

Implementasi *Naïve Bayes Classifier* Dan *Information Gain* Pada Analisis Sentimen Penggunaan *E-Wallet* Saat Pandemi

*Amelia Isnanda¹⁾, Yuyun Umaidah²⁾, Jajam Haerul Jaman³⁾

¹²³Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

Correspondence author: Amelia Isnanda, amelia.isnanda17049@student.unsika.ac.id,

DOI: <https://doi.org/10.37012/jtik.v7i2.648>

Abstrak

Abstrak. *World Health Organization* (WHO) pada Maret 2020 mengategorikan *Coronavirus* sebagai pandemi setelah mewabah ke berbagai negara belahan dunia. Indonesia yang terdampak *pandemi* menetapkan peraturan Pembatasan Sosial Berskala Besar (PSBB) seperti *physical distancing*, wajib memakai masker, peraturan jam malam hingga sekolah dan belajar dari rumah untuk mengurangi kemungkinan penyebaran virus. Sejak kebijakan PSBB diberlakukan, penggunaan *e-wallet* meningkat karena diyakini mengurangi kontak fisik antar sesama sehingga mengundang berbagai tanggapan masyarakat mengenai kehadirannya. Tanggapan tersebut dituangkan melalui media sosial yang salah satunya adalah Twitter. Terdapat tanggapan yang menyetujui kehadiran *e-wallet* karena merasakan manfaatnya, namun ada yang beranggapan bahwa penggunaannya sulit atau tidak aman. Berbagai tanggapan tersebut dapat dianalisis untuk mengetahui berbagai tanggapan masyarakat mengenai penggunaan *e-wallet* di masa pandemi. Maka, dilakukan analisis sentimen Twitter mengenai penggunaan *e-wallet* pada masa pandemi dengan *Naïve Bayes* dan dikombinasikan menggunakan seleksi fitur *Information Gain*. Hasil klasifikasi menggunakan *Naïve Bayes Classifier* menghasilkan *accuracy* 84%, *precision* 91% dan *recall* 91%. Hasil klasifikasi menggunakan *Naïve Bayes* dan *Information Gain* adalah *accuracy* 92%, *precision* 92% dan *recall* 100%.

Kata Kunci: *text mining, pandemi, Naïve Bayes Classifier, Chi-Square, Information Gain*

Abstract

Abstract. *The coronavirus that emerged in 2019 was declared in March 2020 as a pandemic by World Health Organization (WHO) after it spread to various countries. Indonesia which has been affected by coronavirus has established Large-Scale Social Restrictions regulations called PSBB, such as social distancing, the obligatory wearing of a mask, curfew regulations to school, and study from home to decrease the positive case. Since the PSBB policy was implemented, the use of e-wallet has increased because it is believed to reduce physical contact between people, thus inviting various public responses regarding its presence. The responses were poured through social media such as Twitter. Some responses agree with the presence of e-wallet because they feel the benefits, but some think that its use is difficult or unsafe. Those responses can be analyzed to find out various public responses regarding the use of e-wallet during the pandemic. Thus, a Twitter sentiment analysis regarding the use of ewallet during the pandemic using Naïve Bayes Classifier with Information Gain for selecting the features. The result of classification using Naïve Bayes is accuracy 84%, precision 91% and recall 91%. The result of Naïve Bayes and Information Gain is accuracy 92%, precision 92% and recall 100%.*

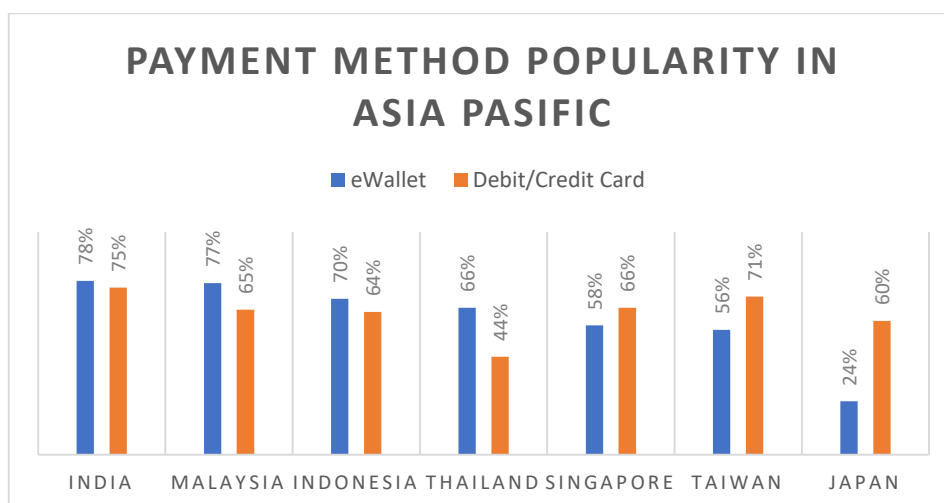
Keyword: *text mining, pandemi, Naïve Bayes Classifier, Chi-Square, Information Gain*

PENDAHULUAN

Presiden Indonesia, Joko Widodo mengumumkan bahwa pada 2 Maret 2020 terdapat kasus positif virus Covid-19 untuk pertama kalinya di Indonesia yang menjangkit dua warga asal Depok (Rachman & Pramana, 2020). Dilakukanlah upaya pencegahan penyebaran virus yang salah satunya adalah pemberlakuan Pembatasan Sosial Berskala Besar (PSBB) seperti

physical distancing, *work from home*, wajib memakai masker dan mencuci tangan, pemberlakuan jam malam, hingga wajib isolasi mandiri bagi masyarakat yang bepergian antar kota dan negara selama 14 hari (Olivia et al., 2020).

Upaya lainnya yaitu mengurangi penggunaan uang tunai sebagai alat transaksi dan diganti menjadi transaksi non-tunai (*cashless*) yang salah satunya adalah *e-wallet*. *E-wallet* atau dompet elektronik dapat diartikan sebagai alat pembayaran non-tunai dengan menggunakan media elektronik (Maulana & Fauzan, 2020). Terdapat data statistik hasil survey yang dilakukan oleh *Rapyd* pada tahun 2020 untuk membandingkan metode pembayaran antara *e-wallet* dan *debit/credit card* di negara-negara Asia Pasifik. Survey dilakukan dengan melibatkan 3500 responden pada tahun 2020 dan Gambar 1 menunjukkan hasil bahwa metode pembayaran di Indonesia lebih didominasi oleh *e-wallet*.



Gambar 1. Statistik penggunaan *e-wallet*

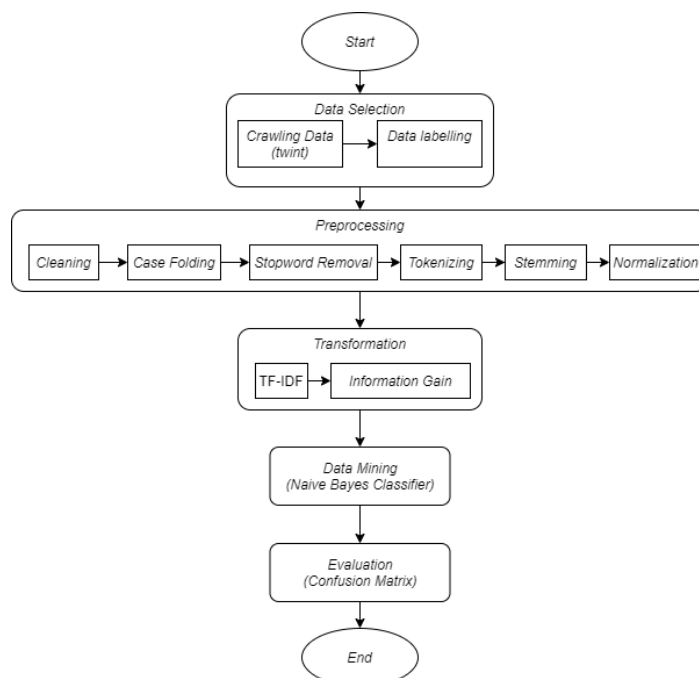
Trend penggunaan *e-wallet* ketika pandemi Covid-19 memicu berbagai tanggapan dari masyarakat Indonesia yang diutarakan melalui media sosial Twitter. Masyarakat yang menyambut baik kehadiran *e-wallet* sebagai transaksi non-tunai berpendapat bahwa *e-wallet* dapat membantu menekan penyebaran Covid-19 di Indonesia, sementara itu terdapat masyarakat yang berpendapat bahwa penggunaan *e-wallet* memberikan kesulitan bagi masyarakat yang tidak terlalu mengerti teknologi. Penyampaian pendapat melalui media sosial Twitter tersebut dapat dianalisis untuk menentukan subjektivitas opini sehingga hasil analisis dari *tweet* tersebut disebut sebagai sentimen analisis (Wongkar & Angdresy, 2019).

Analisis sentimen telah banyak digunakan sebagai penelitian menggunakan berbagai algoritma *machine learning* seperti penelitian mengenai analisis *review* film menggunakan KNN dan *Information Gain* untuk menyeleksi fitur, hasil dari penelitian tersebut

menyatakan bahwa performa *K-Nearest Neighbor* (KNN) memiliki akurasi 60% dan akurasi meningkat menjadi 96.8% dengan menggunakan seleksi fitur *Information Gain* (Adiwijaya, 2020). Berdasarkan beberapa permasalahan yang sudah dipaparkan tersebut, Maka penelitian ini akan menggunakan algoritma *Naïve Bayes Classifier* dan *Information Gain* algoritma seleksi fitur untuk mengoptimalkan performa *Naïve Bayes Classifier*.

METODE PENELITIAN

Penelitian dilakukan menggunakan metode KDD (*Knowledge Discovery in Database*) yang meliputi beberapa langkah seperti pengumpulan data, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam dataset berukuran besar yang hasilnya dapat digunakan untuk memperbaiki pengambilan keputusan di masa depan (Sari et al., 2020). Tahapan penelitian dengan metode KDD dapat dilihat pada Gambar 2.



Gambar 2. Tahapan Metode KDD

1. *Data Selection*

Data selection merupakan tahap awal penelitian dimana pada tahap ini data hasil seleksi yang berasal dari sekumpulan data operasional disimpan dalam suatu berkas. Jenis data pada penelitian berupa data teks, yaitu *tweet* pengguna Twitter mengenai sentimen atau opini penggunaan *e-wallet* pada masa pandemi Covid-19 dari Agustus 2020 sampai Mei 2021. Data diambil menggunakan *tool twint* yang dapat diakses menggunakan bahasa pemrograman *python*.

2. *Preprocessing*

Preprocessing dilakukan dengan tujuan menghilangkan *noise*, menyeragamkan bentuk kata dan mengurangi volume kata. Pada tahap ini terdiri dari proses *case folding*, *cleaning*, *stemming*, *tokenizing*, *stopword removal* dan *normalization*.

3. *Transformation*

Transformation merupakan proses yang pengaplikasiannya dilakukan dalam bentuk *coding*, selain itu juga menyediakan fungsi lain untuk menemukan beberapa pola dalam proses *data mining*. Pada tahap ini dilakukan pembobotan kata dengan TF-IDF. *Term Frequency Inverse Document frequency* (TF-IDF) ialah algoritma ekstraksi teks dengan memberi nilai pada teks tersebut. TF mengkalkulasi kemunculan kata pada dokumen sedangkan IDF mencari tingkat kepentingan kata pada dokumen (Syakuro, 2017). Nilai TF diketahui dari frekuensi kemunculan fitur t pada dokumen d . Nilai IDF dihasilkan dari logaritma jumlah dokumen N dibagi jumlah dokumen DF_t yang terdapat fitur t . Nilai TF-IDF (w_t) didapatkan dari hasil perkalian TF dan IDF (Hormansyah & Utama, 2018).

$$TF_{t,d} = F(t, d)$$

$$IDF_t = \log \frac{N}{DF_t}$$

$$w_t = TF_{t,d} \times IDF_t$$

Setelah TF-IDF dilakukan penyeleksian fitur dengan algoritma *Information Gain*. Teknik ini dilakukan untuk memilih fitur terpenting untuk mengidentifikasi data. Persamaan *information Gain* dapat dilihat dibawah ini:

$$Info(D) = - \sum_{i=1}^m (P_i) \log_2(P_i)$$

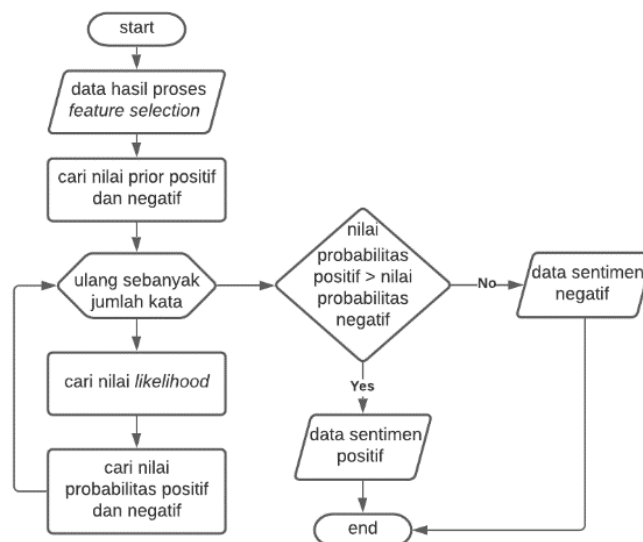
m merepresentasikan banyaknya kelas dan P_i merupakan probabilitas bahwa sampel acak di partisi D termasuk dalam kelas C_i .

4. *Data Mining*

Proses menganalisis sentimen pada penelitian menggunakan algoritma *Naïve Bayes Classifier*. Teorema Bayes memiliki konsep *supervised learning* (Putra Nuansa, 2017). *Naive Bayes Classifier* merupakan metode pembelajaran probabilistik, konsepnya adalah setiap kata memiliki probabilitas kemunculan yang independen yang artinya pengaruh nilai atribut kategori tidak akan terpengaruh juga tidak akan mempengaruhi nilai atribut lainnya. Teorema Bayes merupakan teorema dengan konsep probabilitas bersyarat, yang dapat diekspresikan dengan rumus berikut:

$$P(A|B) = \frac{P(A|B)P(A)}{P(B)}$$

Pada penelitian ini, terdapat label kelas sentimen positif dan negatif. Setiap *tweet* yang sudah melewati hasil *preprocessing* dan seleksi fitur selanjutnya akan dicek arti kata yang mengandung kata negatif dan positif. Langkah pada tahap *data mining* disajikan dalam diagram alir berikut.



Gambar 3. Proses Data Mining

5. Evaluation

Tahap ini merupakan penentuan dari performa hasil pengolahan *data mining* apakah sesuai dengan fakta atau hipotesa yang sebelumnya atau malah bertentangan. Pengukuran validitas klasifikasi dilakukan dengan metode *Confussion Matrix* yang melakukan perhitungan nilai *accuracy*, *recall* dan *precision*. *Confussion Matrix* menggunakan data berupa matriks untuk mengukur performa menggunakan data yang berisi informasi hasil klasifikasi data sebenarnya yang dilakukan oleh algoritma klasifikasi (Santra & Christy, 2012).

HASIL DAN PEMBAHASAN

Data Selection

Data yang didapatkan setelah proses pengumpulan data menggunakan *library twint* berjumlah 1235. Selanjutnya dilakukan penyeleksian data dengan menghapus *tweet* yang tidak berbahasa Indonesia, menghapus *tweet* yang sama dan menghapus *tweet* yang tidak membahas topik *e-wallet* (*out of topic*). Jumlah data setelah dilakukan penyeleksian adalah

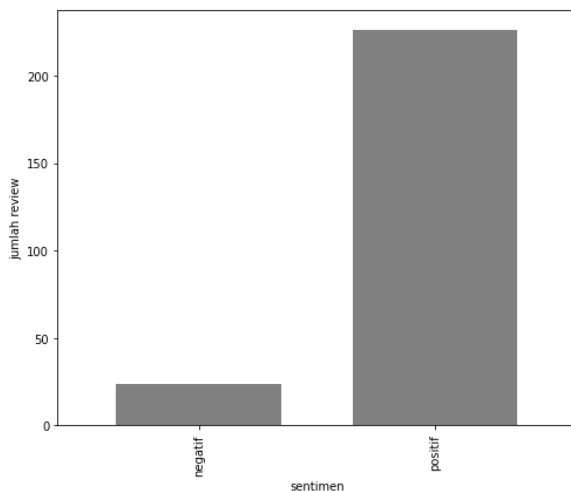
250 data. Data diberikan label positif dan negatif secara manual dengan bantuan seorang pakar bahasa Indonesia. Data yang dilabeli positif merupakan data yang menyetujui dan memberikan komentar baik mengenai penggunaan *e-wallet* saat pandemi Covid-19. Sebaliknya, data yang dilabeli negatif merupakan data yang tidak menyetujui atau mengeluh pada penggunaan *e-wallet* saat pandemi. Gambar 4 menampilkan data yang sudah diberi label dengan nama kolom *value*.

	tweet	value
0	@linkaja Di pandemi gini buat keluar kemana-ma...	positif
1	@gowjek Selama masa pandemi ini saya sudah mal...	positif
2	@maricyaaahhh pandemi tidak menghalangi thr fr...	positif
3	Masyarakat beralih ke layanan digital saat pan...	positif
4	@tiket dan @ShopeePay_ID bekerja sama memberi ...	positif
...
246	Transaksi Gopay Melonjak Selama Pandemi Covid-...	positif
247	@BerburuSale GOPAY 6□*□□□¥*6□□□□□□¥*6□□□ ud...	positif
248	Dah dibalesin juga tadi pagi sama CSnya @triin...	negatif
249	@kangenbilang @PenjahatGunung @metahillry Eman...	positif
250	@subtanyarI konsul ke psikiater kak aku udh 2x...	positif

251 rows x 2 columns

Gambar 4. Hasil *Data Selection*

Diketahui bahwa sentimen positif dengan jumlah 226 data lebih banyak daripada sentimen negatif dengan jumlah 24 data. Perbandingan sentimen positif dan negatif divisualisasikan dalam bentuk diagram pada Gambar 5.



Gambar 5. Perbandingan Sentimen

Preprocessing

Data hasil pengumpulan tidak akan langsung digunakan pada proses *data mining* dengan menggunakan algoritma karena masih berupa data kotor yang terdapat banyak gangguan (*noise*). Maka proses *preprocessing* sangat diperlukan dengan tujuan membersihkan data

dan menyeleksi atribut yang kurang berpengaruh pada proses klasifikasi. Tabel berikut menunjukan contoh tahap *preprocessing* yang telah dilakukan.

Tahap preprocessing	Hasil
Dataset	@linkaja Di pandemi gini buat keluar kemana-mana itu susah mau ga mau tetep harus keluar.Tapi setelah tau fitur ewallet salah satunya @linkaja akhirnya buat beli perlengkapan sekolah lain #lengkapnyauntuksemua
Cleaning	linkaja Di pandemi gini buat keluar ke mana mana itu susah mau ga mau tetep harus keluar Tapi setelah tau fitur ewallet salah satunya linkaja akhirnya buat beli perlengkapan sekolah lengkap nya untuk semua
Case Folding	linkaja di pandemi gini buat keluar ke mana mana itu susah mau ga mau tetep harus keluar tapi setelah tau fitur ewallet salah satunya linkaja akhirnya buat beli perlengkapan sekolah lengkap nya untuk semua
Stemming	linkaja di pandemi gini buat keluar ke mana mana itu susah mau ga mau tetep harus keluar tapi setelah tau fitur ewallet salah satu linkaja akhir buat beli lengkap sekolah lengkap nya untuk semua
Stopword Removal	linkaja pandemi keluar mana susah mau keluar tau fitur ewallet linkaja akhir beli lengkap sekolah lengkap
Tokenizing	linkaja, pandemi, keluar, mana, susah, keluar, tau, fitur, ewallet, linkaja, beli, lengkap, sekolah, lengkap
Normalization	linkaja, pandemi, keluar, mana, <u>sulit</u> , keluar, tahu, fitur, ewallet, linkaja, beli, lengkap, sekolah

Transformation

Transformation dilakukan menggunakan metode TF-IDF untuk mengubah struktur data. Data *tweet* yang berbentuk *text* diolah menggunakan TF-IDF sehingga menghasilkan nilai. Pertama akan dilakukan *splitting data* dengan persentase perbandingan 90% *data training* dengan jumlah 225 data dan 10% *data testing* sejumlah 25 data. Setelah *splitting data*, selanjutnya perhitungan *term* pada setiap *tweet* dan dilanjutkan dengan pembobotan TF-IDF. Gambar 6 merupakan hasil dari pembobotan TF-IDF.

	abis	ada	adain	admin	agar	air	ajabanyak	ajar	aje	akal	akhir	akibat	akses	aktif	aktifitas	aktivasi	aktivitas	aku	akun	alam	alas	alat	alfaeapres		
0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	
1	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.201413	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
...
220	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
221	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
222	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
223	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.253908	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.341236	0.0	0.0	0.306767	0.0	0.0	0.0	0.0	0.0
224	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0

225 rows x 885 columns

Gambar 6. Hasil TF-IDF

Selanjutnya dilakukan proses penyeleksian fitur. Proses seleksi fitur dengan algoritma *Information Gain* dilakukan menggunakan *threshold* sebesar 0,0008 dan menghasilkan

sejumlah 885 fitur berhasil dihilangkan karena tidak terlalu memberikan pengaruh pada tahap klasifikasi sehingga tersisa 452 fitur. Gambar 6 menunjukkan hasil penyeleksian fitur.

```

Threshold: 0.0008.
Previous number of features: 885
Number of features after information gain: 452
    
```

Gambar 4. Hasil *Information Gain*

Data Mining dan Evaluation

Pada *data mining* dilakukan klasifikasi sentimen penggunaan *e-wallet* saat pandemi dengan rasio 90% *data training* dan 10% *testing*. Gambar 9 menampilkan hasil dari klasifikasi menggunakan *Naïve Bayes*. Berdasarkan hasil tersebut, dapat diketahui bahwa *accuracy* 84%, *precision* 91% dan *recall* 91%.

```

Confusion Matrix:
[[ 1  2]
 [ 2 20]]
    
```

	precision	recall	f1-score	support
negatif	0.33	0.33	0.33	3
positif	0.91	0.91	0.91	22
accuracy			0.84	25
macro avg	0.62	0.62	0.62	25
weighted avg	0.84	0.84	0.84	25

AKURASI Before Fitur Selection: 0.840

Gambar 9. Evaluasi Klasifikasi *Naïve Bayes*

Gambar 10 menampilkan hasil dari klasifikasi menggunakan *Naïve Bayes* dan seleksi fitur *Information Gain*. Berdasarkan hasil tersebut, dapat diketahui bahwa *accuracy* 92%, *precision* 92% dan *recall* 100%.

Hasil Klasifikasi Naive Bayes + Information Gain (90:10):

	precision	recall	f1-score	support
0	1.00	0.33	0.50	3
1	0.92	1.00	0.96	22
accuracy			0.92	25
macro avg	0.96	0.67	0.73	25
weighted avg	0.93	0.92	0.90	25

```

Confusion Matrix:
[[ 1  2]
 [ 0 22]]
    
```

Gambar 10. Evaluasi Klasifikasi *Naïve Bayes* dan *Information Gain*

Sehingga masing-masing performa dari hasil klasifikasi dapat dibandingkan dalam bentuk Tabel 10.

Tabel 10. Hasil *Data Mining*

Hasil Klasifikasi 90:10		Actual			
		Naïve Bayes		Naïve Bayes + IG	
		Positif	Negatif	Positif	Negatif
prediction	Positif	1	2	1	2
	Negatif	2	20	0	22

Disimpulkan bahwa *data testing* 10 % dan 90% *data training*, hasil klasifikasi menggunakan *Naïve Bayes* adalah 20 data kelas negatif benar diprediksi sebagai kelas negatif, namun 2 kelas negatif diprediksi sebagai kelas positif. Sedangkan untuk 1 kelas positif berhasil diprediksi sebagai kelas positif dan 2 kelas positif diprediksi sebagai kelas negatif. Klasifikasi *Naïve Bayes* dengan kombinasi *Information Gain* menghasilkan 22 kelas negatif diprediksi dengan benar, 2 kelas negatif diprediksi sebagai kelas positif. Sedangkan untuk kelas positif, 1 data diprediksi dengan benar.

KESIMPULAN DAN SARAN

Sentimen pada aplikasi Twitter mengenai penggunaan *e-wallet* saat pandemi dapat dianalisis menggunakan *Naïve Bayes* menjadi kelas positif dan negatif. Evaluasi dan performa proses *data mining* menggunakan *Naïve Bayes* menghasilkan *accuracy* 84%, *precision* 91% dan *recall* 91%. Evaluasi dan performa proses *data mining* menggunakan *Naïve Bayes* dan *Information Gain* adalah *accuracy* 92%, *precision* 92% dan *recall* 100%. Sehingga dapat disimpulkan bahwa penambahan *Information Gain* sebagai untuk menyeleksi fitur dapat membantu meningkatkan performa *Naïve Bayes Classifier*. Pada penelitian selanjutnya disarankan data yang digunakan dapat dilakukan dengan kuantitas yang lebih banyak lagi dan pembagian kelas positif dan negatif yang seimbang dan membandingkan lebih dari satu rasio *splitting data* antara data training dan data testing.

REFERENSI

- Adiwijaya, N. O. F. D. (2020). Sentiment Analysis on Movie Reviews Using Information Gain and K-Nearest Neighbor. *Journal of Data Science and Its Applications*, 3(1), 1–007. <https://doi.org/10.34818/JDSA.2020.3.22>
- Hormansyah, D. S., & Utama, Y. P. (2018). Aplikasi Chatbot Berbasis Web Pada Sistem

- Informasi Layanan Publik Kesehatan Di Malang Dengan Menggunakan Metode Tf-Idf. *Jurnal Informatika Polinema*, 4(3), 224. <https://doi.org/10.33795/jip.v4i3.211>
- Maulana, A., & Fauzan, D. N. A. (2020). *Penggunaan E-Money Dalam E-Commerce di Indonesia*. 1–9.
- Olivia, S., Gibson, J., & Nasrudin, R. (2020). Indonesia in the Time of Covid-19. *Bulletin of Indonesian Economic Studies*, 56(2), 143–174. <https://doi.org/10.1080/00074918.2020.1798581>
- Putra Nuansa, E. (2017). Analisis Sentimen Pengguna Twitter Terhadap Pemilihan Gubernur Dki Jakarta Dengan Metode Naïve Bayesian Classification Dan Support Vector Machine. *Institut Teknologi Sepuluh Nopember Surabaya*, 1–101.
- Rachman, F. F., & Pramana, S. (2020). Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter. *Health Information Management Journal ISSN*, 8(2), 2655–9129. <https://inohim.esaunggul.ac.id/index.php/INO/article/view/223>
- Santra, A. K., & Christy, C. J. (2012). Genetic Algorithm and Confusion Matrix for Document Clustering. *International Journal of Computer Science Issues*, 9(1), 322–328.
- Sari, D., Panggabean, O., Buulolo, E., & Silalahi, N. (2020). *Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda*. 7(1), 56–62. <https://doi.org/10.30865/jurikom.v7i1.1947>
- Syakuro, A. (2017). Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG) Halaman Judul Skripsi Oleh : Abdan Syakuro. *Analisis Sentimen Masyarakat Terhadap E-Commerce Pada Media Sosial Menggunakan Metode Naive Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG)*, 1–89.
- Wongkar, M., & Angdresey, A. (2019). Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter. *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*, 1–5. <https://doi.org/10.1109/ICIC47613.2019.8985884>