

NBC BERBASIS EKSTRAKSI FITUR UNTUK ANALISIS SENTIMEN MASYARAKAT TERHADAP FLUKTUASI HARGA SAHAM

Tazki Yatun Niyah¹, Yuni Yamasari²

^{1,2}S1 Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya

tazki.17051204039@mhs.unesa.ac.id¹

yuniyamasari@unesa.ac.id²

Abstrak— Saat ini, banyak orang memulai investasi karena ingin mencapai kebebasan finansial. Investasi juga penting untuk melindungi aset dari penurunan nilai akibat inflasi. Investasi saham termasuk salah satu jenis investasi yang populer di Indonesia. Karena harganya yang terjangkau, investasi saham juga sangat mudah dilakukan dan sifatnya *fleksibel*. Berkaitan dengan investasi saham, banyak orang mencari informasi melalui media sosial, yang salah satunya adalah instagram. Banyak akun instagram yang membuat konten tentang edukasi saham yang sangat bermanfaat. Akun tersebut biasanya melakukan update tentang fluktuasi Indeks Harga Saham Gabungan (IHSG). Keadaan inilah yang menjadi penguat bahwa perlu dilakukannya penelitian ini. Penelitian ini fokus pada analisis sentimen masyarakat terhadap fluktuasi IHSG berdasarkan komentar instagram. Komentar pada media sosial ini perlu diklasifikasi karena beberapa unggahan yang sedang trend memantik ratusan komentar. Teknik klasifikasi yang diterapkan adalah teknik *Naïve Bayes Classifier* (NBC) dengan ekstraksi fitur N-Gram. Pengujian dilakukan menggunakan 2 cara yaitu pengujian menggunakan *k-fold cross validation* dan parameter dengan *unigram*, *bigram* dan *trigram*. Hasil ujicoba memperlihatkan *cross-validation* menghasilkan akurasi tertinggi sebesar 97%. Hasil ini dicapai pada penerapan *bigram* dengan $k=8$ dan *trigram* dengan $k=2$, $k=4$, $k=5$. Untuk pengujian parameter, level akurasi tertinggi sekitar 97% dicapai pada penerapan *trigram* dengan $\alpha = 0,2$ dan $\alpha = 0,3$.

Kata Kunci: Saham, Naïve Bayes Classifier, N-Gram, Instagram

I. PENDAHULUAN

Kemajuan inovasi data berkembang pesat saat ini. Perkembangan teknologi dapat dilihat dengan tingginya pengguna internet khususnya di wilayah Indonesia. Semua aktivitas yang kita lakukan di zaman

komputerisasi ini tidak bisa terhindar dengan internet dan media sosial. Berdasarkan laporan terbaru *We Are Social*, pada tahun 2020 tercatat terdapat 175,4 juta pengguna internet di Indonesia. Dibandingkan dengan tahun lalu, ada peningkatan 17% atau 25 juta pengguna internet di negeri ini. Berdasarkan jumlah penduduk Indonesia yang mencapai 272,1 juta jiwa, itu membuktikan bahwa 64% setengah dari penduduk Indonesia sudah pernah mengakses internet[1]. Media sosial berperan sebagai panggung yang sangat efektif untuk dimanfaatkan sebagai penyebaran data, yang bisa menjadi tempat menuangkan karya, pemikiran, reaksi, pendapat, pertunjukan bahkan media dapat digunakan dalam mengekspresikan apa yang sedang berlangsung.

Salah satu media sosial yang terkenal saat ini adalah instagram. Popularitas instagram telah berkembang sejak diluncurkannya pada tahun 2010 lalu. Dengan pengguna aktif lebih dari 500 miliar, media terpopuler kedua setelah facebook adalah instagram[2].

Saham sedang menjadi sorotan publik belakangan ini. Banyak *public figure* atau *influencer* sekaligus investor membagikan kondisi portofolio saham mereka, entah saham yang menghasilkan keuntungan ataupun sebaliknya. Posisi sebagai *public figure* menuntut mereka untuk berperilaku baik sekaligus menjadi panutan bagi para fans atau pengikut instagramnya. Sehingga hal ini tidak mengherankan banyak pengikut instagram mereka yang ikut-ikutan atau istilah zaman sekarang *FOMO* (*Fear of missing out*) ingin investasi saham. Hal lain yang menarik adalah banyak akun instagram yang membuat konten tentang edukasi seputar saham yang sangat bermanfaat. Sehingga akun instagram tersebut juga banyak diikuti oleh banyak pengguna instagram lainnya dan akun tersebut biasanya *update* mengenai naik turunnya IHSG.

Didalam dunia saham, investor sudah tidak asing lagi dengan sebutan IHSG. IHSG merupakan

kependekan dari Indeks Harga Saham Gabungan yaitu indeks yang mencakup pergerakan harga saham perusahaan yang tercatat di Indonesia.

Menurut Jogiyanto (2013: 47), "Indeks Harga Saham Gabungan merupakan angka indeks harga saham yang sudah disusun dan dihitung sehingga menghasilkan trend. Dimana, angka indeks adalah angka yang diolah sedemikian rupa sehingga angka tersebut dapat digunakan untuk membandingkan kejadian berupa perubahan harga saham dari waktu ke waktu". Zulbiadi Latief (Analisis.co.id, 2008) berpendapat "IHSG adalah indeks untuk seluruh saham yang diperdagangkan di BEI, yang mencerminkan kecenderungan perkembangan dan nilai rata-rata keseluruhan saham dari emiten yang ada di Indonesia.

Pentingnya IHSG bagi investor adalah sebagai alat analisis kondisi pasar terkini, sebagai penanda arah pergerakan pasar modal dan sebagai bahan pertimbangan yang cukup akurat untuk melakukan transaksi jual beli saham perusahaan.

Banyak hal menyebabkan naik turunnya harga saham yang akan berpengaruh terhadap naik turunnya IHSG. Variabel penyebab naik turunnya harga saham berasal dari dalam atau luar perusahaan. Komponen dalam adalah sebagai berikut: 1). Kegiatan korporasi perusahaan: semua bentuk pengaturan yang ditetapkan oleh administrasi perusahaan untuk akuisisi, merger, *stock split* dan *right issue*. 2). Hal-hal penting perusahaan yang baik maupun yang buruk. 3). Eksekusi perusahaan di masa depan. Sedangkan, komponen luar meliputi : 1). Masalah keuangan skala besar ; tingkat ekspansi, kolaps dan kenaikan suku bunga Bank Indonesia merupakan komponen yang dapat mempengaruhi naik turunnya harga saham yang nantinya akan mempengaruhi IHSG. 2). Nilai tukar rupiah terhadap dollar AS. 3). Serangan panik ; informasi selingkung politik dan ekonomi tertentu yang membuat investor menjadi panik[3].

Penelitian ini melakukan klasifikasi sentimen masyarakat terhadap fluktuasi Indeks Harga Saham Gabungan berdasarkan komentar instagram. Komentar instagram hendaknya dikelompokkan menggunakan *machine learning*. Proses ini perlu dilakukan karena dalam waktu yang tidak sebentar dan dalam unggahan yang sedang populer memiliki ratusan atau ribuan komentar pada media sosial instagram. Hal ini mengakibatkan proses klasifikasi membutuhkan waktu lama dan sulit secara manual. Penelitian ini menerapkan teknik klasifikasi NBC dengan ekstraksi fitur N-Gram

Terpilihnya perhitungan algoritma NBC dalam penelitian ini adalah karena sesuai untuk informasi dan konten yang singkat. Keuntungan lainnya adalah fitur ini memiliki tingkat ketepatan dan eksekusi yang tinggi dalam klasifikasi teks [4].

Ekstraksi fitur N-Gram digunakan sebagai parameter mempercepat proses ekstraksi teks. Hal ini

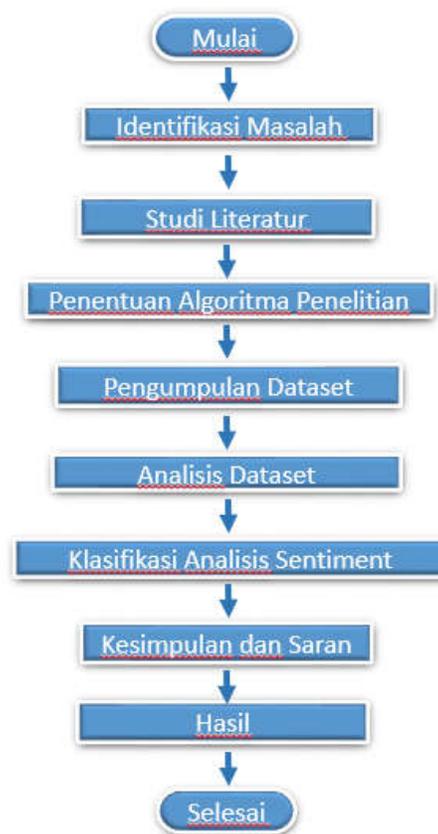
didasarkan pada penelitian sebelumnya yang menerapkan NBC untuk klasifikasi Berita Lokal Radar Malang dengan karakteristik N-Gram. Tingkat akurasi yang didapat adalah 78.66% [5].

Sesuai masalah yang telah digambarkan maka penulis akan melakukan penelitian dengan menerapkan teknik NBC dengan ekstraksi fitur N-Gram untuk klasifikasi sentimen masyarakat terhadap fluktuasi harga saham berdasarkan komentar instagram. Dengan dilakukannya penelitian ini, kita diharapkan bisa memperhatikan seberapa luas transmisi atau dampak investasi saham melalui instagram.

II. METODOLOGI

Tahapan adalah suatu bentuk yang dicapai untuk melakukan penelitian agar tahapan penelitian bisa diatur dengan tepat dan efisien agar mencapai harapan sesuai yang diinginkan. Tahapan metodologi penelitian dapat dilihat pada Gambar 1.

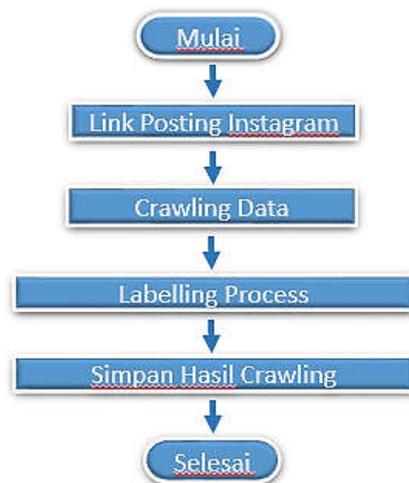
Teknik NBC dengan ekstraksi fitur N-Gram digunakan dalam penelitian ini. Ekstraksi yang dilakukan adalah unigram, bigram dan trigram. Untuk teknik pengujian, penelitian ini menerapkan pengujian *k-fold cross validation* dan pengujian parameter.



Gbr 1. Diagram Alur Penelitian

A. Pengumpulan Data

Pengumpulan data dilakukan bertujuan untuk mengklasifikasi analisis sentimen, pada penelitian ini akan digunakan dataset yang berdasar dari komentar media sosial instagram, komentar didapatkan dari unggahan instagram terkait fluktuasi indeks harga saham gabungan yang berdasar dari beberapa akun yang dapat diandalkan antara lain yaitu `idx_channel`, `ngertisaham`, `kuliahsaham` lalu komentar instagram akan di *crawling*. Proses *crawling* data dapat diuraikan seperti pada Gambar 2.



Gbr 2. Tahapan *Crawling* Data

Informasi komentar diambil dari media sosial instagram menggunakan link unggahan beberapa akun terkait dengan fluktuasi harga saham dan selanjutnya dilakukan proses *crawling* menggunakan bahasa pemrograman *python* kemudian disimpan dalam bentuk file format `.csv` dari data tersebut kemudian dilanjutkan tahap pra-pemrosesan atau *pre-processing*.

Pra-pemrosesan adalah suatu proses dalam pembersihan informasi yang dilakukan dalam penelitian. Dataset diambil dari komentar postingan di instagram masih mempunyai noise sehingga pra-pemrosesan dilakukan untuk membersihkan dataset tersebut supaya dataset yang digunakan pada saat training nantinya merupakan data yang masuk ke dalam kategori yang baik. Berikut proses yang terdapat pada pra-pemrosesan :

1. *Case Folding*

Case Folding adalah metode pengubahan semua huruf besar menjadi huruf kecil. Dalam hal ini, perubahan dilakukan pada kumpulan informasi komentar instagram yang diubah menjadi huruf kecil semua.

2. *Cleaning*

Cleaning adalah metode pembersihan kata yang tidak berdampak sama sekali terhadap munculnya klasifikasi asumsi. Dalam hal ini, informasi dibersihkan dengan cara menghapus informasi atau atribut yang tidak berguna, misalnya menghilangkan username (@), hastag (#), link atau URL ('http', 'bit.ly'), karakter symbol (~!#\$%^&*()_+?<>.,:;{}[])) dan lain lain.

3. *Tokenizing*

Tokenizing berupa penentuan konten yang dibagi menjadi unit-unit kecil yang disebut sebagai *token* atau *term* seperti kata atau angka[6].

4. *Stopword removal*

Stopword removal merupakan metode menghilangkan kata yang tidak sesuai dengan subjek laporan. Jika kata tersebut menghilangkan makna kalimat secara total, dapat diartikan kata tersebut tidak mempunyai arti. Kata ini lebih sering berbentuk kata sambung atau kata umum yang tidak sesuai. Dalam hal ini, perlu menghilangkan kata yang termasuk dalam *stopwords* sehingga tahapan pengelompokkan lebih cepat karena banyaknya kata yang diproses tidak banyak.

5. *Stemming*

Stemming adalah metode pengubahan kata gabungan menjadi kata esensial sesuai dengan KBBI. Susunan ini menghilangkan *prefix*, *postfix* dan kombinasi.

B. N-Gram

N-Gram bisa berupa *substring* dengan panjang *n* karakter dari sebuah string. N-Gram dapat menjadi prosedur yang terkait untuk membuat kata atau karakter. N-Gram digunakan dalam memulihkan potongan karakter numerik dari kata yang koherensinya dianggap dari substansi sumber hingga kesimpulan catatan.

N-Gram mencirikan kombinasi kata-kata tetangga atau panjang huruf *n* dalam pengaturan tertentu. N-Gram mewakili karakter *n* atau kelompok kata (dicirikan sebagai gram yang menunjukkan struktur bahasa) yang mengikuti satu sama lain. N-Gram banyak digunakan untuk mengantisipasi kata lain berdasarkan N-1 kata lampau[7]. N-Gram sering digunakan dalam pencarian informasi dan tugas penyusunan kata. Diambil contoh, jika "Harga saham sedang naik" itu artinya $n=4$ (jumlah kata dalam kalimat. Kata ini digunakan untuk mencatat seberapa sering kata-kata mengikuti satu sama lain. N-Gram bisa di simbolkan seperti persamaan berikut:

$$N\text{-Grams}_k = X - (N-1) \quad (1)$$

Pada tahap *text processing*, fitur N-Gram juga digunakan dalam melakukan analisis sentimen. N-Gram digunakan untuk menggabungkan kata deskriptif yang terjadi secara umum atau pembobotan kata-kata penting dalam sebuah catatan untuk memunculkan perkiraan. Dalam hal ini, dilakukan teknik N-Gram yang menangkap representasi bahasa dari huruf atau kata yang diberikan dalam struktur masukan (judul). Penelitian ini menggunakan tiga macam teknik N-Gram yaitu *Unigram*, yaitu token yang jumlahnya 1 kata ($n=1$), yang kedua *Bigram*, yaitu token yang jumlahnya 2 kata ($n=2$) dan yang terakhir *Trigram*, yaitu token yang jumlahnya 3 kata ($n=3$)[8]

N-Gram dikenal berdasarkan banyaknya potongan karakter sebesar n , untuk membantu dalam mengambil potongan kata berupa karakter huruf tersebut maka dilakukan *padding* dengan *blank* diawal dan diakhir suatu kata. Sebagai contoh : kata “TEXT” dapat digambarkan dalam beberapa N-Gram sebagai berikut :

Uni-grams : T,E,X,T

Bi-grams : _T,TE,EX,XT,T_

Tri-grams : _TE,TEX,EXT,XT_

N-Gram dapat digunakan untuk berbagai hal, salah satu yang paling sering digunakan adalah untuk mengantisipasi kata yang akan digunakan, contohnya pada kemampuan khusus koreksi otomatis pada konsol ponsel atau pada pencarian kata kunci tertentu dalam tampilan *google* untuk membantu pengguna. N-Gram adalah salah satu model pemrosesan bahasa alami yang paling populer yang sering digunakan, N-Gram terbukti memberikan kinerja terbaik dibandingkan model lainnya[9].

Sistem kerja N-Gram dapat ditunjukkan dengan *pseudocode* berikut ini :

```
count_vectorizer=
CountVectorizer(ngram_range=(1,2))

final_vectorized_data=
count_vectorizer.fit_transform(commentsdf['processed_comment'])
```

C. Naïve Bayes Classifier

Naive Bayes Classifier merupakan suatu teknik pengelompokan yang ditetapkan dalam hipotesis bayes. Teknik pengelompokan menggunakan teknik probabilitas dan statistik yang ditemukan oleh ilmuwan inggris Thomas Bayes. Teknik ini memperkirakan peluang di masa yang akan datang berdasarkan keterlibatan di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Keistimewaan dari Naïve

Bayes Classifier ini adalah kecurigaan otonomi yang sangat kuat (percaya) dari setiap kondisi atau peristiwa.

Naïve Bayes Classifier adalah konsep yang masuk akal untuk penentuan kelompok kelas dokumen. Perhitungan klasifikasi ini dapat mengolah informasi yang luas dengan ketepatan yang tinggi (Rossi, A. dkk). Klasifikasi NBC didasarkan pada kemungkinan dan hipotesis teorema Bayesian dengan anggapan bahwa setiap variable X sifatnya bebas (independence), dengan harapan bahwa keberadaan suatu kualitas (variabel) tidak ada hubungannya dengan atribut (variabel) yang lainnya. Kinerja NBC yang kompetitif dalam tahapan pengelompokan meskipun menggunakan asumsi ke-independen-an atribut (tidak ada hubungan antar atribut).

D. Pengujian Menggunakan K-Fold Cross Validation

Dalam hal ini dilakukan pengujian menggunakan *K Fold Cross Validation*.

E. Confusion Matrix

Confusion Matrix adalah suatu teknik yang sering digunakan untuk menghitung akurasi, *recall*, *precision*, dan *error date*. Yang artinya, akurasi adalah kekuatan sistem untuk menemukan peringkat yang paling relevan dan ditandai sebagai presentase dokumen yang di *retrieve* dan benar-benar relevan terhadap query. *Recall* adalah kerangka kerja untuk menemukan semua hal penting dari koleksi dokumen dan ditandai sebagai presentase dokumen yang relevan terhadap query. *Accuracy* adalah perbandingan kasus yang dikenali secara akurat dengan jumlah keseluruhan kasus dan tingkat kesalahan dapat berupa kasus yang dikenali dengan jumlah seluruh kasus[10]. Matrix tersebut dijelaskan pada table 1 :

TABEL 1
 CONFUSION MATRIX

	Predicted Negative	Predicted Positive
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

1. *True Positive*, menunjukkan seberapa banyak data yang actual kelasnya positif dan model juga memprediksi positif.
2. *True Negative*, menunjukkan seberapa banyak data yang actual kelasnya negatif dan model juga memprediksi negatif.

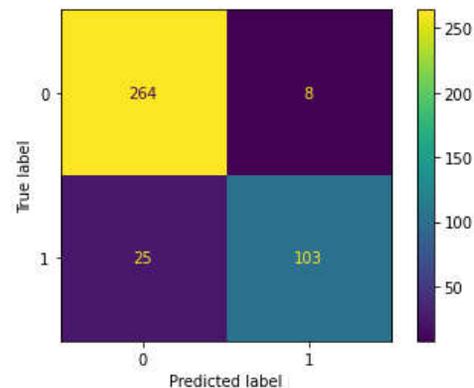
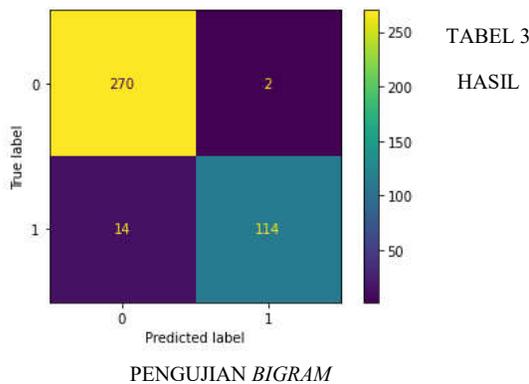
k-fold	Accuracy	Precision	Recall	F1-Score
2	96%	98%	90%	94%
3	96%	98%	89%	93%
4	96%	98%	89%	93%
5	96%	98%	89%	93%
6	96%	98%	89%	93%
7	96%	98%	89%	93%
8	96%	98%	91%	94%
9	96%	98%	89%	93%
10	96%	98%	89%	93%
5	97%	98%	92%	95%
6	96%	98%	91%	94%
7	96%	98%	91%	94%
8	96%	98%	91%	94%
9	96%	98%	91%	94%
10	96%	98%	91%	94%

Penerapan Trigram menggunakan algoritma NBC menghasilkan tabel dengan hasil perbandingan nilai akurasi dari setiap variasi nilai *k-fold cross validation* sebagai berikut :

TABEL 4.
 HASIL PENGUJIAN TRIGRAM

Berdasarkan table hasil perbandingan diatas, nilai akurasi tertinggi didapatkan pada *k-fold* = 2, *k-fold* = 4 dan *k-fold* = 5 dengan nilai akurasi sebesar 97%. Pada *k-fold* = 2, *k-fold* = 4 dan *k-fold* = 5 nilai *precision*, *recall*, dan *f1-score* nya adalah 98%, 92%, 95%.

Dapat disajikan menggunakan *confusion matrix* sebagai berikut :



B. Pengujian Menggunakan Parameter

Berdasarkan hasil pengujian menggunakan parameter yang menerapkan *unigram*, *bigram* dan *trigram* untuk menghasilkan level akurasi dari setiap nilai α yang digunakan dalam pengujian ini dapat diperhatikan tabel hasil perbandingan nilai akurasinya sebagai berikut :

TABEL 5.
 HASIL PENGUJIAN PARAMETER

N-Gram	α	Accuracy	Precision	Recall	F1-score
	0.1	95%	95%	89%	91%
	0.2	94%	94%	86%	90%

c. Trigram

Gbr 7. Confusion matrix bigram

unigram n = 1	0.3	92%	92%	82%	87%
	0.4	91%	92%	80%	86%
	0.5	91%	94%	78%	85%
	0.6	91%	95%	75%	84%
	0.7	90%	94%	72%	82%
	0.8	90%	94%	72%	82%
	0.9	85%	94%	71%	81%
bigram n = 2	1.0	88%	93%	69%	79%
	0.1	97%	97%	93%	95%
	0.2	97%	97%	92%	95%
	0.3	97%	98%	92%	95%
	0.4	97%	98%	92%	95%
	0.5	97%	98%	92%	95%
	0.6	97%	98%	92%	95%
	0.7	96%	98%	91%	94%
	0.8	96%	98%	91%	94%
	0.9	96%	98%	90%	94%
trigram n = 3	1.0	96%	98%	89%	93%
	0.1	97%	97%	95%	96%
	0.2	97%	98%	94%	96%
	0.3	97%	98%	94%	96%
	0.4	97%	98%	93%	96%
	0.5	97%	98%	93%	96%
	0.6	97%	98%	92%	95%
	0.7	97%	98%	92%	95%
	0.8	97%	98%	92%	95%
	0.9	97%	98%	92%	95%

Berdasarkan table diatas, pada pengujian parameter yang menerapkan unigram diperoleh akurasi terbesar yaitu pada $\alpha = 0.1$ dengan nilai akurasi sebesar 95%. Pada $\alpha = 0.1$ nilai *precision*, *recall*, dan *f1-score* nya adalah 95%, 89%, 91%. Pada pengujian parameter yang menerapkan bigram diperoleh akurasi terbesar yaitu pada $\alpha = 0.3$, $\alpha = 0.4$, $\alpha = 0.5$, $\alpha = 0.6$ dengan nilai akurasi sebesar 97%. Pada $\alpha = 0.3$, $\alpha = 0.4$, $\alpha = 0.5$, $\alpha = 0.6$ nilai *precision*, *recall*, dan *f1-score* nya adalah 98%, 92%, 95%. Pada pengujian parameter yang menerapkan trigram diperoleh akurasi terbesar yaitu pada $\alpha = 0.2$, $\alpha = 0.3$, dengan nilai akurasi sebesar 97%. Pada $\alpha = 0.2$, $\alpha = 0.3$ nilai *precision*, *recall*, dan *f1-score* nya adalah 98%, 94%, 96%.

IV. KESIMPULAN

Berdasarkan hasil pembahasan menunjukkan bahwa metode *NBC* dengan ekstraksi fitur N-Gram dapat digunakan untuk klasifikasi analisis sentimen yang dalam hal ini setiap data diklasifikasi menjadi dua yaitu sentimen positif dan sentimen negatif dengan menggunakan dataset sebanyak 400.

Berdasarkan hasil pengujian *k-fold cross validation* yang menerapkan ekstraksi fitur N-Gram diperoleh hasil akurasi tertinggi diperoleh pada penerapan *bigram* dengan *k-fold* = 8 dengan nilai akurasi sebesar 97%. Pada *k-fold* = 8 nilai *precision*, *recall*, dan *f1-score* nya

adalah 98%, 92%, 95% dan pada penerapan *trigram* akurasi tertinggi didapatkan pada *k-fold* = 2, *k-fold* = 4 dan *k-fold* = 5 dengan nilai akurasi sebesar 97%. Pada *k-fold* = 2, *k-fold* = 4 dan *k-fold* = 5 nilai *precision*, *recall*, dan *f1-score* nya adalah 98%, 92%, 95%.

Sedangkan pada pengujian parameter diperoleh hasil akurasi tertinggi pada penerapan *trigram* yaitu pada $\alpha = 0.2$, $\alpha = 0.3$, dengan nilai akurasi sebesar 97%. Pada $\alpha = 0.2$, $\alpha = 0.3$ nilai *precision*, *recall*, dan *f1-score* nya adalah 98%, 94%, 96%.

V. SARAN

Penulis menyadari bahwa masih banyak kekurangan dan kelemahan pada penelitian ini sehingga perlu dilakukan pengembangan lagi untuk mendapatkan hasil yang akurat. Beberapa saran penulis untuk penelitian selanjutnya adalah sebagai berikut :

1. Dataset yang digunakan pada penelitian ini masih terbatas, akan lebih presisi jika menggunakan sumber data yang lebih banyak lagi.
2. Percobaan variasi pada penggunaan kategori N-Gram seperti menggabungkan *uni+bi gram* atau *uni+tri gram* atau menambahkan teknik lain sebagai bahan perbandingan agar hasil pengujian menjadi lebih bervariasi.

UCAPAN TERIMA KASIH

Dengan memanjatkan puji syukur kehadiran Allah SWT, atas rahmat dan hidayah-Nya, syafaat tarbiat Rasulullah SAW, barokah nadroh beliau Ghoutsu Hadzaz Zaman RA dan juga shalawat selalu tercurahkan kepada junjungan kita Nabi Muhammad SAW yang telah diberikan kepada saya, sehingga saya dapat menyelesaikan penelitian ini. Terimakasih juga kepada kedua orang tua dan orang di sekitar saya yang selalu memberikan dorongan dan semangat dalam menuntut ilmu, terimakasih juga diucapkan kepada dosen pembimbing saya yang sudah membimbing penelitian ini dari awal sampai akhir. Sebagai penutup, saya ingin menyampaikan ribuan terimakasih kepada semua pihak yang telah membantu jalannya penelitian ini.

REFERENSI

- [1] Agus Tri Haryanto, "Riset: Ada 175,2 Juta Pengguna Internet di Indonesia," 2020. .
- [2] Dwi, "Mengapa Instagram Menjadi Social Media Terpopuler," 2020. .
- [3] Fiki Ariyanti, "Ini Nih Biang Kerok Penyebab Fluktuasi Harga Saham," 2020. .
- [4] E. Putra, "Klasifikasi sentimen masyarakat terhadap trasngender berdasarkan komentar di Instagram menggunakan metode Naive Bayes Classifier," p. 143, 2019.

- [5] D. N. Chandra, G. Indrawan, and I. N. Sukajaya, "Klasifikasi Berita Lokal Radar Malang Menggunakan," vol. 10, no. 2, pp. 11–19, 2019.
- [6] E. Mulyani, F. Pralienka, B. Muhamad, and K. A. Cahyanto, "Pengaruh SOC terhadap Klasifikasi Buku menggunakan Ekstraksi dan Seleksi Fitur pada Multinomial Naïve Bayes," vol. 5, pp. 264–272, 2021, doi: 10.30865/mib.v5i1.2672.
- [7] S. S. Kumar and A. Rajini, "An Efficient Sentimental Analysis for Twitter Using Neural Network based on Rmsprop," *IOSR J. Eng.*, no. Iccids, pp. 17–25, 2018.
- [8] P. Pratama, E. Indarbensyah, and N. Rochmawati, "PENERAPAN N-GRAM MENGGUNAKAN ALGORITMA RANDOM FOREST DAN NAÏVE BAYES CLASSIFIER PADA ANALISIS SENTIMEN KEBIJAKAN PPKM 2021," vol. 2, 2021.
- [9] S. S. Aljameel *et al.*, "A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent covid-19 outbreaks in Saudi Arabia," *Int. J. Environ. Res. Public Health*, vol. 18, no. 1, pp. 1–12, 2021, doi: 10.3390/ijerph18010218.
- [10] A.- Arini, L. K. Wardhani, and D.- Octaviano, "Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Pada Tweet Hashtag #2019gantipresiden," *Kilat*, vol. 9, no. 1, pp. 103–114, 2020, doi: 10.33322/kilat.v9i1.878.
- [11] T. Kurniawan, "Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naïve Bayes Classifier Dan Support Vector Machine Media Mainstream Menggunakan Naïve Machine," *IT J.*, vol. 23, p. 1, 2017.
- [12] F. Kusuma Wardani, R. Valentinus Hananto, and V. Nurcahyawati, "Analisis Sentimen Untuk Peningkatan Popularitas Situs Belanja Online Di Indonesia Menggunakan Metode Naïve Bayes," *Jsika*, vol. 08, no. 01, pp. 1–9, 2019.
- [13] V. Rahmayanti, S. Basuki, and H. Hilman, "Klasifikasi sinopsis novel menggunakan metode naïve bayes classifier," *J. Repos.*, vol. 1, no. 2, p. 125, 2019, doi: 10.22219/repositor.v1i2.799.

