

Klasifikasi Berita Kriminal Menggunakan Algoritma Naïve Bayes Berbasis PSO

Dzaffa 'Ulhaq ^{1*}, Nana Suarna ¹, Gifthera Dwilestari ²

¹ Teknik Informatika; STMIK IKMI CIREBON; Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45131, 0231-490480; e-mail:

dzaffaulhaq11mia3@gmail.com, nana.ikmi@gmail.com

² Sistem Informasi; STMIK IKMI CIREBON; Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45131, 0231-490480; e-mail: gqdwilestari@gmail.com

* Korespondensi: e-mail: dzaffaulhaq11mia3@gmail.com

Diterima: 7 Juni 2022 ; Review: 14 Juni 2022; Disetujui: 27 Juni 2022

Cara sitasi: Dzaffa 'Ulhaq, Nana Suarna, Gifthera Dwilestari 2022. Klasifikasi Berita Kriminal Menggunakan Algoritma Naïve Bayes Berbasis PSO. Informatics for Educators and Professionals. 6 (2) : 12 - 21.

Abstrak: Perkembangan teknologi informasi yang begitu cepat memicu penyebaran informasi kriminal melalui internet menjadi tidak terkontrol. Sehingga diperlukan suatu sistem cerdas yang dapat melakukan klasifikasi konten berita kriminal yang tersebar melalui media internet. Penelitian ini mengklasifikasikan berita kriminal berdasarkan subkategorinya yang terbagi menjadi 2 yaitu pembunuhan dan narkoba. Berdasarkan klasifikasi berita kriminal tersebut maka akan dapat menekan terjadinya tindakan kriminal yang lebih buruk lagi. Tujuan dari penelitian ini adalah ingin melakukan klasifikasi berita kriminal menggunakan algoritma Naïve Bayes berbasis Particle Swarm Optimization. Metode yang digunakan dalam penelitian ini adalah Naïve Bayes yang merupakan salah satu algoritma klasifikasi yang sederhana akan tetapi Naïve Bayes memiliki kekurangan yaitu sangat sensitive dalam pemilihan fitur maka dari itu dibutuhkan metode Particle Swarm Optimization untuk meningkatkan hasil akurasi. Proses klasifikasi kriminal dapat dilakukan melalui tahap preprocessing kemudian pembobotan kata dan dilakukan klasifikasi menggunakan naïve bayes. Hasil akurasi dari klasifikasi berita kriminal diperoleh sebesar 93.33%. Berdasarkan hasil klasifikasi tersebut maka dapat digunakan sebagai dasar penetapan berita kriminal yang valid. Sehingga masyarakat lebih mudah untuk mengetahui informasi yang terdapat dalam suatu berita dan dapat digunakan sebagai dasar sumber informasi yang dipercaya dalam bermasyarakat.

Kata kunci: kriminal, text minning naïve bayes, klasifikasi,

Abstract: The rapid development of information technology has triggered the spread of criminal information through the internet to be out of control. So we need an intelligent system that can classify criminal news content spread through internet media. This study classifies criminal news based on its subcategories which are divided into 2, namely murder and drugs. Based on the classification of criminal news, it will be able to suppress the occurrence of even worse criminal acts. The purpose of this study is to classify criminal news using the Naïve Bayes algorithm based on Particle Swarm Optimization. The method used in this study is Nave Bayes which is a simple classification algorithm, but Naive Bayes has a disadvantage, namely it is very sensitive in the selection of features, therefore Particle Swarm Optimization method is needed to improve accuracy results. The criminal classification process can be carried out through the preprocessing stage, then word weighting and classification using nave Bayes. The results of the accuracy of the classification of criminal news obtained by 93.33%. Based on the results of the classification, it can be used as a basis for determining valid criminal news. So that it is easier for people to find out the information contained in a news and can be used as a basis for trusted sources of information in society.

Keywords: crime, text mining naïve bayes, classification,

1. Pendahuluan

Menurut Kamus Besar Bahasa Indonesia (KBBI) kriminal mempunyai arti segala hal yang berkaitan dengan kejahatan (pelanggaran hukum) yang kemudian dapat dihukum menurut undang – undang yang berlaku. Tindak kriminal atau pidana merupakan sesuatu yang melanggar hukum atau sebuah tindak kejahatan. Pelaku kriminalitas dinamakan seorang kriminal. Terpidana atau narapidana adalah sebutan bagi seorang pelaku kriminal yang dinyatakan mempunyai kesalahan oleh pengadilan dan harus menerima hukuman. Pertumbuhan teknologi informasi yang begitu cepat, mengakibatkan penyebaran informasi melalui internet menjadi tidak terkontrol termasuk didalamnya berita – berita kriminal. Menurut data yang dihimpun internetworldstat, Indonesia menjadi Negara nomor 3 di Asia sebagai negara pengguna internet terbanyak dengan mencapai 212,35 juta jiwa pengguna internet pada maret 2021[1]. Oleh karenanya dibutuhkan metode klasifikasi dokumen agar dapat membedakan jenis berita kriminal yang tersebar di internet.

Berdasarkan jurnal yang berjudul “Klasifikasi Berita Kriminal Menggunakan Naïve Bayes Classifier (NBC) Dengan Pengujian K – Fold Cross Validation” Pada penelitian ini, metode yang dipakai ialah melakukan klasifikasi dengan menggunakan algoritma naïve bayes serta melakukan pengujian k - fold cross validation dan menghasilkan total rata - rata nilai precision bernilai 98,53%, nilai accuracy bernilai 99,38% dan nilai recall bernilai 98,44%. [2]

Menurut jurnal “Penerapan Algoritma Naive Bayes Dan Particle Swarm Optimization Untuk Klasifikasi Berita Hoax Pada Media Sosial”. Algoritma yang digunakan pada penelitian ini yaitu algoritma naïve bayes dan juga algoritma naïve bayes berbasis Particle Swarm Optimization. Hasil yang didapat yaitu tingkat akurasi menggunakan hanya algoritma naïve bayes bernilai 74.67 %. Di sisi lain, dengan menggunakan algoritma naïve bayes berbasis Particle Swarm Optimization menghasilkan akurasi bernilai 85.19%. Tujuan peneliti ialah untuk mengetahui seberapa besar pengaruh PSO terhadap keakuratan klasifikasi yang dihasilkan menggunakan algoritma naïve bayes[3].

Pada jurnal lain yang berjudul “Penerapan Naïve Bayes Classifier Untuk Klasifikasi Berita Kriminal Di Kalimantan Selatan” Peneliti menggunakan naïve bayes sebagai metode klasifikasi berita kriminal dengan 5 kategori yaitu : pembunuhan, pencurian, narkoba, korupsi dan pemerkosaan.. Hasil yang didapat penulis dari 180 data yang diuji ialah nilai rata –rata akurasi sebesar 85.56% dengan nilai akurasi tertinggi pada kategori korupsi sebesar 94.44% dan akurasi terendah pada kategori pemerkosaan yaitu sebesar 63.89%[4]

. Berita kriminal selalu hangat diperbincangkan baik secara tidak langsung (media sosial) ataupun langsung (media cetak). Tujuan dari penelitian ini adalah agar dapat mengklasifikasikan berita kriminal berdasarkan subkategorinya yaitu pembunuhan dan narkoba. Dengan demikian, dapat memberikan informasi yang tepat mengenai berita yang diterima oleh masyarakat.

2. Metode Penelitian

Dalam penelitian ini, penulis menggunakan jenis penelitian kuantitatif. Metode penelitian kuantitatif merupakan sebuah cara yang dipakai peneliti untuk menyelesaikan permasalahan yang ada dalam penelitian yang berhubungan dengan data baik itu berupa program statistik maupun berupa angka. Peneliti menggunakan Algoritma naïve bayes classifier untuk mengklasifikasikan berita criminal yang diperoleh dari media online berdasarkan subkategori yaitu narkoba dan pembunuhan. Kemudian dioptimasi menggunakan particle swarm optimization agar dapat meningkatkan nilai akurasi data. Data yang diperoleh kemudian diolah menggunakan aplikasi RapidMiner.

Secara umum teorema Bayes dapat dinotasikan sebagai berikut :

$$P(A|B) = \frac{P(A)P(A|B)}{P(B)} \dots\dots\dots(1)$$

Keterangan :

P(A) = robabilitas terjadinya A

P(B) = probabilitas terjadinya B

P(A|B) = probabilitas terjadinya A dengan syarat B

Dalam tahap klasifikasi, dari semua dokumen yang dijadikan testing algoritma akan melakukan pencarian berdasarkan probabilitas yang paling tinggi. Berikut persamaannya adalah :

$$V_{\text{map}} = \frac{\text{argmax}}{v_j \in V} P(v_j) \prod_i P(a_i | v_j) \dots\dots\dots(2)$$

Penjelasan :

V_{MAP} = Probabilitas tertinggi dari semua kategori

$P(v_j)$ = Probabilitas terjadinya kategori tertentu dari sekumpulan data

$P(a_i | v_j)$ = Probabilitas terjadinya kata a_i pada suatu dokumen dengan kategori v_j

$P(v_j)$ atau probabilitas terjadinya kategori tertentu dari sekumpulan data (Prior probability) yang diinput dapat dihitung menggunakan persamaan :

$$P(v_j) = \frac{|doc_j|}{|training|} \dots\dots\dots(3)$$

Penjelasan :

doc_j = Jumlah dokumen atau text yang memiliki kategori j

training = Jumlah dokumen atau text dalam contoh yang dipakai untuk dijadikan training.

Untuk hasil dari suatu probabilitas kata yang menunjukkan kecenderungan pada kategori tertentu dalam sebuah dokumen atau text dapat dihitung menggunakan persamaan :

$$P(a_i | v_j) = \frac{|n_i| + 1}{|n| + |kosakata|} \dots\dots\dots(4)$$

Penjelasan :

n_i = Total kemunculan kata a_i pada dokumen yang berkategori v_j

n = Banyaknya seluruh kata dalam dokumen yang berkategori v_j

kosakata = Banyaknya kata dalam contoh pelatihan

Teknik pengumpulan data yang digunakan peneliti yaitu dengan melakukan studi dokumen. Peneliti melakukan studi dokumen dengan mengunjungi situs atau website yang menyediakan berbagai jenis berita kriminal.

Data yang dibutuhkan peneliti yaitu berasal dari media online dan diperoleh peneliti melalui website <https://jpnn.com/kriminal>. Data yang digunakan yakni berupa 120 jenis berita kriminal. Dari 120 berita tersebut, terbagi menjadi 55 berita kriminal dengan kategori pembunuhan serta 65 berita kriminal dengan kategori narkoba.

Dalam penelitian ini, penulis menerapkan metode eksperimen sebagai metode analisis data. Metode ini dipakai dikarenakan kesamaan data dengan metode pengklasifikasian teks yang paling relevan dan sudah dilakukan oleh banyak peneliti sebelumnya. Metode yang dipakai ialah algoritma Naïve Bayes serta algoritma Particle Swarm Optimization sebagai optimasi untuk meningkatkan akurasi. Peneliti menggunakan aplikasi RapidMiner untuk melakukan eksperimen. RapidMiner dipilih sebagai media pengolah data sebagai alat bantu untuk mengukur tingkat akurasi dari eksperimen.

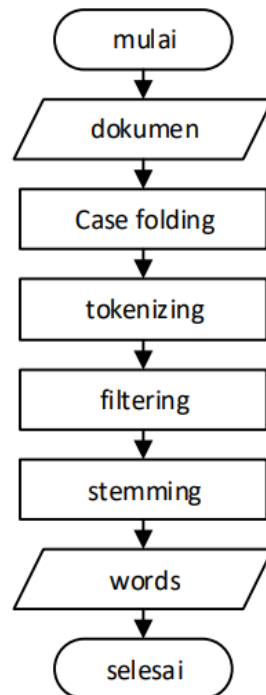
Dalam penerapan text mining pada penelitian ini ada beberapa tahap yang harus dilewati, tahapan tersebut ialah : Document, Text Preprocessing, Data Mining, dan Evaluation :

Document

Document adalah kumpulan text yang dibutuhkan sebelum dilakukannya proses atau langkah penggalan informasi dalam text mining. Document dari penelitian ini diambil dari salah satu website penyedia berita kriminal yaitu <https://jpnn.com/kriminal>. Dalam penelitian ini, sebanyak 120 dataset digunakan dimana 60 jenis berita kriminal subkategori pembunuhan dan 60 jenis berita kriminal subkategori narkoba.

Text Preprocessing

Text preprocessing merupakan sebuah proses dalam text mining dimana ia berfungsi untuk merubah sebuah data tidak terstruktur menjadi data terstruktur sesuai kebutuhan data mining [5].



Sumber : Hasil Penelitian (2021)

Gambar 1 Tahap *Preprocessing*

Dalam text preprocessing terdapat 4 langkah seperti terlihat pada gambar 1 yaitu : Case Folding, Tokenizing, Stopword Removal, dan Stemming [6].

Data Mining

Algoritma naïve bayes adalah algoritma yang sering digunakan untuk kebutuhan klasifikasi atau pengelompokan data karena sifatnya sederhana[7]. (NBC) atau Naive Bayes Classifier merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dicetuskan oleh Thomas Bayes seorang ilmuwan dari Inggris, yang mana metode tersebut memprediksi probabilitas di masa depan berdasarkan pengalaman di masa lalu atau sebelumnya. Dalam metode NBC, terdapat 2 tahapan untuk proses klasifikasi teks. yang pertama adalah tahap training atau pelatihan dan yang kedua adalah tahap testing atau klasifikasi. Algoritma naïve bayes juga dinilai dapat memberikan hasil yang lebih akurat jika dibandingkan dengan algoritma lainnya[8]. Wawasan berfikir atau kecerdasan diterapkan pada tahap ini yaitu untuk mengolah pola informasi dari data yang dipilih yang diharapkan dapat berguna dan memperoleh informasi baru dari yang sudah tersedia atau ditampilkan[9].

Evaluation

Dalam tahap evaluasi, kita dapat melihat apakah hasil dari proses data mining dapat menyelesaikan tujuan yang telah ditetapkan atau tidak. Maka dari itu, dibutuhkan suatu cara yang sistematis agar dapat mengevaluasi kinerja dalam sebuah metode. Evaluasi dalam klasifikasi dapat dilihat berdasarkan pengujian pada objek yang benar dan objek yang salah atau biasa disebut dengan confusion matrix. Confusion Matrix berisi informasi aktual berupa hasil yang telah diprediksi oleh sebuah sistem klasifikasi. Performa dari sebuah sistem tersebut biasanya dievaluasi menggunakan data kembali yang berbentuk matriks.

Tabel 1 model *Confusion Matrix*

#	Relevant	Not Relevant
<i>Retrieved</i>	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
<i>Not Retrieved</i>	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Sumber : Hasil Penelitian (2021)

Penjelasan :

True Positive (TP) = Total kelas positif dan diklasifikasikan dengan benar sebagai kelas positif oleh sistem.

False Positive (FP) = Total kelas bukan positif yang diklasifikasikan sebagai kelas positif oleh sistem

False Negative (FN) = Total kelas positif yang diklasifikasikan sebagai kelas bukan positif oleh sistem

True Negative (TN) = Total kelas bukan positif (negatif) dan diklasifikasikan dengan benar sebagai kelas negatif oleh sistem.

Pengukuran tingkat performansi sistem akan dilakukan dengan melihat nilai akurasi, precision, recall, dan F – measure[10].

Akurasi

Akurasi merupakan ketepatan suatu sistem dalam melakukan proses klasifikasi dengan benar.

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \dots\dots\dots(5)$$

Precision

Precision merupakan jumlah rasio atau dokumen yang sering muncul dengan total jumlah dokumen yang ditemukan (classifier).

$$\text{Precision} = \frac{TP}{TP+FP} \dots\dots\dots(6)$$

Recall

Recall merupakan jumlah rasio atau dokumen yang sering muncul yang ditemukan kembali oleh classifier dengan total jumlah dokumen yang relevan.

$$\text{Recall} = \frac{TP}{TP+FN} \dots\dots\dots(7)$$

F - measure

F - measure merupakan penggabungan dari rata - rata harmonic dari recall dan precision yang selaras dengan nilai keduanya.

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \dots\dots\dots(8)$$

3. Hasil dan Pembahasan

Dataset yang digunakan dalam penelitian ini terdiri dari 55 berita kriminal berkategori pembunuhan dan 65 berita kriminal berkategori narkoba. Data tersebut diambil dari situs online <https://ipnn.com/kriminal>. Data kemudian disimpan dalam sebuah folder dengan file ekstensi .txt.



Sumber : Hasil Penelitian (2021)

Gambar 2 Pemanggilan Operator *Process Document From Files & Validation*

Langkah selanjutnya terlihat pada Gambar 2 yaitu melakukan preprocessing atau data cleaning yang terbagi menjadi 4 tahapan yaitu :

Case Folding

Case Folding ialah proses merubah semua huruf dalam sebuah dokumen menjadi huruf kecil yang mana hanya huruf 'a' sampai dengan huruf 'z' yang akan dapat diproses. Karakter selain itu akan dihilangkan[11]. Proses case folding bisa kita lihat pada table 2 berikut.

Tabel 2 Tabel Case Folding

Sebelum	Sesudah
Ketiganya hendak pesta narkoba jenis sabu-sabu. Kapolres Sukabumi : "Mereka kami tangkap karena kedapatan memiliki narkoba jenis sabu - sabu".	ketiganya hendak pesta narkoba jenis sabu sabu kapolres sukabumi mereka kami tangkap karena kedapatan memiliki narkoba jenis sabu sabu

Sumber : Hasil Penelitian (2021)

Tokenizing

Tahap Tokenization merupakan tahap pemecahan string input berdasarkan tiap - tiap kata penyusunnya[12]. Contoh proses Tokenizing bisa kita lihat pada table 3.

Tabel 3 Tabel Tokenizing

Sebelum	Sesudah
Ketiganya hendak pesta narkoba jenis sabu-sabu. Kapolres Sukabumi : "Mereka kami tangkap karena kedapatan memiliki narkoba jenis sabu - sabu".	Ketiganya hendak pesta narkoba jenis sabu sabu kapolres sukabumi mereka kami tangkap karena kedapatan memiliki narkoba jenis sabu sabu

Sumber : Hasil Penelitian (2021)

Stopword Removal

Stopword Removal ialah tahap penghilangan text - text umum yang tidak mempunyai arti atau informasi yang dibutuhkan[13]. Berikut contoh dari proses Stopword Removal seperti pada table 4.

Tabel 4 Tabel Stopword

Sebelum	Sesudah
Ketiganya hendak pesta narkoba jenis sabu-sabu. Kapolres Sukabumi : "Mereka kami tangkap karena kedapatan memiliki narkoba jenis sabu - sabu".	Ketiganya hendak pesta narkoba sabu kapolres sukabumi mereka kami tangkap kedapatan memiliki narkoba sabu

Sumber : Hasil Penelitian (2021)

Stemming

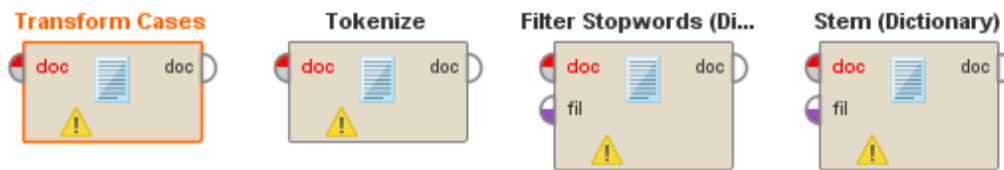
Stemming ialah proses untuk merubah token yang memiliki kata imbuhan menjadi kata dasar. Dengan cara menghapus seluruh kata imbuhan yang terdapat pada token tersebut. Proses stemming penting dilakukan agar dapat menghapus atau meniadakan kata imbuhan baik itu pada awalan maupun akhiran[14]. Proses stemming bisa kita lihat pada table 5 di bawah ini.

Tabel 5 Tabel Stemming

Sebelum	Sesudah
Ketiganya hendak pesta narkoba jenis sabu-sabu. Kapolres Sukabumi : "Mereka kami tangkap karena kedapatan memiliki narkoba jenis sabu - sabu".	tiga hendak pesta narkoba sabu kapolres sukabumi mereka kami tangkap dapat milik narkoba sabu

Sumber : Hasil Penelitian (2021)

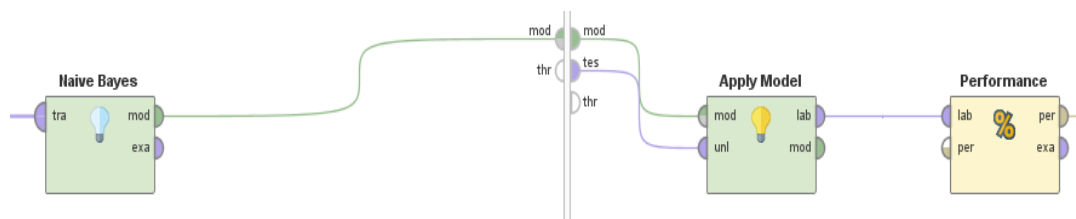
Tahapan menggunakan RapidMiner dapat kita lihat pada gambar berikut.



Sumber : Hasil Penelitian (2021)

Gambar 3 Tahap *Preprocessing*

Pada Gambar 3 yaitu proses penambahan operator untuk melakukan tahap preprocessing yaitu operator transform case, tokenize, filter stopwords dan stem.



Sumber : Hasil Penelitian (2021)

Gambar 4 Tahapan Validasi *Naive Bayes*

Setelah menambahkan operator preprocessing, selanjutnya kita tambahkan operator Naive bayes pada validation seperti yang terlihat pada gambar 4.

Dengan data yang telah ada, untuk berita narkoba sebanyak 57 berita diklasifikasikan sebagai berita narkoba serta sisanya yaitu 8 berita yang diprediksikan sebagai berita narkoba namun hasil dari prediksinya adalah bukan berita narkoba. Sedangkan untuk berita pembunuhan sebanyak 41 berita diklasifikasikan sebagai berita pembunuhan serta sisanya yaitu 14 berita yang diprediksikan sebagai berita pembunuhan namun hasil dari prediksinya adalah bukan berita pembunuhan. Berikut terlampir pada gambar 5 berikut hasil dari klasifikasi menggunakan Naive Bayes.

accuracy: 81.67% +/- 12.25% (mikro: 81.67%)

	true Narkoba	true Pembunuhan	class precision
pred. Narkoba	57	14	80.28%
pred. Pembunuhan	8	41	83.67%
class recall	87.69%	74.55%	

Sumber : Hasil Penelitian (2021)

Gambar 5 Tahap Metode *Naive Bayes* dengan *Confusion Matrix*

Pengujian menggunakan algoritma naive bayes kemudian ditambahkan optimasi PSO (Particle Swarm Optimization) dalam penelitian ini, penentuan nilai training dilakukan dengan uji coba yaitu memasukkan parameter nilai population size dan juga nilai inertia weight. Terlampir pada tabel 6 merupakan hasil dari percobaan untuk menentukan nilai training.

Tabel 6 Tabel pengujian Population dan Inertia Weight PSO

Population Size (Q)	Inertia Weight (W)	Accuracy
4	1.0	89,17%
5	1.0	93,33%
6	1.0	85,83%
7	1.0	88,33%
8	1.0	90,83%

Sumber : Hasil Penelitian (2021)

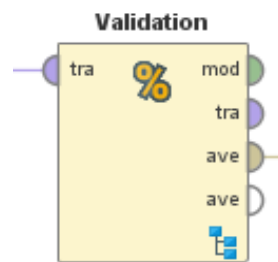
Hasil yang paling signifikan dari percobaan naïve bayes yang ditambahkan optimasi PSO (Particle Swarm Optimization) yaitu dengan memasukkan sebesar 5 nilai untuk parameter population size serta sebesar 1.0 nilai untuk parameter inertia waight. Dengan inputan tersebut, akurasi yang dihasilkan untuk percobaan diatas yaitu di angka 93.33 %. Dalam percobaan yang dilakukan peneliti, hasil klasifikasi berita kriminal dengan subkategori pembunuhan dan narkoba pada media online dengan algoritma naïve bayes yang ditambahkan optimasi PSO (Particle Swarm Optimization) pada aplikasi Rapidminer 7.1 bisa kita lihat pada gambar di bawah ini :



Sumber : Hasil Penelitian (2021)

Gambar 6 Tahap Percobaan Validasi Naïve Bayes bagian 1

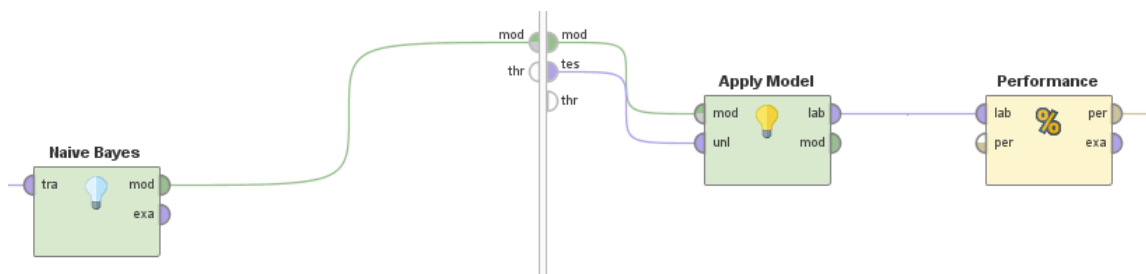
Berbeda dengan ketika pengujian hanya dengan naïve bayes, tahap pengujian untuk klasifikasi berbasis PSO operator langsung ditambahkan di awal model seperti terlihat pada gambar diatas. Yang kemudian, didalam operator PSO tersebut terdapat operator Validation.



Sumber : Hasil Penelitian (2021)

Gambar 7 Model Pengujian Validasi Naïve Bayes bagian 2

Di dalam operator PSO terdapat operator validation yang kemudian dilanjutkan penambahan operator naïve bayes untuk memproses data training dan data testing. Dan di dalam operator Validation terdapat algoritma yang kita gunakan yaitu naïve bayes seperti terlihat pada gambar 8 di bawah ini :



Sumber : Hasil Penelitian (2021)

Gambar 8 Tahap Percobaan Validasi naïve bayes Berbasis PSO di dalam cross validation

Dengan menggunakan data yang telah tersedia, untuk berita narkoba sebanyak 61 berita diklasifikasikan sebagai berita narkoba dan sisanya yaitu sebanyak 4 berita yang diprediksi sebagai berita narkoba namun hasil prediksinya adalah bukan berita narkoba. Sedangkan untuk

berita pembunuhan sebanyak 51 berita diklasifikasikan sebagai berita pembunuhan serta sisanya yaitu sebanyak 4 berita yang diprediksi sebagai berita pembunuhan namun hasil dari prediksinya adalah bukan berita pembunuhan. Berikut terlampir pada gambar 9 hasil dari klasifikasi menggunakan algoritma naïve bayes yang ditambahkan optimasi PSO :

accuracy: 93.33% +/- 7.26% (mikro: 93.33%)

	true Narkoba	true Pembunuhan	class precision
pred. Narkoba	61	4	93.85%
pred. Pembunuhan	4	51	92.73%
class recall	93.85%	92.73%	

Sumber : Hasil Penelitian (2021)

Gambar 9 Tahap Metode Naïve Bayes Berbasis PSO pada Confusion Matrix

4. Kesimpulan

Cara pengolahan berita yang tersebar di media online dapat dilakukan dengan cara pemilihan jenis berita kemudian data yang telah diperoleh diolah menggunakan aplikasi Rapidminer dan dengan metode klasifikasi akan dapat diketahui tingkat akurasi jenis suatu berita. Penggunaan naïve bayes dalam penelitian ini menunjukkan hasil yang cukup tinggi, yaitu dari total 120 dokumen (65 berita narkoba dan 55 berita pembunuhan) yang diperoleh peneliti dari website <https://ipnn.com/kriminal> menghasilkan nilai akurasi sebesar 81.67%. Ini menunjukkan bahwa metode naïve bayes dapat digunakan dalam klasifikasi berita kriminal subkategori narkoba dan subkategori pembunuhan. Hasil akurasi yang didapat setelah penambahan PSO cukup signifikan. Dimana sebelum penambahan PSO, hasil akurasi menunjukkan di angka 81.67%. Sedangkan setelah penambahan PSO hasil akurasi meningkat dan menunjukkan di angka 93.33%. Dengan demikian, dalam penelitian ini penggunaan PSO menambah tingkat akurasi klasifikasi berita kriminal subkategori narkoba dan subkategori pembunuhan sebesar 11.66%.

Referensi

- [1] A. P. J. I. I. APJII, "Peluang Penetrasi Internet dan Tantangan Regulasi Daerah," *BULETIN APJII EDISI 79 Januari 2021*. p. 15, 2021, [Online]. Available: <https://apjii.or.id/downfile/file/BULETINAPJIIEDISI79Januari20211.pdf>.
- [2] H. Rhomadhona and J. Permadi, "Klasifikasi Berita Kriminal Menggunakan Naïve Bayes Classifier (NBC) dengan Pengujian K-Fold Cross Validation," *J. Sains dan Inform.*, vol. 5, no. 2, pp. 108–117, 2019, doi: 10.34128/jsi.v5i2.177.
- [3] R. Wati, "Penerapan Algoritma Naive Bayes Dan Particle Swarm Optimization Untuk Klasifikasi Berita Hoax Pada Media Sosial," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 2, pp. 159–164, 2020, doi: 10.33480/jitk.v5i2.1034.
- [4] H. Rhomadhona, J. Permadi, J. T. Informatika, P. Negeri, and T. Laut, "Penerapan Naïve Bayes Classifier Untuk Klasifikasi Implementation of Naive Bayes Classifier for Classification," in *2nd SEMINASTIKA, 2019*, pp. 18–24.
- [5] H. Mustofa and A. A. Mahfudh, "Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes," *Walisono J. Inf. Technol.*, vol. 1, no. 1, p. 1, 2019, doi: 10.21580/wjit.2019.1.1.3915.
- [6] Y. D. Pramudita, S. S. Putro, and N. Makhmud, "Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, p. 269, 2018, doi: 10.25126/jtiik.201853810.
- [7] steven roy Simanjuntak, "Text Mining Untuk Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes (NB)," *J. Telemat.*, vol. 12, no. 01, pp. 7–12, 2017, [Online]. Available: <http://journal.ithb.ac.id/telematika/article/view/154>.

- [8] H. Muhamad, C. A. Prasajo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin, "Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, p. 180, 2017, doi: 10.25126/jtiik.201743251.
- [9] W. Dwi Septiani Program Studi Manajemen Informatika AMIK BSI Jakarta Jl Kramat Raya No and J. Pusat, "Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis," *J. Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 76–84, 2017, [Online]. Available: <http://archive.ics.uci.edu/ml/>.
- [10] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 1, 2020, doi: 10.24076/citec.2019v6i1.178.
- [11] E. Junianto and D. Riana, "Penerapan PSO Untuk Seleksi Fitur Pada Klasifikasi Dokumen Berita Menggunakan NBC," *JURNAL INFORMATIKA*, vol. 4, no. 1, pp. 38–45, 2017.
- [12] M. Hengki and M. Wahyudi, "Klasifikasi Algoritma Naïve Bayes dan SVM Berbasis PSO Dalam Memprediksi Spam Email Pada Hotline-Sapto," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 1, pp. 61–67, 2020, doi: 10.31294/p.v22i1.7842.
- [13] A. T. Jaka, "Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining," *Inform. UPGRIS*, vol. 1, pp. 1–9, 2015.
- [14] 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.