ANALISIS SENTIMEN APLIKASI JASA KURIR DI *PLAY STORE* MENGGUNAKAN ALGORITMA *NAIVE BAYES*

Adhitia Erfina[1], Muhamad Fani Al-shufi[2]
Program Studi Sistem Informasi, Universitas Nusa Putra
Cibolang Kaler, Kecamatan Cisaat, Kabupaten Sukabumi
Email : [1] adhitia.erfina@nusaputra.ac.id,[2] muhamad.fani_si18@nusaputra.ac.id

Abstract

Along with the passage of sales through the marketplace. The courier service industry has also experienced an increase. There are many choices of courier services to choose from, each courier service has different qualities depending on how the service provided by the courier service to customers. To find out the quality of courier services, this research was carried out, this research was conducted on courier service applications on Playstore, usually the best application predicate is for applications with a high number of downloads and ratings, while comments from users must also be considered to get results which is relevant. In this study, the Naïve Bayes algorithm is used because it has a high level of accuracy to determine the best courier service application. Based on the research that has been done, the highest accuracy value is obtained by the JNT application of 100% but has a positive sentiment of 50 reviews. The two JNE applications with an accuracy value of 98% with a positive sentiment of 46 reviews, the three ninjaxpres applications with an accuracy value of 97.87% with a positive sentiment of 49 reviews, the fourth by the Sicepat application with an accuracy value of 97.85% with a positive sentiment of 47 reviews and finally by the Idexpress application with an accuracy value of 94% with a positive sentiment of 49 reviews.

Keywords : *Naive Bayes, Playstore, Sentiment Analysis*

PENDAHULUAN

Efek dari semakin memburuknya keadaan karena meluasnya infeksi COVID-19, menjadikan hampir semua sektor terkena dampaknya. Tak terkecuali dalam sektor ekonomi. Pengurangan aktivitas memiliki dampak pada *sector* bisnis dan berimbas pada perekonomian. WFH (*work from home*) merupakan langkah kebijakan pemerintah daerah untuk mengurangi penyeberan virus ini dengan meminimalisir kegiatan orang banyak. Tetapi dengan pesatnya pertambahan kemajuan teknologi informasi dan komunikasi sekarang ini menjadi pemucu meningkatnya aktivitas digital masyarakat disaat masa pandemi termasuk didalamnya aktivitas belanja online. Aktivitas ini yang dapat mendongkrak perekonomian. Akibat Penyebaran virus COVID-19 memaksa banyak pelaku usaha untuk membuat strategi penjualan yang baru yaitu dengan melakukan transaksi penjualan secara daring. Selain meningkatnya penggunaan *marketplace* ada sektor lain yang diuntungkan dengan adanya sistem penjualan secara daring ini, yaitu sektor logistik dan jasa pengiriman barang.

Sejak penyebaran virus COVID-19 di Indonesia, beberapa perusahan jasa pengiriman mengalami lonjakan pendistribusian barang. Pada kondisi ini berkisar 60%-70% transaksi pengiriman berasal dari *marketplace*. Pada saat ini pengiriman dengan layanan 1 hari sampai mngalami kenaikan terutama pada barang komoditas pangan, karena membludaknya jumlah pemesanan dalam *marketplace*, didukung pula layanan jasa pengiriman yang semakin bagus. Dengan adanya sistem *tracking*, *e-wallet*, dan *multidrop*. Sistem ini dapat memantau jasa kurir yang mengantar paketnya, sementara *e-wallet* dapat mempermudah proses pembayaran (*cashless*). Lalu *multidrop* menjadikan penjual untuk mengirim banyak barang ke berbagi tujuan dalam satukali pengiriman.

Kajian Fajar Ratnawati tentang penerapan algoritma *Naive Bayes* dalam menganalisa opini sentimental pada film dengan data dari Twitter menghasilkan klasifikasi sentimen data tweet film Indonesia menggunakan algoritma *Naive Bayes Classifier* dimana kumpulan data

dibagi menjadi 5 kali yang disebut validasi silang. Semakin banyak data latih yang digunakan, maka akan semakin mempengaruhi kinerja sistem, dan semakin akurat hasil akan menunjukkan bahwa sistem berhasil mendapatkan peringkat yang baik. Akurasi tertinggi dicapai pada fold kedua yaitu 90%, *precision* 92%, *Recall* 90% dan *f-measure* 90% (Ratnawati, 2018).

Kajian lain dalam resensi film pada cuitan Twitter berbahasa Indonesia, dimana kajian ini memakai Naive Bayes dengan koreksi kata yang tidak baku. Dalam pengujian ini dilakukan dua langkah. Secara khusus, pengujian tahap 1 dengan pra-pemrosesan dan tanpa koreksi kata dan pengujian tahap 2 dengan pra-pemrosesan untuk mendapatkan hasil terbaik dalam hal akurasi dari analisis sentimen film pada dokumen Twitter dalam bahasa Indonesia menggunakan Naive Bayes yang tidak diedit, dengan hasil sebesar 98.33%. Dan untuk *precision* 96.77%, *recall* 100%, dan *f-measure* 98.36% (Antinasari et al., 2017).

Penelitian lebih lanjut terkait dengan analisis sentimen sedang berlangsung yang dilakukan oleh Nugraha (2014). Dalam penelitian ini, menggunakan metode K-Nearest Neighbor (KNN) untuk mengklasterkan emosi positif dan negatif dalam kaitannya dengan ulasan film berbahasa Inggris. Penelitian ini membandingkan metode KNN dan metode naive Bayes. Hasil akurasi yang diperoleh menunjukkan bahwa metode naive Bayes lebih unggul dari metode KNN, dimana metode naive Bayes memiliki nilai akurasi paling tinggi sebesar 81% dibandingkan saat menggunakan metode KNN dengan nilai akurasi sebesar 71%.

Dudih Gustian, Inda Suziati, dan Saepudin meneliti tentang sistem pakar dimana segmentasi menggunakan algoritma naive Bayesian untuk memprediksi hasil produksi plasma ayam pedaging (misalnya Pt.Sekawan Sinar Surya) menunjukkan bahwa algoritma naive Bayesian dapat memprediksi klasifikasi. Sebelum menghitung implementasi di area produksi, kami menghitung untung dan rugi. Pengujian dilakukan dengan menggunakan software Weka untuk menentukan nilai akurasi. Rata-rata akurasi dari algoritma yang diperoleh, Naive Bayes 96,36%, mendekati 100%, memberikan hasil yang maksimal, dan receiver operating curve (ROC) adalah 0,9995, yang semakin mendekati 1 maka model yang dihasilkan semakin baik (Gustian et al., 2019).

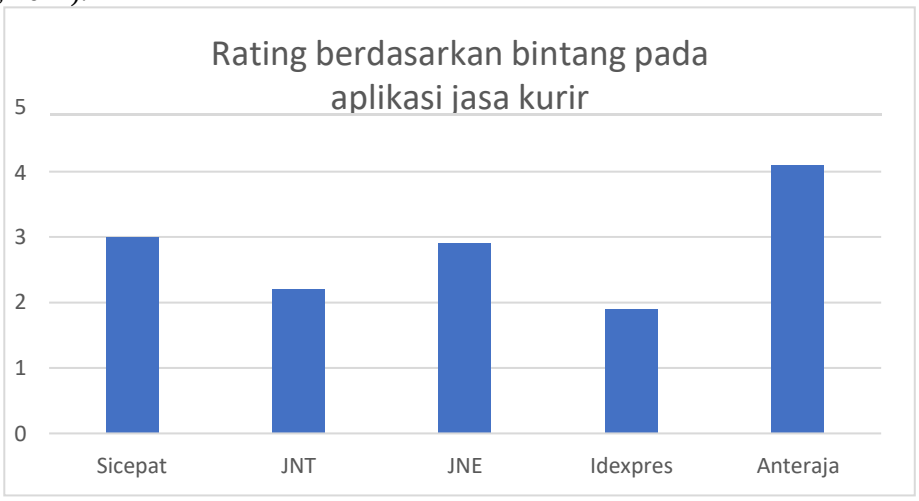
Penelitian yang dilakukan oleh Kalingara dkk, (2021). Mengenai "Analisis Sentimen Review Customer Terhadap Layanan Ekspedisi JNE dan J&T Express Menggunakan Metode Naïve Bayes", berdasarkan penelitian tersebut klasifiksi menggunaka algoritma naïve bayes memiliki akurasi yang baik dengan perbandingan 75:25, didapatkan nilai akurasi untuk JNE sebesar 79% dengan performasi precision 80%, recall 79% dan f1-score 79% dan untuk jnt sebesar 76% dengan performasi precision 76%, recall 76% dann f1-score 76% yang menunjukan nilai keseluruhan masuk ketegori cukup baik.

Devita dkk, (2018) dengan judul "Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia". Hasil penelitian ini menunjukan bahwa kinerja dari algoritma naïve bayes lebih baik daripada algoritma K-Nearest Neighbor dengan hasil dari 40 data uji dengan algoritma naïve bayes mampu mengklasifikasi artikel sebanyak 28 dokumen, sedangkan jika menggunakan algoritma K-Nearest Neighbor hanya dapat mengklasifiksi 26 dokumen saja.

Berdasarkan Penelitian yang dilakukan oleh Lorosae dkk, (2018) mengenai "Analisis Sentimen Berdasarkan Opini Masyarakat Pada Twitter Menggunakan Naïve Bayes" dapat disimpulkan jika penggunaan algoritma naïve bayes mendapat nilai akurasi sebesar 84%. Nilai akurasi dapat meningkat jika dta training yang digunakan lebih banyak.

Analisis sentimen, umumnya dikenal sebagai jajak pendapat (ekstraksi opini) adalah bidang studi yang menganalisis pendapat, perasaan, evaluasi, penilaian, gambar, dan emosi orang untuk suatu entitas, seperti produk, layanan, organisasi, masalah, peristiwa, topik , dan atributnya (Liu, 2012). Maka dengan itu, penelitian ini akan menganalisa opini tentang aplikasi jasa kurir di *Google Playstore* dimana sering digunakan oleh *marketplace* di Indonesia yakni JNT, JNE, ID Express, Sicepat dan Anteraja. Data ulasan didapat dari kolom komentar pada *Google Playstore*, dimana dapat memberikan informasi kepada pembaca tentang jasa kurir mana yang dapat digunakan secara baik. Analisis sentimen mempelajari tentang sudut pandang, sikap dan sifat

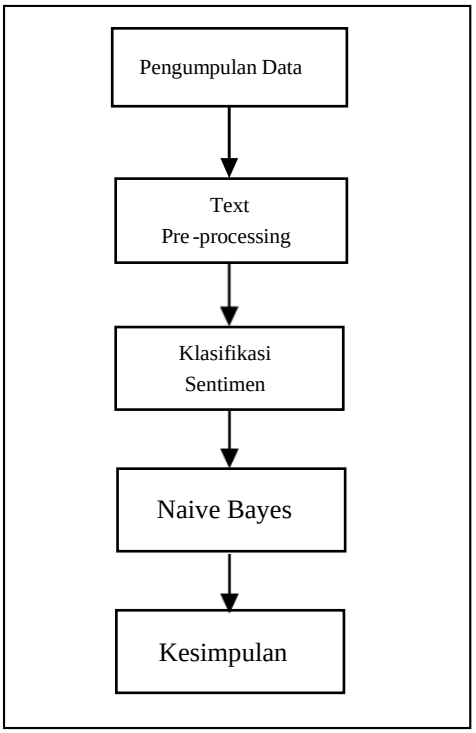
maupun emosi seseorang terhadap sebuah individu, masalah, aktifitas, subjek (Rustiana & Rahayu, 2017).



Gambar 1. Gambar rating bintang aplikasi jasa kurir di *Google play store*

Dengan menggunakan analisis sentimen kita dapat menganalisis review suatu barang dari sebuah website lalu dilakukan proses pengklasifikasian komentar pada barang tersebut menjadi opini positif ataupun opini negatif, algoritma naïve bayes digunakan sebagai teknik klasifikasi pada penelitian ini. Berdasarkan hal tersebut, penelitian ini bertujuan untuk menganalisis aplikasi jasa kurir di Google *play store* berdasarkan ulasan komentar dan mendapat peringkat aplikasi jasa kurir terbaik menurut nilai akurasi yang didapatkan oleh masing-masing aplikasi.

METODE PENELITIAN



Gambar 2. Alur Penelitian

A. Pengumpulan data

Data dikumpulkan dengan menyortir data pencarian dari Google Play Store menggunakan metode web scraping, metode ekstraksi data yang digunakan untuk menguraikan sejumlah besar data yang diambil dari Internet di Google Play Store. Saat Anda menggunakan metode pengikisan web (penambang data) untuk mendapatkan teks ulasan, pengulas, dan waktu pengiriman, data yang dipilih diubah menjadi serangkaian tabel (spreadsheet) yang dikumpulkan di Microsoft Excel untuk langkah berikutnya.

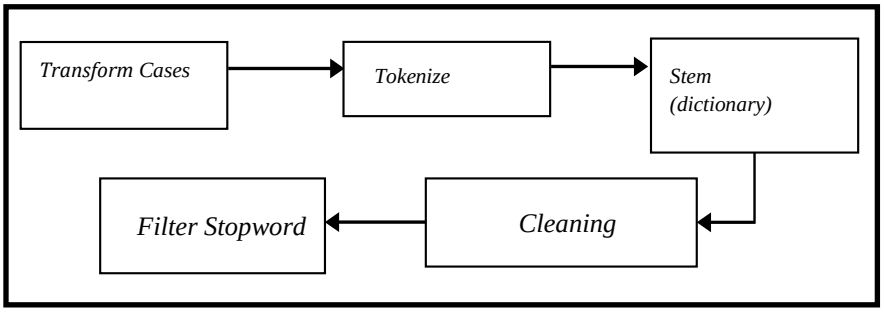
Machine learning adalah bidang ilmu yang mempelajari perkembangan algoritma komputer yang dikembangkan untuk dapat belajar dengan mandiri dengan mempelajari data yang diperoleh untuk melakukan tugas tertentu.

B. Text Mining

Merupakan proses pemrosesan dan analisis data dalam jumlah besar untuk menemukan informasi dan tren baru yang sebelumnya belum ditemukan. Saat menganalisis sebagian atau seluruh teks yang tidak terstruktur, penambangan teks mencoba menggunakan aturan tertentu untuk menautkan sebagian teks ke teks lainnya. Hasil yang diharapkan adalah informasi baru atau "wawasan" yang sebelumnya tidak diungkapkan secara eksplisit (Adiwijaya, 2006).

C. Text Preprocessing

Adalah tahap awal dari text mining. Word preprocessing adalah proses pemilihan data yang akan diproses pada setiap dokumen. Untuk memasukkan metode naive Bayesian classifier dan support vector machine, diperlukan langkah preprocessing yaitu menghapus data dari kata-kata yang tidak perlu, sehingga dapat diperoleh hasil perhitungan yang optimal. Langkah-langkah preprocessing teks meliputi case folding, tokenization, filtering, root extraction, dan stopwords.



Gambar 3. Tahap *preprocessing*

Data yang diperoleh dari google playstore menggunakan teknik scrapping adalah data teks berupa file csv tidak terstruktur (*unstructured*). Karena perubahan perlu dilakukan untuk mengonversi data menjadi data terstruktur. Data teks perlu dibersihkan sebelum masuk ke model machine learning, proses pembersihan data ini disebut text preprocessing. Tahap prerocesing meliputi :

1. Case folding merupakan proses perubahan teks menjadi huruf kecil (lower text).
2. Cleansing merupakan cara yang bertujuan untuk membersihan kalimat dari spasi, tab, hastag dan tanda lainnya.
3. Remove punctuation merupakan proses dengan tujuan untuk membersihan teks dari tanda baca.
4. Tokenizing adalah tahapan yang bertujuan untuk pemisahan teks menjadi potongan kata yang disebut sebagai token. Ditujukan agar dapat diproses pada tahap *spell normalization* dan filtering.
5. Filtering Adalah tahapan *text preprocessing* yang bertujuan untukpembersihan teks.

D. Klasifikasi Sentimen
Klasifikasi sentimen merupakan proses pelabelan pada data teks ulasan komentar yang sudah melalui tahap pembersihan noise pada *preprocessing* sehingga data yang akan diberi label sudah bersih dari *noise* sehingga lebih memudahkan dalam pelabelan. Setelah data ulasan komentar diberi label sentimen yaitu sentimen positif, negatif atau netral, kemudian data yang sudah diberi label tersbut diklasifikasikan sesuai dengan sentimennya masing-masing untuk mengetahun berapa jumlah klasifikasi tiap komentar berdasarkan sentimennya.

E. Naïve Bayes Classifier
Naive Bayesian classifier adalah salah satu metode yang banyak digunakan untuk tujuan data mining karena kemudahan penggunaan dan waktu pemrosesan yang cepat, kemudahan implementasi, struktur yang agak sederhana dan efisiensi yang tinggi. Klasifikasi Naive Bayesian juga menunjukkan akurasi dan kecepatan yang tinggi ketika digunakan pada kumpulan data yang besar.

HASIL DAN PEMBAHASAN

A. Pengumpulan data
Ulasan yang didapatkan yakni sedari dua bulan terakhir dimana menghasilkan 200 ulasan JNE, 200 ulasan JNT, 200 ulasan Sicepat, 200 ulasan IDexpres, dan 200 ulasan ninjaexpres di *Google Play Store*. Dengan total yang dimilik yakni 1000 ulasan. Data diambil meggunakan teknik *scraping* data dengan menggunakan ekstensi bawaan dari *Google Chrome*, data diambil secara otomatis dan dirubah kedalam bentuk excel

B. Tahap *pre-processing*

1. *Case folding*
   Diproses ini data yang berupa teks akan dirubah kedalam model untuk menjadi huruf kecil.

2. *Tokenize*
   Dalam tahapan ini dilangsungkan pembagian kata pada ulasan yang memiliki pemisah spasi. Yaitu proses pemisahan kata-kata pada kalimat.
3. *Stemming*
   Proses ini menghilangkan kalimat imbuhan yang melekat dalam teks supaya nanti teks menjadi kata dasar.
4. *Cleansing*
   Merupakan proses penghapusan *noise* seperti *emoticon* dan karakter yang kurang penting dalam review.
5. *Stopword Removal*
   yaitu penghapusan kata yang dianggap tidak memiliki arti. kata yang termasuk *stopword* adalah “dia”, “aku”, “ialah”, dan lainnya.

C. Klasifikasi sentimen
Setelah melalui tahap *pre-processing* terhadap apliksi jasa kurir maka didapat hasil sebagai berikut:

Gambar 4. klasifikasi sentimen ninja xpres

Pada gambar tersebut didapat hasil sentimen negatif sebanyak 113, netral 39 dan positif 49 untuk aplikasi ninjaexpres.

```
Negatif  Netral Positif
    110      40      50
```

Gambar 5. klasifikasi sentimen JNT

Pada gambar tersebut didapat hasil sentimen negatif sebanyak 110, netral 40 dan positif 50 untuk aplikasi JNT

```
Negatif  Netral Positif
     98      56      46
```

Gambar 6. klasifikasi sentimen JNE

Pada gambar tersebut didapat hasil sentimen negatif sebanyak 98, netral 56 dan positif 46 untuk aplikasi JNE

```
Negatif  Netral Positif
    106      45      49
```

Gambar 7. klasifikasi sentimen IDexpres

Pada gambar tersebut didapat hasil sentimen negatif sebanyak 106, netral 45 dan positif 49 untuk aplikasi Idexpres

```
Negatif  Netral Positif
     89      64      47
```

Gambar 8. klasifikasi sentimen Sicepat

Pada gambar tersebut didapat hasil sentimen negatif sebanyak 106, netral 64 dan positif 47 untuk aplikasi Idexpres

D. Implementasi *Naïve Bayes*

Dengan menerapkan algoritma *naïve bayes data diolah* sehingga menghasilkan nilai akurasi, yang sebelumnya telah melalui tahap klasifikasi.

```
Overall Statistics

               Accuracy : 0.9787
                 95% CI : (0.9252, 0.9974)
    No Information Rate : 0.5532
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9643

 Mcnemar's Test P-Value : NA
```

Gambar 9. Nilai akurasi aplikasi ninjaxpress

```
Overall Statistics

               Accuracy : 1
                 95% CI : (0.9611, 1)
    No Information Rate : 0.5591
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 1
```

Gambar 10. Nilai akurasi aplikasi JNT

```
Overall Statistics

               Accuracy : 0.9892
                 95% CI : (0.9415, 0.9997)
    No Information Rate : 0.4624
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9833

 Mcnemar's Test P-Value : NA
```

Gambar 11. Nilai akurasi aplikasi JNE

```
Overall Statistics

               Accuracy : 0.9462
                 95% CI : (0.879, 0.9823)
    No Information Rate : 0.4731
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.9157

 Mcnemar's Test P-Value : 0.1718
```

Gambar 12. Nilai akurasi aplikasi Idexpres

```
Overall Statistics

               Accuracy : 0.9785
                 95% CI : (0.9245, 0.9974)
    No Information Rate : 0.4516
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9665

 Mcnemar's Test P-Value : NA
```

Gambar 13. Nilai akurasi aplikasi Sicepat

Setelah melalui proses pengumpulan data tahap *pre-processing*, pengklasifikasian sentimen dan penerapan algoritma *naïve bayes* makan didapat hasil sebagai berikut :

Tabel 1. Hasil nilai akurasi dan Klasifikasi Sentimen

Aplikasi Jasa Kurir	Nilai Akurasi	Sentimen Positif	Sentimen Netral	Sentimen Negatif
Ninjaexpress	97%	49	39	113
JNT	100%	50	40	110
JNE	98%	46	56	98
Idexpress	94%	49	56	106
Sicepat	97%	47	45	89

KESIMPULAN DAN SARAN

Belandaskan pada temuan dan pemrosesan data yang selesai dilakukan nilai akurasi tidak dapat digunakan sebagai acuan untuk menetapkan peringkat aplikasi jasa kurir terbaik. Karena pada aplikasi JNT yang memiliki nilai akurasi tertinggi sebesar 100% memiliki ulasan negatif yang lebih banyak daripada ulasan positif, hal ini juga berlaku untuk jasa kurir yang lain, itu berarti pada aplikasi ini mendapat respon negatif dari para pengguna aplikasi jasa kurir. Terlihat dari data yang diolah kesalahan umum yang dilakukan oleh jasa kurir yaitu keterlambatan

pengiriman paket yang tidak sesuai dengan perhitungan waktu yang ditentukan. Jadi untuk pihak jasa kurir untuk memperbaiki manajemen pengirimannya.

Berdasarkan hasil dari implementasi algoritma naïve bayes dapat disimpulkan bahwa algoritma ini memiliki nilai akurasi yang tinggi, yaitu mendapat nilai keseluruan diatas 90%, dari hasil pengujian yang telah dilakukan mendapat kesimpulan bahwa algoritma *naïve bayes* dapat digunakan untuk mengklasifikasi opini mengenail jasa kurir.

SARAN

Saran untuk penelitian selanjutnya diharapkan data yang digunakan dalam penelitian bisa lebih banyak, karena semakin banyak data yang digunakan untuk *training* akan berpengaruh pada kinerja dari sistem tersebut.

DAFTAR PUSTAKA

Adiwijaya, I. (2006). Text Mining dan Knowledge Discovery. *Kolokium Bersama Komunitas Datamining Indonesia & Soft-Computing Indonesia*, 1–9. http://web.ipb.ac.id/~ir-lab/pdf/tm (text summarization).pdf

Antinasari, P., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(12), 1718–1724. http://j-ptiik.ub.ac.id

Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan kinerja metode naive bayes dan k-nearest neighbor untuk klasifikasi artikel berbahasa indonesia. *JTIIK (Jurnal Teknologi Informasi Dan Ilmu Komputer) Vol*, *5*.

Gustian, D., Suciati, I., & Saepudin, S. (n.d.). *Sistem Pakar Dengan Algoritma Naive Bayes Untuk Prediksi Hasil Produksi Ayam Broiler Plasma (Studi Kasus: Pt. Sekawan Sinar Surya)*. no.

Kalingara, N. I. P., Pratiwi, O. N., & Anggana, H. D. (2021). Analisis Sentimen Review Customer Terhadap Layanan Ekspedisi Jne Dan J\&t Express Menggunakan Metode NaÏve Bayes. *EProceedings of Engineering*, *8*(5).

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Lorosae, T. A., Prakoso, B. D., Saifudin, S., & Kusrin, K. (2018). Analisis Sentimen Berdasarkan Opini Masyarakat pada Twitter Menggunakan Naive Bayes. *Semnasteknomedia Online*, *6*(1), 1–10.

Nugraha, M. W. A. (2014). *Sentiment Analysis Pada Review Film Dengan Menggunakan Metode K-Nearest Neighbor*. Universitas Widyatama.

Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, *3*(1), 50. https://doi.org/10.35314/isi.v3i1.335

Rustiana, D., & Rahayu, N. (2017). Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naïve Bayes. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, *8*(1), 113–120. https://doi.org/10.24176/simet.v8i1.841