JNATIA Volume 1, Nomor 1, November 2022
Jurnal Nasional Teknologi Informasi dan Aplikasinya

# Klasifikasi Hoax Menggunakan Algoritma *Naïve Bayes*

Riana Pramesti Putri[a1], Ngurah Agus Sanjaya ER[a2]

[a]Program Studi Informatika, Universitas Udayana
Bukit Jimbaran, Bali, Indonesia
[1]rianaprmst22@gmail.com
[2]agus_sanjaya@unud.ac.id

***Abstract***

*The use of social media which is so mushrooming today, has many positive impacts but does not cover the negative impacts, one of which is the misuse of information. Hoax is one of the causes of disinformation and public unrest. The speed of spread, which sometimes cannot be controlled, is one of the reasons why hoax news is still being spread every day. Therefore, it is necessary to classify hoax news with the aim of helping the public in separating the news that is being spread. This study uses the Naive Bayes algorithm as a classification model with the addition of hyperparameter tuning. The best model is produced with an alpha of 0.01 which has an accuracy of 87.9%*

***Keywords:*** *Hoax, Classification, Naïve Bayes, Hyperparameter Tuning*

## 1. Pendahuluan

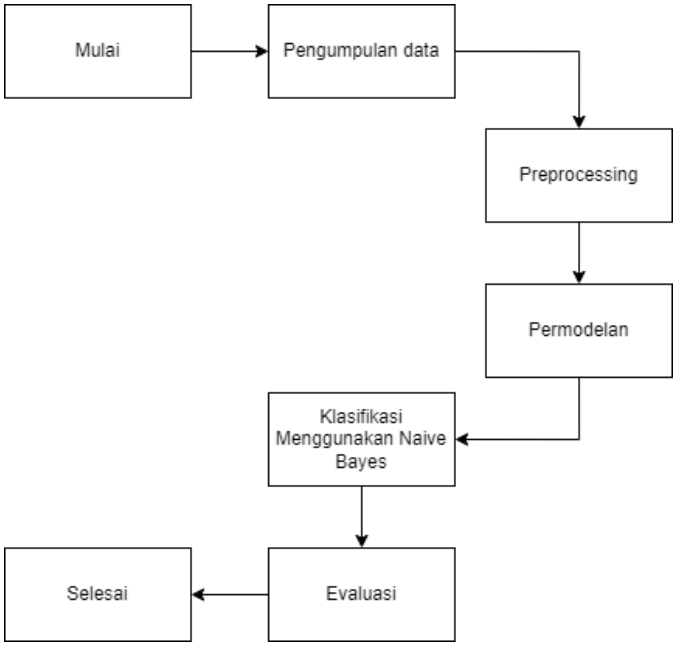
Berada pada era *society* 5.0 yang berarti bahwa masyarakat saat ini diharuskan untuk hidup berdampingan, menguasai, bahkan memanfaatkan teknologi yang tersedia. Ditambah lagi setelah melewati situasi pandemi Covid-19 yang lalu, menjadikan masyarakat dari segala rentan usia mulai terbiasa berkomunikasi menggunakan teknologi dimana komunikasi yang mulanya dapat dilakukan secara langsung antar individu, menjadi terbatas hanya melalui media online. Meningkatnya penggunaan serta kebutuhan akan teknologi dan kemampuannya untuk berkomunikasi, menjadikan tersedianya begitu banyak pilihan. Mulai dari TV, handphone, pc, dan masih banyak lagi dengan tambahan berbagai media sosial yang dapat mempermudah komunikasi. Menurut Nasrullah (2015), media sosial adalah sebuah medium di internet yang dapat memungkinkan penggunanya merepresentasikan diri sendiri maupun berinteraksi, bekerja sama, berbagi, maupun berkomunikasi dengan pengguna lain yang bertujuan untuk membentuk ikatan sosial secara virtual [1]. Media sosial juga merupakan salah satu sarana komunikasi tanpa batasan jarak yang kini penggunaan terus mengalami peningkatan seiring dengan berkembangnya alat komunikasi maupun ilmu pengetahuan manusia. Oleh karena itu, semakin banyak dan mudahnya juga setiap informasi yang belum pasti dapat masuk maupun keluar terhadap satu masyarakat dengan yang lainnya. Arus informasi yang begitu cepat ini, dapat menyebabkan terjadinya disinformasi atau dapat disalahgunakan untuk kepentingan sepihak melalui penyebaran hoax.

*Hoax* atau dengan arti lainnya yaitu berita bohong, dapat diartikan yaitu sebuah berita yang tidak memiliki sumber yang pasti atau berita palsu yang sengaja disebarluaskan untuk menciptakan kondisi dan situasi yang ricuh di masyarakat dengan tujuan tertentu [2]. Menurut hasil survey pada yang dilakukan oleh tim AIS Kementrian Komunikasi dan Informasi sampai bulan Juli 2021, telah ditemukan sebanyak 1400 *hoax* atau berita palsu yang tersebar dimasyarakat [3]. Penyebaran yang tidak terkoreksi, bahkan masyarakat terpelajar pun kesulitan untuk membedakannya, akhirnya akan berdampak pada hukum dan mampu memecah belah publik. Banyak dampak yang ditimbulkan akibat penyebaran *hoax* yang begitu cepat, selain disinformasi dan ricuhnya masyarakat. Terjadinya salah persepsi terhadap sesuatu dan memicu tindakan lanjutan, seperti kerusuhan, demo, dan masih banyak lagi bentuk perlawanan. Maka dari itu, penyebaran *hoax* ini harus diminimalisir, setidaknya dengan mempermudah masyarakat dengan membantu membedakan mana berita palsu atau berita asli yang dapat dipercaya.

Klasifikasi *Hoax* ini adalah salah satu cara untuk membantu masyarakat membedakan serta memisahkan sebuah berita, apakah berita tersebut masuk ke dalam golongan berita palsu *(hoax*) atau berita asli dengan sumber terpercaya. Penelitian sebelumnya, melakukan klasifikasi menggunakan algoritma Naïve Bayes dengan komponen library PHP-Machine Learning menghasilkan akurasi sebesar 82,6% [4]. Penelitian kali ini menggunakan logika *Naive Bayes* dengan menyempurnakan hasilnya menggunakan *hyperparameter tuning* dimana *dataset* yang digunakan adalah data berupa berita-berita yang tersebar dimasyarakat. Tujuan utama penelitian ini adalah untuk mengetahui hasil akurasi berupa persentase yang dapat dihasilkan dari klasifikasi *hoax* menggunakan metode yang berbeda yaitu permodelan *Naïve Bayes*.

.

## 2. Metodologi Penelitian




*Gambar 2.1 Alur Metode Penelitian*

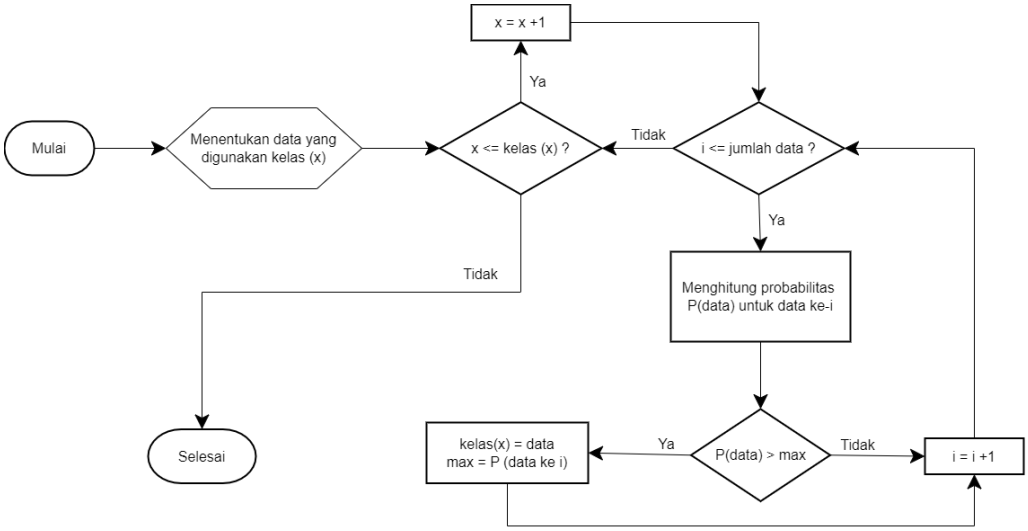
Pada gambar diatas dapat dilihat merupakan alur dari penelitian yang akan dilakukan. Diawali dengan mengumpulkan data. *Dataset* yang digunakan pada penelitian ini merupakan data kumpulan berita Indonesia yang berjenis supervised data yang diperoleh dari website bernama *Kaggle* yang diakses pada 25 September 2022. *Dataset* ini berjumlah 4701 data dengan jumlah label 5 diantaranya yaitu ID, tanggal, judul, narasi, dan nama file gambar sebagai berikut.

| | | | | |
|---|---|---|---|---|
| 238057 | 13-Jul-20 | Narasi Tito Karnavian Berideologi Komunis Karena Pernah Disekolahkan Partai Komunis China di Beijing | TITO KARNIVAN ITU BERIDIOLOGI KOMUNIS DIA BISA DI KATAKAN PKI KARENA DI PERNAH DI SEKOLAHLAH OLEH PA... | 238057.jpg |

*Gambar 2.2 Contoh label pada dataset*

Sesuai dengan alur penelitian diatas, penelitian akan diawali dengan melakukan *preprocessing* data yang bertujuan untuk menjadikan data yang digunakan lebih bersih dan siap digunakan untuk proses selanjutnya. Pada penelitian ini tahapan preprocessing yang digunakan adalah menghilangkan stopword, stemming, dan melakukan tokenisasi. Setelah melewati tahapan preprocessing, dilanjutkan dengan pembuatan model berupa vector kemudian dilakukan

pembentukan model Naïve Bayes. Algoritma pemodelan Naïve Bayes dapat dilihat pada diagram alir berikut.




*Gambar 2.3 Flowchart Naïve Bayes*

Setelah dibentuk model, akan dilakukan pengujian menggunakan data pembentuknya dan data testing untuk didapat hasil akurasi model Naïve Bayes. Setelah itu, untuk meningkatkan kualitas model klasifikasi dilakukan hyperparameter tuning yaitu dengan mencari nilai alpha terbaik.

## 3. Hasil dan Pembahasan

Penelitian ini menggunakan dataset yang berjumlah 4701 data dalam bentuk supervised data yang siap digunakan. Selanjutnya akan masuk kedalam tahapan preprocessing yang kemudian dilanjutkan dengan permodelan Naïve Bayes dan evaluasi untuk menghasilkan model terbaik yang dapat digunakan untuk klasifikasi. Untuk proses dari tahapan tersebut dapat dilihat sebagai berikut

### 3.1. Preprocessing data

Pada tahap ini data akan diolah lebih lanjut agar siap digunakan, setelah mengambilkan data dari library dan menampilkannya, dilanjutkan dengan melakukan Drop data yang bertujuan untuk memisahkan data yang tidak penting didapat hasilnya berupa emosi dan teks, selanjutkan akan dilakukan tahap preprocessing dimana menghilangkan Stopword serta melakukan stemming yaitu mengembalikan sebuah kata ke kata dasarnya. Dilanjutkan dengan melakukan tokenisasi yaitu memberikan token sesuai jumlah kalimat, kemudian dikembalkan lagi menjadi bentuk teks agar bisa digunakan kedalam model Naïve Bayes. Data pada penelitian ini menunjukkan jumlah yang tidak sama, maka disesuaikan agar jumlah data negative maupun positifnya sama. Data yang telah mengalami preprocessing, kemudian dibagi menjadi data training dan data testing.

### 3.2. Permodelan Naïve Bayes

Pemodelan dilakukan menggunakan data yang sebelumnya telah di preprocessing pada tahap sebelumnya. Sebelum data digunakan dalam pelatihan model, data teks akan dirubah ke bentuk vektor menggunakan prinsip bag of word. Model selanjutnya dibentuk menggunakan library sklearn seperti pada gambar berikut.

```
from sklearn.feature_extraction.text import CountVectorizer
vectorizer = CountVectorizer(max_features= 10000, min_df = 1, max_df = 0.1)
# fit the model for training data
X_train_data = vectorizer.fit_transform(X_latih)
X_train_data.shape
```

*Gambar 3.1 Permodelan Awal*

```
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
X1_train, X1_test, y1_train, y1_test = train_test_split(X_train_data, Y_latih, test_size=0.33,random_state = 2)
#X1_train_scratch, X1_test_scratch, y1_train_scratch, y1_test_scratch = train_test_split(X_latih, Y_latih, test_size=0.20,random_state = 15)
```

*Gambar 3.2 Permodelan Naïve Bayes*

Setelah pembentukkan model, model akan diuji untuk mengetahui akurasi dari model menggunakan confussion matrix. Setelah dilakukan pengujian didapat hasil akurasi pelatihan model Naïve Bayes sebesar 92% dan akurasi pengujian sebesar 83,9%. Selanjutnya setelah model dibentuk, dilakukan upaya untuk meningkatkan hasil akurasi klasifikasi model dengan melakukan Hyperparameter tuning untuk mencari nilai alpha terbaik.

```
from sklearn.model_selection import GridSearchCV
params = {'alpha': [0.01,0.1,0.2,0.3,0.5,0.6,0.7,0.8,0.9,1],
         }

multinomial_nb_grid = GridSearchCV(MultinomialNB(), param_grid=params, n_jobs=-1, cv=10, verbose=1)
multinomial_nb_grid.fit(X1_train,y1_train)
```

*Gambar 3.3 Hyperparameter tuning*

Setelah dilakukan hyperparameter tuning, model kembali diuji dengan data training dan data testing yang berjumlah masing-masing 70% dan 30% dari data total. Didapat hasil peningkatan terhadap akurasi model, dengan ini model Naïve Bayes menjadi lebih baik untuk digunakan saat klasifikasi nantinya. Perbandingan hasil akurasinya dapat dilihat pada gambar berikut.

```
Multinomial Naive Bayes model train accuracy(in %): 92.34088457389427
Multinomial Naive Bayes model test accuracy(in %): 83.39903635567237
```

*Gambar 3.4 Hasil Akurasi Sebelum Hyperparameter Tuning*

```
Fitting 10 folds for each of 10 candidates, totalling 100 fits
Train Accuracy : 0.960
Test Accuracy : 0.879
```

*Gambar 3.5 Hasil Akurasi Setelah Hyperparameter Tuning*

Berdasarkan penelitian yang telah dilakukan, didapatkan hasil akurasi model dengan data training menjadi 96% dan dengan data latih sebesar 87,9%. Melihat hasil penelitian sebelumnya yang telah dilakukan, berjudul Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia buatan Faisal Rahutomo, Inggrid Yanuar Risca Pratiwi, dan Diana Mayangsari Ramadhani tahun 2019, diketahui hasil akurasi sistemnya sebesar 82,6%, sehingga dapat disimpulkan bahwa model yang telah dibuat pada penelitian kali ini lebih baik dari penelitian sebelumnya [4]. Sedangkan berdasarkan penelitian sebelumnya juga yang berjudul Klasifikasi Berita Hoax Menggunakan Algoritma Naïve Bayes Berbasis PSO buatan Hegarmanah Muhabatin, Candi Prabowo, Irfan Ali, Cep Lukman Rohmat, dan Dita Rizki Amalia tahun 2021 mendapatkan hasil sebesar 91,82% yang ternyata jauh lebih besar dari penelitian ini [5]. Dapat disimpulkan bahwa model pada penelitian ini belum dapat lebih baik dari penelitian sebelumnya yang menggunakan PSO. Hal ini dapat disebabkan karena perbedaan metode yang digunakan.

## 4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa hyperparameter tuning nyatanya dapat meningkatkan akurasi model Naïve Bayes pada mulanya 83,3% menjadi 87,9%. Hal ini diperoleh setelah mendapatkan nilai alpha terbaik. Melalui peningkatan tingkat akurasi model Naïve Bayes ini, diharapkan dapat mempengaruhi setiap klasifikasi yang dilakukan menggunakan model tersebut. Selain itu, melalui penelitian ini juga, dapat diketahui bahwa masih banyak terdapat metode yang lebih baik untuk klasifikasi selain menggunakan algoritma *Naïve Bayes*.

## References

[1] A. Setiadi, "Pemanfaatan Media Sosial Untuk Efektifitas Komunikasi," *Jurnal Humaniora,* vol. 16, no. 2, pp. 1-7, 2016.

[2] J. E. Latupeirissa, J. D. Pasalbessy, E. Z. Leasa and C. Tuhumury, "Penyebaran Berita Bohong (HOAX) Pada Masa Pandemi Covid-19 dan Upaya Penanggulangannya di Provinsi Maluku," *Jurnal Belo,* vol. 6, no. 2, pp. 1-16, 2021.

[3] Direktorat Jenderal Aplikasi Informatika, "Laporan Isu Hoaks," Penjabat Pengelola Informasi Dan Dokumentasi Kementerian Komunikasi Dan Informatika, Jakarta, 2021.

[4] F. Rahutomo, I. Y. R. Pratiwi and D. M. Ramadhani, "Eksperimen Naive Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia," *Jurnal Penelitian Komunikasi dan Opini Publik,* vol. 23, no. 1, pp. 1-15, 2019.

[5] H. Muhabatin, C. Prabowo, I. Ali, C. L. Rohmat and D. R. Amalia, "Klasifikasi Berita Hoax Menggunakan Algoritma Naïve Bayes Berbasis PSO," *Informatics For Educators And Professionals,* vol. 5, no. 2, pp. 156-165, 2021.

Halaman ini sengaja dibiarkan kosong