

ANALISIS SENTIMEN ULASAN PRODUK TOKO ONLINE RUBYLICIOUS UNTUK PENINGKATAN LAYANAN MENGGUNAKAN ALGORITMA NAIVE BAYES

Humairah¹, Irfan Darmawan², Oktariani Nurul Pratiwi³

^{1,2,3} S1 Sistem Informasi, Fakultas Rekayasa Industri, Universitas Telkom

¹humairah@student.telkomuniversity.ac.id, ²dirfand@telkomuniveristy.co.id,

³onurulp@telkomuniversity.ac.id

Abstrak

Berbelanja online merupakan solusi bagi masyarakat yang tidak punya waktu banyak dan tenaga untuk mengunjungi pusat perbelanjaan. Berbagai keuntungan jual beli online tidak hanya dirasakan oleh para pembeli saja tapi juga dirasakan oleh para penjual yang berdagang secara online karena cakupan calon pembeli lebih luas dan mereka tidak perlu mengeluarkan biaya untuk membangun atau menyewa tempat sebagai toko fisik. Namun tidak dapat dipungkiri bahwa berbelanja secara online juga memiliki kekurangan seperti kesalahan produk yang tidak sesuai dengan pesanan. Rubylicious merupakan salah satu toko online yang saat ini melakukan penjualan melalui Shopee. Penelitian ini dilakukan untuk mengategorikan dan menganalisis pandangan pelanggan terhadap produk dengan memanfaatkan data ulasan produk Rubylicious yang diambil dari kolom komentar pada aplikasi Shopee. Jumlah data ulasan yang sangat banyak menyebabkan proses analisis sulit dan kurang efektif. Untuk mengatasi masalah tersebut maka digunakan metode analisis sentimen terhadap ulasan barang menggunakan algoritma Naive Bayes Classifier (NBC). Hasil analisis berupa sentimen positif, negatif dan netral. Dari hasil analisis sentimen menggunakan NBC diperoleh nilai akurasi sebesar 82,39%, precision 91,3%, recall 51% dan f1-measure 51,6%. Hasil analisis sentimen kemudian divisualisasikan menggunakan wordcloud untuk melihat kata-kata yang paling sering muncul pada masing-masing sentimen. Dari hasil penelitian ini, sentimen positif dapat dijadikan sebagai acuan untuk mempertahankan kualitas layanan yang juga mencakup penjualan produk dan hasil sentimen negatif dapat digunakan sebagai bahan evaluasi dalam peningkatan layanan yang mencakup penjualan produk pada toko online Rubylicious.

Kata kunci: Toko online, ulasan, analisis sentimen, Naive Bayes.

Abstract

Shopping online is a solution for people who don't have a lot of time and energy to visit the spending center. The various advantages of buying and selling online are not only accepted by buyers but are also accepted by sellers who trade online because they support prospective buyers more broadly and they don't need to pay for building or buying a place as a physical store. However, it cannot be denied because buying online also has shortcomings such as products that are not in accordance with the order. Rubylicious is an online store that currently sells through Shopee. This research was conducted to categorize and analyze customer views on products using Rubylicious product review data taken from the comments column in the Shopee application. The large amount of review data makes the analysis process difficult and less effective. To overcome these problems, the sentiment analysis method for goods is used using the Naive Bayes Classifier (NBC) algorithm. The results of the analysis are positive, negative and neutral sentiments. From the results of sentiment analysis using NBC obtained an accuracy value of 82.39%, precision 91.3%, recall 51% and f1-measure 51.6%. The results of sentiment analysis are then visualized using WordCloud to see the words that most often appear on each sentiment. From the results of this study, positive sentiments can be made as a reference to maintain service quality which also includes product sales and negative sentiment results can be used as evaluation material in improving services that involve product sales at Rubylicious online stores.

Keywords: Online shop, review, sentiment analysis, Naive Bayes.

1. Pendahuluan

Di era digital ini, aktivitas belanja sudah dapat dilakukan tanpa harus mendatangi toko atau pasar yang menjual barang-barang yang kita butuhkan seperti makanan, alat elektronik, pakaian dan lain-lain. Hal ini dikarenakan sudah banyaknya toko *online* yang menyediakan barang untuk kebutuhan sehari-hari sehingga aktivitas belanja dapat dilakukan hanya dengan mengandalkan gadget. Kini beberapa toko *online* sudah menyediakan kolom komentar pada situs toko *online* mereka dengan tujuan agar para pembeli bisa memberikan ulasan terhadap produk dan layanan yang mereka berikan baik berupa ulasan positif maupun negatif. Salah satu toko *online* yang sering mendapatkan ulasan positif ataupun negatif adalah Rubylicious.

Rubylicious merupakan sebuah bisnis menengah yang bergerak di industri *fashion retail* yang berpusat di kota Bandung yang kini juga menjual produk mereka secara online melalui aplikasi Shopee. Rubylicious dapat memanfaatkan ulasan-ulasan pada kolom komentar untuk mengevaluasi pendapat dan tingkah laku dari pelanggan terhadap produk mereka. Hal ini dapat dilakukan dengan mengategorikan dan menganalisis pandangan seseorang terhadap sebuah barang sehingga bisa meningkatkan daya guna serta penjualan barang tersebut dengan mengetahui kelemahan barang dari sudut pengguna [1]

Masalah yang muncul adalah ulasan pada kolom komentar toko *online* Rubylicious terlalu banyak sehingga cukup sulit dalam mengategorikan dan menganalisis ulasan para pelanggan. Untuk mengetahui kecenderungan komentar serta informasi yang terdapat dalam *review* atau ulasan bukan hal yang mudah, karena jumlah data ulasan yang terlalu banyak dan memakan waktu lama untuk menyimpulkan jika dibaca secara manual [2]. Oleh karena itu dibutuhkan suatu metode untuk mengolah data ulasan pelanggan secara cepat dan otomatis untuk mengategorikan ulasan-ulasan tersebut baik ulasan positif maupun ulasan yang bersifat negatif. Metode yang dapat digunakan untuk mengetahui kecenderungan komentar dalam ulasan tersebut adalah dengan melakukan analisis sentimen. Analisis sentimen membantu seorang penjual untuk mengevaluasi pendapat dan tingkah laku dari klien terhadap barang mereka, sehingga penjual tersebut mendapatkan *review* barang mereka secara langsung dari klien dengan menggunakan media sosial seperti Twitter, Instagram atau komentar pada sebuah toko *online* [1]

2. Dasar Teori

2.1. Text Mining

Text mining adalah sebuah teknik berbasis komputer untuk mendapatkan suatu pengetahuan baru yang tersembunyi dari sekumpulan teks. Text mining merupakan bagian dari keilmuan information retrieval yang bekerja pada data bertipe teks yang cenderung tidak terstruktur [3]. text mining merupakan pendekatan algoritmik yang secara sistematis memproses data teks melalui beberapa tahapan. Secara umum, tahapan besar dalam *text mining* terdiri dari tiga bagian utama yakni teks *preprocessing*, *feature selection*, dan *text analytic* [4]

2.2. Analisis Sentimen

Analisis sentimen adalah tugas menemukan opini dari penulis tentang suatu entitas tertentu. Analisis sentimen pada *review* adalah proses menyelidiki *review* produk di internet untuk menentukan opini atau perasaan terhadap suatu produk secara keseluruhan. Menurut Thelwall dalam Haddi, analisa sentimen diperlakukan sebagai suatu tugas klasifikasi yang mengklasifikasikan orientasi suatu teks ke dalam positif atau negatif [5]

Salah satu hal yang paling penting dalam melakukan analisis sentimen adalah tahap *preprocessing*. Di dalam tahap *preprocessing* inilah kita melakukan *case folding* yaitu mengubah semua huruf dalam dokumen menjadi huruf kecil [6], *punctuation removal* yaitu proses penghapusan angka dan tanda baca pada tiap kalimat yang bertujuan agar angka dan tanda baca tersebut tidak akan mengganggu proses perhitungan dalam penerapan algoritma *text mining* [7], selanjutnya *tokenization* yaitu pemotongan urutan karakter dan sebuah set dokumen yang diberikan menjadi potongan-potongan kata atau karakter yang sesuai dengan kebutuhan sistem. Potongan-potongan tersebut dikenal dengan istilah token [7], *stopwords removal* yaitu menghilangkan kata-kata umum yang tidak memiliki makna atau informasi yang dibutuhkan [8], dan *stemming* yaitu proses mengembalikan suatu kata ke bentuk dasar dengan menghilangkan imbuhan pada awalan dan akhiran [9].

2.3. TF-IDF

TF-IDF (*Term Frequency - inverse Document Frequency*) adalah salah satu algoritma yang digunakan untuk menganalisa hubungan antara sebuah kalimat dengan sekumpulan dokumen dan metode untuk menghitung bobot setiap kata yang paling umum di gunakan [7]. Penggunaan TF-IDF pada analisis sentimen bertujuan untuk mengekstrak kata ke dalam bentuk numerik agar data dapat diolah menggunakan *machine learning* pada tahap selanjutnya. Hal ini disebabkan oleh metode *machine learning* yang tidak bisa mengolah data dalam bentuk selain numerik sehingga data yang berupa kata atau teks perlu melalui tahap ekstraksi yang pada umumnya dapat dilakukan menggunakan metode pembobotan kata TF-IDF (*Term Frequency - inverse Document Frequency*) [10]. Nilai TF-IDF bisa didapatkan dengan formula berikut:

$$W(tk) = tfk \cdot \log (N/dfk)$$

$W(tk)$: nilai TF-IDF
 tfk : jumlah kata tk pada dokumen
 N : total dokumen
 dfk : jumlah dokumen yang didalamnya terdapat term tk

2.4. Naive Bayes Classifier

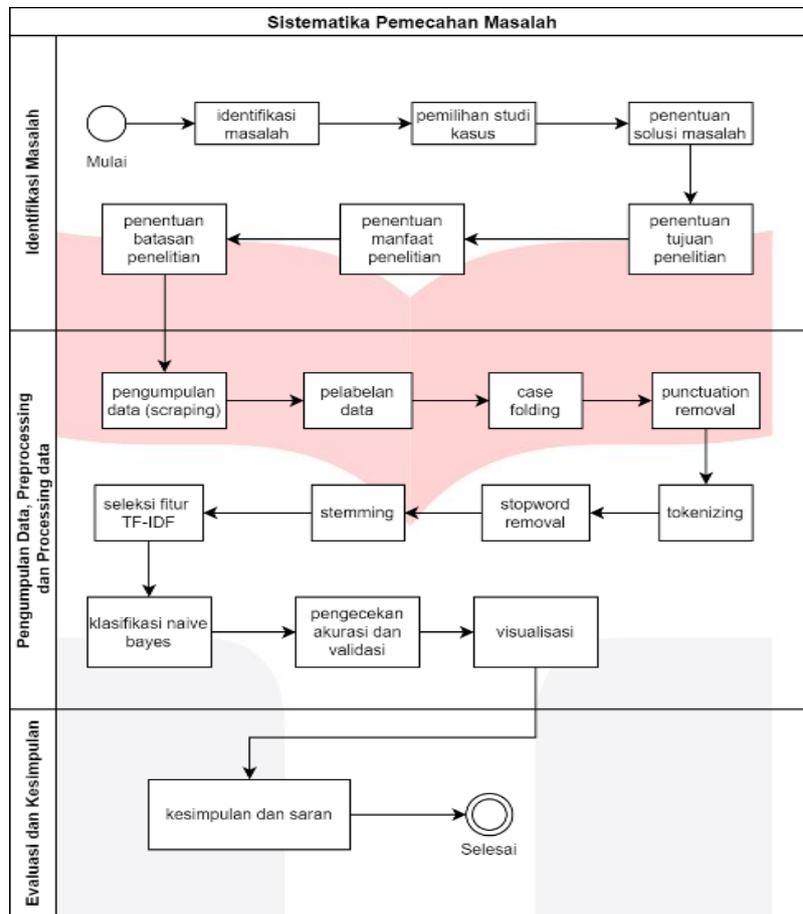
Naive Bayes adalah salah satu algoritma dalam text mining yang digunakan untuk melakukan klasifikasi teks yang menggunakan perhitungan probabilitas dan statistik yang dikemukakan oleh Thomas Bayes. Naive Bayes memungkinkan klasifikasi berdasarkan asumsi kondisi tersendiri antara prediksi atribut diberikan class. Untuk itu Naive Bayes Classifier adalah klasifikasi yang benar-benar kompeten, bekerja cukup baik dalam tugas-tugas klasifikasi sehingga banyak peneliti yang mencoba untuk meningkatkan performa Naive Bayes [11]. Rumus naive bayes classifier adalah sebagai berikut :

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

$P(C|X)$: Probabilitas suatu kelas terhadap dokumen
 $P(X|C)$: Probabilitas suatu dokumen terjadi pada suatu kelas
 $P(C)$: Probabilitas kelas klasifikasi
 $P(X)$: Probabilitas kemunculan dokumen

3. Metodologi

Pada penelitian ini akan dilakukan analisis sentimen ulasan toko online Rubylicious yang terdapat pada e-commerce sebagai media penjualan online mereka yaitu Shopee. Proses analisis akan menghasilkan klasifikasi teks ulasan produk berdasarkan label yang tercantum pada dataset. Bahasa pemrograman yang digunakan untuk memproses data pra-pemrosesan adalah Python, sedangkan untuk mengklasifikasikan dataset peneliti menggunakan algoritma *Naive Bayes*. Berikut ini adalah sistematika pemecahan masalah yang dilakukan oleh peneliti untuk penelitian ini:



Gambar 1 Sistematika Pemecahan Masalah

4. Pengumpulan dan Pengolahan Data

4.1 Dataset

Data yang digunakan diambil dari ulasan customer Rubylicious pada platform jual beli Shopee. Pengambilan data ulasan dilakukan dengan teknik *scraping* menggunakan bahasa pemrograman Python. Jumlah data yang terkumpul yaitu sebanyak 7208 data yang merupakan ulasan dari produk yang terdapat pada toko Rubylicious hingga 27 Mei 2020 dan kemudian disimpan dalam bentuk file CSV.

Data ulasan yang telah terkumpul kemudian diberi label dengan angka 1, 2 atau 3. Angka 1 mewakili sentimen positif, angka 2 mewakili sentimen negatif, sedangkan angka 3 mewakili sentimen netral. Sedangkan angka 3 mewakili sentimen netral. Berikut contoh ulasan yang telah diberi label:

Tabel 1 Data ulasan yang telah diberi label

No.	Text	Label
1	Kualitas produk sangat baik, pengiriman produk sangat cepat	1
2	Kenapa warna navy kesannya seperti barang lama ya, mungkin efek bahan atau apa, tapi okelah.	3

3	Bagus.. enak banget dipakai, habis lahiran di bb 65 kg.. mantap	1
4	Pengiriman sangat lama	2
5	Bahannya cakep banget. tebal. Halus tapi adem. Beli buat teman semoga dia suka dan pas di dia. Bakalan beli lagi buat aku juga deh.	1

Berikut ini adalah tabel jumlah data ulasan yang telah diberi label:

Tabel 2 Jumlah Ulasan Produk

Label	Jumlah Ulasan
1	5276
2	1600
3	332
Total	7208

4.2 Preprocessing

Case folding dilakukan untuk meyetarakan huruf pada dokumen menjadi huruf kecil. Tabel 3 menggambarkan hasil case folding dari beberapa data ulasan:

Tabel 3 Case Folding

No	Sebelum	Sesudah
1	Kualitas produk sangat baik, pengiriman produk sangat cepat	kualitas produk sangat baik, pengiriman produk sangat cepat
2	Kenapa warna navy kesannya seperti barang lama ya, mungkin efek bahan atau apa, tapi okelah.	kenapa warna navy kesannya seperti barang lama ya, mungkin efek bahan atau apa, tapi okelah.
3	Bagus.. enak banget dipakai, habis lahiran di bb 65 kg.. mantap	bagus.. enak banget dipakai, habis lahiran di bb 65 kg.. mantap
4	Pengiriman sangat lama	pengiriman sangat lama

Punctuation removal dilakukan untuk menghapus angka dan tanda baca pada tiap kalimat yang bertujuan agar angka dan tanda baca tersebut tidak mengganggu proses perhitungan dalam penerapan algoritma *text mining*.

Tabel 4 Punctuation Removal

No	Sebelum	Sesudah
1	Kualitas produk sangat baik, pengiriman produk sangat cepat	kualitas produk sangat baik pengiriman produk sangat cepat
2	Kenapa warna navy kesannya seperti barang lama ya, mungkin efek bahan atau apa, tapi okelah.	kenapa warna navy kesannya seperti barang lama ya mungkin efek bahan atau apa tapi okelah
3	Bagus.. enak banget dipakai, habis lahiran di bb 65 kg.. mantap	bagus enak banget dipakai habis lahiran di bb kg mantap
4	Pengiriman sangat lama	pengiriman sangat lama

Tokenizing dilakukan untuk memisahkan kalimat menjadi potongan-potongan kata yang disebut sebagai token untuk kemudian dianalisa.

Tabel 5 Tokenizing

No	Sebelum	Sesudah
1	Kualitas produk sangat baik, pengiriman produk sangat cepat	[kualitas, produk, sangat, baik, pengiriman, produk, sangat, cepat]
2	Kenapa warna navy kesannya seperti barang lama ya, mungkin efek bahan atau apa, tapi okelah.	[kenapa, warna, navy, kesannya, seperti, barang, lama, ya, mungkin, efek, bahan, atau, apa, tapi, okelah]
3	Bagus.. enak banget dipakai, habis lahiran di bb 65 kg.. mantap	[bagus, enak, banget, dipakai, habis, lahiran, di, bb, kg, mantap]
4	Pengiriman sangat lama	[pengiriman, sangat, lama]

Stopword removal dilakukan untuk menghilangkan kata-kata yang sering muncul namun tidak penting seperti kata ganti atau kata penghubung.

Tabel 6 Stopword Removal

No	Sebelum	Sesudah
1	Kualitas produk sangat baik, pengiriman produk sangat cepat	kualitas produk baik pengiriman produk cepat
2	Kenapa warna navy kesannya seperti barang lama ya, mungkin efek bahan atau apa, tapi okelah.	warna navy kesannya seperti barang lama efek bahan
3	Bagus.. enak banget dipakai, habis lahiran di bb 65 kg.. mantap	bagus enak dipakai habis lahiran mantap
4	Pengiriman sangat lama	pengiriman lama

Stemming dilakukan untuk menghilangkan imbuhan kata dengan tujuan mentransformasikan kata ke bentuk dasarnya sehingga data siap diolah untuk proses selanjutnya.

Tabel 7 Stemming

No	Sebelum	Sesudah
1	Kualitas produk sangat baik, pengiriman produk sangat cepat	kualitas produk baik kirim produk cepat
2	Kenapa warna navy kesannya seperti barang lama ya, mungkin efek bahan atau apa, tapi okelah.	warna navy kesan seperti barang lama efek bahan
3	Bagus.. enak banget dipakai, habis lahiran di bb 65 kg.. mantap	bagus enak pakai habis lahiran mantap
4	Pengiriman sangat lama	kirim lama

5. Analisis Sentimen

Sebelum melakukan analisis sentimen menggunakan *Naive Bayes* perlu dilakukan ekstraksi fitur menggunakan TF-IDF. Data akan dibagi menjadi data *training* dan data *testing*. Dalam hal ini data *training* dan data *testing* dibagi dengan rasio 80:20 yang berarti 80% dari jumlah data akan digunakan sebagai data training dan 20% dari jumlah data akan digunakan sebagai data testing. Rincian untuk jumlah komentar pada data testing dapat dilihat pada tabel berikut:

Tabel 8 Data Testing

No.	Label	Ulasan
1	Positif	1056
2	Negatif	320
3	Netral	66

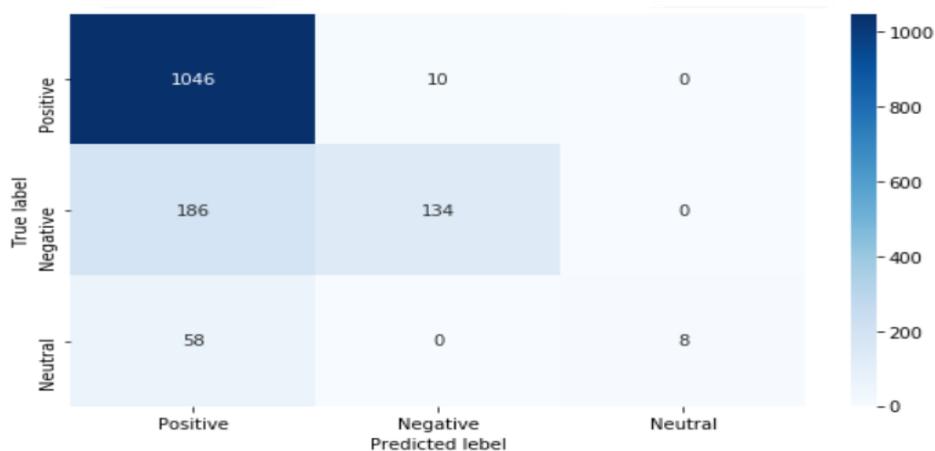
Setelah melakukan klasifikasi menggunakan *naive bayes*, didapatkan hasil yang berbeda dengan pelabelan manual. Berikut perbandingan hasil pelabelan manual dengan pelabelan prediksi sistem:

Tabel 9 Perbandingan Hasil Pelabelan manual dan prediksi sistem

Label	Ulasan	
	Manual	Prediksi Sistem
Positive	1056	1290
Negative	320	144
Neutral	66	8

6. Akurasi

Dalam penelitian kali ini dilakukan uji akurasi terhadap algoritma dengan menggunakan confusion matrix. Berikut tabel confusion matrix dari hasil pengujian data testing dengan ekstraksi fitur TF-IDF



Gambar 2 Confusion Matrix

Berdasarkan gambar di atas dapat dijabarkan bahwa:

1. True Positive atau ulasan yang diprediksi benar berjumlah 1046 ulasan sedangkan False Positive atau ulasan positif yang di prediksi salah berjumlah 10 ulasan.
2. True Negatif atau ulasan negatif yang diprediksi benar berjumlah 134 sedangkan False Negative atau ulasan negatif yang di prediksi salah oleh sistem berjumlah 186 ulasan.
3. True Neutral atau ulasan netral yang diprediksi benar berjumlah 8 ulasan sedangkan False Neutral atau ulasan netral yang di prediksi salah berjumlah 58 ulasan.

Berdasarkan penjabaran confusion matriks maka diperoleh nilai akurasi sebesar 82% dengan perhitungan:

$$\begin{aligned}
 \text{Akurasi} &= ((1046 + 134 + 8) / (1046 + 10 + 134 + 186 + 8 + 58)) 100 \\
 &= (1188 / 1442) 100 \\
 &= 82,39\%
 \end{aligned}$$

- [3] M. Inzalkar and J. Sharma, "A Survey of Text Mining Techniques and Applications.," *International Journal of Research In Science & Engineering.*, 2015.
- [4] Priyato and Ma'arif, "Implementasi Web Scraping dan Text Mining untuk Akuisisi dan Kategorisasi Informasi Laman Web Tentang Hidroponik.," *Indonesian Journal of Information Systems (IJIS)*, pp. 27-28., 2018.
- [5] D. A. Mutia, "ANALISIS SENTIMEN PADA REVIEW RESTORAN DENGAN TEKS BAHASA INDONESIA MENGGUNAKAN ALGORITMA NAIVE BAYES. JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER," *JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER*, pp. 40-41, 2017.
- [6] I. Sunni and D. H. Widyantoro, "Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik.," *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika*, p. 202, 2012.
- [7] P. E. Utomo, K. U. Manaar and T. Suratno, "ANALISIS SENTIMEN ONLINE REVIEW PENGGUNA BUKALAPAK MENGGUNAKAN METODE ALGORITMA TF-IDF.," *Jurnal Sains dan Sistem Informasi*, p. 36, 2019.
- [8] A. P. Syah, Adiwijaya and S. H. Faraby, "ANALISIS SENTIMEN PADA DATA ULASAN PRODUK TOKO ONLINE DENGAN METODE MAXIMUM ENTROPY.," *e-Proceeding of Engineering*, p. 4635, 2017.
- [9] J. Ling, I. Kencana and T. B. Oka, "ANALISIS SENTIMEN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DENGAN SELEKSI FITUR CHI SQUARE.," *E-Jurnal Matematika*, p. 93, 2014.
- [10] Y. Setiowati and Helen Afrida, "KLASIFIKASI ANALISIS SENTIMEN MENGENAI HOTEL DI YOGYAKARTA," p. 4, 2018.
- [11] A. M. Hidayat and M. Syafrullah, "ALGORITMA NAÏVE BAYES DALAM ANALISIS SENTIMEN UNTUK KLASIFIKASI PADA LAYANAN INTERNET PT.XYZ," *Jurnal ELEMATIKA MKOM*, p. 93, 2017.
- [12] N. Anggraini, Jasmir and P. A. Jusia, "Penerapan Metode K-Means Clustering untuk Menentukan Persediaan Stok Barang Pada Toko Pensmart Jambi.," *Jurnal Ilmiah Mahasiswa Teknik Informatika*, pp. 64-65, 2019.
- [13] H. Annur, "Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering.," *JURNAL INFORMATIKA UPGRIS*, p. 42, 2019.