

# TEXT MINING SENTIMEN ANALISIS PENGGUNA APLIKASI MARKETPLACE TOKOPEDIA BERDASAR RATING DAN KOMENTAR PADA GOOGLE PLAY STORE

Saiful Ulya<sup>a,\*</sup>, Achmad Ridwan<sup>b</sup>, Widya Cholid Wahyudin<sup>c</sup>, Fida Maisa Hana<sup>d</sup>

<sup>ab</sup>Sistem Informasi, Universitas Muhammadiyah Kudus

<sup>cd</sup>Ilmu Komputer, Universitas Muhammadiyah Kudus

Purwosari, Kudus, Indonesia

Email : [saifululya@umkudus.ac.id](mailto:saifululya@umkudus.ac.id),

## Abstrak

*Text Mining* didefinisikan sebagai kegiatan intensif pengetahuan di mana pengguna berinteraksi dengan kumpulan dokumen dari waktu ke waktu dengan menggunakan sekumpulan alat analisis. Teknik penambangan data text atau *text mining* digunakan untuk menambang data text yang berupa kalimat, data yang tidak terstruktur tersebut kemudian diproses sehingga jadi sebuah dataset yang terstruktur. Dalam sektor bisnis apapun, *feedback* dari konsumen merupakan salah satu hal penting dari sebuah organisasi untuk memperoleh informasi dan keinginan dari konsumen sehingga tujuan melayani konsumen dengan baik akan tercapai. Suatu perusahaan teknologi tentunya memiliki banyak cara untuk memperoleh feedback dari konsumennya, salah satunya adalah dengan cara menganalisa komentar dan rating dari aplikasi yang mereka sediakan lewat platform store. Umpan balik dari konsumen akan mempengaruhi calon pengguna untuk menentukan keputusan akan menginstall atau tidak aplikasi tersebut. Pada penelitian ini kami menggunakan salah satu aplikasi yang disediakan oleh perusahaan teknologi penyedia jasa layanan jual beli atau *marketplace* Tokopedia pada *platform store PlayStore*. Kami menggunakan data komentar dan rating dari pengguna aplikasi tersebut untuk menganalisa sentiment dari para penggunanya dengan menggunakan beberapa cara. Cara pertama dengan model pengumpulan data dengan menambang data text dari komentar yang ada pada aplikasi yang Tokopedia yang berada di *Playstore*, data tersebut kami kumpulkan dengan mengambil 1000 *record* berdasar kelompok komentar paling relevan. Setelah data terkumpul data tersebut mengalami beberapa penyesuaian sehingga yang diambil hanya dua variable yaitu komentar dan rating sedangkan data seperti tanggal dan nama pengguna kami exclude dari data. Data kemudian mengalami preprosesing sehingga rating dibawah 3 dianggap sebagai sentiment negatif dan rating 3 sampai dengan 4 dianggap sentiment positif dan kemudian dijadikan sebagai label data. Setelah preprosesing data tersebut diproses dengan menggunakan beberapa metode (tokenisasi dan filter) sehingga data text tersebut berubah menjadi sebuah dataset terstruktur yang kemudian akan dapat digunakan untuk pemodelan *machine learning* dan *data mining sentiment analitic*.

**Kata Kunci:** *Text Mining, Tokenisasi, Dataset, Analisa Sentimen.*

## Abstract

*Text Mining is defined as a knowledge-intensive activity in which a user interacts with a collection of documents over time using a group of analytical tools. Text data mining techniques or text mining are used to mine text data in the form of sentences, the unstructured data is then processed so that it becomes a structured dataset. In any business sector, feedback from consumers is one of the important things for an organization to obtain information and wishes from consumers so that the goal of serving consumers well will be achieved. A technology company certainly has many ways to get feedback from its customers, one of which is by analyzing the comments and ratings of the applications they provide through the platform store. Feedback from consumers will influence potential users to determine whether to install the application or not. In this study, we used an application provided by a technology company that provides buying and selling services or the Tokopedia marketplace on the PlayStore platform store. We use data on comments and ratings from users of the application to analyze the sentiments of its users in several ways. The first way is with a data collection model by mining text data from comments on the Tokopedia application that is on the Playstore, we collect this data by taking 1000 records based on the most relevant group of comments. After the data was collected, the data underwent several adjustments so that only two variables were taken, namely comments and ratings, while data such as dates and usernames were excluded*

from the data. The data then undergoes preprocessing so that ratings below 3 are considered negative sentiment and ratings 3 to 4 are considered positive sentiment and are then used as data labels. After preprocessing the data is processed using several methods (tokenization and filters) so that the text data is turned into a structured dataset which can then be used for machine learning modeling and sentiment analytical data mining.

**Keywords:** Text Mining, Tokenisasi, Dataset, Sentiment Analiticss.

## I. PENDAHULUAN

Data merupakan salah satu bagian terpenting dalam proses pemodelan baik itu *Data Mining* maupun *Machine Learning*, karena data tersebut akan digunakan sebagai pelatihan dalam menentukan suatu model, terlebih lagi pada pemodelan *data mining* dan *machine learning* yang fokus pada sebuah data teks, misalnya untuk sentiment analisis. Proses pelatihan yang dilakukan dalam sentiment analisis cenderung lebih sulit apabila dibandingkan dengan bidang *Machine Learning* lainnya. Ini terjadi karena data yang digunakan dalam proses pelatihan merupakan data yang bersifat subjektif yang tidak memiliki nilai konkret, salah satu contoh dari data subjektif ini adalah opini. Karena data ini bersumber dari manusia sudah tentu setiap manusia memiliki selera dan cara yang berbeda-beda dalam menyampaikan pendapatnya[1].

*Zero Moment of Truth* (ZMOT) menunjukkan bahwa model pengambilan keputusan yang dilakukan oleh konsumen diperkenalkan oleh Google pada tahun 2011[2], kegiatan membaca sebuah informasi suatu produk serta *feedback* atau ulasan pengguna sebelumnya dari suatu produk secara *online* akan mempengaruhi pengguna untuk mengambil keputusan. Sebuah penelitian di Amerika menunjukkan bahwa 46% orang yang disurvei menyebutkan bahwa ulasan online sangat membantu meyakinkan mereka dalam mengambil keputusan untuk membeli dan 45% lainnya berusaha menemukan ulasan online menggunakan gadget mereka bahkan Ketika mereka sudah berada di toko fisik[3].

Di masa lalu, saat konsumen menghadapi produk mereka cukup kesulitan saat menentukan keputusan untuk membeli suatu produk. Sekarang zaman sudah berubah,

karena saat ini proses keputusan penggunaan suatu produk sangat dipengaruhi oleh ulasan online. Banyak penelitian yang telah dilakukan tentang rating dan kegunaan dari sebuah *feedback* suatu produk, karena *feedback* diyakini dapat menambah promosi dari produk tersebut[4][5]. Karena penelitian telah menunjukkan bahwa ulasan produk atau *feedback* dapat mempengaruhi kinerja penjualan, maka banyak perusahaan mendorong pembeli untuk menulis ulasan.

Perusahaan dapat menganalisa komentar pengguna untuk dijadikan sebagai bahan evaluasi perusahaan dalam upaya selalu memberi layanan terbaik kepada pengguna. Proses menganalisa suatu data text yang tidak teratur dapat menggunakan metode *text mining*. *Text mining* merupakan Kegiatan menambang data berupa text yang bersumber dari data sebuah dokumen, dengan tujuan untuk mencari kata-kata yang dapat mewakili isi dari dokumen tersebut sehingga keterhubungan antar dokumen dapat dipetakan dan dianalisa[6].

Data berkualitas yang tersedia saat ini terbilang masih minim terlebih data yang berkaitan dengan sentiment analisis berbahasa Indonesia. Dari problem tersebut mendorong kami untuk melakukan kegiatan penambangan data teks berbahasa Indonesia untuk dapat digunakan dalam pemodelan sentiment analisis. Dalam penelitian ini penambangan text dilakukan pada *Platform Google Play Store* berfokus pada ulasan aplikasi *marketplace* Tokopedia untuk memetakan sentiment analisis berdasar komentar positif dan negatif. Tokopedia merupakan perusahaan teknologi di Indonesia yang focus awalnya adalah menemukan penjual dan pembeli dalam sebuah sistem berbasis IT.

Salah satu alasan terkuat kami menggunakan Tokopedia sebagai objek penambangan data adalah karena Tokopedia terpilih menjadi *Marketeers of the Year* untuk sektor *e-commerce* pada acara *Markplus Conference* pada tahun 2014 yang diselenggarakan oleh *Markplus Inc.* Tokopedia dianggap telah memberikan kontribusi dalam mengembangkan bisnis online di Indonesia[7].

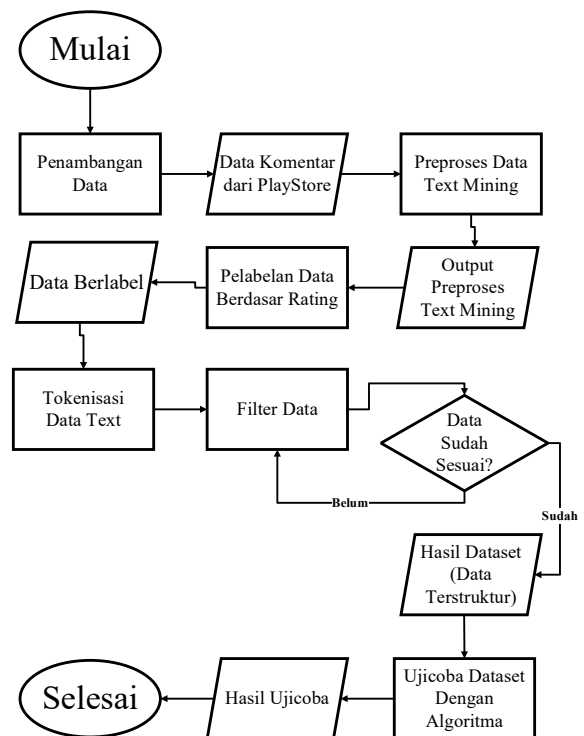
## II. LANDASAN TEORI

*Text Minig* didefinisikan sebagai kegiatan intensif pengetahuan di mana pengguna berinteraksi dengan kumpulan dokumen dari waktu ke waktu dengan menggunakan sekumpulan alat analisis[8]. Secara umum terdapat dua model pendekatan dalam melakukan sentiment analisis, kedua model tersebut yaitu secara *learning-based* (berbasis *machine learning*) dan *lexion-based* (berbasis *laksikal*). Pendekatan dengan *machine learning* harus menggunakan sebuah dataset terstruktur atau disebut *dataset* yang telah diklasifikasikan secara manual sebelumnya kemudian digunakan sebagai data latih untuk menghasilkan hasil klasifikasi teks secara otomatis. *Lexion-based* memiliki cara yang berbeda, yaitu bergantung pada kamus opini (*lexion*) dalam menentukan klasifikasi. Kamus opini berisi kumpulan kata yang akan digunakan untuk mengidentifikasi jenis opini suatu kalimat[9].

Terdapat beberapa penelitian yang berkaitan dengan konteks sentiment analisis dalam dan juga sudah tersedia beberapa dataset yang dipublikasikan secara umum. Salah satunya seperti yang publikasikan oleh Ferdiana R, Jatmiko F.[10], terdapat data dari penambangan data text twitter berbahasa - *General-Sentiment-Analysis-Dataset*, data tersebut memiliki 10.806 tweet yang kemudian 10% dari data tersebut dikategorikan sebagai tiga jenis yang pertama adalah tweet positif kemudian negative dan yang terakhir adalah netral, ada juga dataset yang diberi judul *Indonesian-Emotion-Twitter-Dataset*, data tersebut terdapat empat ribu empat ratus tiga record tweet dengan pengelompokan lebih spesifik yaitu *love, joy, anger, sadness, dan fear*[11].

## III. METODE PENELITIAN

Dalam penelitian ini, prosedur yang dilakukan terdiri dari beberapa proses. Proses pertama dimulai dengan penambangan data teks dari *google play store*, kemudian masuk pada praproses data, kemudian polaritas dan pelabelan data, tokenisasi, filtering, hingga data tersebut berubah menjadi data yang terstruktur. Diagram alir tahapan ditampilkan pada Gambar 1.



Gambar 1 Prosedur Penelitian

Setiap tahapan yang ada pada diagram alir di atas akan diproses menggunakan Python dan Rapid Miner. Pada Python menggunakan Jupiter Notebook dengan menggunakan beberapa library diantaranya *google-play-scraper, pandas, numpy, sastrawi, vaderSentiment, matplotlib.pyplot*. Sedangkan pada Rapid miner menggunakan *nominal to text*, dan beberapa extension *text processing* seperti *process document from data, filter*.

## IV. PEMBAHASAN DAN HASIL

### A. Penambangan Data dan Preprocessing

Proses penambangan data dilakukan dengan menggunakan *Google Colab* dengan menggunakan sebuah *Script Scrape* yang mengarah pada link platform Tokopedia dengan link store '`com.tokopedia.tkpd`'.

Data yang diambil sejumlah 1000 record komentar dengan model *sort.MOST\_RELEVANT* atau data yang paling relevan.

Dari proses tersebut diperoleh sebuah data dengan jumlah 10 atribut dan 1000 hasil

**Tabel 1** Hasil Penambangan Data

<b>reviewId</b>	581812d8-52ea-4b6d-8a74-ea9a12313b0d	b4f21ab2-a7a4-4288-b59c-47cf24785406
<b>userName</b>	iroiroSENNA.	Muhammad Alfi Himawan A. A
<b>userImage</b>	<a href="https://play-lh.googleusercontent.com/a-/AD5-WCkID83FLtBFdlbPyLcscySixlKKszuJXRzvCmp44w">https://play-lh.googleusercontent.com/a-/AD5-WCkID83FLtBFdlbPyLcscySixlKKszuJXRzvCmp44w</a>	<a href="https://play-lh.googleusercontent.com/a-/AD5-WC17vbw69_cSreDQJNOtvEGUvZpJsLcHJuk829fAwA">https://play-lh.googleusercontent.com/a-/AD5-WC17vbw69_cSreDQJNOtvEGUvZpJsLcHJuk829fAwA</a>
<b>content</b>	Quick review, jujur ini apk enteng di luar berat di dalam. Cuma Tokped yang buat HP hang bisa berimbas ke UI. Karena masih butuh ya pake dahh. Sering crash apknya. Untuk tampilannya sendiri agak ribet dan kesannya sempit. Kebanyakan fitur ndak tau navigasinya. Estimasi mempelajari ini apk sekitar 45 menit. Untuk promo cukup gratis ongkir yang ngebantu banget. Sisanya kayak gopay coins ndak terlalu butuh bagiku. Ada yg hrg miring juga. Keseluruhan, layak ndak layak masih banyak PRnya.	Tokopedia skrg sering banget error saat menambahkan produk yg mau dibeli, ini ganggu banget jadi ga bisa transaksi. Sering banget begini, padahal aplikasi terupdate, jaringan ga ada kendala, udh coba logout-login.
<b>score</b>	3	2
<b>thumbsUpCount</b>	43	20
<b>reviewCreatedVersion</b>	3205	3206
<b>at</b>	44965,40662	44967,41722
<b>replyContent</b>	Makasih ratingnya, Toppers. Ke depannya kami akan meningkatkan performa aplikasi, fitur serta layanan agar bisa memberikan pengalaman bertransaksi yang lebih baik lagi untuk semua pengguna.	
<b>repliedAt</b>	44965,43208	

Setelah tahapan penambangan data selesai dan diperoleh data berjumlah 1000 *record*, tahap selanjutnya adalah tahapan preprosesing, tujuan dari preprosing ini adalah mengambil data yang diperlukan dan mengeluarkan data

ulasan pengguna. Berikut tampilan dari Sebagian data hasil penambangan ditunjukkan pada Tabel 1.

yang tidak diperlukan. Data yang diperlukan adalah *content* dan *score*. Hasil preprosesing tahap satu dapat dilihat pada Tabel 2.

**Tabel 2** Preprosesing Tahap Satu.

<b>content</b>	<b>score</b>
Saat gunakan promo dan tagihan sudah keluar, Pembayaran tidak berhasil. Waktu tidak gunakan promo pembayaran langsung berhasil. Hoak	1
Penarikan saldo terlalu berbelit belit suruh ini lah suruh itu lah.. Gw tuh butuh banget dan itu harus menunggu 2 - 3 hari bahkan lebih.. Gak seperti toko sebelah 1 - 2 hari kerja langsung bisa cair.. Klo ini hadeh.. Melanggar ketentuan juga enggak.. Cuma baru sekali ini masukan rekening yang sebelumnya gak pernah tautkan rekening aja di persulit minta ampun.. Kapok pake jasa tokopedia.. Besok ke orange aja lah.. Payah banget pelayanannya	1
Di tahun 2023. Pengiriman semakin amburadul. Estimasi semakin tidak tepat, toko dan pembeli berada di satu pulau, bisa pula barangnya dibawa ke luar pulau. 10 hari tak sampai juga.	2

Sering sekali blanja tp pengiriman kurang profesional.. sering telat g sesuai estimasi... Mending kalau gratis ongkir udah toko gak menyediakan gratis ongkir masih aja pengiriman jauh melebihi estimasi..	2
Jelek banget. Masa barang yang dipesan nggak sampai-sampai. Ditungguin juga. Pengiriman nggak sesuai estimasi produk sampai. Sangat mengecewakan.	1
Ada masalah pembelian karena penjual ga amanah, udh lapor CS udah ngikutin alurnya tapi ga ada kejelasan, bener2 ga ada kelanjutan. Nyoba report seller tapi selalu gagal karena "alasan tidak valid" padahal pilih alasan juga yang tertera di pilihannya dan akhirnya nyoba 1 per 1 tapi tetep "alasan tidak valid" aneh. Ga bakal lagi pake e-com ini. Selalu ada notif kalau punya voucher gratis ongkir dll, tapi giliran mau dipake ga bisa dengan alasan "sudah melanggar" padahal ga pernah aneh2.	1
Tokopedia selalu ada selalu bisa, trima kasih untuk semuanya selama ini bertransaksi di e-commerce yaitu Tokopedia memang baik dan tepat untuk di recommended dari dulu hingga sekarang.	5

Pada tahap berikutnya yaitu pelabelan data yang diperoleh dari hasil preprosesing tahap satu, pelabelan data komentar menggunakan dua tipe yaitu Positif dan Negativ. Label Positif dan Negativ diperoleh dari rating yang disematkan pada komentar. Nilai rating >3 dianggap sebagai label Positif dan selain itu dianggap sebagai label Negativ. Parameter data yang sebelumnya adalah *score* dan *content* disesuaikan menjadi *content* dan **Tabel 3** Pelabelan Data

*sentiment* dimana *content* sebagai parameter dan *sentiment* sebagai label. Pada tahapan ini data record tinggal 996 *record* karena telah mengalami cleaning dengan jumlah label Positif sebanyak 166 baris dan label Negativ 828 baris. Hasil preview pelabelan data pada tahap ini dapat dilihat pada Tabel 3.

<i>content</i>	<i>sentiment</i>
Saat gunakan promo dan tagihan sudah keluar, Pembayaran tidak berhasil. Waktu tidak gunakan promo pembayaran langsung berhasil. Hoak	Negativ
Penarikan saldo terlalu berbelit belit suruh ini lah suruh itu lah.. Gw tuh butuh banget dan itu harus menunggu 2 - 3 hari bahkan lebih.. Gak seperti toko sebelah 1 - 2 hari kerja langsung bisa cair.. Klo ini hadeh.. Melanggar ketentuan juga enggak.. Cuman baru sekali ini masukan rekening yang sebelumnya gak pernah tautkan rekening aja di persulit minta ampun.. Kapok pake jasa tokopedia.. Besok ke orange aja lah.. Payah banget pelayanan nya	Negativ
Di tahun 2023... Pengiriman semakin amburadul. Estimasi semakin tidak tepat, toko dan pembeli berada di satu pulau, bisa pula barangnya dibawa ke luar pulau. 10 hari tak sampai juga.	Negativ
Sering sekali blanja tp pengiriman kurang profesional.. sering telat g sesuai estimasi... Mending kalau gratis ongkir udah toko gak menyediakan gratis ongkir masih aja pengiriman jauh melebihi estimasi..	Negativ
Jelek banget. Masa barang yang dipesan nggak sampai-sampai. Ditungguin juga. Pengiriman nggak sesuai estimasi produk sampai. Sangat mengecewakan.	Negativ
Ada masalah pembelian karena penjual ga amanah, udh lapor CS udah ngikutin alurnya tapi ga ada kejelasan, bener2 ga ada kelanjutan. Nyoba report seller tapi selalu gagal karena "alasan tidak valid" padahal pilih alasan juga yang tertera di pilihannya dan akhirnya nyoba 1 per 1 tapi tetep "alasan tidak valid" aneh. Ga bakal lagi pake e-com ini. Selalu ada notif kalau punya voucher gratis ongkir dll, tapi giliran mau dipake ga bisa dengan alasan "sudah melanggar" padahal ga pernah aneh2.	Negativ
Tokopedia selalu ada selalu bisa, trima kasih untuk semuanya selama ini bertransaksi di e-commerce yaitu Tokopedia memang baik dan tepat untuk di recommended dari dulu hingga sekarang.	Positif

### B. Tokenizing

Pada bagian ini, content akan dipecah dengan menggunakan tokenizing. Tokenizing merupakan proses menguraikan atau memecah yang sebelumnya string menjadi

satuan kata. Pada dasarnya tokenizing ini merupakan proses pemenggalan kalimat menjadi kata. Hasil tokenizing ditampilkan pada Tabel 4.

Tabel 4 Hasil Tokenizing

content	Hasil Tokenizing
Saat gunakan promo dan tagihan sudah keluar, Pembayaran tidak berhasil. Waktu tidak gunakan promo pembayaran langsung berhasil. Hoak	'Saat', 'gunakan', 'promo', 'dan', 'tagihan', 'sudah', 'keluar', 'Pembayaran', 'tidak', 'berhasil.', 'Waktu', 'tidak', 'gunakan', 'promo', 'pembayaran', 'langsung', 'berhasil.', 'Hoak'
banyak bug pas verifikasi nomor hp suruh masukan kode verifikasi pas udh di masukan malah nalik lagi dari awal, terus pas masukan lagi nomor hpnya malah ada bacaan, nomor hp terverifikasi padahal belum pernah di daftarin. apk seterkenal ini masih banyak bug	'banyak', 'bug', 'pas', 'verifikasi', 'nomor', 'hp', 'suruh', 'masukan', 'kode', 'verifikasi', 'pas', 'udh', 'di', 'masukan', 'malah', 'nalik', 'lagi', 'dari', 'awal, terus', 'pas', 'masukan', 'lagi', 'nomor', 'hpnya', 'malah', 'ada', 'bacaan, nomor', 'hp', 'terverifikasi', 'padahal', 'belum', 'pernah', 'di', 'daftarin. apk', 'seterkenal', 'ini', 'masih', 'banyak', 'bug'
Sering sekali blanja tp pengiriman kurang profesional.. sering telat g sesuai estimasi... Mending kalau gratis ongkir udah toko gak menyediakan gratis ongkir masih aja pengiriman jauh melebihi estimasi..	'Sering', 'sekali', 'blanja', 'tp', 'pengiriman', 'kurang', 'profesional..', 'sering', 'telat', 'g', 'sesuai', 'estimasi...', 'Mending', 'kalau', 'gratis', 'ongkir', 'udah', 'toko', 'gak', 'menyediakan', 'gratis', 'ongkir', 'masih', 'aja', 'pengiriman', 'jauh', 'melebihi', 'estimasi'
Jelek banget. Masa barang yang dipesan nggak sampai-sampai. Ditungguin juga. Pengiriman nggak sesuai estimasi produk sampai. Sangat mengecewakan.	'Jelek', 'banget.', 'Masa', 'barang', 'yang', 'dipesan', 'nggak', 'sampai-sampai.', 'Ditungguin', 'juga.', 'Pengiriman', 'nggak', 'sesuai', 'estimasi', 'produk', 'sampai.', 'Sangat', 'mengecewakan'

### C. Filtering

Filtering merupakan proses mengambil kata penting dari hasil tokenizing. Filtering dapat dengan menggunakan model *stop list* atau *word list*. Filtering juga bisa disebut sebagai proses mengambil kata-kata penting dari hasil proses tokenizing atau menghapus *stopword*. *Stopword* adalah kumpulan kosakata yang tidak merupakan ciri dari dokumen seperti ada, adalah, adanya, adapun, agak, agaknya. Data *Stopword Bahasa Indonesia* tersedia di kumpulan dataset terbuka Kagle. Preview *Stopword* ditampilkan pada Tabel 5.

Tabel 5 Preview Stopword

ada
adalah
adanya
adapun
agak
agaknya
agar
akan
akankah
akhir

### D. Stemming

*Stemming* disebut sebagai bagian mencari root kata dari hasil filtering. *Stemming* adalah proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya (*stem*).

### E. Proses Penghitungan

Algoritma TF/IDF digunakan dalam proses perhitungan bobot (TF/IDF) terminologi kata. Algoritma ini digunakan untuk menghitung bobot setiap kata yang paling umum digunakan pada information retrieval. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat [12]. Metode ini akan menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. Metode ini akan menghitung bobot setiap token *t* di dokumen *d* dengan rumus:

$$W_{dt} = tf_{dt} * IDF_t$$

Dimana

- $d$  : dokumen ke-d
- $t$  : kata ke-t dari kata kunci
- $w$  : bobot dokumen ke-d terhadap kata ke-1
- $tf$  : term frequency (banyaknya kata yang dicari pada sebuah dokumen)
- IDF : Inversed Document Frequency
- $D$  : Total Dokumen
- $df$  : banyak dokumen yang mengandung kata yang dicari

nilai  $tf$  diperoleh dari

$$tf_d = \frac{\text{Jumlah munculnya kata } t \text{ dalam dokumen}}{\text{Total jumlah seluruh kata dalam dokumen}}$$

Setelah bobot ( $W$ ) masing-masing dokumen diketahui, maka dilakukan proses pengurutan dimana semakin besar nilai  $W$ , semakin besar tingkat similaritas dokumen tersebut terhadap kata kunci, demikian sebaliknya.

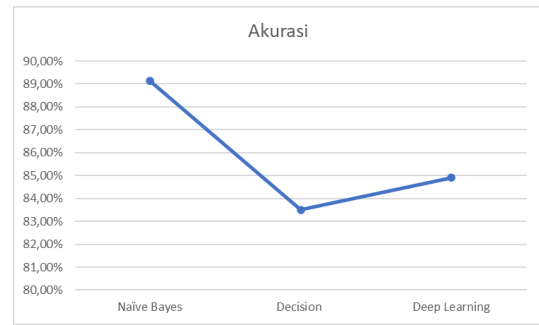
**F. Hasil**

Dari berbagai tahapan yang telah dilakukan, didapatkan sebuah data terstruktur dari hasil text mining yang telah dilakukan sebelumnya. Untuk menguji kelayakanh dataset yang diperoleh dalam penelitian berikutnya, maka dilakukan 3 kali percobaan dengan menggunakan algoritma klasifikasi.

Pemodelan yang pertama dilakukan dengan Naïve Bayes dengan hasil akurasi mencapai 89.13%, kemudian pengujian kedua dengan Decision Tree dengan akurasi mencapai 83.51%, dan pengujian ketiga dengan Deep Learning dengan akurasi mencapai 84.91%. Tabel perbandingan dari ketiga algoritma yang telah digunakan ditunjukkan pada Tabel 6.

**Tabel 6.** Perbandingan Hasil Pengujian

Algoritma	Akurasi
Naïve Bayes	89,13%
Decision	83,51%
Deep Learning	84,91%



**Gambar 2** Grafik Perbandingan Hasil Pengujian

**V. KESIMPULAN**

Pada makalah ini, dipublikasikan sebuah dataset berbahasa Indonesia, yaitu Feedback-Sentiment-AnalysisDataset, yang merupakan dataset berupa ulasan dari pengguna aplikasi marketplace Tokopedia pada Google Play Store untuk analisis sentimen. Data yang diambil ini berjumlah 1000 ulasan dan telah diberi label dengan dua polaritas sentimen, yaitu Positif dan Negatif. Kedua polaritas ditentukan berdasar rating pada komentar, nilai >3 dianggap sebagai sentimenPositif dan selain itu dianggap sentimenn Negativ. Selain itu, juga disampaikan hasil simulasi pengujian berupa perbandingan pemodelan dataset tersebut dengan menggunakan 3 algoritma yaitu Naïve Bayes, Decision Tree dan Deep Learning. Dari ketiga simulasi tersebut diperoleh hasil yang cukup baik, dan memungkinkan dataset tersebut dapat digunakan sebagai bahan untuk penelitian-penelitian analisis sentimen di waktu yang akan datang.

Namun, pada paper ini, dan pada Dataset yang telah dihasilkan masih terdapat beberapa beberapa hal yang belum sempurna, diantaranya belum dimilikinya label polaritas emosi yang spesifik, seperti senang, sedih, amarah, dan lain-lain. Dataset ini dapat diperoleh dan dapat dikembangkan untuk penelitian berikutnya dengan mengunduh tautan <https://s.id/1zzGD>.

**DAFTAR PUSTAKA**

G. Vinodhini dan R. M. Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey," Int. J. of Advanced Research in Computer Science and



- Software Engineering, vol. 2, no. 6, pp. 282–292, 2012.
- J. Lecinski, “ZMOT WINNING THE ZERO MOMENT OF TRUTH.”
- Smith & Anderson, “<https://www.pewresearch.org/internet/2016/12/19/online-shopping-and-e-commerce/>,” 2016.
- C. Dellarocas, X. Zhang, and N. F. Awad, “Exploring the value of online product reviews in forecasting sales: The case of motion pictures,” *Journal of Interactive Marketing*, vol. 21, no. 4, pp. 23–45, 2007, doi: 10.1002/dir.20087.
- W. Duan, B. Gu, and A. B. Whinston, “Do online reviews matter? - An empirical investigation of panel data,” *Decis Support Syst*, vol. 45, no. 4, pp. 1007–1016, Nov. 2008, doi: 10.1016/j.dss.2008.04.001.
- R. J. Mooney, “CS 391L: Machine Learning Text Categorization Text Categorization Applications • Web pages- Recommending-Yahoo-like classification • Newsgroup/Blog Messages-Recommending-spam filtering-Sentiment analysis for marketing • News articles-Personalized newspaper • Email messages-Routing-Prioritizing-Folderizing-spam filtering-Advertising on Gmail Text Categorization Methods.”
- Tokopedia, “<https://www.tokopedia.com/about/our-story/>,” 2014.
- O. Y. Findawati, M. M. Muhammad, A. Rosid, S. Kom, and M. Kom, *BUKU AJAR TEXT MINING* Diterbitkan oleh UMSIDA PRESS.
- Y. Azhar, “METODE LEXICON-LEARNING BASED UNTUK IDENTIFIKASI TWEET OPINI BERBAHASA INDONESIA,” 2017.
- R. Ferdiana, F. Jatmiko, D. D. Purwanti, A. Sekar, T. Ayu, and W. F. Dicka, “Dataset Indonesia untuk Analisis Sentimen,” 2019.
- M. Dong, Universitas Telkom, Chinese and Oriental Languages Information Processing Society, Institute of Electrical and Electronics Engineers. Indonesia Section. Computer Society Chapter, and Institute of Electrical and Electronics Engineers, Proceedings of the 2018 International Conference on Asian Language Processing (IALP): 15-17 November 2018, Telkom University, Bandung, Indonesia.
- Abdul Azis Maarif, “PENERAPAN ALGORITMA TF-IDF UNTUK PENCARIAN KARYA ILMIAH,” 2015.