahwa waktu proses pada KKN lebih cepat dari pada waktu proses SVM maka dapat di silmpulkan bahwa KNN merupakan algoritma yang terbaik. Dari penjelasan yang sudah dipaparkan sebelumnya dapat di lihat menggunakan *clickbait* pada judul berita memberikan pengaruh besar terhadap situs berita. Tetapi untuk para pembaca hal itu sangat membuat para pembaca resah karena para pembaca tidakdapat membedakan berita *clickbait* atau bukan *clickbait*.

Pada penelitian ini menggunakan metode text mining dimana pada metode text mining, dengan 3 skenario untuk pembagian dan algoritma K-Nearest Neighbor ini diharapkan dapat membantu pembaca atau pengunjung situs dalam memilih berita *clickbait* atau bukan *clickbait*.

2. METODOLOGI PENELITIAN

Metode yang digunakan adalah metode text mining, dimana metode ini terdapat beberapa proses diantaranya adalah, 1) *data Selection,* 2) *pre-processing*, 3) *tranformation*, 4) *data mining*, dan 5) *evaluation,* berikut alur penelitian.



Gambar 1 Alur Penelitian

2.1 *Data Selection*

Pada tahap ini merupakan tahap persiapan dan pemilihan data. Pemilihan data dilakukan dengan mengumpulkan 1000 berita, artikel berita yang akan menjadi data training dan data testing didapat dari situs berita dengan data berita terbit pada bulan januari 2020. Artikel berita yang telah dikumpulkan akan di kelompokkan dan dilabeli secara manual sesuai dengan kategorinya masing-masing yakni clickbait dan bukan clickbait. Pembagian data digunakan 3 skenario pembagian antara data training berbanding dengan data testing yaitu, 80%:20%, 50%:50%, 20%:80%. Dari 3 skenario pembagian data bertujuan untuk mendapatkan nilai performa sistem jika jumlah data training lebih banyak dari data testing, jumlah data training sama dengan jumlah data testing dan jika jumlah data training lebih sedikit dibandingkan dengan data testing Irfa & Mubarok (2018).

2.2 *Pre-processing*

Pada tahap *preprocessing* terdapat bebeapa tahapa untuk mentransformasikan data ke dalam bentuk representasi format data lain, berikut tahapan-tahapan pada *pre-processing*:

1. *Case folding* merupakan transformasi huruf pada teks, dimana semua huruf diubah menjadi huruf kecil. Karakter huruf yang dihasilkan harus berupa 'a' sampai pada karakter huruf 'z'. Selain karakter huruf tersebut harus dianulir [7].
2. *Tokenizing,* merupakan proses memecah yang semula berupa kalimat menjadi kata-kata atau memutus urutan *string* menjadi potongan-potongan seperti kata-kata berdasarkan tiap kata yang menyusunnya [8].
3. *Stopwords Removal atau Filtering, stoplist* atau *stopword* adalah kata-kata yang tidak penting yang dapat dibuang dengan pendekatan *bag-of- words*. Hasil dari *stoplist* adalah *wordlist* yang berisi kata penting. [9].
4. *Stemming*, merupakan proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya (*stem*) [10].

2.3 *Tranformation*

Pada tahap ini metode dipresentasikan dalam dokumen yang berisi bobot kata. Bobot tersebut menyatakan kepentingan atau kontribusi kata terhadap suatu dokumen dan kumpulan dokumen. Tahap pembobotan kata pada dokumen menggunakan metode TF-IDF. Metode TF-IDF ini menggabungkan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut [11].
Rumus dalam menentukan pembobot dengan TF-IDF adalah sebagai berikut:

$$IDF\ (Word) = \log \frac{td}{tf} \quad (1)$$

IDF *(word)* adalah nilai IDF dari setiap kata yang akan di cari, *td* adalah jumlah keseluruhan dokumen yang ada, *df* adalah jumlah kemuculan kata pada semua dokumen.

2.4 *Data Mining*

Pada tahap ini menggunakan *K-Nearest Neighbor (KNN),* Dalam pengklasifikasian terdapat 2 proses yang dilakukan yakni proses *training* dan proses *testing*. Proses pengklasifikasian dilakukan dengan menentukan nilai K terlebih dahulu, menghitung kuadrat jarak euclid(*query instance*) masing-masing objek terhadap *training* data yang diberikan, lalu mengumpulkan label class Y (klasifikasi *Nearest Neighbor*).

$$D\ (x_i, y_i) = \sqrt{\sum_{i=1}^{n} (x_i, y_i)^2} \quad (2)$$

Keterangan: Xi = *data training*
Yi = *data testing*
D (xi,, yi) = jarak
i = variabel data

n = Dimensi data

2.5 *Evaluation*

Pada tahap terakhir ini dilakukan evaluasi nilai *akurasi, precision* dan *recall* yang merupakan tahapan dalam penelitian dengan tujuan untuk mengetahui pengaruh model yang dihasilkan dari pengklasifikasian yang telah dilakukan. Pada penelitian ini hasil yang disajikan berupa grafik dengan metode validasi yang digunakan adalah *Confusion matrix* yang memberikan keputusan berdasarkan hasil dari *training* dan *testing* serta memberikan penilaian performansi klasifikasi berdasarkan data dengan benar atau salah.

Tabel 1 *Confusion Matrix*

Classification	*Predicted Class*	
	Class = Yes	*Class = No*
Class = Yes	a (*True Positive* – TP)	b (*False Negative* – FN)
Class = No	c (*False Positive* – FP)	d (*True Negative* - TN)

Terdapat beberapa cara untuk mengukur performa, beberapa cara yang sering digunakan adalah dengan menghitung akurasi, *recall, dan precission*[8].

$$Akurasi = \frac{TP + TN}{Tp + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{Tp + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

Evaluasi dan validasi hasil dihitung dengan menggunakan rumus *accuracy, precision, sensitifity, dan spesificity*. Nilai akurasi dinyatakan dalam persentase. Jika akurasi mencapai angka 100%, hal tersebut berarti bahwa semua kasus yang diprediksi diklasifikasikan seluruhnya sengan benar [12].

3. HASIL DAN DISKUSI

Bagian ini memaparkan hasil dari proses klasifikasi berita *clickbait*. Proses dimulai dengan membagi data berita yang dikumpulkan pada situs berita dengan data terbit pada bulan januari 2020 ke dalam data latih dan data uji. Kedua, melaksanakan setiap langkah *pre-processing*, kemudian menghitung bobot dengan TF-IDF dan implementasi algoritma *K-Nearest Neighbor (KNN)*. Terakhir melakukan evaluasi dan uji validasi dengan *Confusion Matrix*

3.1 *Data Selection*

Pada penelitian ini data berita yang digunakan adalah 1000 berita yang telah dilakukan labelling dengan 2 kategori yaitu kategori clickbait dan bukan clickbait. Data berita dilakukan pelabelan dengan cara manual tiap berita berita diberi label *clickbait* (C) atau bukan *clicbait* (BC).berita yang dilabeli *clickbait* (C) mengandung kata yang bersifat hiperbola pada judul dan ketidaksesuaian isi berita dengan judul, sedangkan berita yang dilabeli bukan *clickbait* (BC) berita yang tidak menganduk kata hiberbola dan berita sudah sesuai antara judul dan isi berita. Data yang digunakan sudah periksa kembali dengan membaca isi berita, berita yang sesuai dengan judul atau yang tidak sesuai dengan judul, berita yang tidak sesuailah yang nantinya akan mendapatkan label dengan kategori *clickbait*. Proses pelabelan memakan waktu 2 bulan, mulai dari tanggal 16 Maret sampai 16 Mei 2020.

Judul	Kategori
Yasonna Copot Dirjen Imigrasi Ronny Sompie Terkait Kasus Harun Masiku, Ini Respons KPK	C
2 Bukti Ditemukan, Polisi Bekuk Pegawai Kedai Kopi Lempar Susu ke Driver Ojol sampai Berdarah	BC
2 Orang Kembali Dilaporkan Terinfeksi Virus Corona di Thailand	BC
2 Pengasuh PAUD Jadi Tersangka Kasus Mayat Balita Tanpa Kepala, Ayah Korban_ Kami Tak Terlalu Senang	BC
3 Hari Cari Lutfi Alfiandi, Ibunda Ungkap Telpon Terakhir Saat Anak Ditangkap_ Ampun Ampun ke Polisi	BC
3 Minuman Tradisional di Bandung yang Bisa Menghangatkan Tubuh Saat Musim Hujan	BC
3 Restoran Legendaris di Jakarta Pusat Sejak 1960-an, Cocok Bersantap Bersama Keluarga	BC
3 Zodiak Pendiam yang Tidak Suka Membuka Obrolan dengan Orang Baru, Kamu Termasuk_	BC
4 Cara Mudah untuk Mengurangi Kebiasaan Menunda Pekerjaan, Coba Yuk!	BC
4 Fakta Jelang Hasil Autopsi Lina Diumumkan_ Ada Video CCTV Sebelum Lina Meninggal, Kekecewaan Teddy	BC
5 Fakta Pasukan Quds, Pasukan Elit Iran yang Dipimpin Mayjen Qasem Soleimani, Pahlawan atau Teroris	C
5 Fakta Penangkapan Kurir Narkoba di Serpong, dari Kode 'Rahasia' hingga Sabu Senilai 864 Miliar	BC
5 FAKTA TERBARU Kasus Medina Zein_ Akui Tak Beri ASI pada Anaknya, hingga Dukungan Zaskia Sungkar	BC
5 Hantu Populer yang Dikenal Masyarakat Indonesia, dari Tuyul hingga Kuntilanak	BC
5 Wanita Cantik 'Korban' Valentino Rossi, Ada Mantan Istri Musuh Bebuyutannya di MotoGP	BC

Gambar 2 Data Berita Hasil Pelabelan

Setalah dilakukan pelabellan pembagian data data training dan data testing dilakukan dengan 3 skenario rasio yaitu 80%:20%, 50%:50%, dan 20%:80%. Data input pada penelitian menggunakan format ekstensi xlsx.

3.2 *Pre-processing*

Pada tahan *pre-procesing* terdapat beberapa tahap yang harus dilakukan yaitu tahap *case folding, tokening, filtering, dan stemming*. Tahapan ini menggunakan bahasa pemrograman Python pada Google Colab dengan memanfaatkan libray Sastrawi.

1. *Case folding,* tahap ini data berita dirubah menjadi huruf kecil, menghapus angka, menghapus tanda baca dan menghapus spasi.
2. *Tokennizing, tahap ini dilakukan pemecahan kalimat judul berita menjadi kata sehingga setelah tokenizing judul berita terpisah-pisah perkata.*
3. *Filtering*, tahap ini dilakukan penghilangan kata sambung atau kata yang tidak penting dengan menggunakan teknik *removes stopword*. Pada penelitian *library* yang digunakan untuk teknik *removes stopword* adalah *library* Sastrawi.
4. *Stemming, Pada tahap ini mengubah kata imbuhan menjadi kata dasar. proses ini juga menggunakan library yang sama yaitu library Sastrawi*

Tabel 1 Contoh proses *pre-processing*

Data	3 Zodiak Pendiam yang Tidak Suka Membuka Obrolan dengan Orang Baru, Kamu Termasuk
Case Folding	zodiak pendiam yang tidak suka membuka obrolan dengan orang baru kamu termasuk
Tokennizing	zodiac, pendiam, yang, tidak, suka, membuka, obrolan, dengan, orang, baru, kamu, termasuk
Filtering	zodiac, pendiam, suka, membuka, obrolan, orang, baru, kamu, termasuk
Stemming	zodiac, diam, suka, membuka, obrolan, orang, baru, kamu, termasuk

3.3 *Transformation*

Pada tahap ini dilakukan proses *Term Frequency* dilakukan untuk pembobotan kata dengan menggunakan TF-IDF. Pemberian bobot nilai pada tiap kata akan dihitung probabilitasnya, perhitungan pembobotan kata dilakukan dengan cara menentukan nilai *Term Frequency* (TF).

	tfidf
sompie	0.345173
ronny	0.345173
respons	0.312015
dirjen	0.312015
copot	0.301340
...	...
intip	0.000000
investasi	0.000000
ipb	0.000000
iphone	0.000000
zombie	0.000000

3242 rows × 1 columns

Gambar 3 Hasil Pembobotan Kata TF-IDF

3.4 *Data Mining*

Pada tahap ini dilakukan pemodelan dengan *K-Nearest Neighbor* (KNN). Pada pemodelan ini pembagian *data training* dan *data testing* dibagi menjadi 3 skenario yaitu pembagian antara data training berbanding dengan data testing yaitu, 80%:20%, 50%:50%, 20%:80%.

Tabel 2 Pembagian *Data Training* dan *Testing*

Skenario 1	Data Training (80%)	Data Testing (20%)
	800	200
dSkenario 2	Data Training (50%)	Data Testing (50%)
	500	500
Skenario 3	Data Training (20%)	Data Testing (80%)
	200	800

Pengujian KNN pada penelitian ini dilakukan pemodelan dengan menggunakan K-Nearest Neighboar dengan nilai k pada KNN yaitu 1,3,5,7,9,11,13,15. Berikut merupakan hasil dari modek klasifikasi KNN dengan menggunakan data training.

Tabel 3 Tabel Hasil Nilai Akurasi

	Skenario 1	Skenario 2	Skenario 3
1-NN	58%	55%	57%
3-NN	57%	54%	60%
5-NN	66%	58%	58%
7-NN	68%	56%	59%
9-NN	69%	59%	59%
11-NN	71%	57%	56%
13-NN	69%	55%	56%
15-NN	68%	56%	55%

Berdasarkan Tabel 3 dapat dilihat bahwa hasil model KNN untuk nilai akurasi dengan nilai k=11 mendapatkan nilai terbaik pada skenario 1 yang menghasilkan nilai akurasi sebesar 71%.

Tabel 4 Tabel Hasil Nilai *Precision*

	Skenario 1	Skenario 2	Skenario 3
1-NN	58%	55%	57%
3-NN	57%	54%	60%
5-NN	66%	58%	58%
7-NN	68%	56%	59%
9-NN	69%	59%	59%
11-NN	72%	57%	56%
13-NN	69%	55%	56%
15-NN	68%	56%	55%

Berdasarkan Tabel 4 dapat dilihat bahwa hasil model KNN untuk inlai *Precision* dengan nilai k=11 mendapatkan nilai terbaik pada skenario 1 yang menghasilkan nilai akurasi sebesar 72%.

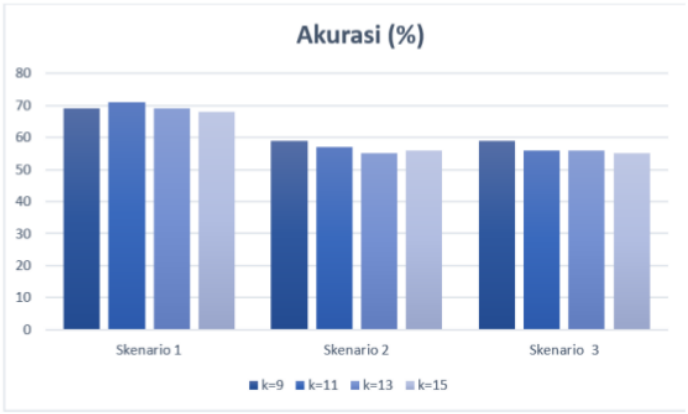
Tabel 5 Tabel Hasil Nilai *Recall*

	Skenario 1	Skenario 2	Skenario 3
1-NN	58%	55%	57%
3-NN	57%	54%	60%
5-NN	66%	58%	58%
7-NN	68%	56%	59%
9-NN	69%	59%	59%
11-NN	71%	57%	56%
13-NN	69%	55%	56%
15-NN	68%	56%	55%

Berdasarkan Tabel 5 dapat dilihat bahwa hasil model KNN untuk inlai *Recall* dengan nilai k=11 mendapatkan nilai terbaik pada skenario 1 yang menghasilkan nilai akurasi sebesar 71%.

3.5 *Evaluation*

Hasil yang diperoleh keseluruhan semua tes yang dilakulukan pemodelan KNN.menghasilkan nilai tingkat akurasi, precision, dan recall. Berikut ini merupakan hasil grafik pemodelan KNN menggunakan 3 skenario.



Gambar 4 Grafik Akurasi Pemodelan KNN

Pada Gambar 4 menunjukkan perbandingan nilai akurasi dengan skenario pembagian dataset. Dari hasil tersebut menunjukkan bahwa jumlah data training mempengaruhi tingkat performa, terlihat pada skenario pembagian dataset 80%-20% dengan nilai k=9 sampai k=15

mendapatkan nilai akurasi tertinggi dengan nilai mencapai 71% dan pada splitting dataset 20%:80% mendapatkan nilai akurasi terendah. Hal ini disebabkan karena semakin banyak data training akan menambah informasi dari karakteristik setiap kelas berita.

4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan dapat disimpulkan Pengklasifikasian berita clickbait dilakukan beberapa proses seperti case folding, tekonizing, filtering, semming, term weight (TF-IDF) dan melakukan pemodelan dengan menggunakan algoritma K-Nearest Neighboar dan melakukan pemilihan nilai K untuk mendapatkan model terbaik. Rasio perbandingan data latih dan uji pada skenario 1 dengan rasio perbandingan sebesar 80%:20% memiliki performa paling tinggi, sehingga dapat disimpulkan bahwa jumlah data latih yang semakin banyak akan menambah ketepatan klasifikasi dikarenakan sistem akan banyak mendapatkan informasi dari data latih. Berdasarkan hasil uji coba parameter pada model pengklasifikasian dengan menggunakan algoritma K-Nearest Neighboar (KNN) menunjukan bahwa model pengklasifikasian terbaik adalah model dengan nilai k=11 pada skenario 1 yang menghasilkan nilai akurasi sebesar 71%, dengan di dapatkannya nilai akurasi terbaik maka dapat diketahui bahwa pengaruh *clickbait* terhadap tingkat pengunjung pada situs berita sangat berperan sehingga situs berita mendapatkan traffic cukup bagus untuk setiap berita-berita nya

5. SARAN

Ada beberapa saran yang perlu diperhatikan untuk dapat melakukan penelitian lebih lanjut agar mendapatkan hasil yang lebih baik lagi. Diantaranya menambahkan jumlah dataset yang digunakan, dengan menambahkan jumlah dataset diharapkan mampu menambah jumlah dan keragaman informasi sehingga dapat menambah informasi dari metode k-nearest neighbors dan meningkatkan performa sistem. Menggunakan algoritma lainya untuk mendapatkan model pengklasifikasian terbaik yang dapat memungkinkan meningkatkan nilai akurasi yang rendah.

DAFTAR PUSTAKA

[1] A. F. Yavi, "Klasifikasi Artikel Berbahasa Indonesia untuk Mendeteksi Clickbait menggunakan Metode Naïve Bayes," *J. Inf. Technol.*, vol. 06, no. 01, pp. 141–147, 2018.

[2] M. Rizky and M. R. Kertanegara, "Penggunaan Clickbait Headline pada Situs Berita dan Gaya Hidup Muslim Dream . co . id," *Mediat. J. Komun.*, vol. 11, no. June 2017, pp. 31–43, 2018.

[3] Y. Yamlean, "Clickbait Journalism Dan Pelanggaran Etika Jurnalistik (Studi Kasus Pelanggaran Etika Jurnalistik Dalam Praktik Clickbait Pada Media Online Jogja.Tribunnews.Com Periode 1 Maret 2019 - 30 April 2019)," Universitas mercu buana yogyakarta, 2019.

[4] L. A. Utami, "Analisis Sentimen Opini Publik Berita Kebakaran Hutan Melalui Komparasi Algoritma Support Vector Machine Dan K-Nearest Neighbor Berbasis Particle Swarm Optimization," *J. Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 103–112, 2017.

[5] S. kartika Lidya, O. S. Sitompul, and S. Efendi, "Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (Svm) Dan K-Nearest Neighbor (K-Nn)," *Semin. Nas. Teknol. Inf. dan Multimed. 2015*, no. maret, pp. 1–8, 2015.

[6] A. A. Irfa and M. S. Mubarok, "Klasifikasi Topik Berita Berbahasa Indonesia Menggunakan k-Nearest Neighbor," *e-Proceeding Eng.*, vol. 5, no. 2, pp. 3631–3640, 2018.

[7] F. N. Rozi and D. H. Sulistyawati, "Klasifikasi Berita Hoax Pilpres Menggunakan

Metode Modified K-Nearest Neighbor Dan Pembobotan Menggunakan Tf-Idf," *KONVERGENSI*, vol. 15, no. Januari, 2019.

[8] S. Nur and K. Fithriasari, "Klasifikasi Berita Online Menggunakan Metode Support Vector Machine dan K- Nearest," *J. Sains dan Seni ITS*, vol. 5, no. 2, pp. 2337–3520, 2016.

[9] A. R. Prasetyo, Idriati, and P. P. Adikara, "Klasifikasi Hoax Pada Berita Kesehatan Berbahasa Indonesia Dengan Menggunakan Metode Modified K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 12, pp. 7466–7473, 2018.

[10] Z. U. Siregar, R. R. A. Siregar, and R. Arianto, "Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor," *J. Kilat*, vol. 8, no. 1, pp. 81–92, 2019.

[11] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, "Klasifikasi Berita Online d engan menggunakan Pembobotan TF-IDF dan Cosine Similarity," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 306–312, 2018.

[12] A. Khoirunnisa, B. Irawan, and R. M Rumani, "Analisis dan implementasi perbandingan algoritma c.45 dengan naïve bayes untuk prediksi penawaran produk," *e-Proceeding Eng.*, vol. 3, no. 3, pp. 5029–5035, 2016.

Klasifikasi Berita Clickbait Menggunakan K-Nearest Neighbor (KNN)

ORIGINALITY REPORT

31%	30%	11%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	id.123dok.com Internet Source	7%
2	eprints.mercubuana-yogya.ac.id Internet Source	3%
3	jurnal.untag-sby.ac.id Internet Source	3%
4	media.neliti.com Internet Source	2%
5	docplayer.info Internet Source	1%
6	stt-pln.e-journal.id Internet Source	1%
7	eprints.uns.ac.id Internet Source	1%
8	www.neliti.com Internet Source	1%
9	sistemasi.ftik.unisi.ac.id	

	Internet Source	1%
10	Riris Aulya Putri, Siti Sendari, Triyanna Widiyaningtyas. "Classification of Toddler Nutrition Status with Anthropometry Calculation using Naïve Bayes Algorithm", 2018 International Conference on Sustainable Information Engineering and Technology (SIET), 2018 Publication	1%
11	jurnal.una.ac.id Internet Source	1%
12	eprints.umm.ac.id Internet Source	1%
13	Made Sudarma, Juli Sulaksono. "Implementation of TF-IDF Algorithm to detect Human Eye Factors Affecting the Health Service System", INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi, 2020 Publication	1%
14	jurnal.iaii.or.id Internet Source	1%
15	slideplayer.info Internet Source	1%

16	etheses.uin-malang.ac.id Internet Source	1%
17	www.journal.unipdu.ac.id:8080 Internet Source	<1%
18	simki.unpkediri.ac.id Internet Source	<1%
19	kimia.fmipa.unand.ac.id Internet Source	<1%
20	ejournal.nusamandiri.ac.id Internet Source	<1%
21	digilib.unila.ac.id Internet Source	<1%
22	prosiding-snastikomti.stth-medan.ac.id Internet Source	<1%
23	jurnalmahasiswa.unesa.ac.id Internet Source	<1%
24	Ruhmi Sulaehani. "PREDIKSI KEPUTUSAN KLIEN TELEMARKETING UNTUK DEPOSITO PADA BANK MENGGUNAKAN ALGORITMA NAIVE BAYES BERBASIS BACKWARD ELIMINATION", ILKOM Jurnal Ilmiah, 2016 Publication	<1%
25	Sukamto Sukamto, Yanti Adriyani, Rizka Aulia. "Prediksi Kelompok UKT Mahasiswa	<1%

Menggunakan Algoritma K-Nearest Neighbor", JUITA: Jurnal Informatika, 2020
Publication

26	Nuranisah, Syahril Efendi, Poltak Sihombing. "Analysis of algorithm support vector machine learning and k-nearest neighbor in data accuracy", IOP Conference Series: Materials Science and Engineering, 2020 Publication	<1%
27	pt.scribd.com Internet Source	<1%
28	khalfiaramadhana2108.blogspot.com Internet Source	<1%
29	Saud Alguwaizani, Byungkyu Park, Xiang Zhou, De-Shuang Huang, Kyungsook Han. "Predicting Interactions between Virus and Host Proteins Using Repeat Patterns and Composition of Amino Acids", Journal of Healthcare Engineering, 2018 Publication	<1%
30	Qurani Dewi Kusumawardani. "Perlindungan Hukum bagi Pengguna Internet terhadap Konten Web Umpan Klik di Media Online", Jurnal Penelitian Hukum De Jure, 2019 Publication	<1%
31	"Pillars, Hypocrites and False Brothers. Paul's Polemic against Jerusalem in Galatians", Walter	<1%

de Gruyter GmbH, 2010

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		