

## **Pengembangan Sistem Informasi Pariwisata Berbasis Web Sebagai Media Promosi Pada Wisata Di Lampung Selatan**

Yohanes Egi Pratama Yudoutomo <sup>1)</sup>, Ahmad Ghifari <sup>2)</sup>, Muhammad Fadli Kusuma <sup>3)</sup>

<sup>1, 2, 3)</sup> Informatika

<sup>1)</sup> egipratama23@gmail.com

### **Abstrak**

Analisis sentimen adalah sebuah bidang mempelajari opini, tanggapan, dan emosi seseorang terhadap suatu entitas yang dalam penelitian ini adalah review produk. Salah satu cara mengetahui kualitas produk adalah dengan membaca review produk konsumen lain. Produk kosmetik yang beredar di pasaran sangat beragam dari segi jenis maupun brand. Namun tidak semua kosmetik memiliki kualitas yang baik dan hal ini harus diperhatikan oleh konsumen. Pada penelitian ini diterapkan analisis sentiment dengan menggunakan algoritma Naïve Bayes dibuat dengan bantuan aplikasi Google Colaboratory. untuk memprediksi rating dalam skala 1 sampai 0. Akurasi yang dicapai dalam penelitian ini precision 91%, recall 96% F1\_score 94% Akurasi 87.

**Katakunci** : Analisis Sentimen, Google Colaboratory, Opini, Naïve Bayes.

---

### **PENDAHULUAN**

Perkembangan teknologi saat ini berkembang sangat pesat. Ditandai dengan perkembangan Big Data. Hal ini menyebabkan arus data dan informasi tumbuh signifikan dalam ukuran jumlah dan media yang bervariasi, sentiment atau opini masyarakat semakin bertambah luas dan bebas diungkapkan diberbagai media (Sedyastuti *et al.*, 2021), (Pajar and Putra, 2021). Sentimen dapat menjadi potensi besar bagi perusahaan yang ingin mengetahui umpan balik (feedback) dalam bentuk rating skala 1 sampai 5 ataupun opini negative dan positif (Ferdiana, 2020), (Hashim *et al.*, 2016). Dalam perkembangannya sentimen semakin mudah ditemukan dalam berbagai media seperti media online. Di era ini sentimen atau opini banyak bermunculan pada website yang memberikan review produk (Sulistiani and Tjahyanto, 2016), (Suryono *et al.*, 2019). Banyaknya website untuk review menyebabkan ketersediaan data review sangat banyak. Ketersediaan data review inilah yang menjadi alasan penelitian ini menggunakan sumber data berupa review produk untuk analisis sentimen. Review produk dapat memberikan manfaat besar bagi konsumen karena dengan membaca review konsumen mendapat informasi yaitu dapat mengetahui kualitas

suatu produk dari review konsumen lain yang membagikan pengalamannya tentang produk terkait. Dalam data review pelanggan tersebut, peneliti akan melakukan pengolahan analisis sentiment, berdasarkan opini yang diberikan oleh pelanggan pada produk pakaian wanita. dalam skala rating 0 dan 1 merupakan opini baik dan buruk suatu ulasan. sentiment merupakan penilaian publik, penilaian mengenai kesukaan atau ketidaksukaan (Anggarini, 2021). Untuk melakukan analisis sentiment ada beberapa algoritma yang dapat digunakan adalah Naïve Bayes (Pasha, 2020).

## **KAJIAN PUSTAKA**

### **Analisis Sentimen**

Analisis sentimen adalah sebuah bidang mempelajari opini, tanggapan, dan emosi seseorang terhadap suatu entitas yang dalam penelitian ini adalah review produk.

### **Text Mining**

Text Mining Merupakan proses mining atau menambang suatu informasi dari data yang tersaji dalam jumlah besar, dalam hal ini adalah teks. Proses ini dilakukan dalam rangka penggalian, pengolahan, serta pengaturan pada informasi dengan menganalisa keterkaitan antara informasi satu dengan yang lainnya (Sucipto *et al.*, 2019), (Sucipto, Adrian and Kencono, 2021).

### **Tokenization**

Tokenization adalah tugas 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.

### **Informasi**

Informasi merupakan olahan data yang telah diolah atau diklasifikasikan atau diinterpretasikan untuk digunakan dalam proses pengambilan keputusan (Fitriyana and Sucipto, 2020), (WING, no date), (Sofa, Suryanto and Suryono, 2020). Suatu sistem tidak akan berjalan dengan lancar tanpa suatu informasi, dan akhirnya bisa mati. Informasi dalam

suatu organisasi dibutuhkan, tanpa adanya suatu informasi maka organisasi tersebut tidak bisa berjalan dan tidak bisa beroperasi (Samsugi, Mardiyansyah and Nurkholis, 2020), (Saputra and Puspaningrum, 2021). Menurut (Anggraini, Pasha and Damayanti, 2020) dan (Suri and Puspaningrum, 2020) Informasi merupakan olahan data menjadi bentuk yang berguna dan menjadi berarti bagi penerimaannya. Informasi merupakan kumpulan data yang diolah menjadi bentuk yang lebih berguna dan lebih berarti bagi yang menerimanya (Yusmaida, Neneng and Ambarwari, 2020), (Yanuarsyah, Muhaqiqin and Napianto, 2021).

### **Stemming**

Stemming merupakan salah satu proses dari mengubah token yang berimbuhan menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada token tersebut (Syambas *et al.*, 2018), (LIA FEBRIA LINA, 2019).

### **METODE**

#### **Study Literatur**

Study Literatur merupakan proses awal sebelum memulai penelitian dengan cara melakukan literasi jurnal penelitian yang sudah ada sebelumnya dengan tujuan untuk membuka wawasan.

#### **Analisis data**

Berdasarkan latar belakang permasalahan yang sudah dipaparkan sebelumnya penelitian ini mengarah kepada proses klasifikasi teks menggunakan algoritma Naïve Bayes Classifier dengan harapan bahwa dataset opini public mengenai ulasan produk dapat digunakan untuk mendapatkan sebuah pengetahuan baru dan manfaat bagi pelanggan lainnya (Alita *et al.*, 2021), (Widodo *et al.*, 2020). Data yang diolah sebagai berikut :

index	Clothing ID	Age	Title	Review Text	Rating	Recommended IND	
0	2	1077	60	Some major design flaws	I had such high hopes for this dress and reall...	3	0
1	3	1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	1
2	4	847	47	Flattering shirt	This shirt is very flattering to all due to th...	5	1
3	5	1080	49	Not for the very petite	I love tracy reese dresses, but this one is no...	2	0
4	6	858	39	Cagrcoal shimmer fun	I aded this in my basket at hte last mintue to...	5	1

Gambar 1. Data Analisis

Text Mining Merupakan proses mining atau menambang suatu informasi dari data yang tersaji dalam jumlah besar, dalam hal ini adalah teks (Ariyanti and Iswardani, 2020), (Indrayuni, 2019). Proses ini dilakukan dalam rangka penggalian, pengolahan, serta pengaturan pada informasi dengan menganalisa keterkaitan antara informasi satu dengan yang lainnya (Sudiantoro et al., 2018), (Mittal and Patidar, 2019). Proses text mining ini sendiri merupakan hal yang berdasarkan kombinasi yang baik serta kata, dan frase yang sesuai, serta melibatkan proses ekstraksi fitur yang menjadi kunci efektif utama dalam tahap pelatihan atau yang sering disebut dengan training (Handayani & Pribadi, 2015), (Aldino *et al.*, 2021). Tokenization adalah tugas 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 (Purwarianti, 2014), (Abidin and Ahmad, 2021).

Stemming merupakan salah satu proses dari mengubah token yang berimbuhan menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada token tersebut. Stopwords dapat diartikan sebagai menghilangkan karakter, tanda baca, serta kata-kata umum yang tidak memiliki makna atau informasi yang dibutuhkan (Abidin, Wijaya and

Pasha, 2021), (Febrian and Hapsari, 2019), (Fikri, Sabrila and Azhar, 2020). Stopwords umumnya digunakan dalam pengambilan informasi salah satu Case Folding dapat diartikan sebagai mengubah teks menjadi lowecase Naive Bayes Metode yang digunakan untuk klasifikasi sentimen pada penelitian ini adalah Naïve Bayes Classifier (Giovani *et al.*, 2020), (Sulistiani, Muludi and Syarif, 2019), (Wibisono, Rizkiono and Wantoro, 2020). Metode Naïve Bayes berdasarkan pada algoritma dengan teknik klasifikasi yang telah melalui uji efisiensi dan efektifitas dalam suatu data base dengan jumlah data yang besar, Naïve Bayes memiliki tingkat akurasi dan efisiensi yang tinggi Naïve Bayes dengan konsep dasar teorema bersyarat Bayes dapat dilihat pada persamaan (Putri and Surahman, 2019), (Gunawan D, 2020), (Setiawansyah, Adrian and Devija, 2021).

$$P(H|X) = \frac{P(H|X)p(h)}{P(X)}$$

Keterangan :

- X = Data dengan Kelas yang belum diketahui  
H = Representasi data X merupakan Kelas yang spesifik  
P(H|X) = Probabilitas H berdasarkan kondisi X P(H) = Probabilitas  
P(X|H) = Probabilitas X terhadapkondisi H P(X) = Probabilitas X

Cofusion Matrix adalah alat yang berguna untuk menganalisis seberapa baik klasifikasi yang telah dibuat, dapat mengenali tuple dari kelas yang berbeda (Gandhi, Megawaty and Alita, 2021), (Isnain, Marga and Alita, no date), (Pintoko and L., 2018).

## HASIL DAN PEMBAHASAN

Data pada penelitian ini berupa opini berbahasa Inggris tentang produk pakaian wanita di E-Commerce Amazon. Sumber data yang digunakan berasal dari Kaggle dengan banyak data adalah 23468 dengan 11 kolom. Pengujian metode klasifikasi terhadap data yang dilakukan menggunakan Google Colab dengan komposisi data yang digunakan adalah `review_text` dan `recommended`. Setelah data dilakukan fit selanjutnya diproses oleh Naïve Bayes dihasilkan akurasi 87%.

```
[59] 1 nb_report = accuracy_score(y_test, nb_predict)
      2 print('Accuracy:', nb_report)
      3 nb_report1 = recall_score(y_test, nb_predict, pos_label='Recommended', average='binary')
      4 print('Recall:', nb_report1)
      5 nb_report2 = precision_score(y_test, nb_predict, pos_label='Recommended', average='binary')
      6 print('Precision:', nb_report2)

Accuracy: 0.8782100177981185
Recall: 0.9321981424148607
Precision: 0.9205136044023234
```

Gambar 2. Hasil Proses Klasifikasi Naïve Bayes

## SIMPULAN

Klasifikasi data ulasan produk dengan sentiment positif dan negatif dilakukan dengan baik menggunakan metode Naïve Bayes Classifier namun beberapa aspek penting perlu ditingkatkan seperti keberagaman stemmer, dan stop word list guna meningkatkan akurasi pengklasifikasian dengan berbagai jenis pola penulisan berbahasa Inggris.

## REFERENSI

- Abidin, Z. and Ahmad, I. (2021) "Effect of mono corpus quantity on statistical machine translation Indonesian–Lampung dialect of nyo," in *Journal of Physics: Conference Series*. IOP Publishing, p. 12036.
- Abidin, Z., Wijaya, A. and Pasha, D. (2021) "Aplikasi Stemming Kata Bahasa Lampung Dialek Api Menggunakan Pendekatan Brute-Force dan Pemograman C," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(1), pp. 1–8.
- Aldino, A.A. *et al.* (2021) "Implementation of K-means algorithm for clustering corn planting feasibility area in south Lampung regency," in *Journal of Physics: Conference Series*. IOP Publishing, p. 12038.

- Alita, D. *et al.* (2021) “Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa,” *Jurnal Data Mining Dan Sistem Informasi*, 2(1), pp. 17–23.
- Anggarini, D.R. (2021) “Kontribusi Umkm Sektor Pariwisata Pada Pertumbuhan Ekonomi Provinsi Lampung 2020,” 9(2), pp. 345–355.
- Anggraini, Y., Pasha, D. and Damayanti, D. (2020) “SISTEM INFORMASI PENJUALAN SEPEDA BERBASIS WEB MENGGUNAKAN FRAMEWORK CODEIGNITER,” *Jurnal Teknologi dan Sistem Informasi*, 1(2), pp. 64–70.
- Ariyanti, D. and Iswardani, K. (2020) “Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma Naïve Bayes,” *Jurnal IKRA-ITH Informatika*, 4(3), pp. 125–132.
- Febrian, A. and Hapsari, chintia annisa vina (2019) “Strategi Pemasaran Dalam Memengaruhi Keputusan Pembelian Melalui Minat Beli Sebagai Mediasi,” *Buletin Studi Ekonomi*, 24(2), pp. 279–287.
- Ferdiana, R. (2020) “A Systematic Literature Review of Intrusion Detection System for Network Security: Research Trends, Datasets and Methods,” in *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*. IEEE, pp. 1–6.
- Fikri, M.I., Sabrila, T.S. and Azhar, Y. (2020) “Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter,” *Smatika Jurnal*, 10(02), pp. 71–76. doi:10.32664/smatika.v10i02.455.
- Fitriyana, F. and Sucipto, A. (2020) “SISTEM INFORMASI PENJUALAN OLEH SALES MARKETING PADA PT ERLANGGA MAHAMERU,” *Jurnal Teknologi dan Sistem Informasi*, 1(1), pp. 105–110.
- Gandhi, B.S., Megawaty, D.A. and Alita, D. (2021) “Aplikasi Monitoring Dan Penentuan Peringkat Kelas Menggunakan Naïve Bayes Classifier,” *Jurnal Informatika dan Rekayasa Perangkat Lunak*, 2(1), pp. 54–63.
- Giovani, A.P. *et al.* (2020) “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *Jurnal Teknoinfo*, 14(2), p. 115. doi:10.33365/jti.v14i2.679.

- Gunawan D (2020) “Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023,” V(1), pp. 135–138. doi:10.31294/jtk.v4i2.
- Hashim, R. *et al.* (2016) “Estimation of Wind-Driven Coastal Waves Near a Mangrove Forest Using Adaptive Neuro-Fuzzy Inference System,” *Water Resources Management*, 30(7), pp. 2391–2404. doi:10.1007/s11269-016-1267-0.
- Indrayuni, E. (2019) “Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes,” *Jurnal Khatulistiwa Informatika*, 7(1), pp. 29–36. doi:10.31294/jki.v7i1.1.
- Isnain, A.R., Marga, N.S. and Alita, D. (no date) “Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(1), pp. 55–64.
- LIA FEBRIA LINA, B.P. (2019) “KREDIBILITAS SELEBRITI MIKRO PADA NIAT BELI PRODUK DI MEDIA SOSIAL,” 1(2), pp. 41–50.
- Mittal, A. and Patidar, S. (2019) “Sentiment analysis on twitter data: A survey,” *ACM International Conference Proceeding Series*, pp. 91–95. doi:10.1145/3348445.3348466.
- Pajar, M. and Putra, K. (2021) “A Novel Method for Handling Partial Occlusion on Person Re-identification using Partial Siamese Network,” 12(7), pp. 313–321.
- Pasha, D. (2020) “SISTEM PENGOLAHAN DATA PENILAIAN BERBASIS WEB MENGGUNAKAN METODE PIECIES,” *Jurnal Teknologi dan Sistem Informasi*, 1(1), pp. 97–104.
- Pintoko, B.M. and L., K.M. (2018) “Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier,” *e-Proceeding of Engineering*, 5(3), pp. 8121–8130.
- Purwarianti, A. (2014) “Rule based approach for text segmentation on Indonesian news article using named entity distribution,” in *2014 International Conference on Data and Software Engineering (ICODSE)*. IEEE, pp. 1–5.
- Putri, S.E.Y. and Surahman, A. (2019) “PENERAPAN MODEL NAIVE BAYES UNTUK MEMPREDIKSI POTENSI PENDAFTARAN SISWA DI SMK TAMAN SISWA TELUK BETUNG BERBASIS WEB.” Universitas Teknokrat Indonesia.



- Samsugi, S., Mardiyansyah, Z. and Nurkholis, A. (2020) “Sistem Pengontrol Irigasi Otomatis Menggunakan Mikrokontroler Arduino UNO,” *Jurnal Teknologi dan Sistem Tertanam*, 1(1), pp. 17–22.
- Saputra, A. and Puspaningrum, A.S. (2021) “SISTEM INFORMASI AKUNTANSI HUTANG MENGGUNAKAN MODEL WEB ENGINEERING (Studi Kasus: Haanhani Gallery),” *Jurnal Teknologi dan Sistem Informasi*, 2(1), pp. 1–7.
- Sedyastuti, K. *et al.* (2021) “Human Resources Competency at Micro, Small and Medium Enterprises in Palembang Songket Industry,” *Proceedings of the 2nd Annual Conference on Social Science and Humanities (ANCOSH 2020)*, 542(Ancosh 2020), pp. 248–251. doi:10.2991/assehr.k.210413.057.
- Setiawansyah, S., Adrian, Q.J. and Devija, R.N. (2021) “SISTEM PAKAR DIAGNOSA PENYAKIT SELAMA KEHAMILAN MENGGUNAKAN METODE NAIVE BAYES BERBASIS WEB,” *Jurnal Manajemen Informatika (JAMIKA)*, 11(1), pp. 24–36.
- Sofa, K., Suryanto, T.L.M. and Suryono, R.R. (2020) “Audit Tata Kelola Teknologi Informasi Menggunakan Kerangka Kerja COBIT 5 Pada Dinas Pekerjaan Umum Kabupaten Tanggamus,” *Jurnal Teknologi dan Sistem Informasi*, 1(1), pp. 39–46.
- Sucipto, A. *et al.* (2019) “Penerapan Metode Certainty Factor Pada Diagnosa Penyakit Saraf Tulang Belakang.”
- Sucipto, A., Adrian, Q.J. and Kencono, M.A. (2021) “Martial Art Augmented Reality Book (Arbook) Sebagai Media Pembelajaran Seni Beladiri Nusantara Pencak Silat,” *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 10(1), pp. 40–45.
- Sulistiani, H., Muludi, K. and Syarif, A. (2019) “Implementation of Dynamic Mutual Information and Support Vector Machine for Customer Loyalty Classification,” *Journal of Physics: Conference Series*, 1338(1). doi:10.1088/1742-6596/1338/1/012050.
- Sulistiani, H. and Tjahyanto, A. (2016) “Heterogeneous feature selection for classification of customer loyalty fast moving consumer goods (Case study: Instant noodle),” *Journal of Theoretical and Applied Information Technology*, 94(1), pp. 77–83. doi:10.5281/zenodo.579836.

- Suri, M.I. and Puspaningrum, A.S. (2020) “Sistem Informasi Manajemen Berita Berbasis Web,” *Jurnal Teknologi Dan Sistem Informasi*, 1(1), pp. 8–14.
- Suryono, R.R. *et al.* (2019) “Challenges in P2P lending development: Collaboration with tourism commerce,” in *2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)*. IEEE, pp. 129–133.
- Syambas, N.R. *et al.* (2018) “Performance comparison of named data and IP-based network—Case study on the Indonesia higher education network,” *Journal of Communications*, 13(10), pp. 612–617. doi:10.12720/jcm.13.10.612-617.
- Wibisono, A.D., Rizkiono, S.D. and Wantoro, A. (2020) “Filtering Spam Email Menggunakan Metode Naive Bayes,” *Telefortech: Journal Of Telematics And Information Technology*, 1(1), pp. 9–17.
- WING, L.A.S.O.F.F. (no date) “IMPLEMENTASI KENDALI LQR UNTUK PENGENDALIAN SIKAP LONGITUDINAL PESAWAT FLYING WING.”
- Widodo, T. *et al.* (2020) ‘Sistem Sirkulasi Air Pada Teknik Budidaya Bioflok Menggunakan Mikrokontroler Arduino Uno R3’, *Jurnal Teknik dan Sistem Komputer*, 1(2), pp. 1–6.
- Yanuarsyah, M.R., Muhaqiqin, M. and Napianto, R. (2021) “ARSITEKTUR INFORMASI PADA SISTEM PENGELOLAAN PERSEDIAAN BARANG (STUDI KASUS: UPT PUSKESMAS RAWAT INAP PARDASUKA PRINGSEWU),” *Jurnal Teknologi dan Sistem Informasi*, 2(2), pp. 61–68.
- Yusmaida, Y., Neneng, N. and Ambarwari, A. (2020) “Sistem Informasi Pencarian Kos Berbasis Web Dengan Menggunakan Metode Hill Climbing,” *Jurnal Teknologi dan Sistem Informasi*, 1(1), pp. 68–74.