

Komparasi Algoritma *Naive Bayes* dan *Support Vector Machine* Untuk Memprediksi Pengaruh Pembelajaran *Daring* Terhadap Motivasi Belajar

Ariansyah¹, Mira Kusmira²

^{1,2}Department of Information System, Universitas Nusa Mandiri, Indonesia

Article Info

Article history:

Received Sep 3, 2021

Revised Sep 219, 2021

Accepted Sep 28, 2021

Keywords:

Online learning

Learning motivation

Naive Bayes

SVM

ABSTRACT

The COVID-19 pandemic in Indonesia has had a major impact on the education sector. Where currently must adapt the distance learning model. There are pros and cons to public opinion on social media Twitter about the influence of online learning on learning motivation. The study aims to find out if public opinion is positive or negative and to find out which classification algorithms have the highest and best accuracy rates. The data used was tweets in Indonesian with the keywords "online learning", "distance learning" and "motivational learning". With a data set of 455 tweets. Classification uses Naive Bayes classification algorithms and supports vector machines by preprocessing data using tokenize, transform case, filtering, and stemming. The highest and best accuracy result is classification algorithm support vector machine with accuracy 97.22%, precision 94.72%, recall 100%, and error 2.78%.

Copyright © 2021 Universitas Indraprasta PGRI.
All rights reserved.

Corresponding Author:

Ariansyah,

Universitas Nusa Mandiri,

Jl.Ciledug Raya No.168, Ulujami, Kebayoran Lama, Jakarta Selatan.

Email: ansyahrn@gmail.com

1. PENDAHULUAN

Indonesia mengalami situasi genting, dimana bisa dicermati berdasarkan meningkatnya jumlah harian, jumlah kasus aktif dan jumlah kematian COVID 19 Indonesia. Pada akhir bulan Januari 2021, Indonesia mengalami peningkatan jumlah tes (*positive rate*) mencapai 36,18% dan melebihi batas yang telah ditetapkan WHO yaitu 5%. Pada bulan Januari 2021, jumlah kasus kematian di indonesia semakin tinggi yaitu 476 orang. Untuk jumlah kasus kematian di indonesia sangat tinggi yaitu 2,8% dan melebihi jumlah kasus kematian global yaitu 2,3%. Pada awal bulan Februari 2021, jumlah kasus positif COVID-19 indonesia pun semakin tinggi dengan jumlah 175.236 kasus atau sebanyak 15,8% [1]. Kemendikbud mengeluarkan kebijakan mengenai pembelajaran daring dan bekerja dirumah dalam mencegah COVID 19 di indonesia. Berdasarkan data UNESCO sebanyak 850 juta siswa didunia melakukan pembelajaran dirumah [2].

Hambatan yang dihadapi pada proses pembelajaran jarak jauh diantaranya berkaitan dengan kesiapan sumber daya manusia, kurang jelasnya arahan dari Gubernur atau Walikota daerah, informasi kurikulum yang tidak jelas, dan keterbatasan penyediaan sarana pendukung teknologi pembelajaran daring. Kesiapan sumber daya manusia meliputi pendidik, peserta didik, dan dukungan orang tua adalah bagian terpenting pada proses pembelajaran jarak jauh [3]. Pembelajaran daring menjadi permasalahan yang harus dihadapi saat ini karena harus adanya proses penyesuaian dengan sistem belajar ini. Salah satu yang terkena pengaruh yaitu mahasiswa yang mengakibatkan prestasinya turun lantaran motivasi belajar mahasiswa ikut menurun yang diakibatkan pada sistem pembelajaran daring ini. Perubahan mendadak sistem pendidikan yang bermula proses aktivitas belajar mengajar dilakukan secara tatap muka sekarang beralih sebagai pembelajaran daring (*online*).

Dalam memecahkan masalah, penulis dalam penelitian ini menggunakan metode *machine learning* untuk melakukan klasifikasi pendapat pengguna *Twitter* mengenai pengaruh pembelajaran daring terhadap motivasi belajar dan menentukan apakah pendapat pengguna *Twitter* tersebut memiliki sentimen positif atau negatif. Terdapat beberapa algoritma pada metode klasifikasi yaitu Naive Bayes dan Support Vector Machine (SVM). Naïve Bayes adalah salah satu algoritma teknik data mining yang menerapkan teori Bayes dalam klasifikasi dimana algoritma ini mengasumsikan bahwa atribut suatu objek bersifat independen atau bebas. Metode Naïve Bayes telah banyak digunakan dalam penelitian mengenai text mining, beberapa kelebihan Naïve Bayes diantaranya adalah algoritma sederhana tapi memiliki akurasi yang tinggi. Teknik SVM berakar pada teori pembelajaran statistik dan telah menunjukkan hasil empiris yang menjanjikan dalam berbagai aplikasi praktis dari pengenalan digit tulisan tangan sampai kategorisasi teks. SVM juga bekerja sangat baik pada data dengan banyak dimensi dan menghindari kesulitan dari permasalahan dimensionalitas. Berdasarkan penelitian sebelumnya, naïve Bayes memiliki kelebihan yang dapat menangani data dengan jumlah yang besar dan tidak mempengaruhi irrelevant attribute. Naïve Bayes juga memiliki kelebihan dalam kesederhanaannya. SVM memiliki kelebihan dimana tingkat akurasi yang lebih baik dibandingkan Naïve Bayes. Fitur yang berbeda dari pendapat pengguna *Twitter* mengenai pengaruh pembelajaran daring terhadap motivasi belajar dimasa pandemi dengan algoritma *Naive Bayes* dan *Support Vector Machine* untuk memberikan hasil yang terbaik. Selanjutnya kedua algoritma tersebut dibandingkan untuk mendapatkan hasil berdasarkan *accuracy*, *precision*, *recall*, dan *error*. Penelitian terkait mengenai keberlanjutan pembelajaran daring dengan algoritma *Support Vector Machine* menghasilkan *accuracy* sebesar 86%, *precision* 87,38% dan *recall* 85,02% [4].

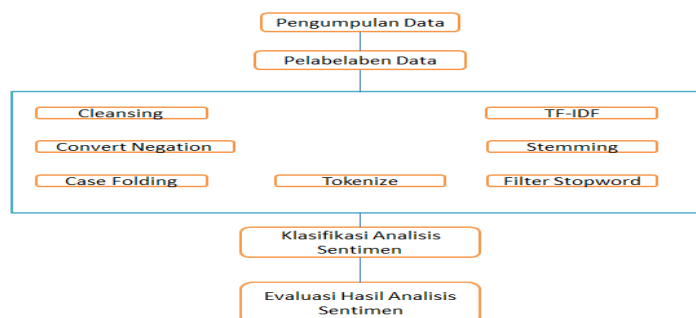
Di sisi lain, *Text mining* merupakan proses mining ataupun menambang sesuatu data dari info yang tersaji dalam jumlah besar, dalam perihal ini ialah bacaan. Proses ini dicoba dalam rangka penggalian, pengolahan, dan pengaturan pada data dengan menganalisa keterkaitan antara data satu dengan yang yang lain [5]. Dengan demikian *text mining* sebagai analisis sentimen karena dapat digunakan untuk memperoleh informasi berupa teks pada *twitter*. Analisis sentimen adalah proses memahami dan mengelompokkan emosi (positif, negatif, dan netral) yang terdapat pada suatu tulisan menggunakan teknik analisis teks [6].

Penelitian ini bertujuan menganalisis opini publik terhadap pembelajaran daring ini yang mengakibatkan motivasi belajar mahasiswa menurun serta menyebabkan prestasi mahasiswa ikut menurun pada masa pandemi COVID-19 khususnya di Indonesia. Penelitian ini menggunakan kata kunci ‘belajar daring’, ‘pembelajaran jarak jauh’, dan ‘motivasi belajar’ yang difilter dengan kata kunci ‘daring’ dan ‘motivasi belajar’ pada *twitter* dalam bahasa Indonesia yang ditarik dari *Twitter* dan diolah menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* untuk menghasilkan analisis sentimen pengaruh pembelajaran daring terhadap motivasi belajar di masa pandemi bersifat positif atau negative dan untuk mengetahui algoritma klasifikasi mana yang memiliki tingkat keakurasian yang tinggi dan terbaik.

2. METODE

Penelitian tentang sentimen analisis dari *tweet* di *Twitter* tentang pengaruh pembelajaran daring terhadap motivasi belajar dimasa pandemi COVID-19. Analisis sentimen dilakukan proses identifikasi berdasarkan suatu opini atau tanggapan masyarakat yang bersifat terbuka didalam media sosial guna menentukan opini tersebut mengandung opini positif atau negatif [7].

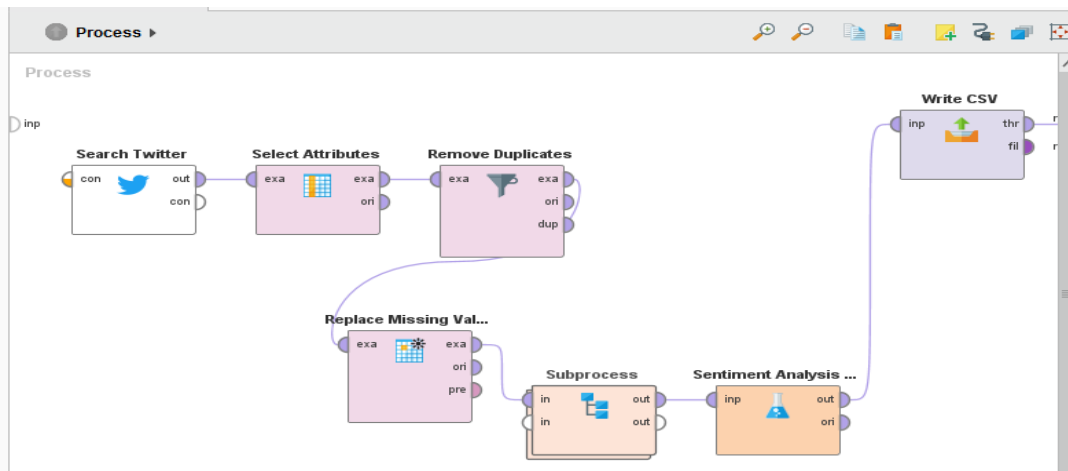
Analisis Sentimen *Twitter* berfokus pada masalah klasifikasi. Klasifikasi adalah proses kumpulan pola ataupun yang berguna untuk mendeskripsikan serta memisahkan kelas data yang satu dengan yang lain untuk dijadikan objek dan dimasukkan kedalam kategori yang telah ditetapkan [8]. Berikut adalah skema alur penelitian yaitu Pengumpulan Data, Pelabelan Data, Pre-processing, Klasifikasi dan Evaluasi Hasil Analisis Sentimen. Gambar 1 menunjukkan langkah-langkah dalam klasifikasi sentimen analisis yang diusulkan.



Gambar 1. Diagram Klasifikasi Analisis Sentimen

A. Pengumpulan Data

Pada tahap ini melakukan proses pengumpulan data dalam analisis sentimen. Data diambil dari Twitter dengan menggunakan kata kunci ‘belajar daring’, ‘pembelajaran jarak jauh’, dan ‘motivasi belajar’ tentang pengaruh pembelajaran daring terhadap motivasi belajar sebanyak 455 tweet menggunakan aplikasi Rapidminer.



Gambar 2. Pengumpulan data dari Twitter

Gambar 2 menunjukkan pengumpulan data dari Twitter menggunakan konten "Search Twitter" dan menyimpan data dalam file excel menggunakan konten "Write Excel" dengan hanya mengambil teks dari tweet, menghapus tweet yang bersifat duplikat dengan operator "Remove Duplicates", operator "Replace Missing Value" untuk mengganti nilai data yang tidak memiliki nilai data dan operator "Sentiment Analysis" berfungsi untuk menentukan suatu kalimat bersifat positif atau negatif.

B. Pelabelan Data

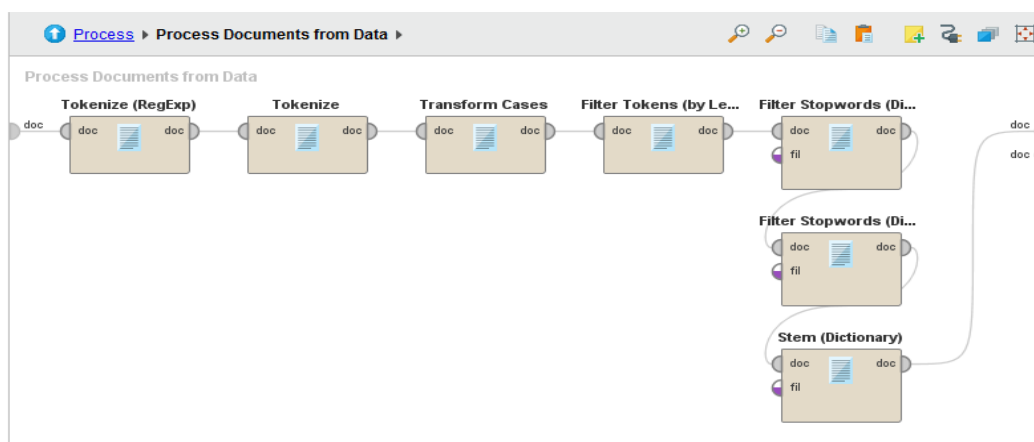
Pada tahap ini adalah pemberian label pada dataset. Pelabelan berfungsi untuk menentukan nilai sentimen terhadap setiap kalimat, yaitu positif atau negatif. Pemberian label pada dataset berguna untuk membagi data latih dan data testing. Data latih berfungsi untuk menguji data dengan model algoritma yang sudah ditentukan, sedangkan data testing berfungsi untuk melakukan pengujian terhadap hasil pelatihan yang telah dilakukan sebelumnya. Tabel 1 salah satu contoh dataset yang telah diberi label:

Tabel 1. Contoh *Label Dataset*

Text	Sentimen
Nadiem Makarim secara gamblang mengakui kegagalan Pembelajaran Jarak Jauh https://t.co/BJjGKolicR Hay manteman □ Bagaimana cara kalian mengurangi tingkat stres dampak belajar daring?. Kalau tips dari saya, seperti di gambar. Memanajemen waktu, agar semuanya bisa seimbang. Tetap patuhi protokol kesehatan, menjaga jarak, sehat fisik dan batin manteman ✨😊 #FikomYai #CyberPR https://t.co/zdIEOXskIs	NEGATIF
	POSITIF

C. Pre- processing

Pada Tahap ini adalah proses pre-processing. Gambar 3 menunjukkan operator “Process Document From Data”, menggunakan konten “Tokenize”, "Transform Cases", “Filter Stopword” dan “Stem” dalam bahasa Indonesia.



Gambar 3. Isi Operator *Process Document*

Berikut ini adalah langkah-langkah dari tahapan pre-processing:

1. *Cleansing*

Pada proses ini adalah proses untuk menghapus atribut yang tidak dibutuhkan [9]. Contoh karakter yang dihilangkan seperti URL, tag (#) dan tanda lainnya. Contoh kalimat pengapusan data, input “Nadiem Makarim secara gamblang mengakui kegagalan Pembelajaran Jarak Jauh <https://t.co/BJjGKoIicR>”, output menjadi “Nadiem Makarim secara gamblang mengakui kegagalan Pembelajaran Jarak Jauh”.

2. *Convert Negation*

Convert negation, yaitu proses mengubah singkatan kata menjadi kata yang utuh [10]. Contoh kalimat, input “Nadiem Makarim secara gamblang mengakui kegagalan Pembelajaran Jarak Jauh”, output menjadi “Nadiem makarim secara gamblang mengakui kegagalan pembelajaran jarak jauh”.

3. *Case Folding*

Dalam menulis tweet, biasanya ada bentuk huruf yang berbeda, Case folding merupakan proses mengubah huruf besar menjadi huruf kecil (*lower case*) [11]. Contoh kalimatnya adalah, input ““Nadiem Makarim secara gamblang mengakui kegagalan Pembelajaran Jarak Jauh””, menjadi output “nadiem makarim secara gamblang mengakui kegagalan pembelajaran jarak jauh”.

4. *Tokenization*

Tokenize merupakan proses pemotongan teks menjadi kata, symbol, karakter ataupun tanda baca, sehingga menjadi token yang dapat di analisa [12]. Ini adalah contoh kalimat tokenization, input “nadiem makarim secara gamblang mengakui kegagalan pembelajaran jarak jauh”, output menjadi “nadiem, makarim, secara, gamblang, mengakui, kegagalan, pembelajaran, jarak, jauh”.

5. *Filtering*

Filtering merupakan proses menyaring kata-kata asing atau singkatan menjadi bahasa Indonesia [13]. Ini adalah contoh kalimat stopwords, input “nadiem makarim secara gamblang mengakui kegagalan pembelajaran jarak jauh”, output menjadi “nadiem makarim mengakui kegagalan pembelajaran jarak jauh”.

6. *Stemming*

Stem adalah bagian kata yang tinggal setelah menghilangkan imbuhan (awalan dan akhiran) [14]. Ini adalah contoh dari kalimat, input “nadiem makarim secara gamblang mengakui kegagalan pembelajaran jarak jauh”, output menjadi “nadiem makarim secara gamblang mengakui kegagalan pembelajaran jarak jauh”, Tabel 2 adalah salah satu contoh kata dari stemming:

Tabel 2. Contoh Kata *Stemming*

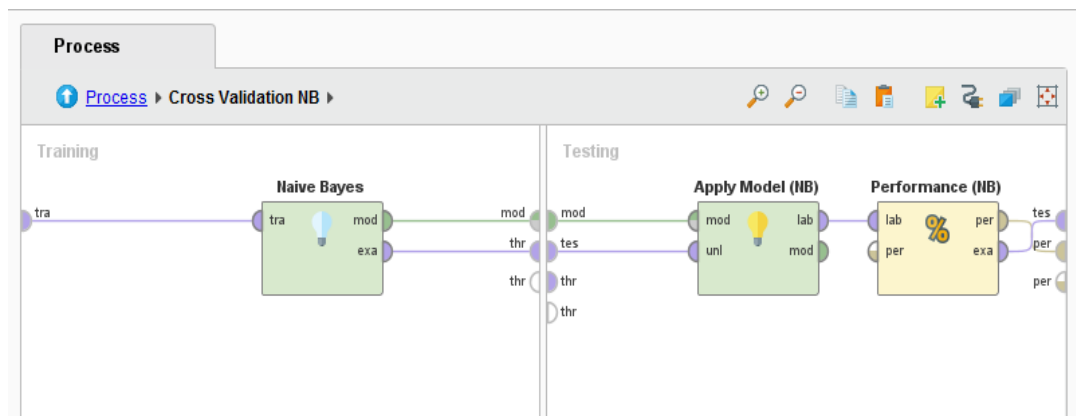
Sebelum	Sesudah
pembelajaran	belajar
kegagalan	gagal
diberikan	beri
sebenarnya	benar

7. *Weighting Word*

Weighting Word adalah mekanisme memberi skor kemunculan kata dalam dokumen teks dan untuk menghitungnya dapat menggunakan Persamaan [15]. Prosedur TF- IDF ialah prosedur guna menghitung bobot tiap kata yang sangat universal digunakan pada information retrieval. Prosedur ini pula populer efektif, gampang serta mempunyai hasil yang akurat. Prosedur ini hendak menghitung nilai Term Frequency(TF) dan Inverse Document Frequency(IDF) pada masing- masing token(kata) di masing- masing dokumen dalam korpus. Metode TF-IDF menggunakan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut [16].

D. Klasifikasi Analisis Sentimen

Setelah semua proses sebelumnya dilakukan. Pada tahap ini adalah proses klasifikasi analisis sentimen. Proses klasifikasi dengan membuat pemodelan algoritma yang akan digunakan. Tahapan ini berfungsi untuk membuat model sesuai dengan dua algoritma klasifikasi, berdasarkan himpunan data yang sudah disiapkan. Gambar 4 menunjukkan konten operator "*Cross Validation*" dalam aplikasi Rapidminer. Dalam hal ini, menggunakan dua operator klasifikasi berbeda untuk perbandingannya, yaitu operator klasifikasi *Naïve Bayes* dan klasifikasi *Support Vector Machine*.



Gambar 4. *Cross Validation Operator*

E. Evaluasi Analisis Sentimen

Pada tahap ini adalah proses evaluasi data analisis sentimen untuk menentukan mutu dan kualitas dari hasil proses pengujian. Kinerja yang akan dijadikan sebagai parameter yaitu accuracy, precision, recall dan error [17].

Akurasi (A) merupakan jumlah dokumen yang diklasifikasikan dengan benar, baik TP ataupun TN. Menghitung tingkat nilai akurasi dengan rumus :

$$Akurasi = \frac{(TP + TN)}{(TP + FP + FN + TN)} \times 100\%$$

Presisi (P) adalah banyaknya hasil pemrosesan cocok dengan data yang mau dicari. Presisi ialah klasifikasi TP dari seluruhnya data diprediksi menjadi nilai kelas positif. Nilai presisi dihitung dengan rumus:

$$P = \frac{TP}{FP + TP} \times 100\%$$

Recall (R) merupakan berapa banyak dokumen yang relevan dalam koleksi dihasilkan oleh sistem. Dengan kata lain, recall merupakan jumlah dokumen yang mempunyai klasifikasi Benar Positif dari seluruh dokumen yang betul- betul positif(tercantum False Negative). Menghitung nilai recall dengan rumus:

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

Variabel seperti TN, TP, FN dan FP berasal dari *confusion matrix*. TN adalah singkatan dari *True Negative*. TP adalah singkatan dari *True Positive*, FN adalah singkatan dari *False Negative*. FP adalah singkatan dari *False Positive* [18]. Untuk penjelasan yang lebih detail pada Tabel 3.

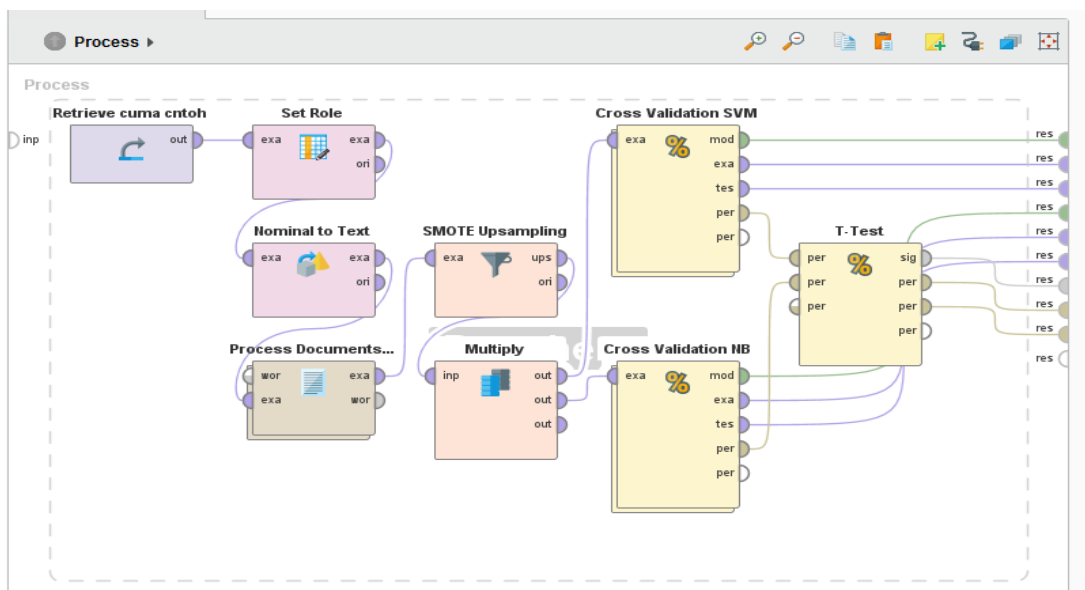
Tabel 3. *Confusion Matrix*

	Prediction Yes	Prediction No
True Yes	TP	FN
True No	FP	TN

Data yang sudah dikumpulkan kemudian diolah dengan aplikasi Rapidminer, kemudian diuji dengan mengkombinasi teknik Smote UpSampling yaitu SMOTE mensintesis instans baru dengan memilah titik acak dari kelas minoritas. Diwaktu yang sama kelas mayoritas diambil guna meenyeimbangkan data dengan menghapus titik data secara acak [19].

3. HASIL DAN PEMBAHASAN

Pada bagian ini menjelaskan proses hasil penelitian dan kinerja analisisnya.



Gambar 5. Proses Utama Pada *Rapidminer*

Name	Type	Missing	Statistics	Filter (2,392 / 2,392 attributes):
Label Sentimen	Binominal	0	Negative AGREEMENT Positive DISAGREEMENT	Values AGREEMENT (413), DISAGREEMENT (413)
Prediction prediction(Sentimen)	Binominal	0	Negative DISAGREEMENT Positive AGREEMENT	Values DISAGREEMENT (436), AGREEMENT (390)
Confidence_AGREEMENT confidence(AGREEMENT)	Real	0	Min 0.160 Max 0.890	Average 0.440
Confidence_DISAGREEMENT confidence(DISAGREEMENT)	Real	0	Min 0.110 Max 0.840	Average 0.560
Text text	Polynomial	0	Least zoom goo [...] tive (1) Most berhenti [...] ive (18)	Values berhenti [...] bjective (18), menempel [...] bjective (
abang	Real	0	Min 0 Max 0.488	Average 0.001
abdullah	Real	0	Min 0 Max 0.340	Average 0.001

Gambar 6. Hasil Prediksi *Rapidminer*

Gambar 5 menunjukkan proses utama dalam aplikasi *Rapidminer*. Konten "Retrieve" digunakan untuk mengakses informasi yang tersimpan di Repositori dalam bentuk Example Set. Konten "Set Role" digunakan untuk membedakan baris penamaan atribut koordinat dan prediksi posisi yang akan dijadikan label data. Konten "Nominal to Text" berfungsi untuk mengubah jenis atribut nominal yang dipilih menjadi teks. Konten "Process Document" digunakan untuk pra-pemrosesan. Konten "Smote UpSampling" berfungsi untuk

menyeimbangkan data pada saat melakukan pengujian. Konten “*Multiply*” digunakan untuk membuat salinan obyek. Konten “*Cross Validation*” digunakan untuk mengklasifikasi dan mengevaluasi analisis sentimen pada eksperimen yang diuji.

Sementara Gambar 6 menunjukkan hasil prediksi dalam aplikasi *Rapidminer*. Melihat sentimen dan juga label prediksi yang sama dengan label sentimen. Berikut ini gambaran hasil dari *confusion matrix* dari masing-masing algoritma pada *Rapidminer*:

Tabel 4 . *Confusion Matrix* Setiap Algoritma

Metode	TP	FP	TN	FN
Naïve Bayes	370	43	413	0
SVM	390	23	413	0

Dari *confusion matrix* pada tabel 4 nilai rata-rata *accuracy*, *precision*, *recall* dan *error* ditampilkan dengan perhitungan menggunakan rumus.

Tabel 5. Nilai *Accuracy*, *Precision*, *Recal*, dan *Error* Menggunakan *Rapidminer*

Metode	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>Error</i>
Naive Bayes	94,79 %	90,57 %	100 %	5,21 %
SVM	97,22 %	94,72 %	100 %	2,78 %

Terlihat dari tabel 5 bahwa algoritma yang memiliki akurasi tertinggi dan terbaik adalah *Support Vector Machine* sebesar 97,22 %. Akurasi menandakan adanya ketepatan antara prediksi yang dihasilkan model dengan nilai sentimen asli pada dataset. Semakin tinggi persentase akurasinya, semakin tinggi ketepatan model dalam memprediksi nilai sesungguhnya. *Precision* menggambarkan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Sedangkan *Recall* menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi dan menandakan adanya relevan antara data dengan model sehingga menghasilkan informasi yang sebenarnya. Dan *Error* menandakan adanya kesalahan pada data dengan model saat melakukan pengujian.

Dalam penelitian ini, model yang dibuat dapat memberikan ketepatan hasil dengan baik dan akurat. Namun, karena kasus pandemi Covid-19 belum berakhir sampai saat ini, tweet terkait pro kontra permasalahan pengaruh pembelajaran daring terhadap motivasi belajar dimasa pandemi Covid-19 akan selalu muncul di Twitter dan analisis dapat berubah karena kebijakan mengenai pembelajaran belum dapat ditentukan oleh pemerintah terkait dengan pandemi Covid-19 yang semakin memburuk dan masih belum bisa dipastikan kapan pandemi Covid-19 ini berakhir. Dan dari hasil yang didapat bahwa algoritma klasifikasi yang memiliki tingkat keakuratan tertinggi dan terbaik adalah algoritma klasifikasi *Support Vector Machine* yaitu sebesar 97,22 % dan opini masyarakat di media sosial twitter mengarah kepada opini negatif karena terjadi pengaruh yang dirasakan masyarakat mengenai pembelajaran daring dimasa pandemi COVID-19 ini.

4. PENUTUP

KESIMPULAN

Dalam penelitian ini, dilakukan klasifikasi analisis sentimen dalam memprediksi opini masyarakat Twitter mengenai pengaruh pembelajaran daring terhadap motivasi belajar dengan menggunakan dua algoritma yaitu *Naive Bayes* dan *Support Vector Machine*. Hasil menunjukkan *accuracy* dari *Naive Bayes* dan *Support Vector Machine* sebesar 94,79% dan 97,22%. Hasil untuk *precision* dari *Naive Bayes* dan *Support Vector Machine* sebesar 90,57% dan 94,72%. Sementara hasil untuk *Recall* dari *Naive Bayes* dan *Support Vector Machine* sama-sama menghasilkan persentase sebesar 100%. Dan untuk hasil *error* dari *Naive Bayes* dan *Support Vector Machine* adalah 5,21% dan 2,78%. Sehingga dapat ditarik kesimpulan bahwa algoritma *Support Vector Machine* adalah algoritma klasifikasi terbaik untuk digunakan dalam perhitungan yang menggunakan dataset media sosial karena memberikan prediksi yang lebih akurat dan tepat serta algoritma *Support Vector Machine* menghasilkan tingkat akurasi tertinggi dan terbaik yaitu sebesar 97,22%. Dan dapat disimpulkan juga bahwa opini masyarakat mengarah kepada opini negatif karena banyak faktor yang mempengaruhi mengenai pembelajaran daring ini baik mengenai sarana dan prasarana teknologi serta sumber daya manusia dan faktor lainnya.

SARAN

Opini masyarakat mengenai analisis sentimen mengenai pengaruh pembelajaran terhadap motivasi belajar dimasa pandemi Covid-19 dapat saja berubah dikarenakan masih terjadi pro kontra mengenai kebijakan pembelajaran daring ini karena situasi Covid-19 yang semakin memburuk dan masih belum bisa dipastikan kapan pandemi Covid-19 ini dapat berakhir. Di masa mendatang, kita harus update berita mengenai kebijakan pembelajaran daring ini karena telah banyak problematika yang dirasakan masyarakat baik untuk orangtua, guru, murid ataupun mahasiswa terkait dengan pembelajaran daring ini karena dapat menimbulkan dampak yang negatif salah satunya adalah penurunan motivasi belajar. Serta dalam penelitian lainnya sebaiknya menggunakan algoritma yang lain agar dapat menghasilkan tingkat keakurasian yang lebih tinggi, baik dan akurat.

DAFTAR PUSTAKA

- [1] N. S. P. Suni, "Tingginya Kasus Aktif dan Angka Kematian Akibat Covid-19 di Indonesia," *J. Pus. Penelit. Badan Keahlian DPR RI*, vol. 13, no. 3, pp. 13–18, 2021.
- [2] Samsir, Ambiyar, U. Verawardina, F. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, pp. 157–163, 2021.
- [3] K. F. Irnanda, D. Hartama, and A. P. Windarto, "Analisa Klasifikasi C4.5 Terhadap Faktor Penyebab Menurunnya Prestasi Belajar Mahasiswa Pada Masa Pandemi," *J. Media Inform. Budidarma*, vol. 5, no. 1, pp. 327–331, 2021.
- [4] A. P. Natasuwarna, "Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring," *Techno.Com*, vol. 19, no. 4, pp. 437–448, 2020.
- [5] A. V. Sudiantoro *et al.*, "Analisis Sentimen Twitter Menggunakan Text Mining Dengan," vol. 10, no. 2, pp. 398–401, 2018.
- [6] E. D. Agustono, D. Sianturi, A. Taufik, and W. Gata, "Analisis Sentimen Terhadap Warga China saat Pandemi Dengan Algoritma Term Frequency-Inverse Document Frequency dan Support Vector Machine," *JIRE (Jurnal Inform. Rekayasa Elektron.*, vol. 3, no. 2, pp. 111–119, 2020.
- [7] E. M. Sipayung, H. Maharani, and I. Zefanya, "Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier," *J. Sist. Inf.*, vol. 8, no. 1, pp. 958–965, 2016.
- [8] A. Novandya, "Penerapan Algoritma Klasifikasi Data Mining C4.5 pada Dataset Cuaca Wilayah Bekasi," *KNiST*, pp. 368–372, 2017.
- [9] A. Amalia, D. Gunawan, Y. Fithri, and I. Aulia, "Automated Bahasa Indonesia essay evaluation with latent semantic analysis," *J. Phys. Conf. Ser.*, vol. 1235, no. 1, 2019.
- [10] L. Jayanti, S. R. Sentinuwo, O. A. Lantang, and A. Jacobus, "Analisa Pola Penyalahgunaan Facebook Sebagai Alat Kejahatan Trafficking Menggunakan Data Mining," *J. Tek. Inform.*, vol. 8, no. 1, 2016.
- [11] E. B. Santoso and A. Nugroho, "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook," *Eksplora Inform.*, vol. 9, no. 1, pp. 60–69, 2019.
- [12] M. F. F. Roji and I. Irhamah, "Topic Discovery pada Dokumen Abstrak Jurnal Penelitian di Science Direct Menggunakan Association Rule," *Inferensi*, vol. 2, no. 2, p. 97, 2019.
- [13] R. Desrianti and H. D. Wijaya, "Implementasi Algoritma Fuzzy C-Means Pada Aplikasi Seleksi Karyawan Digital Talent di PT Telekomunikasi Indonesia," *J. Media ...*, vol. 4, pp. 879–888, 2020.
- [14] L. A. Andika, P. A. N. Azizah, and R. Respatiwulan, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019.
- [15] R. Rasenda, H. Lubis, and R. Ridwan, "Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 369, 2020.
- [16] 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.
- [17] E. Rilvani, A. B. Trisnawan, and P. P. Santoso, "Pelita Teknologi : Jurnal Ilmiah Informatika , Arsitektur dan Lingkungan," *Pelita Teknol. J. Ilm. Inform. Arsit. dan Lingkung.*, vol. 14, no. 2, pp. 103–110, 2019.
- [18] J. Suntoro, *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Elex Media Komputindo, 2019.
- [19] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020.