

Analisis Sentimen Pergerakan Harga Saham Sebuah Perusahaan di Media Sosial *Twitter*

Agung Purwanto Soedarbe^{*1}, Muhammad Rifqi Ma'arif², Aris Wahyu Murdiyanto³, M. Abu Amar Al Badawi⁴

^{1,2,3}Teknologi Informasi, FTI Unjaya, Yogyakarta, Indonesia

⁴Informatika, FTI Unjaya, Yogyakarta, Indonesia

e-mail: ^{*1}soedarbeagung@gmail.com, ²muhammad.rifqi@gmail.com, ³ariswahyumurdiyanto@gmail.com, ⁴abuamar.albadawi@gmail.com

Abstract - *Twitter has become an essential platform for traders and stock investors worldwide, including major countries like America. Traders rely on Twitter to gather information, similar to how they use Bloomberg terminals. While Twitter provides valuable insights, it also contains negative elements such as false information. The sentiment surrounding stocks on Twitter has been growing, and this study aims to analyze the sentiment of Telkom Indonesia's stock price based on tweets. The research involved several stages. First, data was collected from Twitter and labeled manually into positive, neutral, and negative sentiments. The data then underwent pre-processing, including cleaning and dividing it into training and testing datasets using K-Fold Cross Validation. The data was further weighted using the TF-IDF method, and a training process was conducted to develop a model. The final stage involved testing the accuracy of the model. The study successfully implemented the Multinomial Naïve Bayes (MNB) method, achieving an accuracy of 89.0%. The tweet classification results revealed that out of 1000 tweets, 76.5% were classified as positive, 14.3% as negative, and 9.2% as neutral.*

Keywords - *Twitter, Telkom Indonesia, Sentiment Analysis, Multinomial Naïve Bayes (MNB)*

Abstrak - *Twitter telah menjadi platform penting bagi trader dan investor saham di seluruh dunia, termasuk negara-negara besar seperti Amerika. Para trader mengandalkan Twitter untuk mengumpulkan informasi, seperti halnya penggunaan terminal Bloomberg. Meskipun Twitter menyediakan wawasan yang berharga, platform ini juga mengandung elemen negatif, seperti informasi palsu. Sentimen seputar saham di Twitter terus berkembang, dan penelitian ini bertujuan untuk menganalisis sentimen terhadap harga saham Telkom Indonesia berdasarkan tweet. Penelitian ini melibatkan beberapa tahapan. Pertama, data dikumpulkan dari Twitter dan dilabeli secara manual menjadi sentimen positif, netral, dan negatif. Selanjutnya, data mengalami proses pra-pemrosesan, termasuk pembersihan data dan pembagian menjadi set data latih dan uji menggunakan K-Fold Cross Validation. Data kemudian diberi bobot menggunakan metode TF-IDF, dan dilakukan proses pelatihan untuk*

mengembangkan model. Tahap terakhir melibatkan pengujian akurasi model. Penelitian ini berhasil menerapkan metode Multinomial Naïve Bayes (MNB) dengan tingkat akurasi sebesar 89,0%. Hasil klasifikasi tweet menunjukkan bahwa dari 1000 tweet, 76,5% diklasifikasikan sebagai positif, 14,3% sebagai negatif, dan 9,2% sebagai netral.

Kata kunci - *Twitter, Telkom Indonesia, Analisis Sentimen, Multinomial Naïve Bayes (MNB)*

I. PENDAHULUAN

Analisis sentimen, juga dikenal sebagai penambangan opini, adalah studi komputasi yang bertujuan mengenali dan mengekspresikan opini, sentimen, evaluasi, sikap, emosi, subjektivitas, penilaian, atau pandangan yang terkandung dalam suatu teks [1]. Dalam konteks media sosial, sentimen negatif mengacu pada opini yang memberikan kesan kurang baik terhadap pihak yang disebut, sementara sentimen positif menimbulkan kesan baik terhadap pihak yang terlibat atau disebut [2]. Dalam era media sosial yang semakin canggih, pelanggan dapat dengan mudah menyampaikan keluhan terhadap sebuah perusahaan melalui platform media sosial dengan hanya melakukan *mention* kepada perusahaan terkait.

Twitter, sebagai salah satu media sosial yang sangat berguna, menawarkan konsep berinteraksi dan berbagi informasi dengan batasan karakter sebanyak 140 karakter saja. Di *Twitter* seorang pengguna dapat mendaftar dengan bebas, menjalin pertemanan dengan siapa saja, mempublikasikan profil mereka, dan juga termasuk menjalin komunikasi secara langsung (*live*) dengan pengguna lain baik yang dikenal maupun yang tidak dikenal [3]. Keterbatasan ini mengharuskan pengguna untuk berbagi informasi secara singkat, padat, dan jelas. *Twitter* juga memungkinkan berbagi informasi secara *realtime*, membuatnya menjadi alat yang populer di kalangan *trader* dan investor saham. Saham adalah secerik kertas yang menunjukkan hak pemodal yaitu pihak yang memiliki kertas tersebut untuk memperoleh bagian dari prospek atau kekayaan organisasi yang menerbitkan sekuritas tersebut, dan berbagai kondisi yang memungkinkan pemodal tersebut menjalankan haknya [4], sedangkan menurut Fahmi saham adalah kertas yang tercantum dengan jelas nilai nominal, nama perusahaan, dan diikuti dengan hak dan kewajiban yang telah dijelaskan

kepada setiap pemegangnya [5]. Di negara-negara seperti Amerika Serikat, penggunaan *Twitter* oleh investor dan *trader* sudah menjadi kebutuhan sebagaimana mereka memantau informasi melalui terminal Bloomberg. Banyak *trader* yang mencari informasi melalui platform ini dengan melihat *tweet* dari perusahaan sekuritas atau pun *trader* lainnya. Selain itu, para pelaku pasar modal juga memanfaatkan *Twitter* sebagai sarana edukasi melalui kuliah *Twitter* yang banyak diselenggarakan, yang memberikan informasi secara gratis dan mudah dipahami. *Twitter* dapat memberikan banyak manfaat bagi *trader*, namun tidak semua informasi di dalamnya bersifat positif, karena terdapat pula informasi palsu (misinformasi) yang dapat memengaruhi keputusan trading.

Penelitian sebelumnya telah memanfaatkan *Twitter* dalam berbagai konteks analisis sentimen. Misalnya, penelitian telah menganalisis *tweet* berbahasa Indonesia yang membicarakan tentang tokoh publik menjelang pemilihan umum 2014 di Indonesia, menggunakan *Naive Bayes Classifier* yang dikombinasikan dengan fitur deteksi negasi dan pembobotan menggunakan term *frequency* serta *TF-IDF*, dengan akurasi mencapai 90% [6]. Penelitian lainnya menggunakan data opini film yang diambil dari komentar di *Twitter*, dengan hasil analisis sentimen menggunakan algoritma *Naive Bayes* dan mencapai akurasi sebesar 90% dengan nilai *precision* 92%, *recall* 90%, dan *f-measure* 90% [7].

Selain itu, terdapat penelitian yang melakukan analisis sentimen terhadap calon Gubernur DKI Jakarta 2017 di media sosial *Twitter*. Penelitian ini menggunakan metode *preprocessing* data seperti tokenisasi, *cleansing*, dan *filtering*, serta metode *Lexicon Based* untuk menentukan kelas sentimen. Dalam proses klasifikasinya, digunakan metode *Naive Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)*, dengan akurasi tertinggi mencapai 95% pada *NBC* [8].

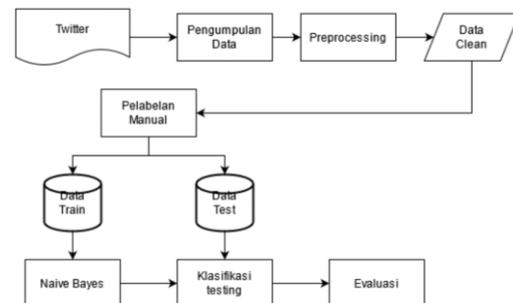
Selanjutnya ada penelitian sebelumnya juga membandingkan dua algoritma klasifikasi, yaitu *K-Nearest Neighbour (KNN)* dan *Naive Bayes Classifier*, pada data aktivitas status gunung berapi di Indonesia. Dalam penelitian tersebut, *Naive Bayes* mencapai akurasi rata-rata sebesar 79,71%, sedangkan *KNN* mencapai 63,68% [9].

Berdasarkan tinjauan terhadap penelitian-penelitian sebelumnya yang telah dijelaskan di atas, penelitian ini akan meneruskan analisis sentimen terhadap pergerakan harga saham sebuah perusahaan dengan menggunakan metode klasifikasi *Multinomial Naive Bayes*. *Multinomial Naive Bayes* merupakan sebuah metode yang bekerja dengan cara menghitung frekuensi setiap *term* pada dokumen [10]. Data yang akan digunakan berasal dari *Twitter* dengan kata kunci "telkomindonesia".

II. METODE PENELITIAN

A. Jalan Penelitian

Penelitian ini menggunakan software Anaconda 3 dan Jupyter notebook untuk mengelola data yang diambil dari *Twitter*. Rancangan penelitian ini tergambar dalam Gambar 1 berikut.



Gambar 1. Alur Penelitian

Tahapan penelitian meliputi pengumpulan data melalui web scraping dari *Twitter* dan penyimpanan hasilnya dalam format CSV atau XLSX. Selanjutnya, dilakukan pelabelan manual pada dataset untuk klasifikasi sentimen menjadi positif, netral, dan negatif. Proses *pre-processing* data dilakukan untuk membersihkan data dari noise, dengan tahapan seperti *case folding*, *cleaning*, *tokenizing*, *filtering*, *stemming*, dan *normalization*. *Training* data melibatkan fitur *extraction* menggunakan algoritma *TF-IDF* untuk memberikan bobot pada setiap kalimat, dan dilanjutkan dengan proses *training* untuk membangun model klasifikasi. *con* data melibatkan evaluasi tingkat akurasi, presisi, dan *recall* dengan menggunakan *confusion matrix*.

Tabel 1. Sample hasil pengambilan data

Waktu	Pengguna	<i>Tweet</i>
2021-09-03	Neo R. Ramadan	"b'RT @TelkomIndonesia: Waktunya #KuisTelkom!\nSobat, ayo bantu tangkap tepat pada logo IndiHome! Menangkan saldo LinkAja masing-masing 100rib\ \"
2021-09-01	Edri Novianto	"b'@IndiHomeCare Saya pelanggan indihome. Memasuki hari ke 11, internet macet. Saya baru tahu inilah perusahaan yang\ \ https://t.co/sxAKIuyBmz"
2021-08-28	Effendy	"b'Telkom Indonesia menyelenggarakan program TeSCA (Telkom Smart Campus Awards) 2021, penghargaan berupa mendapat kese\ \ https://t.co/YCyICF2iAU'"

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah opini *Twitter* tentang perusahaan, yang diperoleh melalui proses *web crawling* dari situs www.api.twitter.com menggunakan aplikasi Jupyter notebook. Pengumpulan data dilakukan dengan menggunakan *library tweepy* pada *Python*, yang digunakan untuk mengakses API *Twitter*. Tahap awal melibatkan pembuatan token autentikasi dengan *consumer key*, *consumer secret*, *access token*, dan *access secret* yang diperoleh melalui akun *Twitter* developer. Setelah autentikasi selesai, data *Twitter* dikumpulkan berdasarkan kriteria tertentu seperti kata kunci, tanggal, dan pengguna. Dalam penelitian ini, kata kunci yang digunakan adalah "telkomindonesia" dengan pengambilan *tweet* dalam Bahasa Indonesia. Data *tweet* tersebut dapat disimpan dalam format CSV atau XLSX. Tabel 1 menunjukkan contoh hasil pengumpulan data yang digunakan dalam penelitian.

2. Preprocessing

Preprocessing dilakukan untuk membersihkan data *tweet* yang masih dalam kondisi kotor agar siap untuk diproses pada tahap selanjutnya. Tahap-tahap *preprocessing* yang dilakukan adalah sebagai berikut:

- *Cleaning*: Pada tahap ini dilakukan penghilangan karakter-karakter seperti *user*, *URL*, *mention*, *hashtag*, angka, emoji, dan lainnya.
- *Case folding*: Tahap ini mengubah semua isi *tweet* menjadi huruf kecil.
- *Filtering*: Tahap ini melibatkan pemilihan kata-kata yang penting dan penghilangan kata-kata yang tidak penting seperti "yang", "dan", "di", "dengan", "dari", dan sebagainya [11].
- *Stopwords*: Tahap ini melibatkan pemotongan string berdasarkan kata-kata yang disusun.
- *Stemming*: Pada tahap ini, digunakan *library* Sastrawi untuk melakukan *stemming* pada Bahasa Indonesia. Kata-kata akan diubah menjadi kata dasarnya menggunakan *stemmer*.
- *Normalization*: Tahap ini melibatkan penggunaan file tambahan yang berisi kata-kata slang dan kata-kata normal. Pada fungsi *normalization*, dilakukan pengecekan antara file tambahan tersebut dengan data teks.

Tahapan-tahapan tersebut dilakukan untuk membersihkan data *tweet* sehingga siap untuk tahap analisis selanjutnya.

3. Pelabelan manual

Pada tahap pelabelan manual, data yang telah terkumpul diberikan label berdasarkan kategori tertentu. Dalam penelitian ini, label yang digunakan adalah angka 1, 0, dan -1. Angka 1 melambangkan sentimen positif, angka 0

melambangkan sentimen netral, dan angka -1 melambangkan sentimen negatif. Proses pelabelan ini dilakukan untuk memberikan klasifikasi pada teks yang telah dibersihkan. Informasi lebih lanjut dapat dilihat pada Tabel 2 yang menunjukkan hasil pelabelan manual beserta teks yang telah dibersihkan.

4. Pembuatan model klasifikasi

Tabel 2. Pelabelan manual

label	Cleaned_Text
1	bindihome telkomindonesia solusi internet kelas cerdas aktivitas batas xexxa
0	btelkomindonesia hasil tangkap logo indihome n nikutan yuk xexxa
-1	btelkonsel telkomindonesia kecewa tekonsel pakai kartu halo makin boros padahal rumah wifiduxexxa

Pada tahap pembuatan model klasifikasi, digunakan Metode *Naive Bayes Multinomial (MNB)* dengan menggunakan *library* scikit-learn atau sklearn pada *Python*. Proses dimulai dengan membagi data menjadi data teks dan data label. Data label akan dienkoder, sedangkan data teks akan diberikan bobot menggunakan metode *TF-IDF*.

Selanjutnya, dilakukan proses pemodelan *MNB* dengan menggunakan data teks dan data label yang telah diproses sebelumnya. Setelah pemrosesan selesai, dibuatlah sebuah model yang dapat digunakan untuk klasifikasi data baru. Model tersebut dapat disimpan dalam format file pickle untuk penggunaan lebih lanjut.

5. Evaluasi dengan *K-fold Cross Validation*

Pada tahap evaluasi, digunakan metode *K-fold Cross Validation* untuk memeriksa tingkat akurasi. *K-fold Cross Validation* merupakan tahapan di *MNB* di mana data akan dievaluasi untuk mengetahui akurasi setiap "fold" atau lipatan data. Pada penelitian ini, dilakukan 10 kali "fold" dan hasilnya ditampilkan menggunakan *library* matplotlib.

Hasil evaluasi dengan *K-fold Cross Validation* dapat divisualisasikan dalam bentuk grafik batang yang menunjukkan tingkat akurasi. Grafik tersebut memberikan informasi tentang sejauh mana model dapat mengklasifikasikan data dengan tepat.

6. Klasifikasi

Pada tahap klasifikasi, dilakukan prediksi label berdasarkan model yang telah dibuat. Model data diberi kelas positif (1) jika data adalah positif, netral (0) jika data adalah netral, dan negatif (-1) jika data adalah negatif. Tahapan *MNB* meliputi *splitting* data, transformasi menggunakan *TF-IDF*, dan proses *MNB*. Proses *splitting* data menggunakan *Shuffle Split* dengan kelipatan *n* (Suku). Evaluasi *MNB* ditampilkan dalam bentuk akurasi dan *confusion matrix*, serta mencakup *precision*, *recall*, *f1-score*, dan *support* untuk

mengukur kinerja model dalam mengklasifikasikan data dengan tepat sementara sentimen netral dan negatif memiliki frekuensi yang lebih rendah. Meskipun sentimen negatif memiliki frekuensi yang lebih tinggi daripada sentimen netral.

III. HASIL DAN PEMBAHASAN

A. Hasil Evaluasi Classifier

Dalam penelitian ini, dilakukan evaluasi sebagai salah satu tahapan *MNB*. Evaluasi bertujuan untuk mencari nilai akurasi dan *F1-Score* dari *MNB*. Evaluasi dilakukan menggunakan *K-fold Cross Validation* dan *confusion matrix*.

Pada *K-fold Cross Validation*, data dibagi menjadi 10 *fold* dan dilakukan pengujian untuk mengetahui tingkat akurasi dari setiap *fold*. Rata-rata akurasi yang diperoleh dari *K-fold Cross Validation* adalah 76,49%. *Fold* keempat memiliki akurasi tertinggi sebesar 86,2%. Hasil *K-fold Cross Validation* ditampilkan menggunakan *matplotlib*.

Confusion matrix digunakan untuk mengukur akurasi dan *F1-Score*. Tabel 3 menampilkan *confusion matrix* yang terdiri dari prediksi dan klasifikasi. Terdapat nilai akurasi sebesar 96,0% dan *F1-Score* sebesar 95,83%. Tabel 4 menyajikan hasil total nilai *precision*, *recall*, dan *F1-Score* dari klasifikasi *MNB*. *Precision*, *recall*, dan *F1-Score* diukur untuk setiap klasifikasi (-1, 0, 1) dan juga dihitung nilai rata-rata (*macro avg* dan *weighted avg*).

Penelitian ini menggunakan data baru sebanyak 1000 data yang diklasifikasikan menggunakan *MNB*.

Tabel 3. *Confusion matrix*

Klasifikasi	Prediksi		
	Negatif (-1)	Netral (0)	Positif (1)
Negatif (-1)	17	0	3
Netral (0)	0	1	2
Positif (1)	2	1	174
Accuracy	96.0%		
<i>F1-score</i>	95.8291846730049%		

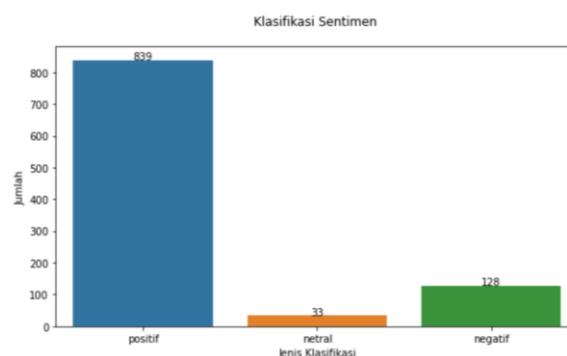
Tabel 4. Hasil klasifikasi *MNB*

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	Supoort
-1	0.89	0.85	0.87	20
0	0.50	0.33	0.40	3
1	0.97	0.98	0.98	177
Accuracy			0.96	200
Macro avg	0.79	0.72	0.75	200
Weighted avg	0.96	0.96	0.96	200

Evaluasi dilakukan untuk mendapatkan tingkat akurasi, *F1-Score*, dan *confusion matrix*. Hasil evaluasi tersebut membantu dalam mengevaluasi kinerja model klasifikasi *MNB*.

B. Hasil Analisis

Penelitian ini bertujuan untuk menganalisis sentimen terhadap perusahaan Telkom Indonesia di *Twitter* melalui klasifikasi data *tweet*. Data *tweet* yang diambil telah diklasifikasikan menjadi tiga label yaitu positif, netral, dan negatif yang mencerminkan sentimen *tweet* tersebut. Sebanyak 1000 data *tweet* terkait perusahaan Telkom Indonesia digunakan sebagai data *training* untuk membangun model *MNB*. Kemudian, model tersebut digunakan untuk mengklasifikasikan 1000 data *tweet* baru. Hasil klasifikasi dapat dilihat pada Grafik 1.



Gambar 1. Grafik hasil klasifikasi

Grafik tersebut menunjukkan hasil klasifikasi otomatis dari aplikasi menggunakan model *MNB*. Data dengan label positif memiliki jumlah tertinggi, yaitu 839 data *tweet* (83,9%). Label negatif memiliki jumlah 128 data *tweet* (12,8%), sedangkan label netral memiliki jumlah terendah, yaitu 33 data *tweet* (3,3%). Dari jumlah data tersebut, dapat disimpulkan bahwa sentimen positif memiliki frekuensi yang tinggi, sementara sentimen netral dan negatif memiliki frekuensi yang lebih rendah. Meskipun sentimen negatif memiliki frekuensi yang lebih tinggi daripada sentimen netral.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan metode *MNB* untuk klasifikasi data sentimen analisis pergerakan saham di *Twitter* dengan tingkat akurasi 96,0% dan *F1-Score* 95,82%. Hasil analisis *tweet* terkait perusahaan Telkom Indonesia menunjukkan 83,9% *tweet* diklasifikasikan sebagai positif, 12,8% sebagai negatif, dan 3,3% sebagai netral. Saran untuk penelitian selanjutnya meliputi penambahan kata-kata penting dalam proses *filtering*, perhatian lebih dalam proses *stemming*, dan penggunaan langkah-langkah *preprocessing* yang lebih lengkap. Pelabelan manual, pengadaan data berkualitas, pemilihan kata kunci yang tepat, dan menghindari penggunaan data duplikat juga disarankan untuk meningkatkan akurasi klasifikasi data sentimen.

DAFTAR PUSTAKA

- [1] J. Sitompul, "Analisa Sentimen Masyarakat Terhadap Calon Legislatif Partai Politik Dengan Menerapkan Algoritma Text Mining," *Informasi dan Teknologi Ilmiah (INTI)*, vol. 10, no. 1, pp. 13–23, 2022.
- [2] M. G. Pradana, A. C. Nurcahyo, and P. H. Saputro, "Pengaruh Sentimen Di Sosial Media Dengan Harga Saham Perusahaan," *Jurnal Ilmiah Educic: Pendidikan dan Informatika*, vol. 6, no. 2, pp. 67–77, 2020.
- [3] N. M. Anggreini, B. M. S. Nasir, I. Noor, and S. Sos, "Pemanfaatan Media Sosial Twitter di Kalangan Pelajar SMK Negeri 5 Samarinda," *eJournal Sosiatri-Sosiologi*, vol. 4, no. 2, pp. 239–251, 2016.
- [4] S. Husnan, "Pengertian dan Konsep-Konsep Dasar Keuangan," *S. Husnan, Modul Manajemen Keuangan. Jakarta: Universitas Terbuka*, 2019.
- [5] R. Linanda and W. Afriyeni, "Pengaruh Struktur modal dan profitabilitas terhadap Harga saham," *Jurnal Ekonomi Dan Bisnis Islam*, vol. 3, no. 1, pp. 135–144, 2018.
- [6] A. F. Hidayatullah and A. S. N. Azhari, "Analisis sentimen dan klasifikasi kategori terhadap tokoh publik pada twitter," in *Seminar Nasional Informatika (SEMNASIF)*, 2015.
- [7] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng-Seri Informatika*, vol. 3, no. 1, pp. 50–59, 2018.
- [8] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *INTEGER: Journal of Information Technology*, vol. 2, no. 1, 2017.
- [9] F. Tempola, M. Muhammad, and A. Khairan, "Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 5, pp. 577–584, 2018.
- [10] S. Fanissa, M. A. Fauzi, and S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 2, no. 8, pp. 2766–2770, 2018.
- [11] M. Saiful Anwar, "Prosiding KONFERENSI ILMIAH MAHASISWA UNISSULA (KIMU) 2 SISTEM Pencarian E-JOURNAL MENGGUNAKAN METODE STOPWORD REMOVAL DAN STEMMING BERBASIS ANDROID," 2019.