

Sentimen Analisis *Customer Review* Produk Shopee Indonesia Menggunakan Algoritma *Naïve Bayes Classifier*

Loemongga Oktaria Sihombing^{*1}, Hannie², Budi Arif Dermawan³
^{1,2,3} Program Studi Teknik Informatika, Universitas Singaperbangsa Karawang
email: loemongga.oktaria17123@student.unsika.ac.id^{*1}, hannie@staff.unsika.ac.id²,
budi.arif@staff.unsika.ac.id³

(Received: 15 Oktober 2021 / Accepted: 17 November 2021 / Published Online: 20 Desember 2021)

Abstrak

Mendapatkan kepuasan serta kepercayaan *customer* telah menjadi tantangan utama dalam mencapai kesuksesan dalam dunia bisnis. Pelaku bisnis perlu mengidentifikasi masalah yang muncul dari *review* yang diberikan oleh para *customer*. Namun, membaca dan mengklasifikasikan setiap *review* membutuhkan waktu lama dan dianggap tidak efektif. Untuk mengatasi hal tersebut, maka penelitian ini bertujuan untuk menganalisis sentimen *customer* produk shopee menggunakan algoritma *naïve bayes classifier*. Data yang digunakan pada penelitian ini adalah *customer review* dari produk *Xiaomi Redmi Note 9* yang dijual pada *website* Shopee Indonesia. Data *review customer* dikumpulkan dengan menerapkan teknik *Web Scraping*. Algoritma yang digunakan pada penelitian ini adalah *Naïve Bayes Classifier* yang dikenal populer dan efektif dalam melakukan klasifikasi data. Penelitian ini juga menerapkan metodologi *Knowledge Discovery in Text* (KDT) untuk menggali informasi dari sebuah data teks. Hasil klasifikasi menggunakan algoritma *Naïve Bayes* mendapati nilai akurasi sebesar 85%. Penelitian ini membuktikan bahwa dengan menerapkan teknik analisis sentimen, para pelaku bisnis mampu mengetahui opini para *customer* sebagai bahan evaluasi yang perlu dilakukan untuk melakukan optimalisasi terhadap produk serta layanan yang diberikan.

Kata kunci: Analisis Sentimen, *Knowledge Discovery in Text*, *Naïve Bayes Classifier*, *Shopee*, *Text Mining*.

Abstract

Gaining customer satisfaction and trust has become the main challenge in achieving success in the business world. Business people need to identify problems that arise from reviews given by customers. However, reading and classifying each review takes a long time and is considered ineffective. To overcome this, this study aims to analyze the customer sentiment of shopee products using the *naïve Bayes classifier* algorithm. The data used in this study is a customer review of the *Xiaomi Redmi Note 9* products which are sold on the *Shopee Indonesia website*. Customer review data is collected by applying the *Web Scraping* technique. The algorithm used in this study is the *Naïve Bayes Classifier* which is known to be popular and effective in classifying data. This study also applies the *Knowledge Discovery in Text* (KDT) methodology to extract information from text data. The results of the classification using the *Naïve Bayes* algorithm found an accuracy value of 85%. This study proves that by applying sentiment analysis techniques, business people are able to find out the opinions of customers as an evaluation material that needs to be done to optimize the products and services provided.

Keywords: *Sentiment Analysis*, *Knowledge Discovery in Text*, *Naïve Bayes Classifier*, *Shopee*, *Text Mining*.

PENDAHULUAN

Kegiatan jual beli barang dan jasa secara *online* atau yang biasa disebut dengan *e-commerce* telah berkembang pesat saat ini. Salah satu *marketplace* dengan model *e-commerce* yang populer khususnya di Indonesia adalah *Shopee*. *E-commerce* menawarkan pengguna

dalam melakukan sebuah transaksi barang maupun jasa tanpa batas jarak dan waktu selama pengguna memiliki akses internet (Išoraitė & Miniotienė, 2018). Namun begitu, tidak jarang *customer* merasa ragu membeli produk secara *online* dengan harga pasar yang tinggi seperti *smartphone*. Hal ini dibuktikan berdasarkan hasil penelitian sebelumnya yang mendapati sebuah kesimpulan bahwa kepercayaan pelanggan mempengaruhi niat pembelian produk *smartphone* di Shopee Indonesia. Pelanggan dinilai masih merasa kurang percaya terhadap para penjual *smartphone* dalam memenuhi keinginan para pelanggan (Picaulty, 2018).

Kurangnya rasa kepercayaan *customer* terhadap penjual menjadi tantangan bagi para pelaku bisnis agar mampu melakukan evaluasi terhadap produk dan layanannya. Pelaku bisnis perlu mengidentifikasi masalah yang muncul dari *review* yang diberikan oleh para *customer*. Namun, membaca dan mengklasifikasikan setiap *review* yang diberikan oleh para *customer* membutuhkan waktu lama, dan dianggap tidak efektif (Wiratama & Rusli, 2019). Untuk mengatasi masalah tersebut, dibutuhkan sistem yang dapat mengklasifikasikan opini ke dalam kelas sentimen positif, netral, atau negatif secara otomatis yang biasa disebut dengan analisis sentimen (Santoso et al., 2017).

Naïve Bayes merupakan salah satu algoritma yang mampu melakukan pengklasifikasian dengan cepat. *Naïve Bayes* juga merupakan salah satu algoritma yang sangat efisien dan efektif bahkan saat digunakan untuk menganalisis data berskala besar (Ahmad et al., 2017). Penelitian sebelumnya yang dilakukan oleh (Pintoko, 2018), mendapatkan hasil akurasi yang cukup tinggi yaitu sebesar 86,80%. Penelitian tersebut dilakukan untuk menganalisa tanggapan masyarakat terhadap jasa transportasi *online* dengan analisis data berupa *tweet* kemudian diklasifikasikan menjadi kelas positif dan negatif menggunakan metode *Naïve Bayes Classifier*.

Data mining merupakan salah satu teknik atau metode yang digunakan dalam proses ekstraksi pengetahuan dari kumpulan data. *Data mining* mampu mempelajari kebiasaan pelanggan sehingga para pelaku bisnis mampu mengembangkan strategi yang lebih efektif (Umarani et al., 2020). Klasifikasi merupakan salah satu teknik yang ada didalam *data mining* dan berguna dalam pengelompokan data berdasarkan keterikatan data dengan data sampel (Harianto et al., 2020; Pramadhana, 2021; Sari et al., 2020). Teknik yang diterapkan pada penelitian saat ini adalah teknik klasifikasi dengan menggunakan algoritma *Naïve Bayes Classifier*. Pendekatan ini dimulai dari proses pelatihan dataset yang kemudian dilakukan pengembangan fungsi yang telah disimpulkan untuk memperkirakan nilai *output* (Saravanan & Sujatha, 2019). Setelahnya, algoritma ini akan membandingkan hasil yang diperoleh dengan hasil aktual atau yang diharapkan untuk mengidentifikasi kesalahan (Saravanan & Sujatha, 2019). Dengan meningkatnya kebutuhan informasi dan pengetahuan dari data teks, algoritma ini sudah semakin banyak digunakan untuk klasifikasi data teks (Kristiyanti et al., 2019).

Penelitian saat ini berfokus pada analisis sentimen menggunakan algoritma *Naïve Bayes* terhadap produk *smartphone* yang diperjual belikan di *website* Shopee. Penjualan *smartphone* di Shopee cukup banyak diminati dibandingkan dengan *marketplace* lainnya seperti Tokopedia dan Lazada. Hal ini dibuktikan dengan hasil pengamatan peneliti terhadap produk *smartphone* yang terjual di beberapa *marketplace*. Produk *smartphone* yang akan menjadi objek penelitian saat ini merupakan *smartphone* keluaran tahun 2020 yaitu *Xiaomi Redmi Note 9* yang telah terjual lebih dari 10.000 produk. Banyaknya produk yang terjual diikuti dengan banyaknya *review* yang masuk terhadap produk tersebut. Tujuan dari penelitian ini adalah menganalisis sentimen untuk mengetahui opini *customer* terhadap sebuah produk serta layanan yang telah diterima oleh para *customer*. Informasi yang didapatkan bisa menjadi bahan evaluasi dalam melakukan optimalisasi guna meningkatkan kepercayaan *customer* terhadap para pelaku bisnis.

METODE

Data mining merupakan salah satu teknik atau metode yang digunakan dalam proses ekstraksi pengetahuan dari kumpulan data untuk mengubah data mentah menjadi sebuah

informasi atau pengetahuan yang berguna (Shu-Hsien et al., 2017). Salah satu penerapan konsep dan teknik *data mining* untuk mencari pola dalam teks adalah *text mining*. Manfaat *text mining* akan sangat terasa di bidang-bidang yang memiliki banyak data berupa teks seperti manajemen risiko, deteksi penipuan, *business intelligence*, serta analisis media sosial (Ferreira-Mello et al., 2019).

Penelitian ini menggunakan teknik analisis sentimen untuk menghasilkan sebuah informasi berupa pandangan seseorang terhadap suatu kejadian atau isu. Teks yang tidak terstruktur mudah diproses dan dirasakan oleh manusia, tetapi secara signifikan lebih sulit untuk dipahami oleh mesin (Allahyari et al., 2017). Maka dari itu, analisis sentimen diperlukan untuk dapat memahami opini seseorang yang tertulis di *platform online* secara otomatis dengan bantuan mesin. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau fitur atau tingkat aspek dan menentukan apakah pendapat yang dikemukakan dalam dokumen, kalimat atau fitur entitas atau aspek bersifat positif, negatif atau netral (Murnawan, 2017).

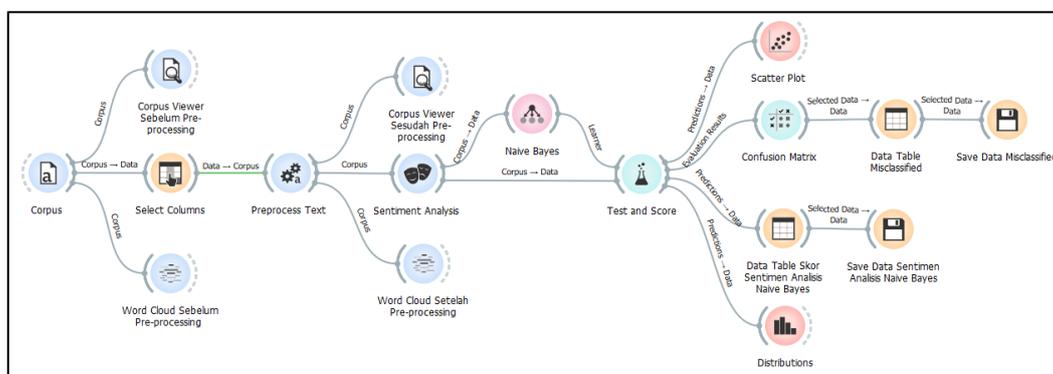
Metodologi yang digunakan pada penelitian ini adalah *Knowledge Discovery in Text* (KDT). *Knowledge Discovery in Text* atau disingkat dengan KDT merupakan bagian dari proses penggalian atau pencarian data teks yang sebelumnya tidak diketahui menjadi sebuah informasi yang dapat dimengerti (Firdaus, 2017). Adapun tahapan yang akan dilakukan pada penelitian ini diantaranya adalah *collecting* dan *labeling data*, dimana data didapat dari hasil *web scraping website* Shopee yang kemudian diberi label *positive*, *neutral*, dan *negative* secara manual. Data yang telah diberi label kemudian masuk ke dalam tahap *preprocessing* untuk membersihkan *noise* dari sebuah data sebelum akhirnya data diklasifikasi menggunakan algoritma *Naïve Bayes Classifier*. Tahap terakhir yang akan dilakukan pada penelitian ini adalah *evaluation* berupa *classification accuracy*.

Penelitian saat ini menggunakan *tools software Orange Data Mining* sebagai perangkat bantu analisis. *Orange Data Mining* merupakan salah satu *software open source* untuk mengolah data, visualisasi data, *machine learning*, dan penambangan data atau *data mining* yang ditulis menggunakan bahasa *Python* (Mohi, 2020).

HASIL DAN PEMBAHASAN

Hasil

Skenario penelitian yang akan dilakukan pada penelitian ini diperlihatkan pada gambar 1 dengan menggunakan bantuan *software Orange Data Mining*.



Gambar 1. Skenario Penelitian

Collection dan Labeling Data

Data yang akan digunakan pada penelitian ini adalah data *review* salah satu produk *smartphone* dengan merek *Xioami Redmi Note 9* yang dijual secara *online* di *website* Shopee Indonesia dan dikelola langsung oleh *xiaomi.official.id*. Pengambilan data dilakukan dengan menerapkan teknik *web scraping* menggunakan bahasa *Python*. Data mentah yang terkumpul

berjumlah 3006 *review* yang kemudian dilakukan pemberian label atau *labeling* secara manual ke dalam beberapa kelas diantaranya *positive*, *neutral*, dan *negative* sesuai dengan sentimen kata yang berada pada *review* produk. *Review* yang tidak mengandung sentimen manapun dan tidak berkaitan dengan produk tersebut akan diberi label *unrelated*. Data dengan label *unrelated* kemudian dihapus sehingga menyisakan kelas *positive*, *neutral*, dan *negative* seperti yang terlihat pada tabel 1.

Tabel 1. Jumlah Data

Dengan Unrelated		Tanpa Unrelated	
Kelas	Jumlah	Kelas	Jumlah
Positive	2475	Positive	2475
Neutral	372	Neutral	372
Negative	102	Negative	102
Unrelated	57		
Total	3006	Total	2949

Pre-processing Data

Data testing yang telah terbentuk akan melalui empat tahap *pre-processing* diantaranya adalah *Transformation*, *Tokenization*, *Filtering*, dan *N-Gram*. *Transformation* merupakan proses mengubah atau mentransformasi data yang sebelumnya mentah menjadi data yang lebih seragam guna mempermudah proses analisis. Beberapa proses transformasi yang digunakan pada penelitian ini diantaranya adalah sebagai berikut:

1. *Lowercase*: Transformasi *lowercase* akan mengubah huruf kapital menjadi huruf kecil secara *default* untuk mempermudah proses analisis. Contoh : Barang bagus, saya suka! → barang bagus, saya suka!.
2. *Remove accents*: berfungsi untuk mengubah data teks yang memiliki aksentuasi menjadi teks tanpa aksentuasi. Contoh : Naïve Bayes → Naive Bayes.
3. *Parse html*: Parse html pada proses transformasi bertujuan untuk mendeteksi *tag html* dan mengurainya menjadi bentuk teks. Contoh : <a href ...> Produk rusak! Kecewa banget. Nyesel belanja disini ... → Produk rusak! Kecewa banget. Nyesel belanja disini ...
4. *Remove urls* : Transformasi ini berfungsi untuk menghapus *urls* yang berada pada data teks. Contoh : Hpnya bagus, warnanya juga cantik. Ga nyesel beli disini. Jangan lupa tonton video aku di <https://www.youtube.com/watch?v=EGKcC2S44Rs> → Hpnya bagus, warnanya juga cantik. Ga nyesel beli disini. Jangan lupa tonton video aku di.

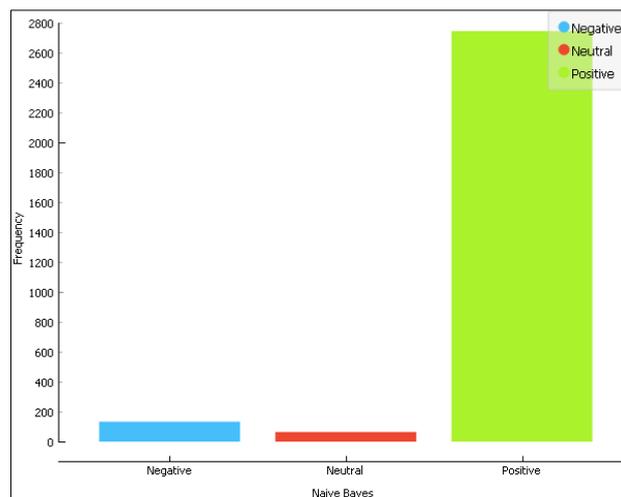
Proses selanjutnya pada tahapan *pre-processing* merupakan *Tokenization* dimana teks akan dipecah menjadi komponen yang lebih kecil. Proses *Tokenization* dilakukan dengan menerapkan *Regex* atau *Regular Expression* pada *software Orange Data Mining* dimana data akan dipisahkan menjadi satuan kata tanpa mengikutsertakan tanda baca seperti titik (.), dan koma (,).

Filtering dilakukan untuk menghapus komponen yang tidak penting dan hanya mempertahankan komponen yang dianggap penting sehingga proses analisis dapat dilakukan dengan lebih mudah. Fitur yang akan digunakan pada penelitian ini diantaranya adalah *Stopwords*, *Regex*, dan *Most frequent tokens*.

1. *Stop Word Removal* berfungsi untuk menghapus kata yang tidak memiliki bobot atau kata tidak penting seperti yang terlihat pada tabel 2.
2. *Regular Expression* atau *Regex* pada *Orange Data Mining* secara *default* diatur untuk menghapus tanda baca yang tidak memiliki bobot atau peran penting dalam proses analisis. Tabel 3 memperlihatkan sejumlah tanda baca dan angka yang akan dihilangkan pada sebuah data.

Tabel 4. Hasil Skoring *Naive Bayes*

Label	Review	Naïve Bayes	NB Negative	NB Neutral	NB Positive
Positive	untuk pengirimannya cepat sekali, dn packingnya juga rapih	Positive	0,016	0,134	0,849
Negative	baterai kapasitas gede.tp gk ada bedanya sama betre yg brkpasitas kcil.jg gk ada jarum pembuka slot.mgkn d jual terpisah x ya...	Negative	0,505	0,262	0,232
Neutral	barang sampai dengan aman dan selamat....tp di deskripsi helio 85g, tp nyatanya helio 65g sedikit mengecewakan	Negative	0,505	0,262	0,232



Gambar 3. Distributions Klasifikasi Naïve Bayes

Pada Gambar 3 terlihat bahwa *review* dengan label *positive* memiliki frekuensi terbanyak yang menunjukkan bahwa produk serta pelayanan yang diberikan sudah cukup baik. Namun begitu, terlihat pada grafik bahwa masih ada *review* dengan label *negative*, yang berarti masih terdapat kekurangan pada produk serta pelayanan yang diberikan. Total masing-masing label hasil klasifikasi *Naïve Bayes* adalah 2750 *review* dengan label *positive*, 65 *review* dengan label *neutral*, dan 134 *review* dengan label *negative*. Data kemudian di filter untuk mengetahui apa saja *review customer* yang memiliki label *negative* sebagai bahan evaluasi pengelola toko.

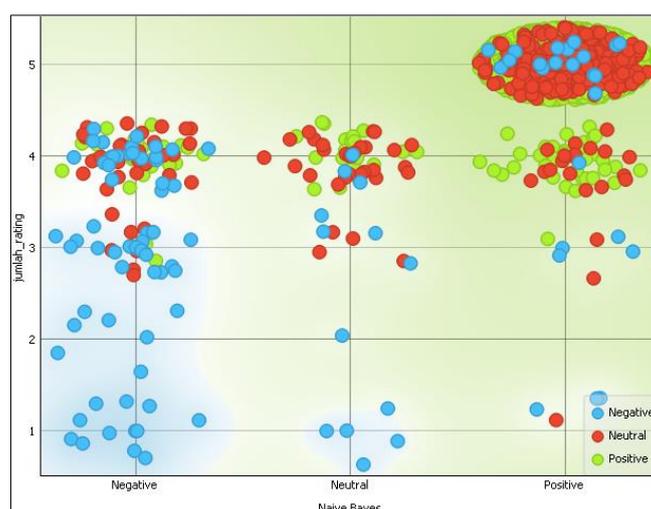
Tabel 5 memperlihatkan beberapa contoh *review customer* yang diklasifikasikan ke dalam label *negative* oleh algoritma *Naïve Bayes*. Sebagian besar data dengan label *negative* menunjukkan kekecewaan pada *customer* terhadap *packing* yang rusak saat dalam perjalanan, kurangnya barang, dan pengiriman yang lama. Dengan mengetahui kekurangan yang ada, pengelola toko diharapkan mampu melakukan perbaikan serta pengoptimalan dalam produk serta pelayanan yang diberikan.

Setelah dilakukan proses skoring dan klasifikasi sentimen terhadap data teks menggunakan algoritma *Naïve Bayes* maka selanjutnya dapat dilakukan visualisasi dengan *widget Scatter Plot* mengenai keterhubungan antara jumlah *rating* yang diberikan para *customer* dengan hasil klasifikasi sentimen menggunakan *Naïve Bayes* dan didapati hasil seperti pada gambar 4.

Tabel 5. *Naïve Bayes Negative*

Review	<i>Naïve Bayes</i>
gua beli hp, bukan beli botol susu yang cuma di packing asal asalan	<i>Negative</i>
Knp gad kartu garansi?	<i>Negative</i>
Pengiriman lama buaangett	<i>Negative</i>
Packingnya uhuu seremm . Kerdusnya penyok”	<i>Negative</i>
Barang sampe sekarang belum tiba	<i>Negative</i>
Pengiriman lama sgt	<i>Negative</i>

Berdasarkan visualisasi *scatter plot* pada gambar 4, terlihat bahwa pemberian jumlah *rating* dan isi *review* tidak begitu relevan dimana terlihat masih ada *review* dengan nilai sentimen negatif berada di *rating* 5 atau *rating* tertinggi dan *review* dengan nilai sentimen *neutral* lebih banyak tersebar pada nilai sentimen *positive*.

Gambar 4. *Scatter Plot Rating dan Naïve Bayes*

Evaluation

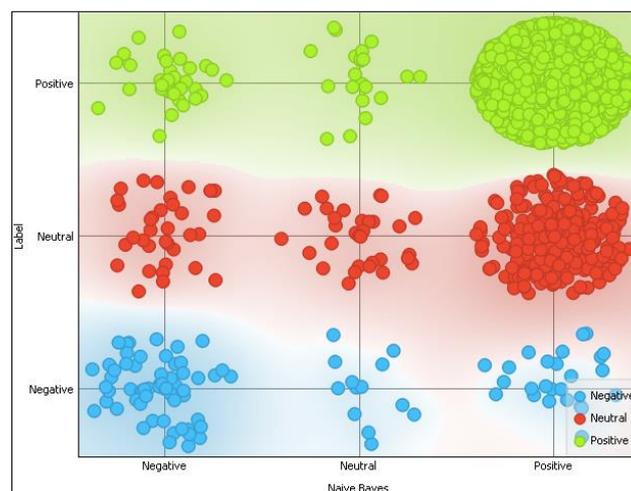
Evaluasi dilakukan untuk mengetahui performa algoritma *Naïve Bayes* dalam melakukan klasifikasi analisis sentimen. Evaluasi pada penelitian ini menggunakan metode *Cross Validation* dengan 10 *fold* atau 10 lipatan. Data dibagi menjadi 10 lipatan dengan ukuran yang kira-kira sama, kemudian 10 *subset* data tersebut yang dipakai untuk mengevaluasi kinerja model atau algoritma. Evaluasi *cross validation* akan menggunakan 9 *fold* sebagai *data training* dan 1 *fold* sisanya sebagai *data testing*. Berdasarkan evaluasi *cross validation* yang dilakukan, didapati hasil akurasi yang cukup baik yaitu sebesar 85%. Dengan hasil akurasi yang didapat, diketahui bahwa ada beberapa *misclassified data* hasil klasifikasi dengan algoritma *Naïve Bayes*. Untuk mengamati contoh spesifik dari sebuah data yang salah diklasifikasi, maka akan dilakukan proses evaluasi lanjutan dengan menerapkan metode *confusion matrix*.

Gambar 5 memperlihatkan bahwa masih terdapat banyak data yang salah diklasifikasi atau *misclassified*. Berdasarkan hasil evaluasi, total data yang mengalami *misclassified* berjumlah 437 data. Dari keseluruhan data *misclassified*, ditemukan bahwa data dengan label *neutral* merupakan data yang paling banyak mengalami kesalahan klasifikasi. Hal ini dikarenakan *Lexicon* atau kamus kata yang terdapat pada *Multilingual Sentiment* hanya memiliki dua kelas yaitu *positive* dan *negative*, sehingga metode tersebut dinilai masih belum akurat untuk memprediksi teks yang memiliki nilai sentimen *neutral*. Visualisasi persebaran data juga dapat dilihat pada gambar 6.

		Predicted			Σ
		Negative	Neutral	Positive	
Actual	Negative	64	13	25	102
	Neutral	34	31	307	372
	Positive	36	21	2418	2475
Σ		134	65	2750	2949

Gambar 5. *Confusion Matrix*

Berdasarkan visualisasi *scatter plot* pada gambar 6, terlihat persebaran data yang mengalami *misclassified* dan terlihat juga bahwa data dengan aktual label *Neutral* adalah data terbanyak yang mengalami *misclassified*. Sebagian besar data dengan aktual label *neutral* saat diklasifikasikan dengan algoritma *Naïve Bayes* menjadi data berlabel *positive*.



Gambar 6. *Scatter Plot*

Pembahasan

Naïve Bayes merupakan salah satu algoritma yang populer dalam melakukan sebuah klasifikasi sentimen. Hal ini dapat dibuktikan dari hasil penelitian saat ini yang memiliki tingkat akurasi yang cukup baik yaitu sebesar 85%. Hasil akurasi yang didapat juga dipengaruhi oleh tahap *preprocessing* yang dilakukan. *Transformation, tokenization, filtering* dan *n-grams* mampu menghilangkan *noise* yang ada pada data sehingga data yang akan memasuki proses analisis menjadi lebih baik. Namun begitu, masih banyak data yang ketika diklasifikasi menghasilkan nilai prediksi yang salah. Data dengan aktual label *neutral* mendominasi *misclassified* data. Hal ini dipengaruhi oleh kata yang sulit dikenali sebagai nilai sentimen *neutral*. Pada umumnya kalimat bersentimen *neutral* memiliki variasi kata yang lebih beragam dan cenderung tidak memiliki kata yang khas seperti nilai sentimen lainnya sehingga sulit dikenali (Santoso et al., 2017). *Lexicon* atau kamus kata yang terdapat pada *Multilingual Sentiment* hanya memiliki dua kelas yaitu *positive* dan *negative*, sehingga metode tersebut dinilai masih belum akurat untuk memprediksi teks yang memiliki nilai sentimen *neutral*.

Klasifikasi sentimen menggunakan *Naïve Bayes* menghasilkan prediksi 2750 *review* dengan label *positive*, 65 *review* dengan label *neutral*, dan 134 *review* dengan label *negative*. Hal ini membuktikan bahwa dengan melakukan analisis sentimen, para pelaku bisnis mampu mengetahui pandangan *customer* terhadap produk serta pelayanan yang telah diberikan. Dari hasil yang didapat, terlihat bahwa hasil penjualan produk *Xiaomi Redmi Note 9* memiliki tingkat kepuasan yang cukup tinggi. Namun begitu, masih terdapat beberapa *customer* yang

tidak begitu puas dengan apa yang didapat. Beberapa ungkapan ketidakpuasan customer berfokus pada *packing* yang rusak dalam perjalanan, kurangnya barang, dan pengiriman yang lama. Hasil temuan kami sesuai dengan hasil penelitian yang dilakukan oleh (Picaulty, 2018). Dimana hasil tersebut menjadi salah satu faktor yang menguatkan hasil penelitian sebelumnya mengenai kepercayaan pelanggan terhadap niat pembelian produk *smartphone* di Shopee Indonesia yang dinilai masih kurang.

SIMPULAN

Algoritma *Naïve Bayes* yang digunakan sebagai model klasifikasi dinilai mampu melakukan sebuah prediksi terhadap sebuah data teks yang mengandung nilai sentimen *positive*, *neutral*, atau *negative*. Masing-masing data teks dianalisis dan melalui proses perhitungan untuk menemukan skor dari masing-masing label atau nilai sentimen. Sentimen dengan skor tertinggi kemudian akan dijadikan hasil sebuah prediksi dari penerapan klasifikasi dengan algoritma *Naïve Bayes*. Penggunaan algoritma *Naïve Bayes* sebagai model klasifikasi sentimen memiliki hasil akurasi yang cukup tinggi yaitu 85%. Walaupun begitu, masih terdapat banyak data yang tidak terklasifikasi dengan baik atau bisa disebut dengan *data misclassified*. Terdapat total 436 data yang tidak terklasifikasi sempurna. Banyak dari data tersebut adalah data dengan aktual label *neutral* yang dinilai sulit untuk dikenali.

REFERENSI

- Ahmad, M., Aftab, S., Muhammad, S. S., & Ahmad, S. (2017). Machine Learning Techniques for Sentiment Analysis : A Review. *International Journal of Multidisciplinary Sciences and Engineering*, 8(3), 27–32.
- Allahyari, M., Pouriye, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). *A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques*. <https://arxiv.org/abs/1707.02919>
- Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(6), e1332. <https://doi.org/https://doi.org/10.1002/widm.1332>
- Firdaus, A. (2017). Reputation Scoring Fake News Using Text Mining. *ACMIT Proceedings*, 4(1), 12–17. <https://doi.org/10.33555/acmit.v4i1.52>
- Hariato, H., Sunyoto, A., & Sudarmawan, S. (2020). Optimasi Algoritma *Naïve Bayes* Classifier untuk Mendeteksi Anomaly dengan Univariate Fitur Selection. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 40–49.
- Işoraitè, M., & Miniotienè, N. (2018). Electronic Commerce: Theory and Practice. *IJBE (Integrated Journal of Business and Economics)*, 2(2), 194–200. <https://doi.org/10.33019/ijbe.v2i2.78>
- Kristiyanti, D. A., Umam, A. H., Wahyudi, M., Amin, R., & Marlinda, L. (2019). Comparison of SVM *Naïve Bayes* Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter. *The 6th International Conference on Cyber and IT Service Management, CITSM*, 1–6. Indonesia: IEEE. <https://doi.org/10.1109/CITSM.2018.8674352>
- Mohi, Z. R. (2020). Orange Data Mining as a tool to compare Classification Algorithms. *Dijlah Journal of Sciences and Engineering*, 3(3), 13–23.
- Murnawan, M. (2017). Pemanfaatan Analisis Sentimen Untuk Peningkatan Popularitas Tujuan Wisata. *Jurnal Penelitian Pos Dan Informatika*, 7(2), 109–120. <https://doi.org/10.17933/jppi.2017.070203>
- Picaulty, M. R. (2018). Pengaruh Kepercayaan Pelanggan Terhadap Niat Pembelian Gadget Di Shopee Indonesia. *Jurnal Manajemen Maranatha*, 18(1), 31–40. <https://doi.org/10.28932/jmm.v18i1.1094>

- Pintoko, B. (2018). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier. *E-Proceeding of Engineering*, 5(3), 8121–8130.
- Pramadhana, D. (2021). Klasifikasi Penyakit Diabetes Menggunakan Metode CFS Dan ROS dengan Algoritma J48 Berbasis Adaboost. *Edumatic: Jurnal Pendidikan Informatika*, 5(1), 89–98. <https://doi.org/10.29408/edumatic.v5i1.3336>
- Santoso, V. I., Virginia, G., & Lukito, Y. (2017). Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine. *Jurnal Transformatika*, 14(2), 72–76. <https://doi.org/10.26623/transformatika.v14i2.439>
- Saravanan, R., & Sujatha, P. (2019). A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification. *Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems, ICICCS 2018, Iciccs*, 945–949. <https://doi.org/10.1109/ICCONS.2018.8663155>
- Sari, V., Firdausi, F., & Azhar, Y. (2020). Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naive Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 1–9. <https://doi.org/10.29408/edumatic.v4i2.2202>
- Shu-Hsien, L., Pei-Hui, C., & Pei-Yuan, H. (2017). Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems with Applications*, 39, 11303–11311.
- Umarani, J., Manikandan, S., Centre, D., & Nadu, T. (2020). Implementation of Data Mining Concepts in R Programming. *International Journal of Trendy Research in Engineering and Technology*, 4(1), 1–7.
- Wiratama, G. P., & Rusli, A. (2019). Sentiment analysis of application user feedback in Bahasa Indonesia using multinomial naive bayes. *The 5th International Conference on New Media Studies, CONMEDIA*, 223–227. Indonesia: IEEE. <https://doi.org/10.1109/CONMEDIA46929.2019.8981850>