

Penerapan Algoritma K-Nearest Neighbor dalam Identifikasi Berita Hoax Pada Media Sosial

Nanda^a, Dewi Nasien^b, Yulvia Nora Marlim^c Irwan^d

^a Institut Bisnis Dan Teknologi Pelita Indonesia, nanda@student.pelitaindonesia.ac.id

^b Institut Bisnis Dan Teknologi Pelita Indonesia, dewinasien@lecturer.pelitaindonesia.ac.id

^c Institut Bisnis Dan Teknologi Pelita Indonesia, yulvia.noramarlum@lecturer.pelitaindonesia.ac.id

^d Institut Bisnis Dan Teknologi Pelita Indonesia, irwan@lecturer.pelitaindonesia.ac.id

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 25 April 2025

Revisi Akhir: 29 April 2025

Diterbitkan Online: 30 April 2025

KATA KUNCI

Data Maining, Berita Hoax

KORESPONDENSI

E-mail: nanda@student.pelitaindonesia.ac.id

A B S T R A C T

Media sosial telah menjadi platform utama bagi individu untuk berbagi informasi, namun, dengan pertumbuhan pesatnya, juga meningkatkan risiko penyebaran berita palsu atau hoax. Identifikasi berita hoax menjadi tantangan yang signifikan dalam upaya menjaga integritas informasi di media sosial. Dalam skripsi ini, kami mengeksplorasi penerapan algoritma K-Nearest Neighbor (KNN) untuk mengidentifikasi berita hoax pada media sosial. Metode ini memanfaatkan pendekatan pembelajaran mesin tanpa pengawasan untuk mengklasifikasikan berita sebagai hoax atau non-hoax berdasarkan fitur-fitur teks yang diekstraksi dari konten berita. Kami menggunakan dataset yang terdiri dari berbagai berita yang telah diidentifikasi sebagai hoax atau non-hoax untuk melatih dan menguji model KNN. Hasil eksperimen terbaik diperoleh pada skenario pertama dengan rasio (90:10) menggunakan nilai $k=3$ dengan hasil akurasi sebesar 84 %, precision sebesar 85 %, recall sebesar 99 % dan f1-score sebesar 91 %.

1. PENDAHULUAN

Teknologi telah berkembang dengan pesat, Akses internet dan media sosial begitu mudah di akses oleh banyak kalangan terutama kalangan pelajar. Disamping itu juga, dibalik pengaruh penggunaan media sosial bagi kalangan pelajar, terdapat juga sejumlah pengaruh buruk sebagai efek negatif dari pemanfaatan inovasi tersebut, semakin banyak berita palsu yang sering disebut dengan hoax dapat diterima masyarakat terutama di dalam penggunaan media sosial. Faktanya, telepon pintar lengkap dengan fasilitas internetnya saat ini tidak lagi sebatas digunakan sebagai sarana pendukung kerja bagi orang dewasa saja, akan tetapi telah dianggap sebagai kebutuhan yang tidak dapat dilepaskan dari kehidupan sehari-hari bagi semua golongan usia.[1]

Media sosial sebagai suatu wadah yang berguna untuk memudahkan manusia dalam berinteraksi sosial. Ia sebagai suatu cara baru dalam membangun hubungan komunikasi antar individu, untuk bekerja sama dan berdiskusi. Manusia selalu membutuhkan adanya informasi dari manusia yang lain, sebab secara alamiah bahwa pengetahuan manusia itu terbatas. Konten-konten yang disebar oleh pengguna internet ke jejaring sosial dapat menjadi sumber informasi bagi orang yang membutuhkan,

lebih pentingnya lagi bahwa media sosial memiliki keunggulannya yaitu dapat digunakan di manapun dan kapanpun.[2]

Penyebaran informasi pada saat ini banyak dilakukan pada media online. Kemudahan yang ditawarkan dalam penyampaian informasi kepada publik yang disediakan dan dimediasi dalam jaringan membuat informasi atau berita tidak dapat difilter dengan baik. Tidak ada redaksi yang dapat bertanggung jawab atas penyebaran informasi yang tersebar di media online, dikarenakan semua orang yang dapat akses untuk melakukan transaksi data media online dapat melakukan penyebaran informasi. Banyaknya informasi yang bersifat anonymous membuat penyebaran hoax pada media online begitu cepat tersebar.[3]

Berita hoax atau palsu telah menjadi masalah besar di era digital, terutama karena banyaknya informasi yang tersebar di media sosial. Berita palsu dapat dengan cepat mencapai ribuan bahkan jutaan orang di platform seperti Facebook, Twitter, dan WhatsApp. Berita Hoax adalah pemberitaan palsu. Usaha untuk menipu atau mengakali pembaca/pendengarnya agar mempercayai sesuatu padahal sang pembuat berita palsu tersebut tahu bahwa berita tersebut palsu [4]. Tujuan pembuatan berita hoax ini adalah untuk membujuk, memanipulasi, mempengaruhi

pembaca berita untuk melakukan hal-hal yang bertentangan atau mencegah tindakan yang sudah benar.[5]

Hoax tidak menyertakan sumber yang kredibel. data penulis dan kapan tulisan tersebut dipublikasikan pun hampir tidak pernah ada. Beda dengan konten di situs kredibel yang hampir selalu netral, konten hoax biasanya cenderung bias dan condong ke satu pihak.[6]

Metode pembelajaran mesin seperti algoritma K-Nearest Neighbors (K-NN) adalah salah satu metode yang dapat digunakan untuk mengatasi masalah penyebaran berita hoaks. Algoritma K-NN adalah metode klasifikasi yang digunakan untuk menganalisis pola data berdasarkan atributnya.

Tujuan dari penelitian ini adalah untuk mempelajari dan menerapkan algoritma K-NN untuk mengidentifikasi berita hoaks di media sosial. Dengan menggunakan data yang ada di platform media sosial, algoritma ini akan dimodifikasi untuk membuat model yang dapat membedakan antara berita yang benar dan berita hoaks. Hal ini diharapkan dapat membantu masyarakat menjadi lebih kritis dan berhati-hati saat berinteraksi dengan konten di media sosial. Fokus penelitian adalah algoritma K-NN karena kemampuan untuk mengklasifikasikan data berdasarkan karakteristik objek, yang sesuai dengan kebutuhan untuk mengidentifikasi berita hoaks.

Diharapkan bahwa penelitian ini akan meningkatkan pemahaman kita tentang keterbatasan dan efektivitas algoritma K-NN dalam mendeteksi berita hoaks di media sosial. Seperti penelitian yang dilakukan oleh [7], yang memanfaatkan metode penelitian ini Untuk Mengatasi penyebaran informasi yang salah di era internet saat ini, hasil penelitian ini dapat membantu masyarakat menjadi lebih melek digital. Akibatnya, masyarakat diharapkan lebih berhati-hati saat menggunakan media sosial untuk memastikan bahwa informasi yang mereka konsumsi adalah akurat, benar, dan dapat dipercaya.

2. TINJAUAN PUSTAKA

2.1. Text Mining

Text mining adalah satu langkah dari analisis teks yang dilakukan secara otomatis oleh komputer untuk menggali informasi yang berkualitas dari suatu rangkaian teks yang terkandung dalam sebuah dokumen [8]. Text mining, atau text analytics, adalah proses analisis teks yang bertujuan untuk mengekstrak informasi yang bermanfaat dari data teks. Tujuannya adalah menemukan pola, tren, dan pengetahuan baru yang terkandung dalam teks yang tidak terstruktur. Beberapa tahapan dalam text mining meliputi preprocessing teks, ekstraksi fitur, seleksi fitur, dan penemuan pola. Teknik-teknik yang digunakan dalam text mining meliputi pemrosesan bahasa alami (NLP), ekstraksi informasi, klasifikasi dokumen, pengelompokan teks, dan lainnya.

2.2. K-Nearest Neighbor

K-Nearest Neighbor (kNN) adalah suatu metode yang menggunakan algoritma supervised dimana data testing yang baru diklasifikasikan berdasarkan mayoritas kelas pada kNN. Tujuan dari algoritma ini adalah mengklasifikasi objek baru berdasarkan parameter dan data training. Prinsip dari kNN

adalah menemukan K objek dari data training yang paling dekat dengan data testing[9].

Sebelum menggunakan algoritma K-Nearest Neighbor, langkah awal yaitu tentukan data latih dan data uji. Setelah itu dilanjutkan dengan proses perhitungan untuk menghitung jarak dengan rumus Euclidean Distance. Algoritma ini sangat simpel dan mudah diterapkan serta menyerupai metode clustering, yaitu penggolongan data baru berdasarkan jaraknya ke beberapa data atau tetangga terdekat. Pertama, tentukan nilai K tetangga sebelum menghitung jarak data ke tetangga. Kemudian, tentukan jarak antara dua titik, satu titik dalam data latih dan satu titik dalam data uji, menggunakan rumus Euclidean Distance.[10]

2.3. Berita Hoax

Berita adalah laporan tentang rangkaian peristiwa atau gagasan yang dapat menarik perhatian pembaca karena hal yang istimewa atau penting, mencakup aspek manusiawi seperti humor, emosi, dan stres. Berita dipublikasikan tentang peristiwa atau perkembangan spesifik, kredibel, menarik, dan terkini yang dianggap penting bagi publik [11]. Berita dapat berupa laporan mengenai peristiwa politik, sosial, ekonomi, budaya, dan lainnya yang relevan untuk diketahui oleh khalayak umum. Berita yang benar dan akurat sangat penting untuk memberikan pemahaman yang tepat kepada pembaca.

Hoax dapat diartikan sebagai sebuah informasi yang belum pasti sebuah fakta, karena pengertian informasi itu adalah kumpulan dari beberapa data yang bersifat fakta[12]. Hoax adalah usaha untuk menipu atau mengakali pembaca/pendengarnya untuk mempercayai sesuatu, padahal sang pencipta berita palsu tersebut tahu bahwa berita tersebut adalah palsu. Salah satu contoh pemberitaan palsu yang paling umum adalah mengklaim sesuatu barang atau kejadian dengan suatu sebutan yang berbeda dengan barang/ kejadian sejatinya. Berita hoax, atau yang sering disebut sebagai berita palsu, merujuk pada informasi yang sengaja dibuat atau disebarluaskan dengan maksud menyesatkan, menipu, atau memanipulasi pembaca.

Ciri khas dari berita hoax adalah informasi yang tidak benar, tidak akurat, dan tidak dapat dipertanggungjawabkan. Berita hoax sering kali disajikan dalam format yang menarik dan provokatif untuk menarik perhatian pembaca tanpa memperhatikan kebenaran informasi yang disampaikan. Sumber informasi yang tidak jelas atau tidak dapat dipercaya juga seringkali menjadi ciri dari berita hoax. Dampak dari penyebaran berita hoax dapat sangat merugikan, seperti memicu konflik sosial, merusak reputasi seseorang atau lembaga, serta mengganggu stabilitas politik dan sosial suatu negara. Oleh karena itu, penting bagi masyarakat untuk meningkatkan literasi digital dan kritisitas dalam menyaring informasi yang diterima agar dapat mengidentifikasi berita hoax dan tidak terjebak dalam penyebaran informasi yang tidak benar.

2.4. Python

Berita Python adalah sebuah bahasa pemrograman tingkat tinggi yang serbaguna, dinamis, dan mudah dipelajari. Python telah menjadi salah satu bahasa pemrograman yang paling populer di dunia. Keberhasilan Python tidak hanya didasarkan pada sintaksisnya yang bersih dan mudah dibaca, tetapi juga pada kesederhanaan dan kekuatannya yang memungkinkan pengembang untuk menangani berbagai macam tugas pemrograman dengan efisien. Salah satu ciri khasnya adalah

pendekatan yang ramah terhadap pemula, yang memungkinkan mereka untuk mulai membuat program dengan cepat tanpa harus menghadapi kompleksitas sintaksis atau konfigurasi yang rumit. Namun demikian, Python juga menyediakan fleksibilitas dan kekuatan yang cukup untuk menangani proyek-proyek pemrograman yang lebih kompleks.

2.5. Media Sosial

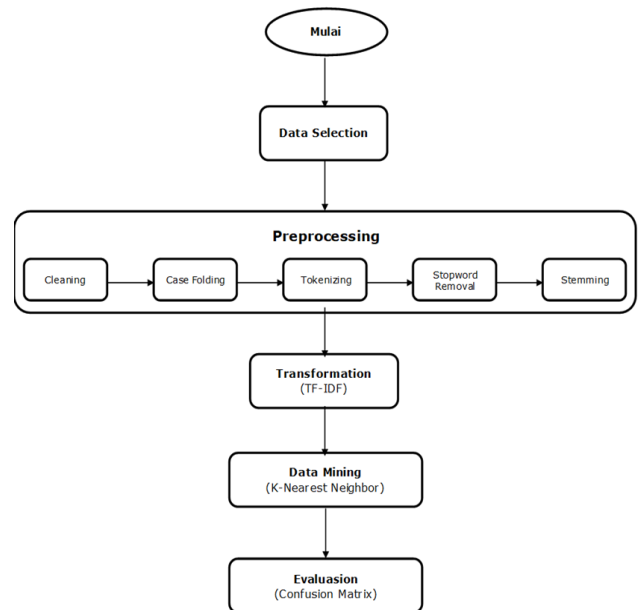
Media sosial adalah sebuah jejaring di internet yang memungkinkan pengguna mengapresiasi dirinya maupun bersosialisasi, bekerja sama, berbagi, berkomunikasi dengan pengguna lain, dan membentuk ikatan sosial atau suatu alat perantara yang digunakan untuk berinteraksi dan berkomunikasi antara seseorang dengan orang lain yang banyak memberikan kontribusi atau manfaat bagi masyarakat.[13]

Media sosial adalah platform online yang memungkinkan pengguna untuk berinteraksi, berbagi informasi, dan menciptakan konten secara kolaboratif. Media sosial memungkinkan pengguna untuk membuat profil, mengunggah foto dan video, serta berkomunikasi dengan pengguna lain melalui pesan pribadi atau komentar. Media sosial memiliki banyak manfaat, seperti memungkinkan pengguna untuk terhubung dengan teman dan keluarga, memperluas jaringan sosial, dan memperoleh informasi terbaru. Media sosial juga memungkinkan pengguna untuk berpartisipasi dalam diskusi publik, mempromosikan bisnis atau produk, serta memperoleh dukungan dari komunitas online. Namun, media sosial juga memiliki beberapa risiko dan tantangan, seperti penyebaran berita hoax atau informasi palsu, penyalahgunaan data pribadi, dan kecanduan media sosial. Selain itu, media sosial juga dapat menjadi tempat berkembangnya perilaku negatif, seperti cyberbullying, trolling, dan hate speech.

3. METODOLOGI

3.1. Kerangka Penelitian

Kerangka penelitian adalah sebuah struktur yang digunakan sebagai landasan untuk pendekatan dalam menyelesaikan masalah penelitian. Biasanya, kerangka penelitian disajikan secara visual melalui diagram atau gambar untuk memberikan gambaran tentang elemen-elemen yang terlibat dalam penelitian dan interaksinya. Seperti yang terlihat pada gambar 3.1.



Gambar 3.1. Kerangka Penelitian

Penjelasan dari gambar 3.1 adalah sebagai berikut:

1. Data Selection.

Data yang sudah didapatkan merupakan dataset yang tersedia dari web “<https://www.kaggle.com/datasets/muhammadghazimuham/indonesiafalsenews>”. Dataset yang digunakan memuat informasi berita hoax dan non hoax dari tahun 2017 sampai dengan 2019. Dataset memuat total sebanyak 4230 record data dengan 6 atribut. Dataset yang diperoleh dari website berupa file berekstensi csv.

2. Preprocessing

Preprocessing data tahap persiapan data mentah sebelum proses selanjutnya. Tujuan dari tahap ini adalah untuk menyatukan struktur kata dan mengurangi jumlah. Pada tahap ini, beberapa proses dilakukan, antara lain :

1). Penyaringan teks

Tahap ini melibatkan penghapusan informasi yang tidak relevan atau tidak diperlukan dalam teks, seperti emotikon, hashtag, URL, dan mention (tag pengguna). Ini membantu fokus pada teks utama yang akan dianalisis.

2). Pengubahan menjadi huruf kecil

Proses ini mengubah semua karakter dalam teks menjadi huruf kecil (lowercase). Tujuannya adalah agar tidak ada perbedaan antara huruf besar dan kecil sehingga kata-kata yang sebenarnya sama dapat dianggap serupa.

3). Pemisahan teks

Proses ini mengubah semua karakter dalam teks menjadi huruf kecil (lowercase). Tujuannya adalah agar tidak ada perbedaan antara huruf besar dan kecil

sehingga kata-kata yang sebenarnya sama dapat dianggap serupa.

4). Penghapusan kata tidak bermakna

Langkah ini melibatkan penghapusan kata-kata umum yang tidak memberikan nilai tambah dalam analisis, seperti kata penghubung, kata depan, dan kata-kata umum lainnya. Misalnya: "ke", "di", "yang", "dari", dsb.

5). Pengubahan kebentuk dasarnya

Proses ini bertujuan untuk mengubah kata-kata ke dalam bentuk dasarnya. Misalnya, kata-kata seperti "berlari", "berlari-lari", dan "lari" diubah menjadi bentuk dasarnya, yaitu "lari". Ini membantu dalam konsistensi analisis karena kata-kata yang bermakna sama diubah ke bentuk yang seragam.

3. Transformation

Pembobotan kata pada penelitian ini menggunakan Metode TF-IDF karena merupakan metode yang terkenal efisien, mudah, dan memiliki hasil yang akurat. Langkah langkah pada proses pembobotan kata menggunakan TF-IDF yaitu Pertama Menghitung kata yang muncul (TF) pada semua dokumen Kedua, Menghitung nilai IDF dengan rumus $IDF = \log N$.

4. Data Mining

Data mining merujuk pada proses ekstraksi pengetahuan yang bermanfaat atau pola yang tersembunyi dari dataset yang besar, kompleks, dan mungkin terstruktur secara acak. KNN (K-Nearest Neighbors) adalah salah satu algoritma dalam data mining yang digunakan untuk klasifikasi dan regresi.

5. Evaluasi

Dalam penelitian ini, Confusion Matrix menggunakan hasil yang mengandung nilai ketepatan, ketepatan, recall, dan skor F1. Pengukuran atau pengujian algoritma dilakukan untuk mengetahui performa dan keoptimalan algoritma.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Pada langkah pemilihan data, judul berita dari media sosial dikumpulkan. Judul berita ini kemudian digunakan sebagai objek penelitian untuk membedakan berita benar dan berita palsu. Data yang diambil merupakan data pada awal tahun 2017 sampai tahun 2020 yang berjumlah 4231 berita yang beredar di media social. Pada awalnya data memiliki empat label yaitu id, tanggal, judul,

dan narasi, tetapi karena yang akan diolah itu hanya 2 atribut agar menjadi lebih efektif dan yang akan diolah yaitu judul dan label.

No	Judul	Label
1	PKS DAN GERINDRA Purwakarta Bantah : 'Pasang Spanduk yang Dinilai Provokatif'	Fakta
2	@#Manajemen Garuda Indonesia Tegaskan Tidak Ada Penjualan atau Penawaran Saham kepada Perusahaan China	Fakta
3	Bukan disebabkan vaksin, siswa MI di Palembang tewas karena radang otak	Fakta
4	Konsumen © Tuntut Produk Quaker Oats Karena Mengandung Pestisida	Fakta
5	Surat suara yang sah harus ada tanda tangan KPPS##	Fakta
6	Modus PenCABulan Dan Pembiusan Oleh PenGamen di Angkutan KoTa Purwakarta @	Hoax
7	Mobil Avanza Masuk Makam di Sukorejo Dikarenakan Hal Mistis	Hoax
8	Segerombolan WNA Cina Berseragam Tentara dan Bersenjata di Kalimantan Utara @	Hoax
9	Mengkonsumsi Sayap dan Ceker AyaM yang Disuntik Hormon Dapat Menyebabkan Kanker	Hoax
10	Makanan Kaleng Produksi Thailand Bisa Tularkan HIV Kembali	Hoax

Tabel 4.1 Dataset

4.2. Preprocessing

Data yang berhasil dikumpulkan masih berbentuk data mentah sehingga dibutuhkan tahap preprocessing agar data tersebut dapat diolah pada tahapan selanjutnya. Pada tahap preprocessing ini dilakukan 5 tahap yaitu Pertama Penyaringan teks, Kedua pengubahan menjadi huruf kecil Ketiga, Pemisahan teks Keempat, Penghapusan kata tidak bermakna dan kemudian pengubahan kebentuk asalnya. hasil dari preprocessing dapat dilihat pada tabel 4.2

No	Judul	Label
1	[pks, gerindra, purwakarta, bantah, pasang, spanduk, nilai, provokatif.]	Fakta
2	[manajemen, garuda, indonesia, tegas, tidak, ada, jual, nawar, saham, pada, usaha, china]	Fakta
3	[Bukan, sebab, vaksin, siswa, mi, Palembang, tewas, karena, radang, otak]	Fakta
4	[konsumen, tuntutan, produk, quaker, oats, karena, mengandung, pestisida]	Fakta
5	[surat, suara, yang, sah, harus, ada, tanda, tangan, kpps]	Fakta

6	[modus, cabul, bius, oleh, ngamen, angkutan, kota, purwakarta]	Hoax
7	[mobil, avanza, masuk, makam, sukorejo, karenakan, hal, mistis]	Hoax
8	[gerombol, wna, cina, rseragam, tentara, senjata, kalimantan, utara]	Hoax
9	[konsumsi, sayap, ceker, ayam, suntik, hormon, dapat, sebabkan, kanker]	Hoax
10	[makanan, kaleng, produksi, thailand, bisa, tularkan, hiv, kembali]	Hoax

Tabel 4.2 Hasil Prerprocessing

4.3. Transformation

Data yang Dalam tahap transformation ini awalnya akan dilakukan split data sesuai dengan skenario yang telah ditentukan, setelah itu dilakukan pembobotan setiap kata menggunakan TF-IDF. Seperti misalnya Yaitu kalimat “PKS dan Gerindra Purwakarta Bantah Pasang Spanduk yang Dinilai Provokatif” jika tidak ada terdapat kalimat yang sesuai dengan Tf-idf maka pembobotannya diberi angka 0.0 Berikut hasil dari transformation dapat dilihat pada gambar 4.3

	Memecah	kompur	kain	layar	kertas	komentar	tumpah	terbuka	jalan	...	Keluar
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
	...	...	...	...	...	...	...	...	...	...	...
120	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
121	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
122	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
123	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
124	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0

Gambar 4.3 Hasil Transformation

4.4. Klasifikasi K-Nearest Neighbor

Dalam klasifikasi ini penulis membagi data uji dan data latih untuk KNN sebanyak 3 percobaan, dan selanjutnya akan mencari hasil terbaik dari ketiga percobaan tersebut. Dalam melakukan klasifikasi dengan KNN digunakan nilai k 1,3,5,7,9 ,dan 11.

1. Percobaan Pertama

Percobaan pertama yaitu membagi data latih dan uji. hasil dari percobaan pertama dengan skenario pembagian data yaitu 90% untuk latih dan 10% untuk uji. Hasil dari pembagian tersebut yaitu didapatkan yaitu 3807 untuk data training dan 424 data testing.

Nilai K	Akurasi	Precesion	Recal	F1-Score
1	76%	85%	86%	86%
3	80%	85%	93%	89%
5	83%	85%	97%	91%
7	83%	85%	97%	91%
9	83%	85%	97%	91%
11	84%	85%	99%	91%

Tabel 4.3 Percobaan Pertama

Pada percobaan pertama ini didapatkan tingkat akurasi klasifikasi KNN dengan tetangga terbaik yaitu 11 dengan akurasi 84%, Precision 85%, Recal 99%, dan F1-Score 91%.

2. Percobaan Kedua

Percobaan kedua yaitu membagi data latih dan uji. hasil dari percobaan pertama dengan skenario pembagian data yaitu 80% untuk latih dan 20% untuk uji. Hasil dari pembagian tersebut yaitu didapatkan yaitu 3384 untuk data training dan 847 data testing.

Nilai K	Akurasi	Precesion	Recal	F1-Score
1	78%	86%	88%	87%
3	82%	85%	95%	90%
5	82%	85%	96%	90%
7	83%	85%	97%	91%
9	84%	85%	98%	91%
11	83%	84%	99%	91%

Tabel 4.3 Percobaan Pertama

Pada percobaan kedua ini didapatkan tingkat akurasi klasifikasi KNN dengan tetangga terbaik yaitu 9 dengan akurasi 84%, Precision 85%, Recal 98%, dan F1-Score 91%.

3. Percobaan Ketiga

Percobaan ketiga yaitu membagi data latih dan uji. hasil dari percobaan pertama dengan skenario pembagian data yaitu 70% untuk latih dan 30% untuk uji. Hasil dari pembagian tersebut yaitu didapatkan yaitu 2961 untuk data training dan 1270 data testing.

Nilai K	Akurasi	Precesion	Recal	F1-Score
1	77%	86%	87%	87%
3	81%	85%	94%	89%
5	82%	85%	96%	90%
7	83%	85%	97%	90%
9	83%	84%	98%	91%

11	83%	84%	98%	91%
----	-----	-----	-----	-----

Tabel 4.3 Percobaan Ketiga

Pada percobaan ketiga ini didapatkan tingkat akurasi klasifikasi KNN dengan tetangga terbaik yaitu 9 dengan akurasi 83%, Precision 84%, Recall 98%, dan F1-Score 91%.

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Untuk mengetahui seberapa baik metode k-nearest neighbor dalam memprediksi berita hoax, pengujian yang telah dilakukan pada bab sebelumnya bahwa hasil terbaik diperoleh pada skenario pertama dengan rasio (90:10) menggunakan nilai k=3 dengan hasil akurasi sebesar 84 %, precision sebesar 85 %, recall sebesar 99 % dan f1-score sebesar 91 %. Sedangkan untuk hasil terendah didapatkan pada skenario (90:10) menggunakan nilai k=1 dengan akurasi sebesar 76 %, precision sebesar 85 %, recall sebesar 86 % dan f1-score sebesar 86 %..

5.2. Saran

Untuk penelitian yang akan dilakukan selanjutnya, ada beberapa saran yaitu:

1. Menambah dataset yang digunakan untuk melakukan penelitian sehingga dapat menghasilkan tingkat akurasi yang lebih akurat.
2. Menggunakan algoritma atau metode klasifikasi yang lain untuk mengetahui apakah ada algoritma yang lebih baik dalam melakukan klasifikasi judul berita hoax.
3. Menerapkan ke dalam sebuah sistem..

DAFTAR PUSTAKA

- [1] Rozi, F. N., & Sulistyawati, D. H. (2019). Klasifikasi Berita Hoax Pilpres Menggunakan Metode Modified K-Nearest Neighbor Dan Pembobotan Menggunakan Tf-Idf. *Konvergensi*, 15(1). <https://doi.org/10.30996/konv.v15i1.2828>
- [2] Singh, G., Yadav, B. K., Singh, B. P., & Sharama, S. (2022). Identification of Fake News Using Machine Learning Approach. *Proceedings - 2022 4th International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2022*, 271–274. <https://doi.org/10.1109/ICAC3N56670.2022.10074374>
- [3] Anita Sindar. (2021). Identifikasi Berita Hoax Menggunakan Fuzzy Logic. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 9(1), 42–46. <https://doi.org/10.21063/jtif.2021.v9.1.42-46>
- [4] Dalle, A., Rantung, V. P., & Munaiseche, C. (2023). Klasifikasi Berita Hoax Pembagian Kuota Internet Menggunakan Algoritma Modified K-Nearest Neighbor. *Linguistik Komputasional*, 6(2), 42–48.
- [5] Hamzah, R. E., & Putri, C. E. (2020). Mengenal dan Mengantisipasi Hoax di Media Sosial pada Kalangan Pelajar. *Jurnal Abdi MOESTOPO*, Vol. 3, No(01), 9–12. <https://journal.moestopo.ac.id/index.php/abdimoestopo/article/viewFile/1361/683>
- [6] Kurniawati, Y. R. (2020). Pertanggungjawaban Pidana Atas Penyebaran Berita Bohong Hoax di media sosial. *Jurnal Ilmiah Ilmu Hukum*, 26, 422–437.
- [7] Nitha Kumala Dewi. (2023). Identifikasi Berita Hoax dengan Menerapkan Algoritma Text Mining. *Journal of Informatics, Electrical and Electronics Engineering*, 2(3), 65–74. <https://doi.org/10.47065/jieec.v2i3.888>
- [8] Nurlaela, L., Dharmalau, A., & Parida, N. T. (2020). Rancangan Sistem Informasi Inventory Barang Berbasis Web (Studi Kasus : CV. Limpoplast). 2(5), 74–90.
- [9] Putra, M. P. R., & Wardani, K. R. N. (2020). Penerapan Text Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial. *JUTIM (Jurnal Teknik Informatika Musirawas)*, 5(1), 63–71. <https://doi.org/10.32767/jutim.v5i1.791>
- [10] Rahmadhany, A., Aldila Safitri, A., & Irwansyah, I. (2021). Fenomena Penyebaran Hoax dan Hate Speech pada Media Sosial. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 3(1), 30–43. <https://doi.org/10.47233/jteksis.v3i1.182>
- [11] Rozi, F. N., & Sulistyawati, D. H. (2019). Klasifikasi Berita Hoax Pilpres Menggunakan Metode Modified K-Nearest Neighbor Dan Pembobotan Menggunakan Tf-Idf. *Konvergensi*, 15(1). <https://doi.org/10.30996/konv.v15i1.2828>
- [12] Singh, G., Yadav, B. K., Singh, B. P., & Sharama, S. (2022). Identification of Fake News Using Machine Learning Approach. *Proceedings - 2022 4th International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2022*, 271–274. <https://doi.org/10.1109/ICAC3N56670.2022.10074374>
- [13] M. Riski Qisthiano, “Perancangan Sistem Informasi Inventaris pada CV Cemerlang Komputer dengan Metode Extreme Programming,” *Din. Inform. J. Ilm. Teknol. Inf.*, vol. 15, no. 1, pp. 1–10, 2023.