

2958-7584-1-SM

by jiko unkhair

Submission date: 30-Mar-2021 09:29PM (UTC-0500)

Submission ID: 1546893151

File name: 2958-7584-1-SM.docx (283.17K)

Word count: 3129

Character count: 19400

Perbandingan Klasifikasi Bahasa Menggunakan Metode *Naïve Bayes Classifier* (NBC) Dan *Support Vector Machine* (SVM)

Deglorians Tuhenay¹, Evangs Mailoa²

^{1,2}Fakultas Teknologi Informasi, Program Studi Teknik Informatika, Universitas Kristen Satya Wacana

Email: ¹lorituhenay12@gmail.com, ²evangs.mailoa@uksw.edu

(Naskah masuk: 27 Maret 2021, diterima untuk diterbitkan: dd mmm yyyy)

Abstrak

Bahasa adalah salah satu aspek penting dalam kehidupan manusia karena melalui bahasa, manusia dapat berkomunikasi dan berinteraksi satu dengan lainnya. Beragam macam suku, budaya, ras, dan agama di Indonesia seringkali menyebabkan kesulitan dalam berkomunikasi karena dialek dan bahasa dari tiap daerah di Indonesia sangatlah banyak dan beragam. Penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan bahasa menurut bahasa dari tiap daerah di Indonesia dan dalam penelitian ini menggunakan bahasa Indonesia sebagai bahasa Ibu dan bahasa daerah yang dipakai adalah dari daerah Ambon dan Jawa. Proses identifikasi dan klasifikasi bahasa menggunakan *Machine Learning* dengan memakai metode *Naïve Bayes Classifier* dan *Support Vector Machine* agar dapat mengetahui nilai probabilitas bahasa dan untuk mengklasifikasi bahasa sesuai data yang sudah dimasukkan di *Ms.Excel*. Penelitian ini juga melakukan perbandingan dari kedua metode tersebut agar dapat mengetahui metode apa yang lebih efektif dipakai untuk mengidentifikasi bahasa. Hasil yang didapatkan dengan memakai metode NBC dan SVM adalah keduanya bagus dalam melakukan identifikasi bahasa karena memperoleh nilai *accuracy* di atas 0,90 hanya saja melalui perhitungan *counfusion matrix*, metode SVM lebih efektif dengan nilai *accuracy* 0,9634 dibandingkan dengan nilai NBC 0,9378.

Kata kunci: Bahasa, NBC, SVM, Prediksi, Perbandingan

Comparison Of Language Classification Using *Naïve Bayes Classifier* (NBC) And *Support Vector Machine* (SVM) Method

Abstract

Language is the one of important aspect of Humans life because through language, humans can communicate and interact with one another. Various ethnic groups, cultures, races, and religions in Indonesia often cause difficulties in communicating because the dialect and languages of each region in Indonesia are very numerous and varied. This study aims to identify and classify languages according to the language of each region in Indonesia. In this study, Indonesian is used as the mother tongue and the regional languages are from Ambon and Java. The process of language identification and classification uses *Machine Learning* using the *Naïve Bayes Classifier* and *Support Vector Machine* method in order to find out the probability value of language and to classify the language according to the data that has been entered in *Ms. Excel*. This study also conducted a comparison of the two methods in order to find out which method was more effective in identifying language. The results obtained by using the NBC and SVM methods are that both of them are good at identifying languages because they get an *accuracy* value above 0,90, only through the calculation of the *counfusion matrix*, the SVM method is more effective with an *accuracy* value of 0.9634 compared with NBC value of 0.9378.

Keywords: Language, NBC, SVM, Prediction, Comparison

1. PENDAHULUAN

Bahasa mempunyai sifat yang arbitrer atau sewenang-wenang yang di gunakan manusia pada zaman dahulu hingga sekarang dan mempunyai peran penting bagi kehidupan umat manusia, karena sifat alami manusia adalah sebagai makhluk sosial maka dengan adanya bahasa, manusia dapat berinteraksi dan berkomunikasi satu dengan lainnya[1]. Bahasa

juga penting bagi kehidupan manusia karena manusia dapat dengan mudah menyampaikan informasi yang ingin di sampaikan baik secara lisan maupun tulisan bahkan melalui simbol-simbol atau kode tertentu.[2]bahasa juga bisa digunakan dalam penyampaian informasi secara satu arah maupun dua arah agar dapat berkomunikasi satu dengan lainnya.[3]

19 Beragam macam suku, budaya, ras dan agama di Indonesia 34 menjadikan negara Indonesia menjadi salah satu negara dengan bahasa daerah terbanyak di dunia yaitu lebih dari 700 bahasa.[4] karena begitu banyak bahasa yang terdapat di Indonesia maka untuk membuat klasifikasi dan mengelompokan bahasa menurut bahasa daerah masing-masing dapat memanfaatkan teknologi komputerisasi *Artificial Intelligence* (AI) atau juga biasa disebut dengan *Machine Learning* dalam cabangnya yaitu *Natural Language Processing* (NLP) dengan memakai algoritma *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) maka dapat memudahkan orang dalam menentukan apakah bahasa daerah yang dipakai adalah bahasa daerah dari daerah mana saja yang ada di Indonesia, karena seringkali banyak terjadi kesalahpahaman pemakaian bahasa dalam dialek atau bahasa daerah, dikarenakan terjadinya percampuran bahasa dengan bahasa Indonesia sebagai bahasa Nasional maka orang akan susah mengetahui apakah bahasa daerah 21 yang dipakai itu berasal dari daerah bagian mana di Indonesia.

14 Bahasa yang digunakan dalam penelitian ini menggunakan bahasa Indonesia sebagai bahasa Nasional dan bahasa daerah menggunakan bahasa Jawa dan Ambon sebagai sampel, maka perlu adanya sistem untuk mengidentifikasi dan mengklasifikasikan bahasa Indonesia, Ambon, dan Jawa agar dapat diterjemahkan menurut bahasa daerah masing-masing dan dapat dipakai sebagai acuan untuk bahasa daerah lain yang ada di Indonesia[5][6].

Dalam dunia komputerisasi, *Natural Language Processing* (NLP) adalah cabang dari *Artificial Intelligence* (AI) atau juga bisa disebut *Machine Learning* yang mempelajari tentang pengelolaan bahasa alami atau bahasa yang biasa digunakan manusia untuk berkomunikasi satu dengan lainnya[7]. Proses penyampaian informasi seringkali mengalami kesalahpahaman artikulasi dan makna dari bahasa, apalagi penyampaian 43 informasi disampaikan dalam bentuk tulisan dan menggunakan bahasa daerah masing-masing maka akan sangat berpotensi menimbulkan kesalahpahaman dalam proses pemaknaan kalimat, maka dari itu perlu adanya sistem untuk menerjemahkan bahasa daerah 41 menurut pengelompokan bahasa daerah dari tiap-tiap daerah di Indonesia. Secara tidak langsung bahasa daerah sudah diperkenalkan melalui 40 em ini dan jika dipakai untuk mengidentifikasi bahasa-bahasa daerah lainnya yang ada di Indonesia maka kemungkinan dapat melestarikan bahasa daerah lebih banyak lagi.

Proses Pengelompokan bahasa daerah menggunakan salah satu algoritma dari *Natural Language Processing* (NLP) yaitu *Classification Naïve Bayes* bertujuan untuk mencari tau nilai kemungkinan bahasa yang diterjemahkan sesuai

dengan klasifikasi bahasa berdasarkan data yang dimasukan sebelumnya[8].

Pentingnya kesadaran untuk melestarikan bahasa di zaman modem ini membuat Farel Fathurrahman, Mayanda Mega 6, antoni, dan Anita Muliawati menulis tentang “Penerapan *Artificial Neural Network* untuk Klasifikasi Citra Teks dalam Penerjemahan Bahasa Daerah” bertujuan agar bahasa daerah dapat dilestarik 61 melalui sistem penerjemah gambar yang berisikan teks bahasa Indonesia menjadi teks bahasa daerah menggunakan metode *Artificial Neural Network* (ANN) yaitu metode pengklasifikasian pada citra atau gambar[9]

Berkembangnya dunia teknologi membuat banyak orang seringkali dalam lingkup *social media* sering melontarkan opini pada suatu postingan tertentu dan terkadang opini tersebut ada yang aktual dan ada yang tidak maka dari itu penelitian yang dilakukan oleh Nico Munasatya dan Sendi Novianto [10] yaitu membahas tentang *Analysis Sentiment* dalam opini publik terhadap Presiden Jokowi yang juga menggunakan *Natural Language Processing* (NLP) untuk proses *preprocessing* dalam mengubah kumpulan struktur *text* menjadi token, karena dengan adanya NLP maka dapat membantu mesin mengenal bahasa alami manusia dengan mengolah sejumlah besar data agar mesin memiliki kemampuan untuk mengerti ucapan dan tulisan dalam bahasa tertentu.

24

2. METODE PENELITIAN

2.1 *Naïve Bayes Classifier*

Naïve Bayes Classifier (NBC) digunakan untuk menentukan nilai probabilitas atau kemungkinan dalam memprediksi peluang berdasarkan data pada pengalaman sebelumnya atau memungkinkan untuk membuat pengelompokan pada suatu sistem. Salah satu contoh *Naïve Bayes Classifier* (NBC) yang dipakai pada penelitian [11] yaitu memprediksi stabilitas lereng pada medan yang rawan terkena bencana longsor untuk mengidentifikasi medan yang rawan terjadi longsor agar dapat dibuat pengelompokan berdasarkan struktur tanah melalui beberapa faktor yang menjadi acuan.

Secara 2 rumus *Naïve Bayes* [12] pada dasarnya adalah sebagai berikut:

$$P(H | X) = \frac{P(H | X) P(H)}{P(X)}$$

Dimana :

X = Data *class* yang belum diketahui

H = Data X merupakan hipotesa *class* yang spesifik

P(H | X) = Probabilitas hipotesa H berdasarkan kondisi X

P(H) = Probabilitas hipotesa H (prior)

P(X | H) = Probabilitas X berdasarkan kondisi

P(X) = Probabilitas dari X

Dalam proses ini metode NBC [38] mengambil petunjuk agar dapat menentukan kelas yang cocok bagi sampel data yang di uji [36] maka dari itu rumus umum NBC di ubah menjadi sebagai berikut :

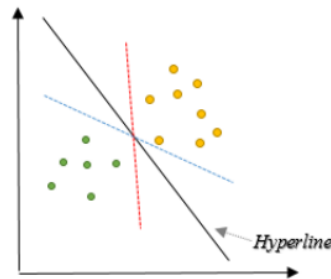
$$P(H | X_1...X_n) = \frac{P(H) P(X_1...X_n|H)}{P(X_1...X_n)}$$

Dimana variabel H merepresentasikan class dan $X_1...X_n$ merepresentasikan class yang belum diketahui dan nantinya akan dibutuhkan untuk proses klasifikasi.

30
2.2 Support Vector Machine

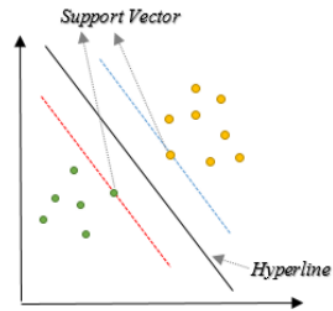
Support Vector Machine (SVM) digunakan untuk membagi data yang sudah diketahui atau sudah dibuat sebelumnya berdasarkan klasifikasinya agar dapat menguji keakuratan data pada sebuah sistem. SVM digunakan untuk pembagian data non linier dan akan membaginya dalam hyperline sebagai pemisah dari titik vector[13]. Dalam penelitian sebelumnya yang membahas tentang perbandingan pemilihan fitur dan pengklasifikasian klasifikasi teks pendek menunjukkan hasil bahwa regresi logistic dan Support Vector Machine mencapai akurasi tertinggi dan paling stabil[14] namun pada penelitian ini masih menggunakan Support Vector machine (SVM) dan Naive Bayes Classifier (NBC) untuk menjadi perbandingan keakuratan sistem.

Cara kerja Support Vector Machine (SVM) secara umum dapat dilihat sebagai berikut:



Gambar 1. Ilustrasi Pembagian SVM 1

Pada Gambar 1. Terdapat 3 garis pembatas, yang di tengah adalah garis hyperline dan 2 garis lainnya adalah pembagi antara titik vector yang ditandai dengan titik warna kuning dan hijau. Metode SVM melakukan pembagian pada kelompok titik vector berwarna kuning dan hijau dengan cara mencari titik vector terdekat dari kelompoknya yang mendekati garis hyperline agar dapat dicari tingkat keakuratannya.



Gambar 2. Ilustrasi Pembagian SVM 2

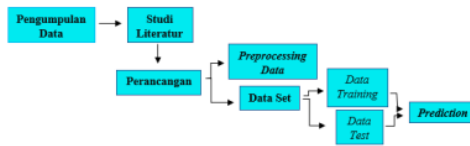
Pada Gambar 2. Titik vector yang terdekat dengan hyperline harus searah lurus dengan hyperline, jika dilihat pada gambar 2 titik vector ini sudah searah lurus dengan hyperline, 2 titik vector ini yang dinamakan dengan Support Vector.

2.3 Performance Evaluation Measure

Dalam pemrosesan data dari data set terdapat suatu tahapan yang dilakukan untuk mengukur evaluasi performa pada suatu sistem yang disebut sebagai Performance Evaluation Measure (PEM) perhitungan yang dipakai dalam sistem ini yaitu mencari informasi perbandingan hasil klasifikasi dengan confusion matrix dengan menghitung nilai accuracy, precision, recall dan F1-score. Accuracy adalah hasil dari data yang terbaca oleh sistem dari keseluruhan informasi, precision adalah tingkat ketepatan antara permintaan pengguna dengan jawaban dari sistem, recall adalah perbandingan ketepatan membaca sistem dengan informasi yang pernah di run sebelumnya dan F1-score adalah harmonic mean atau nilai rata-rata dari precision dan recall. Pada perancangan sistem ini google colab digunakan sebagai coding environment yang bahasa pemrogramannya adalah python, jika dipakai dalam proses perhitungan machine learning maka tidak harus menghitung tingkat accuracy, precision, recall dan F-1-score secara manual karena pada codingan python sudah bisa langsung menghitung secara otomatis.

2.4 Tahapan Penelitian

Proses pengklasifikasian data menggunakan beberapa tahapan dalam penelitian yaitu teknik pengumpulan data, studi literatur, tahap perancangan penelitian dan prediction.



Gambar 3. Alur Tahapan Penelitian

Teknik pengumpulan data di buat pada M.Excel dengan memasukan minimal 3.000 data pada

masing-masing *label* dalam bentuk kalimat dengan menggunakan bahasa Indonesia dan bahasa daerah Ambon dan Jawa untuk menjadi data mentah agar ketika diproses menggunakan metode NBC dan SVM dapat mengetahui nilai probabilitas dan persamaan *hyperline* pada sistem agar dapat membandingkan keakuratan pada kedua metode tersebut.

Studi literatur memakai acuan [12] dari penelitian sebelumnya dengan memanfaatkan metode *Support Vector Machine (SVM)* dan *Naive Bayes Classifier (NBC)* untuk proses pengklasifikasian maupun pengelompokan data [6]jika sudah semakin mudah bagi penelitian ini dalam membuat perbandingan klasifikasi bahasa dengan memanfaatkan kedua metode tersebut[15].

Tahapan perancangan pada sistem yang dibuat yaitu dengan dilakukannya *Preprocessing* data yaitu membuat data sesuai dengan model yang sudah ditentukan seperti, menghilangkan tanda baca pada *cleantext* dan juga kalimat dibuat lebih dari lima kata agar tidak terjadi kesalahan dalam proses *running* dalam sistem.

Contoh :

Kalimat asli : "bia¹⁰su-busu kabaya deng kaeng salele, tetap beta pung mama"
 Preprocessing : "biar busu busu kabaya deng kaeng salele tetap beta pung mama"
 Tanda baca (-) dan (,) dihilangkan

Setelah data *cleantext* di buat pada *Ms.Excel* maka selanjutnya melakukan pengelompokan data berdasarkan *label* yaitu bahasa Indonesia, Jawa dan Ambon. Pengujian dilakukan dengan mengelompokan data untuk menentukan dan membedakan data berdasarkan bahasa daerah, dalam proses ini juga dilakukan *data set* yang dibagi menjadi dua yaitu *data training* 30% dan *data test* 70%. *Data training* dijadikan sebagai basis untuk menguji dan melatih apakah data yang kita pakai sudah bisa berjalan dan dapat melakukan penerjemahan sesuai dengan aturan yang ditentukan atau tidak, dan *data test* yaitu data baru yang digunakan untuk menguji keakuratan dalam sebuah sistem apakah sudah bekerja sesuai dengan standar yang diinginkan atau tidak, setelah itu masuk pada proses *prediction* yaitu untuk mengetahui hasil dari uji *accuracy*, *precision*, *recall* dan *F1-score* pada sistem yang telah dibuat[5].

31
3. HASIL DAN PEMBAHASAN

3.1 Data Set

Data yang dipakai untuk pembuatan sistem ini menggunakan bahasa Indonesia, Ambon dan Jawa yang dimasukan kedalam *Ms.Excel* yang dibagi menjadi *cleantext* sebagai isi dari kalimat dari tiap bahasa dan *label* sebagai identit[35] dari bahasa. Data yang dimasukan di *Ms.Excel* dapat dilihat pada gambar dibawah ini.

Gambar 4. Data Set Bahasa Indonesia,Ambon dan Jawa

Proses penginputan data dilakukan secara manual dengan memasukan data ke *Ms.Excel* dengan membagnya berdasarkan *label* dan *cleantext* yaitu *label* adalah pengelompokan bahasa dan *cleantext* adalah isi dari kalimat sesuai bahasa. Data yang dimasukan dalam penelitian ini tidak memakai tanda baca dan angka. Jumlah data dari keseluruhan data yang diolah adalah ≥ 3.000 bahasa Ambon, ≥ 10.000 , bahasa Jawa, dan ≥ 3.500 bahasa Indonesia. Standar minimum data yang harus dimasukan dalam sistem ini adalah minimal 3.000 data agar sistem dapat membaca peluang kemungkinan dan membagi data sesuai dengan rumus yang sudah ditentukan dalam penelitian ini yaitu *split* data uji sekitar 30%.

3.1 Naive Bayes Classifier

Codingan *Naive Bayes Classifier* untuk dapat menghitung klasifikasi dan prediksi tingkat keakuratan sistem pada *confusion matrix* dapat dilihat sebagai berikut:

```
#Multinomial Naive Bayes
pipeline_mnb = Pipeline([
    ('vect', CountVectorizer()),
    # ('vect', CountVectorizer(ngram_range=(1, 2))),
    ('tfidf', TfidfTransformer(use_idf=True, smooth_idf=True)),
    ('clf', MultinomialNB(alpha=1))
])
```

```
X_train, X_test, y_train, y_test = train_test_split(data['cleanText'], data['label'], test_size=0.33, random_state = 0)
clf_mnb = pipeline_mnb.fit(X_train, y_train)
pred_mnb = pipeline_mnb.predict(X_test)
```

Dengan Perhitungan *count vector n-gram 2,2* dan *split* data uji 0,33 atau sekitar 30% data.

3.3 Naïve Bayes Classifier result

Hasil perhitungan *confusion matrix* dapat dilihat pada tabel berikut:

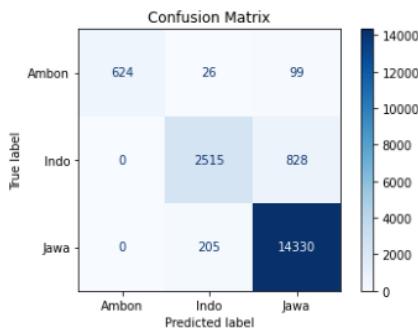
Accuracy: 0.9378

Confusion matrix:

```
[[ 624 26 99]
 [ 0 2515 828]
 [ 0 205 14330]]
```

Tabel 1. Hasil *Confusion Matrix* NBC

	Precision	Recall	F1-score	Support
Ambon	1.00	0.83	0.91	749
Indo	0.92	0.75	0.83	3343
Jawa	0.94	0.99	0.96	14535
Accuracy			0.94	18627
Macro avg	0.95	0.86	0.90	18627
Weighted avg	0.94	0.94	0.94	18627



Gambar 20. Heatmap *Confusion Matrix* NBC

Berdasarkan Tabel 1 dapat dilihat bahwa tingkat keseluruhan perhitun²⁰ dengan *confusion matrix* yang mencakup nilai *precision*, *recall* dan *F1-score* mendapatkan nilai *accuracy* 0,9378.

3.4 Prediction Naïve Bayes Classifier

Penginputan kalimat dari tiap bahasa pada sistem akan diidentifikasi sesuai dengan dialek bahasa daerah yang di input. Gambar 5 dapat dilihat bahwa sistem membaca dengan membedakan bahasa dari bahasa Indonesia, Ambon dan Jawa, yaitu

kalimat “aku telah mendapatkannya”, “maju deng kaeng berang tu saja”, “sapa pung ana par se” diidentifikasi sebagai bahasa “Jawa”, “Ambon”, “Ambon”.

```
predictions = pipeline_mnb.predict(["aku telah mendapatkannya", "maju deng kaeng berang tu saja", "sapa pung ana par se"])
predictions
array(['Jawa', 'Ambon', 'Ambon'], dtype=object)
```

Gambar 6. Prediction Bahasa

3.5 Support Vector Machine

Codingan *Support Vector Machine* untuk menghitung nilai dari *confusion matrix* dapat dilihat sebagai berikut

```
#SVM
pipeline_svc = Pipeline([
    ('vect', CountVectorizer()),
    ('tfidf', TfidfTransformer(use_idf=True, smooth_idf=True)),
    ('clf', LinearSVC())
])
X_train, X_test, y_train, y_test = train_test_split(data['cleanText'], data['label'], test_size=0.33, random_state = 0)
clf_svc = pipeline_svc.fit(X_train, y_train)
pred_svc = pipeline_svc.predict(X_test)
```

Dengan memakai *split* data uji yaitu 0,33 atau sekitar 30% data.

3.6 Support Vector Machine Result

Hasil perhitungan *confusion matrix* dapat dilihat pada tabel berikut:

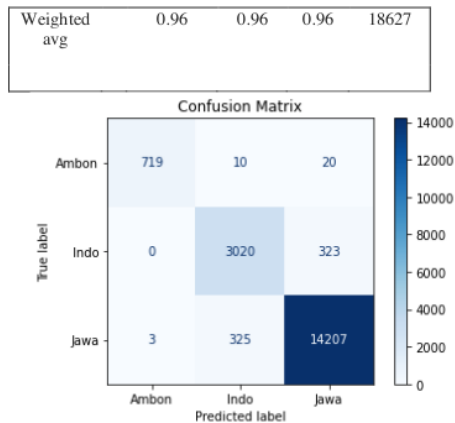
Accuracy: 0.9634

Confusion matrix:

```
[[ 719 10 20]
 [ 0 3020 323]
 [ 3 325 14207]]
```

Tabel 2. Hasil *Confusion Matrix* SVM

	Precision	Recall	F1-score	Support
Ambon	1.00	0.96	0.98	749
Indo	0.90	0.90	0.90	3343
Jawa	0.98	0.98	0.98	14535
Accuracy			0.96	18627
Macro avg	0.96	0.95	0.95	18627



Gambar 7. Heatmap Confusion Matrix SVM

Perhitungan nilai keakuratan secara keseluruhan pada confusion matrix mendapat nilai *accuracy* sebesar 0,9634.

3.7 Prediction Support Vector Machine

Kalimat dimasukan kedalam sistem agar bisa diidentifikasi bahwa kalimat tersebut dapat diterjemahkan sesuai dengan bahasa yang sudah ditentukan atau tidak, dapat dilihat pada gambar 7, kalimat yang dimasukan adalah kalimat “wes aku rapopo”, “saya selalu semangat dengan segala tantangan”, “sapa pung ana par c” diidentifikasi sebagai bahasa “Jawa”, “Jawa”, “Ambon”.

```
[ ] predictions = algoritma_svm.predict(["wes aku rapopo","saya selalu semangat dengan segala tantangan","sapa pung ana par c"])
predictions
array(['Jawa', 'Jawa', 'Ambon'], dtype=object)
```

Gambar 8. Prediction Bahasa

3.8 Perbandingan nilai akurasi

Berdasarkan hasil analisis perbandingan nilai keakuratan sistem yang dihitung dari *confusion matrix* maka dapat dilihat hasil dari metode *Naïve Bayes Classifier* dan *Support Vector Machine* adalah sebagai berikut:

Tabel 3. Perbandingan Nilai Akurasi

Metode	Nilai Akurasi
<i>Naïve Bayes Classifier</i>	0,9378
<i>Support Vector Machine</i>	0,9634

4. KESIMPULAN

Berdasarkan hasil pengujian dan perbandingan yang dilakukan dalam penelitian ini, kedua metode yang dipakai yaitu *Naïve Bayes Classifier* dan

Support Vector Machine sama-sama bagus dalam melakukan penerjemahan bahasa karena memiliki tingkat keakuratan di atas 0,9 atau di atas 90%. Akan tetapi dalam proses perhitungan menggunakan *confusion matrix* pada keseluruhan nilai keakuratan yang mencakup nilai *precision*, *recall* dan *F1-score* maka dapat dilihat bahwa nilai *accuracy* pada metode *Support Vector Machine* lebih tinggi dibandingkan dengan nilai dari *Naïve Bayes Classifier* dengan nilai *accuracy* SVM adalah 0,9634 atau 96,34%. Maka dapat disimpulkan bahwa metode *Support Vector Machine* lebih efektif dibandingkan dengan metode *Naïve Bayes Classifier* dalam hal mengidentifikasi dan mengklasifikasikan Bahasa.

5. DAFTAR PUSTAKA

- [1] Rina Devianty, “Bahasa Sebagai Cermin Kebudayaan,” *J. Tarb.*, vol. 24, no. 2, pp. 226–245, 2017.
- [2] L. Wicaksono, “Bahasa Dalam Komunikasi Pembelajaran,” *J. Pembelajaran Prospektif*, vol. 1, no. 2, pp. 9–19, 2016, [Online]. Available: <http://jurnal.untan.ac.id/index.php/lp3m/article/download/19211/16053>.
- [3] R. P. Suminar, “Pengaruh Bahasa Gaul terhadap Penggunaan Bahasa Indonesia Mahasiswa Unswagati,” *Logika*, vol. 18, no. 3, pp. 114–119, 2016, [Online]. Available: www.jurnal.unswagati.ac.id.
- [4] Berlianty, “Penguatan Eksistensi Bahasa Tana dalam Upaya Perlindungan Hukum Bahasa Daerah sebagai Warisan Budaya Bangsa,” *Bahasa Patrika*, vol. 40, no. 2, p. 99, 2018, doi: 10.24829/tp.2018.v40.i02.p04.
- [5] A. A. Budiman, “Pendeteksi Bahasa Daerah Pada Twitter Dengan Machine Learning,” p. 11523262, 2018.
- [6] L. Flek, “Returning the N to NLP: Towards Contextually Personalized Classification Models,” pp. 7828–7838, 2020, doi: 10.18653/v1/2020.acl-main.700.
- [7] Lalwani, S. Bhalotia, A. Pal, S. Bisen, and Rathod, “Implementation of a Chat Bot System using AI and NLP,” *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 6, no. 3, pp. 26–31, 2018, doi: 10.21276/ijrcst.2018.6.3.2.
- [8] Saputro, M. Aristin, and Dy. L. Tyas, “Berdasarkan Li Menggunakan Metode Tf-,” *J. Teknoloi Inform. dan Terap.*, vol. 4, no. 1, pp. 45–50, 2017.
- [9] F. Fathurrahman et al., “PENERAPAN ARTIFICIAL NEURAL NETWORK UNTUK KLASIFIKASI CITRA TEKS DALAM PENERJEMAHAN BAHASA,” pp. 585–594, 2020.
- [10] N. Munasatya and S. Novianto, “Natural Language processing untuk Sentimen Analisis Presiden Jokowi Menggunakan Multi Layer Perceptron,” *Techno.Com*, vol.

- 19, no. 3, pp. 237–244, 2020, doi: [9.33633/tc.v19i3.3630](#).
- [11] X. Feng, S. Li, C. Yuan, P. Zeng, and Y. Sun, “Prediction of Slope Stability using Naive Bayes Classifier,” *KSCE J. Civ. Eng.*, vol. 22, no. 3, pp. 941–950, 2018, doi: [8.1007/s12205-018-1337-3](#).
- [12] W. Muslehatin, M. Ibnu, and Mustakim, “Penerapan Naive Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau,” *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, p. 7, 2017.
- [13] *et al.*, “Klasifikasi Kabupaten Tertinggal di Kawasan Timur Indonesia dengan Support Vector Machine,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 3, no. 3, pp. 188–195, 2020, doi: [5.10.33387/jiko.v3i3.2364](#).
- [14] Y. Wang, Z. Zhou, S. Jin, D. Liu, and M. Lu, “Comparisons and Selections of Features and Classifiers for Short Text Classification,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 261, no. 1, 2017, doi: [10.1088/1757-899X/261/1/012018](#).
- [15] M. L. Laia and Y. Setyawan, “Perbandingan Hasil Klasifikasi Curah Hujan Menggunakan Metode SVM dan NBC,” vol. 05, no. 2, pp. 51–61, 2020.

ORIGINALITY REPORT

23%

SIMILARITY INDEX

20%

INTERNET SOURCES

9%

PUBLICATIONS

5%

STUDENT PAPERS

PRIMARY SOURCES

1	dspace.uui.ac.id Internet Source	3%
2	Submitted to Universitas Islam Indonesia Student Paper	1%
3	Nittaya Muangnak, Natakorn Thasnas, Thapani Hengsanunkul, Jakkarin Yotapakdee. "The Neural Network Conversation Model enables the Commonly Asked Student Query Agents", International Journal of Advanced Computer Science and Applications, 2020 Publication	1%
4	ejournal.unkhair.ac.id Internet Source	1%
5	Submitted to Qassim University Student Paper	1%
6	conference.upnvj.ac.id Internet Source	1%
7	Submitted to Universitas Brawijaya Student Paper	1%

8	eprints.umm.ac.id Internet Source	1%
9	www.ijitee.org Internet Source	1%
10	www.adadisini.info Internet Source	1%
11	Moch. Fadli Shadiqin Thirafi, Faisal Rahutomo. "Implementation of Naïve Bayes Classifier Algorithm to Categorize Indonesian Song Lyrics Based on Age", 2018 International Conference on Sustainable Information Engineering and Technology (SIET), 2018 Publication	1%
12	openlibrary.telkomuniversity.ac.id Internet Source	1%
13	www.researchgate.net Internet Source	1%
14	adoc.pub Internet Source	1%
15	www.scribd.com Internet Source	1%
16	ojs.unud.ac.id Internet Source	<1%
17	Hendry Cipta Husada, Adi Suryaputra Paramita. "Analisis Sentimen Pada Maskapai	<1%

Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)", Teknika, 2021

Publication

18

core.ac.uk

Internet Source

<1%

19

id.123dok.com

Internet Source

<1%

20

doku.pub

Internet Source

<1%

21

bagawanabiyasa.wordpress.com

Internet Source

<1%

22

publikasi.dinus.ac.id

Internet Source

<1%

23

aemtemite.wordpress.com

Internet Source

<1%

24

Submitted to Universitas Dian Nuswantoro

Student Paper

<1%

25

journal.akprind.ac.id

Internet Source

<1%

26

mafiadoc.com

Internet Source

<1%

27

www.portalgaruda.org

Internet Source

<1%

28	Weiyong Zeng, Mohammed A. S. Khalid, Xiaoye Han, Jimi Tjong. "A Study on Extreme Learning Machine for Gasoline Engine Torque Prediction", IEEE Access, 2020 Publication	<1%
29	help.uui.ac.id Internet Source	<1%
30	link.springer.com Internet Source	<1%
31	media.neliti.com Internet Source	<1%
32	ojs.upy.ac.id Internet Source	<1%
33	usermanual.wiki Internet Source	<1%
34	we-didview.xyz Internet Source	<1%
35	www.slideshare.net Internet Source	<1%
36	Desi Ratna Sari, Dedy Hartama, Irfan Sudahri Damanik, Anjar Wanto. "Penerapan Metode Naive Bayes dalam Memprediksi Kepuasan Mahasiswa Terhadap Cara Pengajaran Dosen", Prosiding Seminar Nasional Riset Information Science (SENARIS), 2019	<1%

37 docobook.com <1%
Internet Source

38 e-journals.unmul.ac.id <1%
Internet Source

39 fauzanafianto.student.telkomuniversity.ac.id <1%
Internet Source

40 id.wikimedia.org <1%
Internet Source

41 mamanruhaman.blogspot.com <1%
Internet Source

42 repository.ar-raniry.ac.id <1%
Internet Source

43 thousands-passed.xyz <1%
Internet Source

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off