

Klasifikasi Pendapat Masyarakat Terhadap Penggunaan Vaksin Dalam Mengantisipasi Covid-19 Menggunakan Teknik Sentiment Analisis Berbasis Naïve Bayes

Widya Damayanti, Bagus Setya Rintyarna*, Wiwik Suharso

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
widyadama61@gmail.com, bagus.setya@unmuhjember.ac.id
wiwiksuharso@unmuhjember.ac.id

Keywords:

Covid-19,
Naïve Bayes,
Multinomial Naïve Bayes,
Chi-Square

ABSTRACT

The Covid-19 pandemic is not over yet. Vaccinations that have been carried out so far are still getting pros and cons among the public. This study contains an analysis of public opinion on vaccination using the Naive Bayes-based method which was collected from the Twitter timeline from February 23 to March 9, 2021. The number of opinions that were successfully collected and used as data in this study were 1000 tweet data which were divided into 200 test data tweets. validation and 800 training data tweets. This study also uses Chi-Square as a feature selection method. The algorithms used in this research are Naive Bayes and Multinomial Naive Bayes. From the measurement results on the tests carried out, the Naive Bayes method obtained an accuracy value of 66%, precision of 79% and recall of 62%. While the Multinomial Naive Bayes method obtains an accuracy value of 71%, a precision of 81% and a recall of 68%. Based on the results obtained, the Multinomial Naive Bayes algorithm has a better performance than Naive Bayes.

Kata Kunci

Covid-19,
Naïve Bayes,
Multinomial Naïve Bayes,
Chi-Square

ABSTRAK

Pandemi Covid-19 masih belum usai. Vaksinasi yang dilakukan sampai saat ini juga masih mendapat pro dan kontra pada kalangan masyarakat. Penelitian ini berisi tentang analisis opini masyarakat terhadap vaksinasi menggunakan metode berbasis Naive Bayes yang dihimpun dari *timeline* Twitter sejak tanggal 23 Februari sampai 9 Maret 2021. Jumlah opini yang berhasil dihimpun dan dijadikan data dalam penelitian ini adalah 1000 data *tweet* yang dibagi menjadi 200 *tweet* data uji validasi dan 800 *tweet* data latih. Penelitian ini juga menggunakan *Chi-Square* sebagai metode seleksi fitur. Algoritma yang digunakan dalam penelitian ini adalah *Naive Bayes* dan *Multinomial Naive Bayes*. Dari hasil pengukuran pada pengujian yang dilakukan, metode *Naive Bayes* memperoleh nilai akurasi sebesar 66%, presisi sebesar 79% dan *recall* sebesar 62%. Sedangkan metode *Multinomial Naive Bayes* memperoleh nilai akurasi sebesar 71%, presisi sebesar 81% dan *recall* sebesar 68%. Berdasarkan hasil yang diperoleh, algoritma *Multinomial Naive Bayes* memiliki performa yang lebih baik daripada *Naive Bayes*.

Korespondensi Penulis:

Bagus Setya Rintyarna
Universitas Muhammadiyah Jember
Jl. Karimata, No. 49, Jember, Jawa Timur
Telepon : +62 857-5512-7381
Email: bagus.setya@unmuhjember.ac.id

1. PENDAHULUAN

Pandemic covid-19 yang melanda di hampir semua negara mengakibatkan banyak permasalahan termasuk Indonesia. Penyebaran virus yang begitu cepat dan tak kunjung usai mengakibatkan masyarakat cemas, paranoid, depresi dan frustrasi. Salah satu upaya dalam menanggulangi pandemic covid-19 ini, pemerintah melaksanakan vaksinasi terhadap masyarakat Indonesia. Namun, pro-kontra mewarnai program vaksin covid-19 yang sedang berlangsung di berbagai wilayah Indonesia. Semua hal dikomentari dan informasi yang belum jelas validasinya disebarluaskan begitu saja melalui media sosial, seperti twitter. Twitter sendiri merupakan salah satu media sosial yang kerap digunakan dalam menuangkan opini oleh masyarakat. Data yang dihasilkan dari twitter memiliki potensi dijadikan sebagai sumber data untuk melakukan pengekstrakan opini atau disebut analisis sentiment. Tentu saja informasi yang terkandung dalam tweet sangat berharga sebagai alat penentu kebijakan dan bisa dilakukan dengan text mining. Sayangnya dalam tweet-tweet belum bisa dijadikan acuan untuk menarik kesimpulan suatu bahasan di twitter, sehingga diperlukan klasifikasi untuk memudahkan pencarian informasi yang dibutuhkan.

Klasifikasi teks memiliki beberapa metode seperti contoh naïve bayes classifier dan multinomial naïve bayes. Pada penelitian ini menggunakan kedua metode tersebut karena metode tersebut hanya membutuhkan jumlah data pelatihan yang kecil dan memiliki tingkat akurasi yang tinggi. Sebelum dilakukan klasifikasi diperlukan pembobotan fitur untuk membantu meningkatkan akurasi klasifikasi. Pembobotan yang digunakan adalah Term frequency-inverse document frequency (TF-IDF). Pada pengklasifikasian teks memiliki masalah yaitu terdapat kata-kata yang tidak relevan atau disebut noise. Banyaknya kata yang mampu merepresentasikan sebuah dokumen terkadang bisa menjadi masalah ketika tingginya dimensi fitur yang ada. Untuk mencegah hal itu terjadi, dilakukan seleksi fitur untuk menyeleksi fitur-fitur yang relevan dalam merepresentasikan dokumen. Dalam penelitian ini memilih chi-square sebagai metode seleksi fitur. Pada seleksi fitur chi-square di dalamnya didasarkan pada teori statistik dari dua proses yaitu term fitur dan term kategori.

Salah satu contoh penelitian yang dilakukan oleh Riandra dkk, mereka melakukan penelitian dengan menggunakan metode survei di kelurahan Pangkalan Jati Kota Depok. Kegiatan yang mereka lakukan yaitu terjun langsung ke masyarakat untuk survei mengenai tanggapan terhadap anjuran vaksin covid-19. Ternyata hasil dari survei banyak masyarakat yang menolak untuk melakukan vaksinasi. Berbagai berita hoax mengenai dampak buruk pasca vaksinasi yang menimbulkan keresahan masyarakat, sehingga Riandra dkk melakukan penelitian dengan memberikan pengertian terhadap dampak hukum apa saja jika menolak melakukan vaksinasi covid-19. Survei yang dilakukan di daerah tersebut juga memperlihatkan bahwa masyarakat belum sepenuhnya memahami dampak hukum penolakan vaksinasi. Setelah diadakan penelitian tersebut, kenaikan jumlah masyarakat yang siap menerima vaksin mencapai 92.3% [3].

Penelitian kedua dilakukan oleh Hendri Naufal dkk. Berbeda dari penelitian Riandra, di penelitian Hendri dkk justru responden malah lebih kritis dalam melihat berita bohong tentang vaksin covid-19. Mereka tidak serta merta menerima semua berita yang beredar di media sosial. Meskipun berita bohong sudah beredar di mana-mana, kebanyakan responden cenderung melakukan pengecekan fakta terlebih dahulu sehingga menghentikan adanya persebaran berita bohong ke orang lain. Hampir setengah dari jumlah responden menyatakan bahwa mereka akan berpikir dua kali untuk menerima vaksinasi covid-19 setelah melihat simulasi pemberian berita bohong yang ada di media sosial. Hal ini juga didukung oleh penurunan jumlah responden yang bersedia menerima vaksin baru dibandingkan dengan jumlah responden yang belum menerima berita bohong mengenai vaksinasi covid-19. Pada akhirnya responden masih mempercayai Lembaga resmi pemerintah sebagai salah satu sumber informasi mengenai vaksinasi covid-19 [4].

Penelitian selanjutnya dilakukan oleh Yuni Lasmita tentang penerimaan vaksin di wilayah Puskesmas Alang-Alang. Dari hasil survei yang didapat, masyarakat yang menerima vaksin covid-19 hanya 43.1% dan yang belum menerima sebanyak 56.9%. Hal itu dipengaruhi oleh hubungan antara pengetahuan, ketersediaan akses informasi dan dukungan keluarga terhadap penerimaan vaksinasi covid-19. Keadaan ini dapat mempengaruhi tingkat pengetahuan responden mengenai program vaksinasi covid-19 [5].

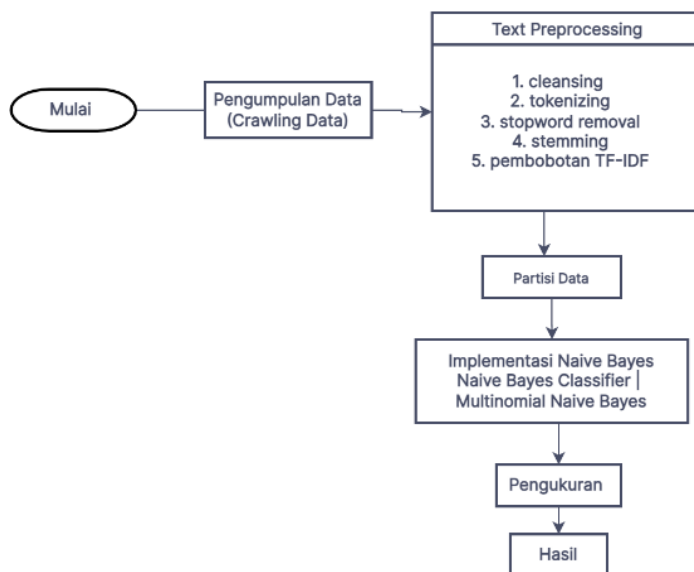
Penelitian terakhir dilakukan oleh Fajar Hidayatullah tentang survei aktivitas dan kebiasaan serta tingkat resikonya dalam menghadapi wabah covid-19. Dari hasil survei yang didapat, responden dengan kebiasaan dan aktivitas yang beresiko tinggi tergolong sedikit. Namun pengumpulan data perlu dilakukan lagi melihat perkembangan keadaan di Indonesia terus berkembang dan jumlah terinfeksi covid terus

meningkat. Peningkatan jumlah terinfeksi covid pada akhir 2019 menunjukkan masih perlu banyak dilakukan sosialisasi terkait pencegahan penyebaran covid-19 di kalangan masyarakat [6].

Berdasarkan fakta yang telah dituliskan di atas, pada penelitian ini akan dibahas tentang sentimen analisis masyarakat Indonesia pada Twitter terhadap vaksinasi covid-19 menggunakan metode *Naive Bayes* dan *Multinomial Naive Bayes* serta dipadukan dengan seleksi fitur menggunakan metode *Chi-Square*. Pada penelitian ini juga digunakan teknik skenario uji menggunakan *K Fold Cross Validation*.

2. METODE PENELITIAN

Berikut ini metode penelitian yang dilakukan dalam penelitian ini :



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data Twitter yang diunduh selama 14 hari dimulai pada tanggal 23 Februari 2021 sampai 9 Maret 2021. Kata kunci yang digunakan dalam *crawling* ini adalah “vaksin” dan “covid-19”.

2.2 Text Preprocessing

Pada tahap *text preprocessing* ini teknik yang digunakan antara lain:

1. *Cleansing*, yaitu menghapus simbol dan karakter selain huruf atau *link*.
2. *Tokenizing*, yaitu mengubah kalimat menjadi per kata.
3. *Sopword removal*, yaitu menghapus kata yang tidak memiliki pengaruh pada sentimen.
4. *Stemming*, yaitu mengembalikan kata yang terdapat imbuhan menjadi kata dasar.
5. *TF-IDF*, yaitu menghitung kemunculan suatu kata pada suatu kalimat sampai keseluruhan dokumen lalu nilai kebalikan atau *invers* pada suatu dokumen.

2.3 Partisi Data

Total data pada penelitian ini adalah 1000 *tweet*. Data akan dibagi menjadi data latih dan data uji dengan persentase 20% data uji dan 80% data latih. Dari pembagian ini diketahui 200 data akan digunakan menjadi data uji validasi dan 800 data akan digunakan menjadi data latih yang di dalamnya akan digunakan teknik *K Fold Cross Validation* sebagai skenario uji dalam pemodelannya.

2.4 Implementasi *Naïve Bayes*

Dalam implementasi *Naive Bayes* terbagi menjadi dua bagian yaitu pemodelan dan pengujian validasi. Pada tahap pemodelan digunakan metode *Naive Bayes* dan *Multinomial Naive Bayes* serta *K Fold Cross Validation* sebagai skenarionya dengan nilai *k fold* adalah 2, 4, 5, 8 dan 10.

3. HASIL DAN ANALISIS

Pada proses ini akan dibagi menjadi dua yaitu pemodelan dan pengujian validasi.

3.1 Pemodelan

Pemodelan dilakukan menggunakan metode *Naive Bayes* dan *Multinomial Naive Bayes*. Data yang digunakan sebanyak 800 *tweet*. Berikut hasil pemodelan yang diperoleh dari metode *Naive Bayes* dan *Multinomial Naive Bayes*.

3.1.1. *Naive Bayes*

Pemodelan ini menggunakan skenario uji *K Fold Cross Validation* dengan nilai *k fold* 2, 4, 5, 8 dan 10. Berikut hasil pengukuran tingkat akurasi, presisi dan *recall*.

Tabel 1. Hasil Pemodelan Pada Metode *Naive Bayes*

<i>K fold</i>	akurasi	presisi	<i>recall</i>
2	73.75	83.71	67.14
4	73.25	83.47	66.51
5	73.87	83.95	67.3
8	73.75	83.85	67.14
10	73.88	84.01	67.29

Pemodelan pada metode *Naive Bayes* yang dipaparkan pada tabel 1 di atas, diketahui bahwa model dengan akurasi terbaik yaitu pada uji *fold* 10 dengan nilai rata-rata 73.88%, dan presisi dengan nilai rata-rata 84.01% dan nilai rata-rata *recall* 67.29%.

3.1.2. *Multinomial Naive Bayes*

Pemodelan ini menggunakan skenario uji *K Fold Cross Validation* dengan nilai *k fold* 2, 4, 5, 8 dan 10. Berikut hasil pengukuran tingkat akurasi, presisi dan *recall*.

Tabel 2. Hasil Pemodelan Pada Metode *Multinomial Naive Bayes*

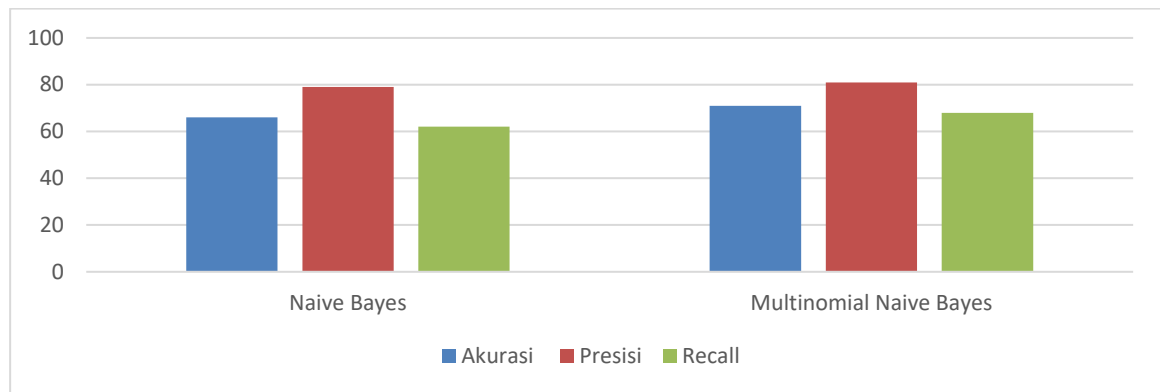
<i>K fold</i>	akurasi	presisi	<i>recall</i>
2	76.25	82.32	70.77
4	76.12	82.62	70.61
5	76.25	82.92	70.77
8	76.12	82.72	70.6
10	76.12	82.84	70.6

Pemodelan pada metode *Multinomial Naive Bayes* yang dipaparkan pada tabel 2 di atas, diketahui bahwa model dengan akurasi terbaik yaitu pada uji *fold* 5 dengan nilai rata-rata 76.25%, dan presisi dengan nilai rata-rata 82.92% dan nilai rata-rata *recall* 70.77%.

3.2 Pengujian

Pengujian dilakukan pada 20% data atau 200 *tweet*. Berikut hasil pengujian pada metode *Naive Bayes* dan *Multinomial Naive Bayes*. Gambar 2 di bawah ini dapat dijelaskan yaitu, hasil pengujian data validasi menggunakan metode *Naive Bayes* memperoleh nilai akurasi sebesar 66%, nilai presisi sebesar 79% dan nilai *recall* sebesar 62%. Sedangkan pengujian menggunakan metode *Multinomial Naive Bayes* memperoleh nilai akurasi sebesar 71%, nilai presisi sebesar 81% dan nilai *recall* sebesar 68%.

Dari hasil pengujian, dapat disimpulkan bahwa metode *Multinomial* lebih baik dalam mengklasifikasi data sentimen walaupun kedua metode ini mengalami penurunan pada tahap pengujian jika dibandingkan dengan hasil pengukuran pada pemodelan.



Gambar 2. Hasil Pengujian Pada Metode *Naïve Bayes* dan *Multinomial Naïve Bayes*

3.3 Implementasi sistem

Tools yang digunakan dalam penelitian ini adalah Rstudio untuk crawling data tweet dan Jupyter Notebook Python untuk proses analisis.

Gambaran data

Membaca data hasil validasi sentimen dalam bentuk csv data yang diambil hanya sentimen positif dan negatif

```
In [2]: dataset = pd.read_csv('data_siap1.csv')
data_fix = dataset[(dataset["sentimen"] == 'p') | (dataset["sentimen"] == 'n')]
```

```
In [3]: data_fix.to_csv('data fix dipakai.csv', index=False)
```

```
In [4]: dataset.shape
```

```
Out[4]: (1286, 3)
```

```
In [5]: data_fix
```

```
Out[5]:
```

	screen_name	text	sentimen
0	Ridwan83079388	*PENCEGAHAN, PENYEBARAN VIRUS CORONA DI MAKO ...	p
1	Ridwan83079388	*OPERASI YUSTISI UNTUK PENDISCIPLINAN PEMUTUSA...	p
2	irzanfajar_	@tsanymu covid kan takut sama presiden weh	n
3	theiconomic	Kolaborasi menjadi kunci penting untuk menyele...	p
5	CibeberPolsek	Personil Polsek Cibeber Polres Lebak Brigadir ...	p
...	...	...	...
1281	kedisneyland	mama stres abis daftar vaksin trs takut kena j...	n
1282	koplopilled	@Slamet_1488_109 @DaiohBot Jangan ambil vaksin...	n
1283	Auji_Zaharudin	Vaksin tak percaya, anggap mental illness ni g...	n
1284	jauharulfh	Ya bebanin ke pelaku lah bosssss, dah dapet va...	n
1285	tapeuyaki	wkwkw bener2. w lagi ngantri vaksin. bisa2nya ...	n

1000 rows x 3 columns

Gambar 3. Data Sentiment Positif dan Negative

Cleaning Result :

Out[13]:

	screen_name	text	sentimen
0	Ridwan83079388	pencegahan penyebaran virus corona di mako po...	p
1	Ridwan83079388	operasi yustisi untuk pendisiplinan pemutusan ...	p
2	irzanfajar_	covid kan takut sama presiden weh	n
3	theiconomic	kolaborasi menjadi kunci penting untuk menyele...	p
5	CibeberPolsek	personil polsek cibeber polres lebak brigadir ...	p
...	...	...	...
1281	kedisneyland	mama stres abis daftar vaksin trs takut kena j...	n
1282	koplopilled	jangan ambil vaksin sinovac	n
1283	Auji_Zaharudin	vaksin tak percaya anggap mental illness ni gi...	n
1284	jauharulfh	ya bebanin ke pelaku lah bosssss dah dapet vak...	n
1285	tapeuyaki	wkwkw bener lagi ngantri vaksin bisa nya atas...	n

1000 rows x 3 columns

Gambar 4. Hasil Proses Cleansing

melakukan tokenizing

```
In [14]: # import word_tokenize & FreqDist from NLTK
from nltk.tokenize import word_tokenize
##nltk.download()

##data_hasil_cleansing

# ----- TOKENIZE -----

# NLTK word tokenize
def word_tokenize_wrapper(text):
    return word_tokenize(text)

data_fix['text_token'] = data_fix['text'].apply(word_tokenize_wrapper)
# data_fix['text_token'] = data_fix['text'].fillna('')
print('Tokenizing Result : \n')
print(data_fix['text_token'])
print('\n\n')
data_fix.to_csv('1_data_hasil_tokenizing.csv', index=False)
data_fix.shape
```

Tokenizing Result :

```
0      [pencegahan, penyebaran, virus, corona, di, m...
1      [operasi, yustisi, untuk, pendisiplinan, pemu...
2      [covid, kan, takut, sama, presiden, weh]
3      [kolaborasi, menjadi, kunci, penting, untuk, m...
5      [personil, polsek, cibeber, polres, lebak, bri...
...
1281   [mama, stres, abis, daftar, vaksin, trs, takut...
1282   [jangan, ambil, vaksin, sinovac]
1283   [vaksin, tak, percaya, anggap, mental, illness...
1284   [ya, bebanin, ke, pelaku, lah, bosssss, dah, d...
1285   [wkwkw, bener, lagi, ngantri, vaksin, bisa, ny...
Name: text_token, Length: 1000, dtype: object
```

Gambar 5. Proses Tokenizing

Multinomial Naive Bayes

```
In [74]: hasil5b = cross_val_score(clf4, X_train2, y_train, cv = 5, n_jobs = -1)
# print('Akurasi Naive Bayes: ', round(pd.DataFrame(hasil5b),4)*100)
# print('Rata-rata akurasi :', round(hasil5b.mean()*100,4))
from sklearn import metrics
scores5b = cross_val_score(
    clf4, X_train2, y_train, cv=5, scoring='precision_macro')
recalls5b = cross_val_score(
    clf4, X_train2, y_train, cv=5, scoring='recall_macro')

# print('Presisi Naive Bayes: ',round(pd.DataFrame(scores5b), 4)*100)
# print('Rata-rata Presisi Naive Bayes: ',round(scores5b.mean()*100,4))
# print('Recall Naive Bayes: ',round(pd.DataFrame(recalls5b), 4)*100)
# print('Rata-rata Recall Naive Bayes: ',round(recalls5b.mean()*100,4))
hasil6 = {
    'akurasi' : hasil5b*100,
    'presisi' : scores5b*100,
    'recall' : recalls5b*100
}

hasil6 = round(pd.DataFrame(hasil6),2)
hasil6
```

```
Out[74]:
```

	akurasi	presisi	recall
0	75.62	83.90	69.33
1	80.00	85.14	75.16
2	76.25	79.75	71.61
3	78.12	81.99	73.70
4	71.25	83.80	64.06

Gambar 6. Implementasi Multinomial Naïve Bayes

Naive Bayes

```
In [76]: hasil8a = cross_val_score(clf3, X_train2, y_train, cv = 8, n_jobs = -1)
# print('Akurasi Naive Bayes: ', round(pd.DataFrame(hasil8a),4)*100)
# print('Rata-rata akurasi :', round(hasil8a.mean()*100,4))
from sklearn import metrics
scores8a = cross_val_score(
    clf3, X_train2, y_train, cv=8, scoring='precision_macro')
recalls8a = cross_val_score(
    clf3, X_train2, y_train, cv=8, scoring='recall_macro')

# print('Presisi Naive Bayes: ',round(pd.DataFrame(scores8a), 4)*100)
# print('Rata-rata Presisi Naive Bayes: ',round(scores8a.mean()*100,4))
# print('Recall Naive Bayes: ',round(pd.DataFrame(recalls8a), 4)*100)
# print('Rata-rata Recall Naive Bayes: ',round(recalls8a.mean()*100,4))
hasil7 = {
    'akurasi' : hasil8a*100,
    'presisi' : scores8a*100,
    'recall' : recalls8a*100
}

hasil7 = round(pd.DataFrame(hasil7),2)
hasil7
```

```
Out[76]:
```

	akurasi	presisi	recall
0	74.0	85.06	66.67
1	74.0	85.06	66.67
2	75.0	85.29	68.75
3	74.0	84.88	67.50
4	75.0	80.54	69.58
5	75.0	85.29	68.75
6	73.0	81.37	66.67

Gambar 7. Implementasi Naïve Bayes Classifier

4. KESIMPULAN

Dari serangkaian proses yang dilakukan pada penelitian ini, berikut poin-poin yang dapat disimpulkan:

Hasil dari pengukuran pada pengujian validasi metode *Naive Bayes* yaitu memiliki tingkat akurasi sebesar 66%, tingkat presisi sebesar 79% dan *recall* sebesar 62%. Sedangkan hasil pengukuran pada pengujian validasi metode *Multinomial Naive Bayes* yaitu memiliki tingkat akurasi sebesar 71%, tingkat presisi sebesar 81% dan tingkat *recall* sebesar 68%.

Hasil perbandingan terhadap metode *Naive Bayes* dan *Multinomial Naive Bayes* yang diperoleh yaitu tingkat akurasi pada metode *Naive Bayes* lebih rendah daripada hasil metode *Multinomial Naive Bayes* dengan selisih 5%. Pada pengukuran tingkat presisi metode *Multinomial* lebih unggul daripada metode *Naive Bayes* dengan selisih 2%. Pada pengukuran tingkat *recall* metode *Multinomial* lebih unggul daripada metode *Naive Bayes* dengan selisih 6%.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada para dosen pembimbing dan Universitas Muhammadiyah Jember yang telah mendukung dan membantu dalam menyelesaikan penelitian ini serta teman-teman yang selalu memberikan bantuan untuk memperlancar penelitian ini.

REFERENSI

Bruno Trstenjaka, Sasa Mikac, Dzenana Donkoc. 2013. KNN with TF-IDF Based Framework for text Categorization, *Procedia Engineering* 69 (2014) 1356 – 1364.

Dirkareshza, Rianda, et al. "PENDAMPINGAN MASYARAKAT MENGENAI DAMPAK HUKUM PENOLAKAN VAKSINASI DI MASA PANDEMI COVID-19." *JMM (Jurnal Masyarakat Mandiri)* 5.5s (2021): 2812-2823.

Hidayatullah, Fajar, Fadibah Setiawan, and Farida Megalini. "Survei Aktivitas Dan Kebiasaan Masyarakat Serta Tingkat Resikonya Dalam Menghadapi Wabah COVID-19 Di Indonesia." *Civic-Culture: Jurnal Ilmu Pendidikan PKN dan Sosial Budaya* 4.1 Extra (2020): 17-31.

Kompas.Com, 2019. Pengguna Aktif Harian Twitter Indonesia Diklaim Terbanyak [Online] (Updated 30 October 2019) Available At: <https://Tekno.Kompas.Com/Read/2019/10/30/16062477/PenggunaAktif-Harian-Twitter-Indonesia-Diklaim-Terbanyak>. (Diakses Pada Tanggal 19 Januari 2021).

Lasmita, Yuni, Misnaniarti Misnaniarti, and Haerawati Idris. "ANALISIS PENERIMAAN VAKSINASI COVID-19 DI KALANGAN MASYARAKAT." *Jurnal Kesmas (Kesehatan Masyarakat) Khatulistiwa* 8.4 (2021): 195-204.

Liu, M., Yang, J., 2012) An Improvement of TFIDF weighting in text categorization. *International Proceedings of Computer Science and Information Technology*, 47, pp. 44 -47

Marbella, Hendry Naufal, et al. "Analisis Pengaruh Berita Bohong di Sosial Media Terhadap Keputusan Masyarakat Indonesia Melakukan Vaksinasi Covid-19." *Jurnal Indonesia Sosial Teknologi* 2.11 (2021): 1951-1966. [4]

Nur, M. Y., & Santika, D. D. (2011). Analisis Sentimen Pada Dokumen Berbahasa Indonesia Dengan Pendekatan Support Vector Machine. *Konferensi Nasional Sistem Dan Informatika*.

RINTYARNA, BAGUS SETYA, et al. "Mapping acceptance of indonesian organic food consumption under COVID-19 pandemic using sentiment analysis of Twitter dataset." *Journal of Theoretical and Applied Information Technology* 99.05 (2021).

Rintyarna, Bagus Setya. "(Dokumen) Evaluating The Performance of Sentence Level Features and Domain Sensitive Features of Product Reviews on Supervised Sentiment Analysis Task." *Journal of Big Data* (2019).

Rintyarna, Bagus Setya. "Peer Review dan Similarity Artikel" Automatic Ranking System of University based on Technology Readiness Level Using LDA-Adaboost. MH". *International Conference on Information and Communication Technology*. AMIKOM Yogyakarta, 2018.