

ANALISIS SENTIMEN SAHAM PTBA MELALUI KOMENTAR YOUTUBE MENGGUNAKAN VADER LEXICON DAN NAÏVE BAYES CLASIFIER

Andi Jumawan¹, Murobbi Aminillah², Muhammad Yasif AbdulKhoiri³,
David Fauzi⁴

^{1,2,3,4}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi,
Universitas Bina Darma

E-mail: ¹andyxjumawan@gmail.com, ²murobbiaminillahrobby@gmail.com,
³urusanyasif@gmail.com, ⁴fauzidavid709@gmail.com

Abstrak

Analisis sentimen saham PTBA (PT Bukit Asam Tbk) dilakukan dengan menggunakan komentar yang ada di platform YouTube. Metode ini menggunakan algoritma *Naive Bayes* dan *Vader Lexicon* untuk membagi sentimen dari komentar menjadi kategori positif, negatif, atau netral. Data komentar diambil dari video YouTube yang berkaitan dengan saham PTBA. Pra-pemrosesan data seperti tokenisasi, penghapusan kata-kata berhenti, dan vektorisasi kata-kata digunakan untuk melakukan analisis. Dengan dataset yang telah disiapkan, model klasifikasi dilatih untuk mengidentifikasi sentimen yang terkandung dalam komentar. Tujuan penelitian ini adalah untuk meningkatkan pemahaman investor tentang pandangan dan persepsi saham PTBA berdasarkan komentar yang ada di platform YouTube. Hasil analisis sentimen diharapkan dapat memberikan informasi yang bermanfaat bagi para investor.

Kata Kunci : Sentimen saham PTBA, Komentar YouTube, Algoritma *Naive Bayes*, Investor saham YouTube

1. PENDAHULUAN

Dalam era digital saat ini, platform media sosial telah menjadi sumber informasi yang signifikan dalam memahami persepsi publik terhadap suatu perusahaan atau produk, termasuk dalam konteks investasi saham. Analisis sentimen terhadap saham menggunakan data dari platform-platform seperti YouTube dapat memberikan wawasan yang berharga bagi investor untuk membuat keputusan investasi yang lebih baik. Penelitian ini akan memfokuskan pada analisis sentimen terhadap saham PTBA (PT Bukit Asam Tbk) dengan mengambil sampel komentar-komentar yang dipublikasikan di platform YouTube. Metode yang akan digunakan adalah algoritma *Naïve Bayes* dan *Veder Lexicon*, yang telah terbukti efektif dalam mengklasifikasikan sentimen dari teks-teks yang ada.

PTBA (PT Bukit Asam Tbk) adalah salah satu perusahaan tambang batubara terbesar di Indonesia. Dalam dunia investasi saham, analisis sentimen terhadap saham suatu perusahaan dapat memberikan wawasan yang berharga kepada para investor. Komentar publik yang disampaikan di platform media sosial seperti YouTube dapat menjadi sumber informasi yang relevan untuk menganalisis sentimen investor terhadap saham PTBA.

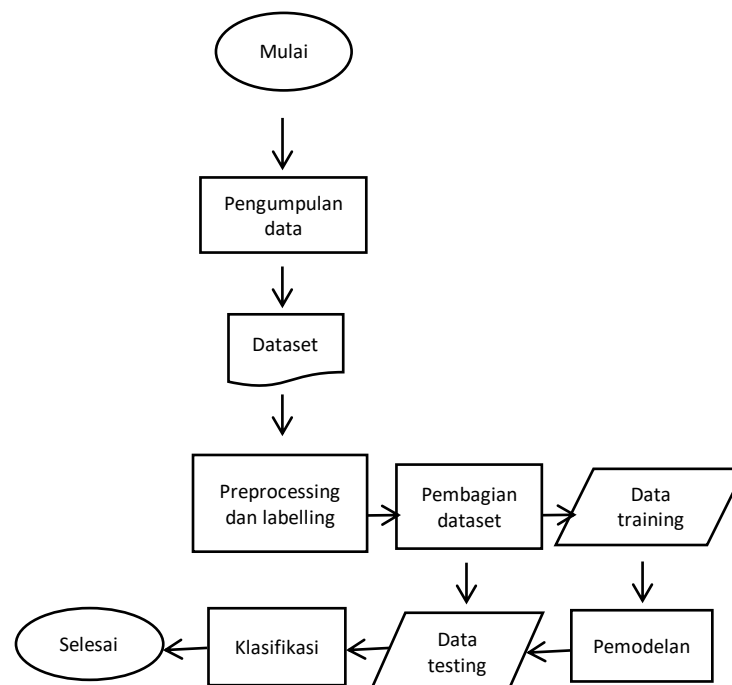
Dalam konteks ini, metode analisis sentimen menggunakan algoritma *Naïve Bayes* dan *Veder Lexicon* dapat menjadi pendekatan yang efektif. Algoritma ini

merupakan salah satu metode klasifikasi yang umum digunakan dalam pemrosesan bahasa alami untuk menentukan sentimen dari teks. Sentimen saham, informasi yang diperoleh dari komentar-komentar di platform media sosial seperti YouTube dapat memberikan gambaran yang lebih komprehensif tentang sikap dan pandangan investor terhadap kinerja perusahaan. Hasil analisis sentimen ini dapat menjadi tambahan informasi yang penting bagi investor untuk memahami dinamika pasar, sikap investor, serta faktor-faktor yang memengaruhi harga saham PTBA.

2. METODOLOGI PENELITIAN

2.1 Diagram Alur Proses

Penelitian ini difokuskan pada analisis sentimen saham dengan mengumpulkan komentar dari satu akun youtube @Mabuk Saham-Willyam Wen. Data yang berhasil dikumpulkan dari komentar-komentar tersebut menjadi fokus utama dalam penelitian ini. Gambaran alur proses penelitian dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur Proses

Gambar tersebut adalah gambaran komprehensif dari langkah-langkah yang dilakukan dalam penelitian ini. Dimulai dengan pengumpulan data berupa teks komentar dari YouTube, proses selanjutnya meliputi tahap *preprocessing* (pra-pemrosesan), *labeling*, pembagian data, pemodelan, dan klasifikasi.

2.2 Pengumpulan Data

Penelitian ini mengumpulkan data melalui pengambilan komentar di youtube menggunakan teknik scrapping dari beberapa video yang berasal dari berbagai akun. Proses scrapping ini dilakukan dengan menggunakan *Google API client* dan bahasa pemrograman Python. Komentar video yang diambil, dikumpulkan, dan disimpan dalam format CSV untuk tahap analisis selanjutnya.

2.3 Preprocessing

Preprocessing adalah langkah krusial dalam memproses teks untuk mempersiapkan dan membersihkan data guna ekstraksi informasi yang lebih efektif. Dalam tahap ini, menggunakan bahasa pemrograman Python dan *library nltk*, beberapa langkah dilakukan, seperti *Cleansing*, *Case Folding*, *Tokenizing*, *Stopword Removal*, dan *Stemming* untuk memproses dan membersihkan data teks.

Cleansing: Membersihkan data dari simbol, emoji, tanda baca, URL, dan angka untuk meningkatkan kebersihan dataset.

Case Folding: Mengubah dataset dari huruf kapital menjadi huruf kecil agar konsistensi teks terjaga.

Tokenizing: Memisahkan teks atau kalimat menjadi bagian-bagian kecil yang disebut "token" untuk analisis lebih lanjut.

Stopword: Menghapus kata-kata hubung atau kata-kata yang kurang penting dalam kalimat untuk fokus pada informasi kunci.

Stemming: Menyederhanakan kata-kata ke bentuk dasarnya agar dianggap sebagai bentuk yang sama, meningkatkan konsistensi dan analisis data yang lebih akurat.

2.4 Labelling

Pada tahap *Labelling*, data diberi label yang dapat berupa positif, negatif, atau netral. Proses pemrograman menggunakan bahasa Python, dengan label ditentukan berdasarkan jumlah kata yang dikategorikan sebagai positif, negative dan netral. Label diberikan sesuai dengan kategori yang memiliki jumlah tertinggi dalam data.

2.5 Klasifikasi

Pada tahapan ini akan dilakukan proses klasifikasi menggunakan algoritma *Naïve Bayes*. Data yang telah melalui tahap *scrapping* dan *labelling* diproses dengan algoritma tersebut. Mesin akan mengenali pola dan mengklasifikasikan data ke dalam tiga kategori: positif, netral, dan negatif..

Naïve Bayes memiliki struktur umum sebagai berikut :

$$P(C|X) = \frac{P(x|c) P(c)}{P(x)}$$

Keterangan :

X : Data dengan class yang belum diketahui

C : Hipotesis data yang merupakan suatu class spesifik

$P(C/X)$: Probabilitas hipotesis C berdasarkan kondisi (*posterior probability*)

$P(c)$: Probabilitas hipotesis C (*prior probability*)

$P(x/c)$: Probabilitas berdasarkan kondisi pada hipotesis

$P(x)$: Probabilitas c

Rumus *Naïve Bayes* menggambarkan probabilitas masuknya data sampel ke dalam kelas C (Posterior), mencerminkan probabilitas munculnya kelas C sebelum data sampel dimasukkan, yang disebut sebagai prior . Penelitian ini menggunakan bahasa pemrograman Python dengan pustaka scikit-learn (*sklearn*), yang merupakan pustaka populer untuk machine learning. Scikit-learn menyediakan algoritma *Naïve Bayes* dan berbagai algoritma lainnya untuk tugas klasifikasi dalam machine learning.

3. HASIL DAN PEMBAHASAN

3.1 *Scrapping*

Data penelitian diperoleh dari komentar pengguna YouTube pada video yang diunggah di akun @Mabuk Saham-Willyam Wen. Proses pengambilan data dilakukan melalui scrapping menggunakan Bahasa Pemrograman Python dengan memanfaatkan *Google API client*. Total 226 komentar diambil dari satu video yang diposting di akun YouTube tersebut, dan komentar-komentar tersebut dikumpulkan pada tanggal 30 November 2023. Berikut adalah hasil dari proses *scrapping* tersebut:

Komentar
Saya sudah pegang ptba lama..msih avg 4650
Mending mana PTBA vs ADRO bro
Mulai serok2 ptba mulai Bln ini
Per hari ini (7 November 2023) saya mines 23% . Awalnya niat dividen tahun kemaren eh kebablasan ðŸ˜… tapi akhirnya belajar dan belajar terus . Dan karena terpaksa akhirnya terbiasa dan mulai nyicil dan jadi jangka panjang ðŸ˜…
Harga ptba aku di 3700 bang berarti hold Aja ya Bang menurut Analisa Abang hari ini thanks bang

3.2 *Text Preprocessing*

Sebelum menggunakan Algoritma *Naïve Bayes* dan *Veder Lexicon* untuk memproses data, langkah preprocessing harus dilakukan terlebih dahulu. Proses ini diperlukan untuk mengatasi inkonsistensi data, mempersiapkan data agar siap untuk tahapan selanjutnya . Langkah-langkah dalam preprocessing melibatkan *Tokenizing*, Normalisasi, *Stemming*, dan penghapusan *Stopword*. Proses ini menggunakan Bahasa Pemrograman Python dengan *library RegEx* untuk menghapus karakter yang tidak diinginkan, serta menggunakan *library nltk* untuk membuat token dan menghapus *stopword*. Berikut adalah hasil dari tahap *preprocessing* teks yang telah dilakukan.:

Komentar	Hasil Preprocessing
Saya sudah pegang ptba lama..msih avg 4650	saya sudah pegang ptba lamamsih avg
Mending mana PTBA vs ADRO bro	mending mana ptba vs adro bro

Mulai serok2 ptba mulai Bln ini	mulai serok ptba mulai bln ini
Per hari ini (7 November 2023) saya mines 23% .Awalnya niat dividen tahun kemaren eh kebablasan ðŸ˜… tapi akhirnya belajar dan belajar terus . Dan karena terpaksa akhirnya terbiasa dan mulai nyicil dan jadi jangka panjang ðŸ˜…	per hari ini november saya mines awalnya niat dividen tahun kemaren eh kebablasan tapi akhirnya belajar dan belajar terus dan karena terpaksa akhirnya terbiasa dan mulai nyicil dan jadi jangka panjang
Harga ptba aku di 3700 bang berarti hold Aja ya Bang menurut Analisa Abang hari ini thanks bang	harga ptba aku di bang berarti hold aja ya bang menurut analisa abang hari ini thanks bang

3.3 Labelling

Data sentimen komentar video youtube dilabeli menggunakan Python dengan kamus kata bahasa Indonesia yang terbagi dalam file kata positif, negatif dan netral. Sistem menghitung jumlah kata positif, negatif dan netral dalam komentar untuk menentukan sentimen. Berikut adalah hasil dari proses labelling dalam penelitian ini:

Komentar	Hasil Preprocessing	Labelling
Saya sudah pegang ptba lama..msih avg 4650	saya sudah pegang ptba lamamsih avg	Netral
Mending mana PTBA vs ADRO bro	mending mana ptba vs adro bro	Netral
Mulai serok2 ptba mulai Bln ini	mulai serok ptba mulai bln ini	Netral
Per hari ini (7 November 2023) saya mines 23% .Awalnya niat dividen tahun kemaren eh kebablasan ðŸ˜… tapi akhirnya belajar dan belajar terus . Dan karena terpaksa akhirnya terbiasa dan mulai nyicil dan jadi jangka panjang ðŸ˜…	per hari ini november saya mines awalnya niat dividen tahun kemaren eh kebablasan tapi akhirnya belajar dan belajar terus dan karena terpaksa akhirnya terbiasa dan mulai nyicil dan jadi jangka panjang	Positif
Harga ptba aku di 3700 bang berarti hold Aja ya Bang menurut Analisa Abang hari ini thanks bang	harga ptba aku di bang berarti hold aja ya bang menurut analisa abang hari ini thanks bang	Negatif

3.4 Klasifikasi Naïve Bayes

Metode klasifikasi *Naïve Bayes* adalah metode klasifikasi yang berbasis pada teorema Bayes dengan asumsi “naive” bahwa setiap fitur dari suatu entitas adalah independen satu sama lain. Dalam klasifikasi ini kami ambil data dari YouTube. Implementasi klasifikasi menggunakan metode *Naïve Bayes* dilakukan dengan Bahasa Pemrograman Python menggunakan pustaka *scikit-learn*. Tahap akhir melibatkan evaluasi kinerja algoritma *Naïve Bayes* dengan mengukur *Precision*, *F1-Score*, *Recall*, dan akurasi. Berikut adalah hasil evaluasi otomatisasi algoritma *Naïve Bayes* yang dilakukan dengan Bahasa Pemrograman Python.

Classification Report:				
	precision	recall	f1-score	support
negatif	1.00	0.25	0.40	4
netral	0.71	0.97	0.82	31
positif	0.50	0.10	0.17	10
accuracy			0.71	45
macro avg	0.74	0.44	0.46	45
weighted avg	0.69	0.71	0.64	45

Gambar 2. Hasil Evaluasi Program

Klasifikasi Naive Bayes adalah metode klasifikasi yang berbasis pada teorema Bayes dengan asumsi “naive” bahwa setiap fitur dari suatu entitas adalah independen satu sama lain. Dalam klasifikasi ini kami ambil data dari YouTube.

```

Hasil Analisis Data:
Positif = 20
Netral = 194
Negatif = 12

Total Data : 226

```

Gambar 3. Hasil Klasifikasi

4. KESIMPULAN

Harga Penelitian ini menginvestigasi sentimen pengguna YouTube terhadap saham PTBA harga fundamental Rp 4.340-Rp 8.680 menggunakan metode klasifikasi *Naïve Bayes* dan *Veder Lexicon*. Komentar-komentar di platform YouTube telah menjadi medan yang memunculkan beragam kontroversi terkait dengan sentimen saham PT Bukit Asam Tbk (PTBA). Beberapa pengguna menyampaikan pandangan optimis terhadap prospek saham perusahaan ini, mengaitkan dengan faktor-faktor seperti kinerja keuangan yang kuat atau proyek-proyek masa depan yang menjanjikan. Namun, kontroversi muncul ketika beberapa komentar mencerminkan ketidakpastian terkait perusahaan, termasuk keraguan terhadap manajemen atau kekhawatiran tentang perubahan pasar yang berdampak pada kinerja saham PTBA.

Metode klasifikasi *Naïve Bayes* dan *Veder Lexicon* digunakan dalam menganalisis sentimen komentar-komentar di YouTube. Langkah-langkahnya melibatkan pengumpulan data, preprocessing teks, *labeling* sentimen, dan penerapan model *Naïve Bayes*. Hasil analisis menunjukkan kompleksitas sikap masyarakat dengan komentar-komentar diklasifikasikan sebagai positif, negatif, atau netral. Evaluasi kinerja model menggunakan metrik seperti *Precision*, *F1-Score*, *Recall*, dan akurasi.

Secara keseluruhan, penelitian ini memberikan pemahaman yang dalam terhadap respons komunitas pengguna YouTube terhadap konten kontroversial di platform tersebut. Sikap masyarakat terhadap isu keagamaan dalam konteks konten digital diungkapkan sebagai fenomena yang kompleks dan beragam. Penelitian ini diharapkan dapat memberikan wawasan tambahan mengenai preferensi dan sikap pengguna terhadap isu-isu sensitif dalam dunia media sosial.

5. REFERENSI

- Hudha, M., Supriyati, E., & Listyorini, T. (2022). Analisis Sentimen Pengguna Youtube Terhadap Tayangan# Matanajwamentiterawan Dengan Metode Naïve Bayes Classifier. *JIKO (Jurnal Informatika dan Komputer)*, 5(1), 1-6.
- Harpizon, H. A. R., Kurniawan, R., Iskandar, I., Salambue, R., Budianita, E., & Syafria, F. (2022). Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes. *Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes*, 5(1), 131-140.