

Klasifikasi Berita Hoax Menggunakan Algoritma Naïve Bayes Berbasis PSO

Hegarmanah Muhabatin¹, Candi Prabowo², Irfan Ali^{3*}, Cep Lukman Rohmat⁴, Dita Rizki Amalia⁵

¹Rekayasa Perangkat Lunak; STMIK IKMI CIREBON; Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45131, 0231-490480; e-mail: info.ikmicirebon@gmail.com.

²Rekayasa Perangkat Lunak; STMIK IKMI CIREBON; Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45131, 0231-490480; e-mail: info.ikmicirebon@gmail.com.

³Rekayasa Perangkat Lunak; STMIK IKMI CIREBON; Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45131, 0231-490480; e-mail: info.ikmicirebon@gmail.com.

⁴Rekayasa Perangkat Lunak; STMIK IKMI CIREBON; Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45131, 0231-490480; e-mail: info.ikmicirebon@gmail.com.

⁵Komputerisasi Akuntansi; STMIK IKMI CIREBON; Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45131, 0231-490480; e-mail: info.ikmicirebon@gmail.com.

* Korespondensi: e-mail: irfanaali0.0@gmail.com

Diterima: 8 April 2021; Review: 12 April 2021; Disetujui: 19 April 2021, 156-165

Cara sitasi: Hegarmanah Muhabatin, Candi Prabowo, Irfan Ali, Cep Lukman Rohmat, Dita Rizki Amalia. 2021. Analisa Kepuasan Mahasiswa Terhadap E-Learning Menggunakan Teknik Machine Learning. *Informatics for Educators and Professionals*. Vol (5) No.2:Hal: 156-165

Abstrak : Perkembangan teknologi informasi yang begitu cepat memicu penyebaran informasi hoax melalui internet menjadi tidak terkontrol. Sehingga diperlukan suatu sistem cerdas yang dapat melakukan klasifikasi konten berita hoax yang tersebar melalui media internet. Naïve Bayes merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki akurasi yang tinggi, akan tetapi Naïve Bayes memiliki kekurangan yaitu sangat sensitive dalam pemilihan fitur maka dari itu dibutuhkan metode Particle Swarm Optimization (PSO) untuk meningkatkan hasil akurasi. Proses klasifikasi hoax dapat dilakukan melalui tahap *preprocessing* kemudian pembobotan kata dan dilakukan klasifikasi menggunakan Naïve Bayes. Setelah dilakukan penelitian dengan metode Naïve Bayes dan metode Naïve Bayes berbasis PSO maka hasil yang didapat adalah Naïve Bayes menghasilkan akurasi sebesar 73.64% sedangkan Naïve Bayes berbasis PSO nilai akurasinya sebesar 91,82%. Tujuan dari penelitian ini adalah untuk melihat seberapa besar pengaruh PSO untuk meningkatkan akurasi pada klasifikasi berita hoax pada media sosial menggunakan pengklasifikasi Naïve Bayes. Setelah menggunakan PSO meningkat sebesar 18,18%.

Kata kunci: Hoax, Klasifikasi, Text Mining Naïve Bayes

Abstract :The rapid development of information technology triggered the spread of hoax information through the internet became uncontrolled. So it is necessary an intelligent system that can classify hoax news content spread through internet media. Naïve Bayes is one of the classification algorithms that are simple but have high accuracy, but Naïve Bayes has a deficiency that is very sensitive in the selection of features, therefore it takes a method of Particle Swarm Optimization (PSO) to improve the accuracy results. Hoax classification process can be done through preprocessing stage then word weighting and classification using Naïve Bayes. After research with Naïve Bayes method and Naïve Bayes method based on PSO, the result obtained is Naïve Bayes produces accuracy of 73.64% while Naïve Bayes based PSO accuracy value is 91.82%. The purpose of this study was to see how much influence PSO to improve accuracy on the classification of hoax news on social media using classifier Naïve Bayes. After using PSO the accuracy result increased by 18.18%.

Keywords : Hoax, Classification, Text Mining Nave Bayes

1. Pendahuluan

Menurut Kamus besar Bahasa Indonesia (KBBI) hoax mempunyai arti berita bohong, berita tidak bersumber. (Nurseha, Aji Teguh, Restu Alamsyah, 2019) Hoax adalah informasi sesat dan berbahaya. Berita hoax berpotensi menyesatkan persepsi atau pandangan manusia melalui penyampaian atau penyebaran informasi palsu sebagai suatu kebenaran. Secara garis besar hoax adalah berita yang menyesatkan karena tidak mempunyai sumber yang bisa dipertanggungjawabkan dan bukti yang jelas. Berita hoax sengaja diciptakan oleh segelintir orang untuk memperoleh keuntungan pribadi demi tujuannya tercapai. [1]

Pada masa ini banyak berita hoax tersebar melalui internet. Perkembangan teknologi informasi begitu sangat cepat, berdampak penyebaran informasi melalui internet yang kurang terkontrol, salah satu di dalamnya terdapat informasi dokumen yang mengandung hoax. Di Indonesia jumlah pengguna internet saat ini semakin terus bertambah. Menurut survei yang telah dilakukan oleh lembaga Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada tahun 2017 dijelaskan bahwa penetrasi pengguna internet Indonesia mencapai 143,26 juta jiwa atau 54,56% dari total populasi penduduk Indonesia 262 juta orang. [2] Oleh karena itu diperlukan metode klasifikasi dokumen baik klasifikasi secara manual maupun secara otomatis oleh sistem Untuk membuktikan informasi hoax atau valid yang tersebar di internet.

Pada jurnal yang berjudul "Perbandingan Metode Klasifikasi Berita Hoaks Berbahasa Indonesia Berbasis Pembelajaran Mesin" ditahun 2020 menjelaskan bahwa algoritma naïve bayes dapat mengklasifikasi berita hoax dengan baik. Menggunakan data sebanyak 200 data yaitu, 140 data latih dan 60 data uji. algoritma naïve bayes menghasilkan nilai akurasi sampai 93%. [3]

Sedangkan pada jurnal berjudul "Klasifikasi Berita Hoax Dengan Menggunakan Metode Naïve Bayes" menggunakan data yang diperoleh dari situs turnbackhoax.com. Data berita hoax diambil dari bulan November 2018 sampai dengan Februari 2019 sejumlah 300 data. Terdapat 297 berita berkategori hoax, dan 63 berkategori fakta. Nilai rata-rata berdasarkan dokumen berita hoax dan dokumen berita benar didapat nilai precision 0,896, recall 0.853 dan FMeasure 0.865. [4]

Hoax dapat berarti berita bohong atau berita yang tidak mempunyai sumber. Dari data temuan yang berasal dari databoks.katadata.co.id jumlah sebaran berita hoax masih tinggi di bulan Januari 2019 sampai bulan Desember 2019 walaupun diagram batang naik turun ditiap pergantian bulan tetapi data masih menunjukkan jumlah berita hoax tinggi. Cendekiawan Muslim, Komaruddin Hidayat, dalam acara tersebut mengatakan bahwa hoax sangat berbahaya bagi masyarakat, ibarat narkoba dan pornografi. masyarakat harus sadar, apabila dibiarkan yang hancur masyarakat itu sendiri. Bisa dikatakan hoax merupakan pembunuhan karakter yang bertujuan ingin menjatuhkan dan memanipulasi, dalam konteks agama disebut dengan fitnah. [5] Semakin berkembangnya sosial media, memicu kenaikan penyebaran hoax setiap hari berita baru muncul kepermukaan dengan status belum jelas benar tidaknya kemudian dengan cepat tersampaikan kepada khalayak umum dan dengan mudah didapat hanya melalui smartphone.

Penyebab penyebaran berita hoax masih tinggi sering terjadi karena beberapa hal yaitu keterbukaan informasi dan tingginya konsumsi media sosial, minimnya sikap kritis terhadap informasi yang didapat, mudah menyebarkan berita tanpa melacak kebenaran berita terlebih dahulu, karena kedekatan emosi dan keyakinan pribadi dengan informasi yang disebarkan, dan juga terdapat konflik secara horizontal (penanaman kebencian, peredaran pesan kebencian, dan kecenderungan pada bullying sosial) yang semua itu berdampak buruk bagi masyarakat. Penyebaran yang sangat cepat dan luas harus segera ditangani sedini mungkin agar tidak terjadi keributan karena adanya berita hoax yang menyesatkan pandangan publik. Penelitian ini akan berfokus kepada ketepatan website turnbackhoax.id dalam mengklasifikasi berita hoax dan non hoax yang kemudian di prosentase menggunakan algoritma naïve bayes.

2. Metode Penelitian

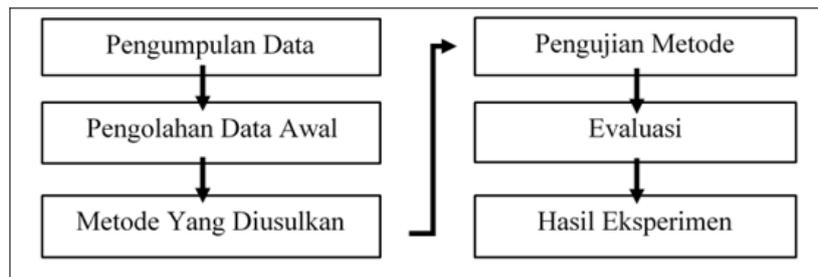
Pada penelitian ini peneliti menggunakan pendekatan deskriptif kualitatif. Pendekatan deskriptif dapat diartikan jenis penelitian dengan tujuan dapat menyajikan keseluruhan gambaran mengenai setting sosial atau dimaksudkan untuk klarifikasi dan eksplorasi berkaitan dengan fenomena atau kenyataan sosial, yaitu menjabarkan variabel yang berkaitan dengan masalah dan sektor yang diteliti antara fenomena yang diuji. dan data kualitatif adalah proses penelitian dan pemahaman berdasar pada metodologi yang menyelidiki suatu fenomena sosial dan masalah masyarakat. Pada pendekatan ini, prosedur penelitian yang menghasilkan data deskriptif berupa perkataan atau text yang tertulis dari orang-orang yang diamati dan perilaku yang diamati.[6]

Adapun teknik pengumpulan data merupakan suatu cara yang dilakukan oleh peneliti untuk memperoleh data-data yang diperlukan. Yang akan dibagi menjadi beberapa bagian yaitu :

Studi dokumen dilakukan dengan mengunjungi website yang menyediakan berbagai macam berita hoax dan non hoax dalam penelitian ini website yang dijadikan studi dokumen adalah <https://turnbackhoax.id/> dan sebagian berita didalamnya akan menjadi dataset didalam RapidMiner.

Jenis data yang digunakan oleh peneliti dalam penyelesaian permasalahan pada penelitian ini ialah data kualitatif.

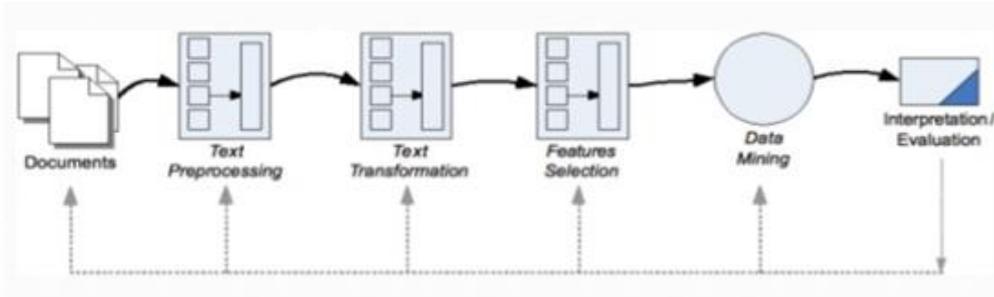
Sumber data dari penelitian ini adalah data kualitatif. Dataset yang digunakan adalah berita dari media sosial yang diperoleh dari <https://turnbackhoax.id/> dimana berita tersebut terdiri dari 55 berita Hoax dan 55 berita non Hoax. Dataset berbentuk sekumpulan teks yang terpisah dan akan diolah dalam bentuk dokumen. Data berita hoax dan non hoax dikumpulkan dan disimpan dengan nama hoax dan non_hoax dalam media teks yaitu notepad dengan ekstensi .txt.



Sumber : Sumber : Han, J. W., & Kamber, M. (2000). Data Mining: Concepts and Techniques
Gambar 1 Langkah Penelitian

Untuk menjelaskan Langkah penelitian, maka terdapat beberapa penjelasan berikut :[1]Pengumpulan Data, Pengumpulan data yang telah diperoleh dari hasil studi dokumen dan menghasilkan data sekunder yang telah dilakukan yaitu meliputi : Telah didapatkan kriteria yang menjadi bahan penelitian dan Telah didapatkan nilai untuk masing – masing kriteria bahan penelitian.[2]Pengolahan data awal, Berdasarkan data yang telah diperoleh, maka dilakukan pengolahan data untuk mendukung penelitian mengenai klasifikasi berita hoax pada media sosial.[3]Metode yang diusulkan, Didasarkan pada hasil pengolahan data menghasilkan analisa data dalam bentuk tabel dan kriteria yang akan digunakan sebagai data pendukung metode yang digunakan di dalam penelitian.[4]Pengujian metode, Berdasarkan data yang diperoleh, maka pengujian metode memiliki mekanisme untuk menentukan data uji yang dapat menguji perangkat lunak secara lengkap dan mempunyai kemungkinan tinggi untuk menemukan suatu kesalahan.[5]Evaluasi, Berdasarkan data yang sudah diperoleh perlu dievaluasi untuk merancang, menyempurnakan, atau menguji jalannya suatu program.[6]Hasil eksperimen, Hasil eksperimen adalah hasil akhir dari sebuah penelitian yang dilakukan, mencakup kekurangan dan kelebihanannya.

Metode text mining yang digunakan untuk menganalisis dalam penerapan data mining menerapkan tahapan yang terdiri dari Document, Text Preprocessing, Text Transformation, Feature Selection, Data Mining, Evaluation



Sumber : Han, J. W., & Kamber, M. (2000). Data Mining: Concepts and Techniques

Gambar 2 Tahapan Text Mining

Variabel penelitian pada dasarnya adalah segala sesuatu yang berbentuk apa saja yang ditetapkan oleh peneliti untuk dipelajari sehingga diperoleh informasi tentang hal tersebut, kemudian ditarik kesimpulannya. Menurut hubungan antara satu variabel dengan variabel yang lain, maka variabel dalam penelitian dapat dibedakan menjadi :

Variabel bebas (independent variabel) adalah variabel yang mempengaruhi atau yang menjadi sebab perubahan atau timbulnya variabel terikat (dependen variabel). Variabel bebas yang digunakan pada penelitian ini adalah text minning (X)

Variabel terikat (dependent variabel), sering disebut sebagai variabel output, kriteria konsekuen. Dalam bahasa Indonesia sering disebut sebagai variabel terikat. Variabel terikat merupakan variabel yang dipengaruhi atau menjadi akibat, karena adanya variabel bebas, dalam penelitian ini yang menjadi variabel terikat adalah berita hoax(Y).

Operasional variabel merupakan indikator yang digunakan dalam penyusunan penelitian. Dalam penelitian ini terdapat dua variabel yang diteliti, yaitu text mining (X), dan beritahoax (Y). Untuk melakukan pengolahan data, diperlukan unsur lain yang berhubungan dengan variabel seperti konsep variabel, dimensi, indikator, ukuran, dan skala dimana variabel penelitian akan diukur dengan skala ordinal. Operasionalisasi variabel untuk penelitian ini adalah sebagai berikut :

Tabel 1 Indikator Variabel

No.	Variabel	Indikator
1	Text mining (X) adalah proses ekstraksi pola berupa informasi dan pengetahuan yang berguna dari sejumlah besar sumber data teks, seperti dokumen Word, PDF, kutipan teks, dan lain-lain.	Dokumen Text processing Text transpormation
2	Berita hoax (Y) berasal dari bahasa Inggris artinya tipuan, menipu, berita bohong ,berita palsu atau kabar burung. Dapat disimpulkan bahwa Berita Hoax adalah kalimat yang berarti ketidak benaran suatu informasi. Hoax bukan singkatan tetapi satu kata dalam bahasa inggris yang punya arti sendiri. (Sutantohadi 2018) Sedangkan definisi Hoax menurut wikipedia adalah: "Sebuah pemberitaan palsu adalah usaha untuk menipu atau mengakali pembaca/pendengarnya untuk mempercayai sesuatu, padahal sang pencipta berita palsu tersebut tahu bahwa berita tersebut palsu."	Mencurigakan Mengandung sara Provokasi Fitnah

Sumber : Hasil Penelitian (2021)

3. Hasil dan Pembahasan

Setelah data dikumpulkan dan disimpan dalam notepad dengan ekstensi .txt. langkah selanjutnya dalam penelitian ini adalah melakukan preprocessing atau pengolahan awal data melalui tiga tahapan proses yaitu sebagai berikut:

Case Folding

Case folding yaitu menjadikan keseluruhan huruf dalam sebuah dokumen menjadi huruf kecil yaitu hanya huruf 'a' sampai huruf 'z' yang akan diproses. Karakter selain huruf dihilangkan dan dianggap delimiter.

Tabel 1 Proses Case Folding

Text sebelum dilakukan Case Folding	Text sesudah dilakukan Case Folding
Wilayah NKRI Dikencingi Drone China. Menhan : China Negara Sahabat !!	wilayah nkri dikencingi drone china menhan china negara sahabat

Sumber : Hasil Penelitian (2021)

Tokenization

Tokenization yaitu mengumpulkan semua kata dan menghilangkan tanda baca maupun simbol yang bukan huruf, seperti “ , . / . ;) dan lain-lain. Contoh hasil pengolahan tokenization pada RapidMiner dapat dilihat pada tabel dibawah ini:

Tabel 2 Perbandingan sebelum dan sesudah dilakukan proses Tokenization

Text sebelum dilakukan Tokenization	Text sesudah dilakukan Tokenization
Wilayah NKRI Dikencingi Drone China. Menhan : China Negara Sahabat !!	wilayah nkri dikencingi drone china menhan china negara sahabat

Sumber : Hasil Penelitian (2021)

Stopword Removal

Stopwords adalah menghilangkan kata-kata umum yang tidak memiliki makna atau informasi yang diperlukan.

Tabel 3 Proses Stop Removal

Text sebelum dilakukan Tokenization	Text sesudah dilakukan Tokenization
Wilayah NKRI Dikencingi Drone China. Menhan : China Negara Sahabat !!	wilayah nkri dikencingi drone menhan china negara Sahabat

Sumber : Hasil Penelitian (2021)

Stemming

Stemming yaitu proses untuk mengubah token yang berimbuhan menjadi kata dasar, dengan meniadakan semua imbuhan yang ada pada token tersebut. Proses stemming penting dilakukan untuk menghilangkan imbuhan pada awalan dan akhiran.

Tabel 4 Proses Stemming

Text sebelum dilakukan Tokenization	Text sesudah dilakukan Tokenization
Wilayah NKRI Dikencingi Drone China. Menhan : China Negara Sahabat !!	wilayah nkri kencing drone menhan china negara Sahabat

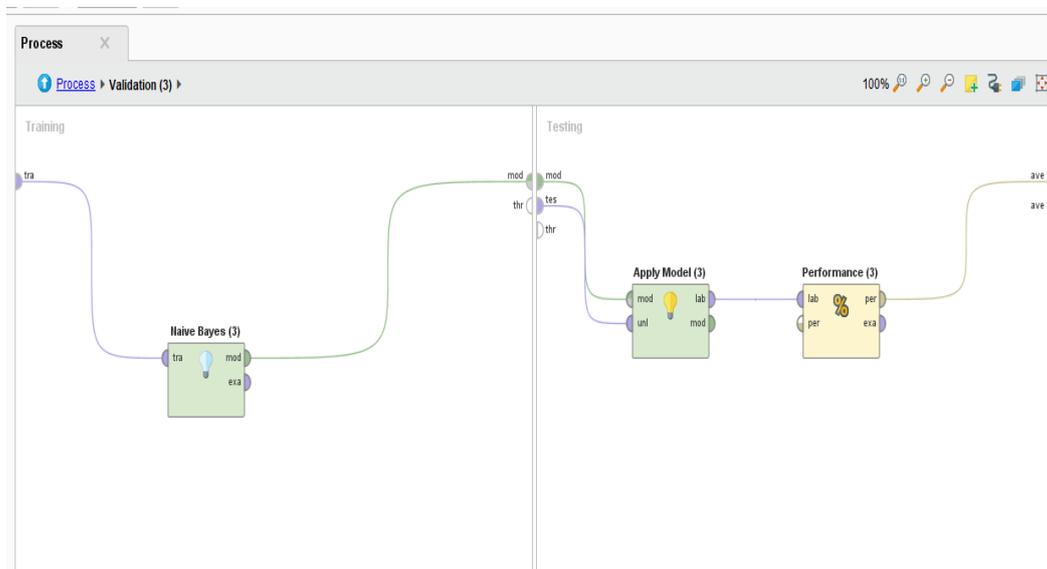
Sumber : Hasil Penelitian (2021)

Preprocessing pada rapid miner dapat dilihat pada gambar 3.



Sumber : Hasil Penelitian (2021)

Gambar 3 Proses pemanggilan operator validation



Sumber : Hasil Penelitian (2021)

Gambar 4 Model Pengujian Validasi Naïve Bayes

Untuk data berita hoax, 40 diklasifikasikan kedalam berita hoax dan 15 data diprediksi berita hoax ternyata hasil prediksinya merupakan berita non hoax. Untuk data berita non hoax, 41 diklasifikasikan sebagai berita non hoax dan 14 data diprediksi kedalam berita non hoax ternyata hasil prediksinya merupakan berita hoax. Hasil klasifikasi menggunakan Naïve Bayes dapat dilihat pada gambar 5.

accuracy: 73.64% +/- 14.35% (mikro: 73.64%)

	true Hoax	true Non_Hoax	class precision
pred. Hoax	40	14	74.07%
pred. Non_Hoax	15	41	73.21%
class recall	72.73%	74.55%	

Sumber : Hasil Penelitian (2021)

Gambar 5 Confix

$$1. \text{ Accuracy} = \frac{\text{TN}+\text{TH}}{\text{TN}+\text{FN}+\text{TP}+\text{FP}} \dots\dots\dots(1)$$

Accuracy = 0,7364 (73,64%)

$$2. \text{ Sensitivity} = \frac{\text{TH}}{\text{TH}+\text{FN}} \dots\dots\dots(2)$$

Sensitivity = 0,7321 (73,21%)

$$3. \text{ Specificity} = \frac{\text{TN}}{\text{TN}+\text{FH}} \dots\dots\dots(3)$$

Specificity = 0,7407 (74,07%)

$$4. \text{ PPV} = \frac{\text{TH}}{\text{TH}+\text{FH}} \dots\dots\dots(4)$$

PPV = 0,7455 (74,55%)

$$5. \text{ NPV} = \frac{\text{TN}}{\text{TN}+\text{FN}} \dots\dots\dots(5)$$

NPV = 0,7273 (72,73%)

Pengujian Algoritma Naïve Bayes berbasis Particle Swarm Optimization (PSO), nilai training dalam penelitian ini ditentukan dengan cara melakukan uji coba yaitu memasukkan nilai parameter Population Size dan Inertia Weight. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk menentukan nilai training.

Tabel 5 Eksperimen Penentuan Nilai Training Naïve Bayes berbasis PSO

Population Size (Q)	Inertial Weight (W)	Accuracy
4	1.0	87,27%
5	1.0	91,82%
6	1.0	86,36%
7	1.0	86,36%
8	1.0	88,18%

Sumber : Hasil Penelitian (2021)

Hasil terbaik dari eksperimen Naïve Bayes berbasis Particle Swarm Optimization (PSO) adalah dengan nilai Population Size sebesar 5 dan nilai Inertia Weight sebesar 1.0, Accuracy yang dihasilkan adalah sebesar 91,82%. Pada penelitian ini hasil klasifikasi berita hoax dan non

hoax pada media sosial menggunakan metode Naïve Bayes berbasis Particle Swarm Optimization (PSO)

Untuk data berita hoax, 48 diklasifikasikan kedalam berita hoax dan 7 data diprediksi berita hoax ternyata hasil prediksinya merupakan berita non hoax. Untuk data berita non hoax, 53 diklasifikasikan sebagai berita non hoax dan 2 data diprediksi kedalam berita non hoax ternyata hasil prediksinya merupakan berita hoax. Hasil Pengujian Model Naïve Bayes berbasis PSO dapat dilihat pada gambar 6.

accuracy: 91.82% +/- 4.90% (mikro: 91.82%)

	true Hoax	true Non_Hoax	class precision
pred. Hoax	48	2	96.00%
pred. Non_Hoax	7	53	88.33%
class recall	87.27%	96.36%	

Sumber : Hasil Penelitian (2021)

Gambar 6 Model Confusion Matrix untuk Metode Naïve Bayes Berbasis PSO

$$1. \text{ Accuracy} = \frac{\text{TN}+\text{TH}}{\text{TN}+\text{FN}+\text{TP}+\text{FP}} \dots\dots\dots(6)$$

$$\text{Accuracy} = 0,9182 \text{ (91,82\%)}$$

$$2. \text{ Sensitivity} = \frac{\text{TH}}{\text{TH}+\text{FN}} \dots\dots\dots(7)$$

$$\text{Sensitivity} = 0,8833 \text{ (88,33\%)}$$

$$3. \text{ Specificity} = \frac{\text{TN}}{\text{TN}+\text{FH}} \dots\dots\dots(8)$$

$$\text{Specificity} = 0,96 \text{ (96\%)}$$

$$4. \text{ PPV} = \frac{\text{TH}}{\text{TH}+\text{FH}} \dots\dots\dots(9)$$

$$\text{PPV} = 0,9636 \text{ (96,36\%)}$$

$$5. \text{ NPV} = \frac{\text{TN}}{\text{TN}+\text{FN}} \dots\dots\dots(10)$$

$$\text{NPV} = 0,8727 \text{ (87,27\%)}$$

Keterangan :

TH: True Hoax

TN: True NonHoax

FH: False Hoax

FN: False NonHoax

PPV : Positive Predictive Value

NPV: Negative Predictive Value

Hasil evaluasi pada penelitian ini menunjukkan bahwa Particle Swarm Optimization dapat meningkatkan nilai akurasi pada Naive Bayes yang merupakan metode yang baik sebagai pengklasifikasi text berita hoax dan non hoax pada media sosial untuk membedakan mana berita yang benar dan mana berita bohong atau hoax.

Untuk memperkuat hasil penelitian, dibuatkan perbandingan dengan algoritma lain yaitu algoritma SVM, bisa dilihat pada gambar dibawah ini:

accuracy: 79.09% +/- 11.54% (mikro: 79.09%)

	true Hoax	true Non_Hoax	class precision
pred Hoax	53	21	71.62%
pred Non_Hoax	2	34	94.44%
class recall	96.36%	61.82%	

Sumber : Hasil Penelitian (2021)

Gambar 7 Model Confusion Matrix untuk Metode SVM dengan penambahan PSO

Untuk lebih mudah dipahami, lihat tabel perbandingan Naïve Bayes dan SVM di tabel 6.

Tabel 7 Tabel Perbandingan Naive Bayes dan SVM

Metode yang dipakai	Akurasi Sebelum PSO	Akurasi Sesudah PSO	Kenaikan Akurasi
Naïve Bayes	73,64%	91,82%	18,18%
SVM	70,00%	79,09%	9,09%

Sumber : Hasil Penelitian (2021)

Dari tabel 6 dapat dilihat untuk metode support vectore machine tanpa penambahan PSO, prosentase yang didapat hanya 70% dan untuk metode support vectore machine dengan penambahan PSO, prosentase yang didapat meningkat sebesar 79,09%, jika dibandingkan dengan metode naïve bayes, eksperimen awal tanpa penambahan PSO dilihat pada tabel memiliki prosentase 73,64% ada selisih 3,64% dengan metode support vectore machine dan untuk eksperimen dengan penambahan PSO didapatkan prosentase sebesar 91,82% ada selisih prosentase sebesar 12,73%.

Hasil penelitian ini menunjukkan bahwa menambahkan Particle Swarm Optimization (PSO) baik pada metode support vectore machine (SVM) ataupun Naïve Bayes dapat meningkatkan nilai akurasi, akan tetapi kenaikan akurasi PSO secara signifikan ditunjukkan pada eksperimen menggunakan metode Naïve Bayes yaitu sebesar 18,18% dibandingkan dengan SVM yaitu sebesar 9,09% hal ini menunjukkan bahwa Algoritma Naïve Bayes dengan penambahan PSO cukup baik digunakan sebagai pengklasifikasi text untuk membedakan berita hoax dan non hoax.

4. Kesimpulan

Penelitian ini peneliti menggunakan pengklasifikasi teks berita hoax dan non hoax pada media sosial dengan klasifikasi Algoritma Naïve Bayes. Penelitian ini menggunakan dataset sebanyak 110 dataset dengan rincian 55 data berita hoax dan 55 data berita non hoax yang diperoleh pada situs turnbackhoax.id. Algoritma Naïve Bayes dapat berfungsi dengan baik sebagai pengklasifikasi text. Akurasi Naïve Bayes sebelum menggunakan Particle Swarm

Optimization (PSO) yaitu sebesar 73.64% setelah menggunakan penggabungan metode Naïve Bayes dan Particle Swarm Optimization (PSO), akurasi meningkat setelah beberapa kali melakukan eksperimen, peningkatannya yaitu menjadi 91,82%. terjadi peningkatan akurasi sebesar 18,18%. Sehingga Algoritma Naïve Bayes berbasis Particle Swarm Optimization (PSO) untuk klasifikasi berita hoax pada media sosial terbukti memberikan nilai akurasi yang lebih akurat.

Referensi

- [1]. O. Nurdiawan, F. A. Pratama, D. A. Kurnia, Kaslani, and N. Rahaningsih, "Optimization of Traveling Salesman Problem on Scheduling Tour Packages using Genetic Algorithms," *J. Phys. Conf. Ser.*, vol. 1477, no. 5, pp. 0–6, 2020, doi: 10.1088/1742-6596/1477/5/052037.
- [2]. DosenPPKN, "https://dosenppkn.com/pengertian-hoax/," *DosenPPKN*, 2018. .
- [3]. H. Mustofa and A. A. Mahfudh, "Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes," *Walisongo J. Inf. Technol.*, vol. 1, no. 1, p. 1, 2019, doi: 10.21580/wjit.2019.1.1.3915.
- [4]. O. Nurdiawan, F. A. Pratama, D. A. Kurnia, Kaslani, and N. Rahaningsih, "Optimization of Traveling Salesman Problem on Scheduling Tour Packages using Genetic Algorithms," *J. Phys. Conf. Ser.*, vol. 1477, no. 5, 2020, doi: 10.1088/1742-6596/1477/5/052037.
- [5]. Fitriani, S., Informatika, T., Teknik, F., & Sidoarjo, U. M. (2019). *Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur*. 3(2), 98–104.
- [6]. R. Firmansyah, "Web Klarifikasi Berita untuk Meminimalisir Penyebaran Berita Hoax," in *Jurnal Informatika*, 2017.
- [7]. M. Athaillah, Y. Azhar, and Y. Munarko, "Perbandingan Metode Klasifikasi Berita Hoaks Berbahasa Indonesia Berbasis Pembelajaran Mesin," *J. Repos.*, vol. 2, no. 5, p. 675, 2020, doi: 10.22219/repositor.v2i5.692.
- [8]. O. Nurdiawan, D. A. Kurnia, D. Solihudin, T. Hartati, and T. Suprpti, "Comparison of the K-Nearest Neighbor algorithm and the decision tree on moisture classification," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012031, 2021, doi: 10.1088/1757-899x/1088/1/012031.
- [9]. Kominfo, "Jumlah Pengguna Internet 2017 Meningkat, Kominfo Terus Lakukan Percepatan Pembangunan Broadband," *Kementrian Komun. dan Infotmatika*, 2018.
- [10] O. Nurdiawan, I. Ali, C. L. Rohmat, and A. R. Rinaldi, "InfoTekJar: Jurnal Nasional Informatika dan Sistem Kendali Sensor Tanah Sebagai Pemonitor Tingkat Kelembaban Media Tanam Padi," *Nas. Inform. dan Teknol. Jar.*, vol. 1, pp. 0–4, 2020.