

Implementasi Algoritma Naïve Bayes Classifier untuk Mendeteksi Berita Palsu pada Sosial Media

Nova Agustina¹, Adrian², Mercy Hermawati³

¹Department of Informatic, Sekolah Tinggi Teknologi Bandung, Indonesia

²Department of Management, Sekolah Tinggi Ilmu Ekonomi Jayakarta, Indonesia

³Department of Informatic, Universitas Indraprasta PGRI, Indonesia

Article Info

Article history:

Received Oct 24, 2021

Revised Dec 20, 2021

Accepted Dec 21, 2021

Keywords:

Naïve Bayes
Classification
Hoax News
Data Mining

ABSTRACT

Hoax news (lie) on the internet has become a global problem that causes turmoil in society. Its presence can disrupt democratic order, the stability of social, cultural, political and economic life. The results of the research of the Indonesian Telematics Society showed that as many as 44.3% of respondents said they received fake news or misinformation every day. According to information released by Kominfo until August 11, 2021, there were 1848 hoax reports regarding the Covid-19 pandemic, 290 hoax reports regarding the Covid-19 Vaccine. Naïve Bayes Classifier is a classification method based on Bayes theorem, which in this paper is used to detect fake news on social media. The analysis was carried out using the Naïve Bayes Classifier algorithm, in this study using the CRISP-DM (Cross-Industry Standard Process for Data Mining) model. Training data sourced from the Kumparan site as much as 700 data and 300 data testing. In the process carried out using the python library for NLP, namely "saratrwi". In testing the model using the confusion matrix method which consists of the number of rows of test data that are predicted to be true and false by the classification model used. The result of our research is that the accuracy of the Naïve Bayes Classifier model is 81% so that it is classified as producing a good model accuracy. At the deployment stage the model is pushed to Heroku so that users can predict news directly through the provided User Interface.

Copyright © 2021 Universitas Indraprasta PGRI.

All rights reserved.

Corresponding Author:

Nova Agustina,
Department of Informatic,
Sekolah Tinggi Teknologi Bandung,
Jl. Soekarno Hatta no. 378, Bandung
Email: nova@sttbandung.ac.id

1. PENDAHULUAN

Berita hoaks (bohong) di internet telah menjadi masalah global yang menimbulkan gejala di masyarakat. Kehadirannya dapat mengganggu ketertiban demokrasi, stabilitas kehidupan sosial, budaya, politik, dan ekonomi [1]. Adanya berita bohong dapat mengganggu ketertiban demokrasi, stabilitas kehidupan sosial, budaya, politik, dan ekonomi. Sudah banyak perusahaan yang sahamnya anjlok karena hadirnya berita palsu[2][3] di kalangan masyarakat tentang perusahaan tersebut. Selain itu, bagi pebisnis apabila tidak melakukan *filter* berita palsu akan mengalami kegagalan dalam mengambil langkah bisnis, salah satunya salah menilai pasar. Hasil penelitian Masyarakat Telematika Indonesia menunjukkan bahwa sebanyak 44,3% responden mengatakan menerima berita bohong atau misinformasi setiap hari. Faktanya, menurut Newman,dkk. [4], lebih dari setengah berita online dari sampel globalnya (55%) mengungkap kekhawatiran

atau kekhawatiran yang kuat tentang “apa yang nyata atau palsu.” Jumlah ini sangat bervariasi di berbagai negara. Misalnya di daftar teratas adalah Brasil (85%), Afrika Selatan (70%), Meksiko (68%), dan Prancis (67%).

Beberapa penelitian mengklasifikasikan postingan *Facebook* dalam klasifikasi *hoax* dan non *hoax* [5]. Berdasarkan jumlah pengguna *Facebook* yang sangat banyak, menjadi media sosial yang subur untuk menyebarkan hoaks dengan cepat. Berbagai berita *hoax* apapun bisa tersebar ke ribuan pengguna media sosial. Mereka juga dikomentari dalam jumlah yang sama besarnya. Dalam laporan “Distribusi Hoax Melalui Platform Digital di Indonesia 2018”, pada survei yang dilakukan terhadap 2.032 orang di Indonesia yang dipimpin oleh DailySocial, ditemukan bahwa tidak semua informasi berkualitas baik, mungkin palsu, dan tidak benar. Informasi palsu dan tidak benar tersebut dapat menyebabkan kekacauan dan kepanikan di antara orang-orang di lapangan [6]. Deteksi dan pencegahan misinformasi dan berita palsu merupakan tantangan untuk menegakkan kebenaran. Penyebarannya bahkan bisa menjadi kekuatan untuk mempengaruhi hasil pemilihan umum suatu negara [7].

Pihak Kominfo hingga saat ini melalui situs layanan aduan konten memang memberikan fasilitas bagi masyarakat, apabila menemukan adanya website/situs, konten media sosial, *game online* yang melanggar aturan perundangan di Indonesia termasuk berita hoaks. Menurut informasi yang dirilis oleh Kominfo hingga 11 Agustus 2021 terdapat 1848 laporan hoaks mengenai pandemik Covid-19, 290 laporan hoaks mengenai Vaksin Covid-19 [8]. Penyebaran berita palsu memiliki potensi dampak yang sangat negatif terhadap individu dan masyarakat. Oleh karena itu, deteksi berita palsu di media sosial baru-baru ini menjadi penelitian baru yang sedang menarik perhatian yang luar biasa [9].

Tentunya selain adanya informasi yang diberikan melalui laporan masyarakat, diperlukan suatu mekanisme untuk mengendalikan dan mengurangi penyebaran berita bohong tersebut. Akhir-akhir ini banyak penelitian dan penelitian mengenai hal ini, bahkan mengadopsi prinsip epidemi untuk memetakan pola penyebaran berita bohong di jejaring sosial. Ditemukan bahwa proses penularan epidemi dan penyebaran informasi memiliki pola teoritis yang sama. Salah satu teknik untuk mengendalikan dan mengurangi berita palsu adalah dengan membuat sistem yang dapat melakukan klasifikasi berita secara otomatis. Dengan klasifikasi sebuah berita, berita tersebut akan diberi label bahwa berita tersebut benar atau palsu. Berdasarkan latar belakang yang sudah dipaparkan, tujuan penelitian kami yaitu:

1. Membangun sistem pendeteksi *hoax* Covid19 berbasis web;
2. Sumber data yang diperoleh diambil dari situs kumparan;
3. Sistem dibangun bertujuan untuk sebagian masyarakat yang memiliki literasi pengetahuan yang kurang akan mudah terprovokasi dengan informasi yang tidak benar.

2. METODE

2.1 Data Mining

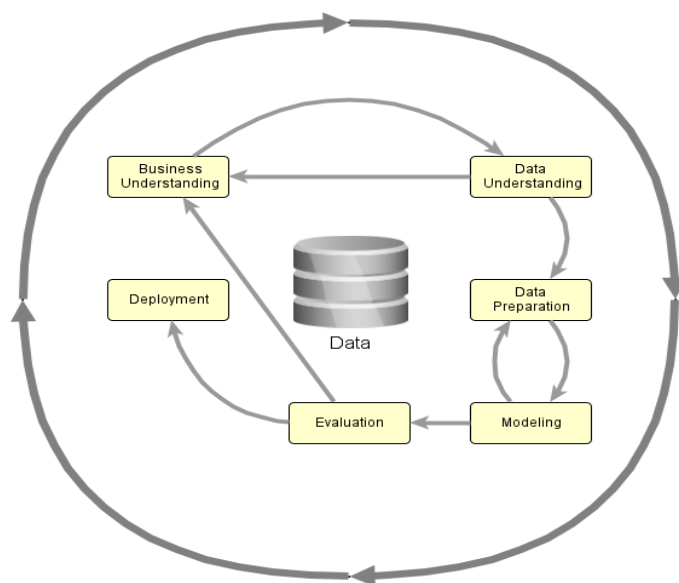
Data mining merupakan proses penemuan pengetahuan yang bermanfaat dan menarik di dalam kumpulan data yang besar, tujuan utama data mining yaitu prediksi (*prediction*) dan uraian (*description*). Data mining juga mempunyai beberapa tugas utama yaitu *classification* (klasifikasi), *regression* (regresi), *clustering* (pengelompokan), *summarization* (ringkasan), *dependency modeling* (pemodelan ketergantungan), *change and deviation detection* (pendeteksian perubahan dan deviasi) [10].

2.2 Algoritma Naïve Bayes Classifier

Naive Bayes Classifier adalah sebuah metoda klasifikasi yang berdasar pada teorema Bayes. Metode pengklasifikasian ini menggunakan metode probabilitas dan statistik yang pertama kali dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, yaitu suatu metode untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga metode ini dikenal sebagai Teorema Bayes. Ciri utama dari Naive Bayes Classifier ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian [11].

2.3 Teknik Analisis Data

Analisis dilakukan menggunakan pengujian algoritma Naïve Bayes Classifier, dalam penelitian ini menggunakan model CRISP-DM (*Cross-Industry Standart Process for Data Mining*) [12].



Gambar 1. Langkah-Langkah Penelitian

1. *Business Understanding*

Mendeteksi berita palsu, untuk memudahkan masyarakat melakukan verifikasi fakta sehingga akan menghindari terjadinya penurunan stabilitas ekonomi. Membantu meningkatkan literasi pengetahuan baik terhadap user yang menggunakan maupun penerapan teknologi kecerdasan buatan.

2. *Data Understanding*

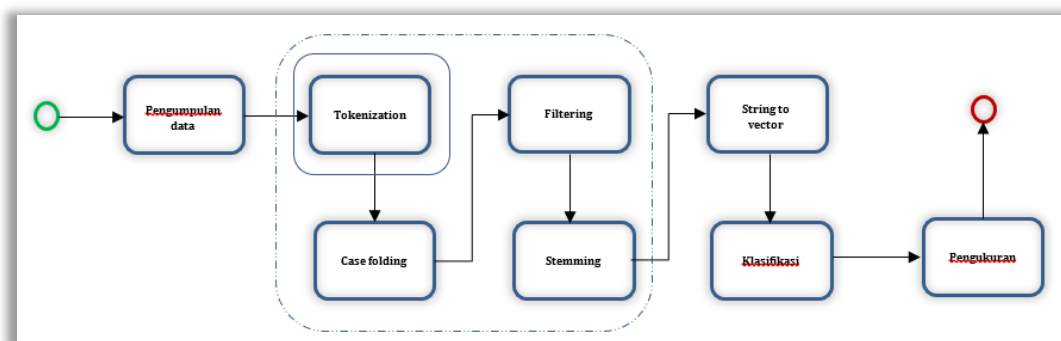
Data dapat berasal dari tabular, SQL, CSV, atau API. Lalu setelah data tersebut didapatkan dilakukan prosedur data engineering terlebih dahulu. Setelah data engineering dilakukan, dilanjutkan dengan algoritma *machine learning (classification)* untuk menilai hoax atau tidak. Data bersumber dari situs Kumbaran sejumlah 1000 Data, dipecah menjadi 700 digunakan untuk data training, dan 300 digunakan untuk data testing.

3. *Data Preparation Phase*

Data berita hoax yang telah dikumpulkan diolah terlebih dahulu, kemudian dilakukan tahap preprocessing merubah format dokumen dalam tahap parsing yaitu melakukan pemotongan dokumen, dilanjutkan dengan tahapan tokenisasi yaitu proses penghilangan tanda baca, spasi dan lain sebagainya. Selanjutnya dilakukan proses stop word, yaitu dokumen yang mengandung kata hoax dihilangkan, supaya tidak mengganggu hasil akurasi. Proses ini dilakukan menggunakan library python untuk NLP, yaitu sastrawi. Sastrawi adalah *library* dari python yang menerapkan Algoritma Nazief dan Adriani untuk melakukan proses stemming bahasa indonesia[13]. Dari hasil *preprocessing* semua dokumen berita dapat terbentuk sejumlah atribut.

4. *Modelling Phase*

Implementasi menggunakan algoritma Naïve Bayes Classifier dengan data masukan berupa dokumen teks. Setelah itu dokumen teks dilakukan *preprocessing*, kemudian dilakukan proses pembobotan kata pada data latih (*data training*). Selanjutnya dilakukan klasifikasi teks berita tersebut dengan menggunakan algoritma Naive Bayes. Pengukuran dilakukan dengan menggunakan akurasi dan F1-Score, karena akurasi akan mengukur performa dari model dan F1-Score dibutuhkan untuk perbandingan rata-rata presisi dan recall yang dibobotkan. Nilai terbaik F1-Score adalah 1 dan nilai terburuknya adalah 0. Secara representasi, jika F1-Score punya skor yang baik mengindikasikan bahwa model klasifikasi kita punya *precision* dan *recall* yang baik.



Gambar 2. Proses *Modelling*

Gambar 2 memvisualisasikan alur proses *modelling* secara umum, yang dimulai dari pengumpulan data. Data yang diambil dari situs kumpulan disimpan ke *spreadsheet* dengan format csv, dan dilakukan proses *tokenization*. *Tokenization* adalah proses memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal yang berdiri sendiri[14]. Kemudian dilakukan *case folding* yang berfungsi untuk konversi text menjadi suatu bentuk yang standar, lalu *filtering* untuk membuang token yang termasuk *stopword* (yang, dan, atau, nya, dsb). Kemudian dilakukan *stemming* pada data *testing*, dan dibuat visualisasi berupa grafik untuk melihat hasil perbandingan data yang sering muncul pada berita hoax dan fakta. Tahap akhir yaitu melakukan klasifikasi menggunakan Naïve Bayes Classifier, pengukuran akurasi model dan F1-Score. Hasil klasifikasi menggunakan Naive Bayes Classifier akan menghasilkan output berupa berita yang sudah terlabeli hoax dan fakta.

5. Evaluation Phase

Kami mengevaluasi model menggunakan metode *Confusion Matrix* untuk mendapatkan akurasi dan F1-Score yang dihasilkan model *Data Mining*[15]. Akurasi digunakan untuk mengidentifikasi keakuratan. Indikator yang dihitung untuk mendapatkan akurasi yaitu : *True Positive* (TP) adalah data berita positif yang terdeteksi dengan benar, *True Negative* (TN) adalah data berita negatif yang terdeteksi dengan benar, *False Positive* (FP) adalah data berita positif tetapi diklasifikasikan salah oleh sistem dan *False Negative* (FN) adalah data berita negatif tetapi diklasifikasikan secara tidak benar oleh sistem. Akurasi dapat dihitung sebagai berikut:

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (1)$$

Presisi digunakan untuk mengidentifikasi kasus positif dengan nilai positif palsu yang tinggi, yang dapat dihitung sebagai berikut:

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

Berbeda dengan presisi, *recall* berfungsi untuk mengidentifikasi kasus positif dengan nilai negatif palsu yang tinggi. *Recall* dapat dihitung dengan cara :

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

F1-Score akan memberikan rata-rata presisi dan recall yang selaras, yang dapat diukur sebagai berikut:

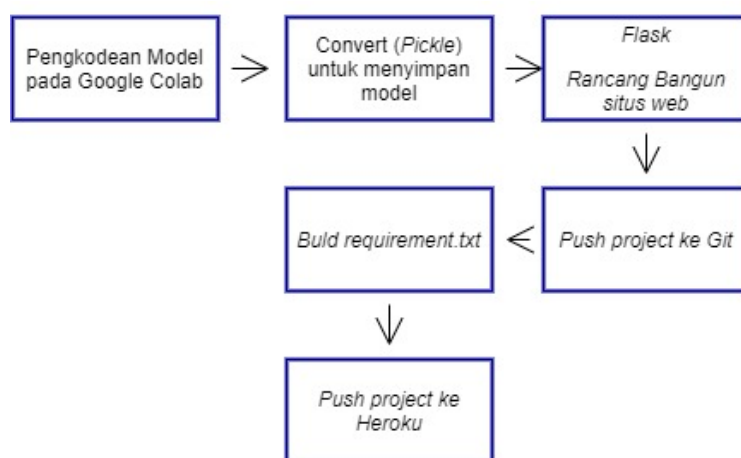
$$F1 - Score = 2x \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (4)$$

Evaluasi model yang dihasilkan digolongkan dalam kelompok berikut[16] :

- a. 0.90-1.00 = klasifikasi sangat baik
- b. 0.80-0.90 = klasifikasi baik
- c. 0.70-0.80 = klasifikasi cukup
- d. 0.60-0.70 = klasifikasi buruk
- e. 0.50-0.60 = klasifikasi salah

6. *Deployment Phase*

Setelah model dibuat pada Google Colab, dilakukan *deploy* model yang akan di-*push* ke Heroku agar pengguna dapat melakukan prediksi berita langsung melalui *User Interface* yang sudah disediakan. Proses ini dilakukan setelah model diuji, dan hasilnya disimpan dengan format *file pickle(.pkl)*. Selanjutnya merancang dan membangun situs web menggunakan framework dari python, yaitu Flask yang memiliki kumpulan kode yang salah satu fungsinya adalah menerapkan model yang sudah disimpan, sehingga tidak perlu lagi menuliskan *script* model Naïve Bayes Classifier pada web yang dibangun. Secara lengkap proses *deploy model* dapat dilihat pada gambar 4.



Gambar 4. *Deploy Model*

Setelah sistem berjalan dengan baik, kami melakukan menyimpan *project* ke git sebagai *repository project* yang kami bangun sebelum dilakukan *hosting*. Kami melakukan *hosting* hasil *deploy* model tersebut dengan memanfaatkan Heroku sebagai salah satu web *hosting* berbasis *cloud* untuk mengembangkan aplikasi web dengan python adalah salah satu Bahasa pemrograman yang didukung oleh Heroku.

3. HASIL DAN PEMBAHASAN

Sumber data (judul berita, isi berita, link berita) yang digunakan dalam penelitian ini adalah data kualitatif. Dataset yang digunakan adalah berita media sosial dari situs berita di Indonesia yang diperoleh dari <https://kumparan.com/> dimana dalam berita tersebut terdiri dari berita Hoax dan berita Fakta. Sumber data Hoax dan Fakta diambil dari data kumparan, dengan *string* pencarian “Cek Fakta” pada kolom pencarian berita di situs kumparan. Pada berita-berita yang mengandung “Cek Fakta”, isi beritanya mengandung berita hoax yang tersebar dan berita fakta yang membantah berita hoax yang sudah tersebar. Penelitian yang kami gunakan untuk pengumpulan data yaitu metode eksperimen. Metode yang dipilih pada penelitian ini berdasarkan kesesuaian data dengan metode pengklasifikasian teks yang paling baik, dan sudah digunakan oleh peneliti sebelumnya. Eksperimen data penelitian menggunakan dataset yang diambil menggunakan BeautifulSoup, *library* dari python. Selanjutnya dataset berita akan dijadikan data *training* dan data *testing* dan sistem sistem dibangun berbasis website menggunakan bahasa pemrograman python.

Untuk membangun model, kami menggunakan salah satu algoritma yang dapat melakukan klasifikasi data, yaitu algoritma Naive Bayes Classifier. Penelitian sebelumnya melakukan deteksi berita hoax Covid19 menggunakan algoritma K-Nearest Neighbor (KNN)[17], dan mendapatkan akurasi model akurasi sebesar 81%. Selain model yang digunakan pada penelitian sebelumnya berbeda, sumber data yang digunakan untuk deteksi hoax berita Covid19 pada penelitian sebelumnya juga berbeda. Penelitian sebelumnya mengumpulkan data yang bersumber dari Jabar Saber Hoaks dan Jala Hoaks.

Model yang sudah kami bangun, sudah dilakukan uji coba model untuk mengklasifikasikan berita menggunakan text yang bersumber dari dataset dengan label berita adalah “true”. Hasil uji coba text berita di dokumentasikan pada Gambar 7. Model yang dibangun sudah dilakukan evaluasi model dengan cara menghitung akurasi model untuk memastikan model naïve bayes sudah menghasilkan prediksi yang baik untuk verifikasi fakta, didapatkan total akurasi sebesar 81%. Proses penyusunan code untuk uji coba dapat dilihat pada alur berikut, dimana fungsi *predict* merupakan class model yang berfungsi untuk menghasilkan prediksi *text* berita :

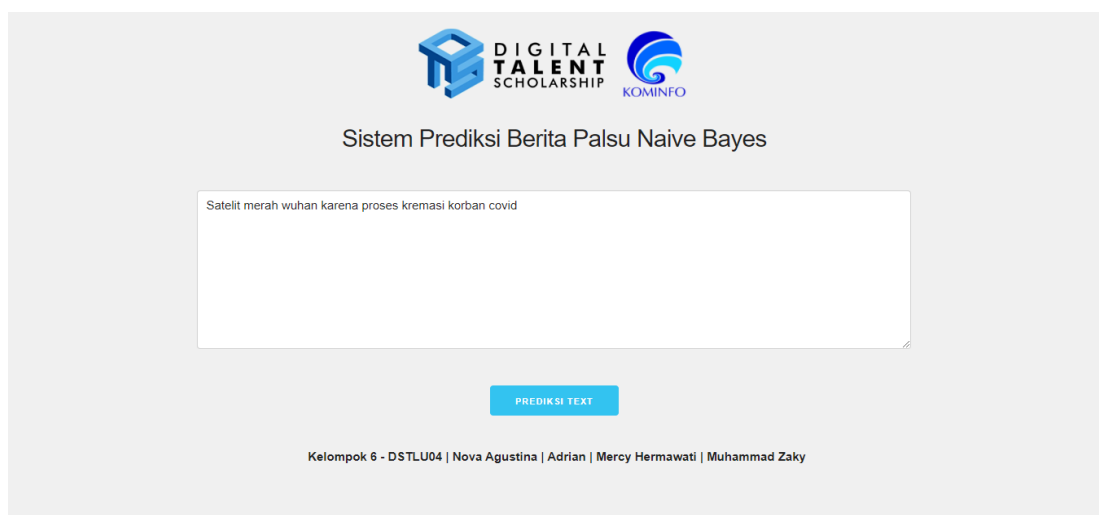
```
test = new_test.iloc[0]
test = "Fakta di Balik Merah Kota Wuhan dari Satelit, Sering Kremasi Hingga Larangan Buka Kantong Jenazah"
pre = pipeline.predict([test])
print(pre)
```

Selain akurasi, juga dihitung F1-Score pada masing-masing berita hoax dan fakta. Hasilnya, pada berita hoax didapatkan hasil F1-Score sebesar 87% dan pada berita fakta mendapatkan hasil F1-Score sebesar 60%. Disimpulkan bahwa model yang sudah dibangun memiliki akurasi yang cukup untuk dilanjutkan ke tahap berikutnya, yaitu *deploy* model dengan platform berbasis website.

	precision	recall	f1-score	support
fake	1.00	0.77	0.87	84
true	0.42	1.00	0.60	14
accuracy			0.81	98
macro avg	0.71	0.89	0.73	98
weighted avg	0.92	0.81	0.83	98

Gambar 5. Akurasi Model Naïve Bayes Classifier

Dibandingkan dengan hasil penelitian sebelumnya yang juga melakukan deteksi berita hoax Covid19 menggunakan algoritma K-Nearest Neighbor (KNN)[17], algoritma Naïve Bayes Classifier yang diterapkan pada penelitian ini menghasilkan akurasi sebesar 81%. Hasilnya 5.1% lebih besar dibandingkan deteksi berita hoax menggunakan KNN pada penelitian sebelumnya yang mendapatkan akurasi sebesar 75.89%.



Gambar 6. Tampilan Dashboard Prediksi Berita Palsu

Tampilan pada Gambar 7 adalah hasil *deployment* model Naïve Bayes Classifier ke Heroku yang sudah diuji. Hasil *deploy* dapat dilihat pada situs <https://hoaxindo-kel6-dstlu04.herokuapp.com/>. Pada tahap ini pengguna harus memberikan *input* berupa text berupa kalimat yang akan diprediksi kebenarannya.



Gambar 7. Tampilan Hasil Prediksi Berita Palsu

Gambar 7 adalah hasil text yang diuji yaitu “Satelit merah Wuhan karena proses kremasi korban covid”, data yang bersumber dari dataset berita tersebut memiliki label “Fake”. Hasil tersebut sesuai dengan prediksi dari sistem. Fakta data yang tersimpan pada dataset adalah Satelit merah di wuhan diakibatkan oleh perkiraan atau prediksi emisi sulfur dioksida di Wuhan berdasarkan data historis dan pola cuaca, bukan data satelit secara real-time.

4. PENUTUP

Algoritma Naïve Bayes Classifier digunakan sebagai metoda untuk melakukan klasifikasi data terutama untuk kebutuhan deteksi terhadap berita berita palsu/fakta. Tahapan yang telah dilakukan dimulai dengan pengumpulan data, proses tokenisasi, fase pemodelan, fase evaluasi, hingga fase *deployment*. Berdasarkan analisis model yang sudah dibuat dapat disimpulkan bahwa model cocok digunakan untuk prediksi berita palsu/ fakta dengan nilai F1-Score model yang diperoleh dengan keakurasian sebesar 81%, sehingga akurasi yang dihasilkan termasuk ke kelompok baik [16]. Model yang di-*deploy* berbasis web juga menghasilkan prediksi yang cukup sesuai. Untuk penelitian lebih lanjut disarankan jumlah dataset yang digunakan untuk membangun model lebih besar agar prediksi yang dihasilkan lebih maksimal.

Tentunya penggunaan metode Algoritma Naïve Bayes Classifier dapat diimplementasikan untuk keperluan lain dengan berbasis klasifikasi. Pada penelitian selanjutnya, kami sarankan melakukan prediksi Hoax dengan membandingkan metode klasifikasi lain, contohnya *Convolutional Neural Network* (CNN) dan *Pre-training of Deep Bidirectional Transformers for Language Understanding* (BERT). Eksplorasi ini dapat dilakukan pada dataset yang sama dalam studi masa depan, sehingga model terbaik dari hasil perbandingan dapat diukur.

DAFTAR PUSTAKA

- [1] A. Qayyum, J. Qadir, M. U. Janjua, and F. Sher, “Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News,” *IT Prof.*, vol. 21, no. 4, pp. 16–24, Jul. 2019, doi: 10.1109/MITP.2019.2910503.
- [2] Siska Amelie F Deil, “Gara-gara Kabar Hoax, Nilai Saham Bank Besar Ini Anjlok Parah,” *liputan6.com*, 2015. <https://m.liputan6.com/bisnis/read/2220228/gara-gara-kabar-hoax-nilai-saham-bank-besar-ini-anjlok-parah?page=2> (accessed Dec. 20, 2021).
- [3] P. Aria, “Terseret Hoax, Harga Saham Sari Roti Melorot Tipis,” *KataData*, 2016. <https://katadata.co.id/pingitaria/finansial/5e9a56bb33961/terseret-hoax-saham-sari-roti-melorot-066-persen> (accessed Dec. 20, 2021).
- [4] N. Newman, R. Fletcher, A. Kalogeropoulos, and R. Nielsen, *Reuters institute digital news report 2019*, vol. 2019. Reuters Institute for the Study of Journalism, 2019.
- [5] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, “Some like it Hoax: Automated fake news detection in social networks,” *CEUR Workshop Proc.*, vol. 1960, Apr. 2017, [Online]. Available: http://www.ghbook.ir/index.php?name=فرهنگ و رسانه های&option=com_dbook&task=readonline&book_id=13650&page=73&chckhashk=ED9C9491B4&I

- temid=218&lang=fa&tmpl=component.
- [6] Z. Saaya and T. W. Hong, "The development of trust matrix for recognizing reliable content in social media," *Int. J. Comput.*, vol. 18, no. 1, pp. 60–66, 2019.
 - [7] W. J. Tee and R. K. Murugesan, "Trust Network, Blockchain and Evolution in Social Media to Build Trust and Prevent Fake News," *Proc. - 2018 4th Int. Conf. Adv. Comput. Commun. Autom. ICACCA 2018*, pp. 1–6, 2018, doi: 10.1109/ICACCAF.2018.8776822.
 - [8] Kominfo, "Data Hoaks Terbaru," *kominfo.go.id*, 2021. <https://trustpositif.kominfo.go.id/pdfhoaks> (accessed Dec. 20, 2021).
 - [9] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media," *ACM SIGKDD Explor. Newsl.*, vol. 19, no. 1, pp. 22–36, 2017, doi: 10.1145/3137597.3137600.
 - [10] U. P. Atri Nurani, Budi Susanto, "Implementasi Naive Bayes Classifier Pada Program Bantu Penentuan Buku Referensi Matakuliah," *J. Inform.*, vol. 3, no. 2, 2011.
 - [11] M. Syukri Mustafa, M. Rizky Ramadhan, and A. P. Thenata, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Citec J.*, vol. 4, no. 2, pp. 151–162, 2017.
 - [12] Sergey Shelpuk, "CRISP-DM: a data science project methodology."
 - [13] F. F. Mailoa, "Metode Term Frequencies untuk penelitian kesehatan di Twitter : Studi pada tweet berbahasa Indonesia terkait Obesitas," *Ber. Kesehat. Masy.*, vol. 35, no. 4, pp. 1–4, 2019, doi: 10.2105/AJPH.2013.301860.Kim.
 - [14] L. Robinson, "Implementasi Metode Generalized Vector Space Model Pada Aplikasi Information Retrieval untuk Pencarian Informasi Pada Kumpulan Dokumen Teknik Elektro Di UPT BPI LIPI," *J. Ilm. Komput. dan Inform. (KOMPUTA)*, 2014.
 - [15] H. Dhika, F. Destiawati, and A. Fitriansyah, "Implementasi Algoritma C4. 5 terhadap Kepuasan Pelanggan," pp. 80–86, 2018, doi: 10.31227/osf.io/fgc7a.
 - [16] F. Gorunescu, *Data mining: Concepts, models and techniques*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
 - [17] E. Utami, A. F. Iskandar, W. Hidayat, and A. B. Prasetyo, "Covid-19 Hoax Detection Using KNN in Jaccard Space," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 15, no. 3, pp. 1–10, 2021, [Online]. Available: <https://journal.ugm.ac.id/ijccs/article/view/67392>.