

## Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes

Muljono<sup>1)</sup>, Dian Putri Artanti<sup>2)</sup>, Abdul Syukur<sup>3)</sup> Adi Prihandono<sup>4)</sup>, De Rosal I. Moses Setiadi<sup>5)</sup>

Fakultas Ilmu Komputer, Universitas Dian Nuswantoro  
Jl. Imam Bonjol 207 Semarang  
e-mail: [muljono@dsn.dinus.ac.id](mailto:muljono@dsn.dinus.ac.id)

### Abstrak

*Situs online marketplace merupakan tempat belanja yang saat ini digemari masyarakat karena menawarkan berbagai kemudahan. Sebagian masyarakat puas dengan pelayanan yang diberikan situs online marketplace yang mereka pilih dan sebagian lagi tidak. Respon yang diberikan konsumen tentang pelayanan situs online marketplace biasanya di ungkapkan melalui media sosial, salah satunya adalah twitter. Pada penelitian ini dilakukan analisis sentimen terhadap postingan opini pelanggan online marketplace di Indonesia pada twitter yang nantinya bisa digunakan untuk menentukan rating online marketplace agar masyarakat tidak salah pilih situs marketplace untuk berbelanja di masa depan. Langkah pertama melakukan koleksi data opini masyarakat di twitter dari situs belanja online. Kemudian dilakukan pre-processing pada data yang meliputi cleansing data, case folding, tokenizing, case normalization, stop word, convert negation dan stemming. Selanjutnya dilakukan proses pemberian label (kelas) pada data tersebut yang dilakukan oleh ahli bahasa. Hasil dari proses clustering tercipta dua kelas data yaitu kelas data positif dan kelas data negatif dengan jumlah total 1200 data. Data yang sudah memiliki kelas data ini, digunakan sebagai data training untuk mesin pengklasifikasi, dalam riset ini menggunakan algoritma mesin pengklasifikasi Naïve Bayes. Terakhir, kami mengukur kinerja dari mesin pengklasifikasi menggunakan 10-fold cross validation. Hasil evaluasi menunjukkan rata-rata akurasi sebesar 93.33%.*

**Kata kunci:** online marketplace, analisis sentimen, twitter, klasifikasi, Naive Bayes

### 1. Pendahuluan

Situs online marketplace atau biasa disebut pasar *e-commerce online* yaitu suatu proses membeli dan menjual produk-produk secara elektronik oleh konsumen dan dari perusahaan ke perusahaan dengan komputer sebagai transaksi bisnis[1]. Pada zaman sekarang situs-situs *online marketplace* menjadi pilihan yang paling digemari masyarakat untuk berbelanja suatu produk karena memiliki banyak kelebihan yaitu pembeli tidak perlu datang langsung ke tempat penjual, wilayah tidak menjadi pembatas untuk membeli, “toko” dapat diakses 24 jam, tidak ada tawar menawar barang, banyak diskon, promo dan bahkan bebas ongkos kirim. Namun karena kelebihan itulah munculah beberapa kekurangan, antara lain tidak jarang terjadi penipuan pada saat melakukan transaksi, penjual tidak mengirim barang yang sudah dipesan, barang yang ada digambar tidak sesuai dengan aslinya, kemasan rusak serta jika terjadi kesalahan warna. Maka dari itu konsumen menggunakan sosial media untuk melampiaskan perasaan terhadap pelayanan yang diberikan oleh perusahaan *online marketplace* tersebut.

Sekarang, banyak orang menggunakan situs media sosial untuk mengungkapkan pendapat mereka tentang sesuatu. Ada banyak situs sosial media populer seperti Facebook, Twitter, Instagram dll. Salah satu media sosial yang paling sering digunakan dalam menyampaikan pendapat adalah Twitter. Dari berbagai opini pelanggan tentang situs *online marketplace* yang ada di Twitter, konsumen tidak dapat menyimpulkan situs online marketplace mana yang memiliki pelayanan paling baik. Hal ini disebabkan banyaknya komentar di twitter bahkan *trending topic* yang ada di twitter pun hanya menampilkan topik yang sedang banyak diperbincangkan tanpa memberikan suatu kesimpulan[2].

Analisis sentimen atau biasa disebut juga dengan *opinion mining* adalah sebuah proses menemukan pendapat pengguna tentang beberapa topik atau teks yang disampaikan pengguna[3]. Dalam pengertian lain, sebuah proses untuk menentukan apakah sepotong tulisan itu bermakna positif, negatif atau netral. Penelitian ini membahas tentang bagaimana melakukan analisis sentimen terhadap data opini pelanggan *online marketplace* pada Twitter yang digunakan untuk menentukan apakah data opini tersebut masuk opini positif atau negatif, sehingga nantinya bisa dimanfaatkan untuk menentukan situs belanja online yang dikategorikan baik.

Beberapa penelitian berkaitan dengan topik analisis sentimen sudah banyak dilakukan, diantaranya : penelitian *An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis*[5] menggunakan enam metode untuk klasifikasi yaitu Lexicon-based classifier, NB, Bayesian Network, SVM (Support Vector Machine), C4.5 (Decision Tree), Random Forest serta satu metode yang disebut dengan Ensemble Classifier yang menggabungkan lima metode (NB, Bayesian Network, SVM, C4.5 dan Random Forest) untuk mendapatkan akurasi yang lebih tinggi. Penelitian ini menggunakan empat kelas yaitu kelas positif (4288 tweet), negatif (35876 tweet), netral (40987 tweet) dan *irrelevant* (26715 tweet). Perolehan akurasi masing-masing saat tidak dikombinasikan dengan dataset dua kelas (menghilangkan kelas netral dan irrelevant) adalah Lexicon Based 67.9%, Naïve Bayesian 90%, Bayesian Network 91.4%, SVM 84.6%, Random Forest 89.8%. Metode Lexicon Based tidak ikut dalam kombinasi karena perolehan akurasinya paling sedikit yaitu 67,9%, perolehan akurasi ensemble dengan dataset dua kelas yaitu 91.7% sedangkan perolehan akurasi ensemble untuk dataset tiga kelas yaitu 84.2%.

Dalam *Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier* [6] digunakan dua metode supervised dengan dua dataset yaitu film dan hotel, semakin banyak data training yang diinputkan semakin bagus akurasi yang diperoleh pada algoritma NB dengan dataset film tetapi untuk metode K-NN akurasi diperoleh secara random.

Penelitian calon presiden, diteliti opini masyarakat terhadap calon presiden Indonesia tahun 2014 [4] yaitu pasangan Prabowo Subianto-Hatta Rajasa dan Joko Widodo-Jusuf Kalla. Penelitian [4] menggunakan NB untuk pengklasifikasian dokumen, selain sederhana metode ini mempunyai akurasi yang tinggi jika diaplikasikan dalam data yang besar dan beragam. Data dalam penelitian ini diambil dalam tiga periode yaitu sebelum pemilu legislatif, saat diadakan pemilu legislatif dan setelah deklarasi pengumuman pemilu legislatif kemudian dari data tersebut penulis mengelompokkan opini masyarakat apakah bersifat positif, negatif atau netral. Hasil dari polaritas sentimen, Prabowo Subianto-Hatta Rajasa mendapatkan 47,7% untuk sentimen positif, 26,4% untuk sentimen negatif dan 25,9% untuk sentimen netral sedangkan Joko Widodo-Jusuf Kalla mendapatkan 37,6% sentimen positif, 34,4% sentimen negatif dan 27,9% sentimen netral dengan tingkat akurasi sebesar 90%.

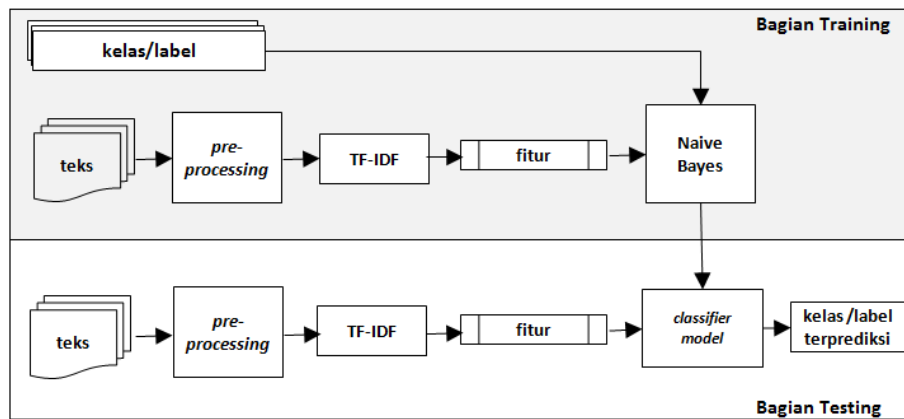
Penelitian Klasifikasi Teks dengan NB untuk Pengelompokkan Teks Berita dan Abstrak Akademis[7]. Dilakukan sebanyak 7 kali percobaan untuk dokumen berita maupun dokumen abstrak akademis pada percobaan pertama dengan jumlah data latih dan data uji 9:1 dari dokumen seluruhnya menghasilkan akurasi yang paling tinggi kemudian data latih dikurangi jumlahnya akurasinya berkurang juga. Penggunaan data latih sebesar 50% dari data seluruhnya memperoleh hasil akurasi lebih dari 75%.

Penelitian Analisis Opini Terhadap Fitur *Smartphone* pada Ulasan *Website* Berbahasa Indonesia[8]. Pengumpulan data dilakukan dengan cara web scraping yaitu mengambil data review dari website target. Yang termasuk dalam target analisa adalah BOLT 4G Powerphone IVO, Samsung Galaxy Grand I9082 White, Apple iPhone 5S 16GB Gold White, Microsoft Lumia Orange dan Nokia Lumia 520 Black. Sebelum dilakukan analisa data yang diperoleh dari web harus melalui tahap preprocessing agar memperoleh review data bersih kemudian Lexicon based, rule bahasa dan score function digunakan untuk menganalisis opini. Dari hasil pengujian diperoleh rata-rata nilai recall dan precision masing-masing sebesar 0.63 dan 0.72 sedangkan akurasinya sebesar 81.76%.

Dalam penelitian ini, untuk melakukan analisis sentimen, digunakan algoritma mesin *learning* Naïve Bayes. Algoritma umum dan mudah digunakan untuk proses klasifikasi dari data opini twitter baru yang belum memiliki kelas label untuk mengklasifikasi kedalam kelas opini positif atau negatif.

## **2. Metode Penelitian**

Penelitian ini dilakukan meliputi beberapa tahapan lihat gambar 1, meliputi : pengumpulan data, pra-pengolahan (*pre-processing*) data, pelabelan data latih dilakukan oleh ahli bahasa, pembangunan sistem pengklasifikasian data menggunakan algoritma Naive Bayes, dan evaluasi sistem.



Gambar 1. Desain sistem pengklasifikasi data opini di twitter pada pelayanan belanja *online*

### 2.1. Pengumpulan Data

Peneliti mengumpulkan data yang berasal dari *tweet* maupun *re-tweet* masyarakat di twitter. Data diambil hanya dari situs *e-commerce* di Indonesia yaitu Lazada, Bukalapak, Blibli dan Tokopedia. Data diperoleh dari proses *crawling* yaitu proses pengambilan kalimat opini dari twitter melalui API twitter. Langkah pertama yang dilakukan untuk memperoleh akses ke API twitter adalah login dengan akun twitter yang kita miliki ke <http://dev.twitter.com> dan <http://twitter.com> setelah mendapatkan API key, API secret, Access token dan Access Token Secret, API Twitter dapat diakses.

Dari data twitter yang telah dikumpulkan sejumlah 1200 data, kemudian dikelompokkan menjadi dua kelas yaitu postingan positif dan postingan negatif. Pengelompokan menjadi dua kelas tersebut dilakukan oleh dua orang ahli Bahasa Indonesia. Hasil pengelompokan tersebut adalah 610 data postingan positif dan 590 sebagai postingan negatif.

Kalimat dalam kelompok data opini positif mengandung kata-kata seperti : acungan, aklamasi, ambisius, asli, adaptif, akomodatif, andal, aspirasi, adil, akurat, aneh, asyik, afinitas, alam mimpi, anggung, afirmasi, alhamdulillah, angin sepoi, bakat, bangga, bantuan, banyak, banyak akal, barang baru, bagus dan lain-lain. Sedangkan kalimat dalam kelompok data opini negatif mengandung kata-kata seperti : abnormal, aggressor, amat panas, anarki, anjlok, absurd, aib, ambigu, anarkis, anomaly, acak, air terjun, ambivalen, anarkisme, antagonis, acak-acakan, akurat, ambivalensi, ancaman, antagonism, acuh, alarm, amoral, aneh, antek, acuh tak acuh, alasan, amoralitas, aneh lagi, anti dan lain-lain.

### 2.3. Pra Pengolahan (*Pre-Processing*) Data

Data *tweet* yang telah diambil dari twitter masih berupa data mentah maka dari itu dilakukan tahap preprocessing untuk mendapatkan data bersih agar dapat diproses ke tahap selanjutnya. Tahapan yang dilakukan adalah :

1) *Cleansing Data* : Dilakukan untuk mengurangi noise pada data tweet. Kata-kata yang tidak penting dihilangkan seperti URL, hashtag (#), username (@username), email, emoticon (:@, :\*, :D), tanda baca seperti koma (,), titik (.) dan juga tanda baca lainnya. Contoh cleansing data, input kalimat : "@LazadaIDCare kecewa dengan pelayanan dari CS Lazada, kemarin saya diinfokan bisa loh utk aktivasi kembali akun.", output kalimat : "kecewa dengan pelayanan dari CS Lazada kemarin saya diinfokan bisa loh utk aktivasi kembali akun."

2) *Case Folding* : Dalam penulisan *tweet*, pasti terdapat perbedaan bentuk huruf, tahapan ini merupakan proses merubah bentuk huruf menjadi huruf kecil (*lower case*) atau dapat disebut juga penyeragaman bentuk huruf. Contoh case folding, input kalimat : "kecewa dengan pelayanan dari CS Lazada kemarin saya diinfokan bisa loh utk aktivasi kembali akun", output kalimat : "kecewa dengan pelayanan dari cs lazada kemarin saya diinfokan bisa loh utk aktivasi kembali akun".

3) *Tokenizing* : Proses *tokenizing* atau *parsing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Pada dasarnya proses *tokenizing* adalah pemenggalan kalimat menjadi kata. Contoh tonkenizing, input kalimat : "kecewa dengan pelayanan dari cs lazada kemarin saya diinfokan bisa loh utk aktivasi kembali akun", output kalimat : "kecewa, dengan, pelayanan, dari, cs, lazada, kemarin, saya, diinfokan, bisa, loh, untuk, aktivasi, kembali, akun".

4) *Stopword Removal* : Stopword Removal merupakan proses menghilangkan daftar kata-kata yang tidak mendeskripsikan sesuatu yang semestinya dihilangkan seperti "yang", "di", "ke", "itu" dan lain sebagainya. Contoh stop word removal, input kalimat : "kecewa, dengan, pelayanan, dari, cs, lazada,

kemarin, saya, diinfokan, bias, loh, untuk, aktivasi, kembali, akun", output kalimat : " kecewa pelayanan cs lazada kemarin saya info bisa aktivasi kembali akun".

5) *Stemming* : Stemming adalah tahapan untuk membuat kata berimbuhan menjadi kata dasar sesuai dengan aturan Bahasa Indonesia yang benar. Contoh stemming, input kalimat : " kecewa pelayanan cs lazada kemarin saya bisa aktivasi kembali akun", output kalimat : " kecewa pelayanan lazada kemarin saya bisa aktif kembali akun".

6) *Convert Negation* : Dalam Bahasa Indonesia terdapat kata "tidak", "nggak", "tak", "kurang", "tanpa" yang disebut kata negasi yaitu kata yang dapat membalikan arti dari kata yang sebenarnya. Contoh *convert negation*, input kalimat : " nggak suka belanja di lazada kualitas barang jelek", output kalimat : " nggak\_suka belanja di lazada kualitas barang jelek".

#### 2.4. Term Frequency Invers Document Frequency (TFIDF)

TFIDF adalah sebuah skema pembobotan kata yang populer (*term weighting scheme*) [11]. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap *term*(kata) di setiap dokumen dalam korpus. Metode ini akan menghitung bobot setiap *term* di dokumen. Dihitung menggunakan rumus :

$$TFIDF(t) = TF * \log \frac{N}{df} \quad (1)$$

Dimana *t* adalah *term*, TF adalah jumlah dari *term* *t* yang muncul dalam dokumen, N adalah total dokumen, dan *df* adalah jumlah dokumen yang memuat *term* *t*.

#### 2.5. Klasifikasi Menggunakan Algoritma Naive Bayes

Klasifikasi adalah proses menemukan model dari sebuah data. Tujuan dari klasifikasi adalah untuk mengambil suatu keputusan dengan memprediksi suatu kasus berdasarkan hasil klasifikasi yang diperoleh[9]. Naive Bayes adalah algoritma klasifikasi yang mudah, sederhana dan sering digunakan untuk klasifikasi dokumen. Dalam proses membangun sistem pengklasifikasi menggunakan NB terdapat 2 tahapan yang dilakukan. Tahap pertama adalah proses pelatihan (*training*) dan tahap yang kedua adalah proses pengujian (*testing*).

- Tahap *Training* (Pelatihan)  
Pada tahap ini dilakukan pelatihan menggunakan data latih yang sudah ditentukan label(kelas) dari dokumen tersebut untuk membangun model[10]. Langkah tahap training menggunakan NB adalah :
  - 1) Menentukan probabilitas kategori dari data latih
  - 2) Menghitung probabilitas kata (*TFIDF*) dari data latih
  - 3) Menghitung probabilitas untuk setiap kelas data latih

$$P(wk) = \frac{pk+1}{n+|vocab|} \quad (2)$$

Dimana :

$P(wk)$  : probabilitas term  
 $pk+1$  : query + angka satu agar hasil perhitungan tidak nol  
 $n$  : semua jumlah kata yang muncul di outcome dokumen  
 $|vocab|$  : total kata keseluruhan

- 4) Menghitung probabilitas kata (*TFIDF*) dari data latih
- Tahap *Testing* (Pengujian)  
Tahap ini untuk mengetahui keakuratan model yang dibangun pada tahap *training*. Pada tahap ini digunakan data yang disebut data uji untuk memprediksi label(kelas) dari data tersebut[10]. Prosesnya adalah sebagai berikut :
    - 1) Menentukan  $V_{map}$  untuk setiap kelas data uji
$$V_{map} = P(wk/c).P(c) \quad (3)$$
    - 2) Menentukan  $V_{map}$  maksimum
    - 3) Diketahui model klasifikasinya
    - 4) Diperoleh kelas prediksi

## 2.6. Evaluasi Sistem

Evaluasi dilakukan dengan menghitung tingkat keakuratan suatu metode untuk menganalisis opini masyarakat, *k-fold cross validation* dipilih untuk menguji keakuratan metode Naïve Bayes sehingga diketahui akurasinya. Prinsip *cross-validation* adalah membagi data menjadi 2 bagian yaitu data pelatihan dan data pengujian. Sebagai contoh *3-fold*. Data dibagi menjadi 3 partisi, partisi satu menjadi data pengujian dan dua-tiga menjadi data pelatihan. Selanjutnya lakukan proses silang, dimana data pengujian dijadikan data pelatihan atau sebaliknya, proses ini diulang sebanyak 3 kali. Dalam penelitian ini menggunakan *10-fold cross validation*.

Dalam proses *10-fold cross validation* terciptalah confusion matrix yang terdiri dari *True Positives* (TP) yaitu jumlah data kelas positif yang diklasifikasikan sebagai kelas positif dan *True Negatives* (TN) yaitu jumlah data kelas negatif yang diklasifikasikan sebagai kelas negatif. Sedangkan *False Positives* (FP) merupakan jumlah data kelas negatif yang diklasifikasikan sebagai kelas positif, dan *False Negatives* (FN) merupakan jumlah data kelas positif yang diklasifikasikan sebagai kelas negatif.

Dari *confusion matrix* tersebut didapatkan pula nilai rata-rata dari *accuracy*, *precision*, *recall*, and *f-measure*. Rumus masing-masing ada di bawah ini [12]

$$accuracy = \frac{\#TP+\#TN}{\#TP+\#TN+\#FP+\#FN} \times 100\% \quad (4)$$

$$precision = \frac{\#TP}{\#TP+\#FP} \times 100\% \quad (5)$$

$$recall = \frac{\#TP}{\#TP+\#FN} \times 100\% \quad (6)$$

$$f - measure = \frac{2 \times \#TP}{2 \times \#TP+\#FP+\#FN} \times 100\% \quad (7)$$

## 3. Hasil dan Pembahasan

Pada bagian ini dijelaskan eksperimen yang dilakukan dan hasil yang diperoleh. Setelah data dipersiapkan, kemudian dilakukan *pre-processing* terhadap data dan dilakukan pelabelan untuk tiap dokumennya. Data yang digunakan sebanyak 1200 dokumen dengan pelabelannya adalah 610 data opini positif dan 590 sebagai opini negatif. Selanjutnya peneliti melakukan proses pelatihan(training) menggunakan algoritma NB. Kemudian dilanjutkan dengan pengujian dengan menggunakan data uji. Eksperimen dilakukan sebanyak sepuluh kali menggunakan teknik validasi silang (*10-fold cross validation*).

Proses Naïve Bayes yang telah dilakukan dengan data pelatihan 1080 data dan data pengujian 120 data yang dilakukan validasi silang sebanyak sepuluh kali, menghasilkan tabel *confusion matrix* seperti terlihat di tabel 1.

Tabel 1. *Overall Confusion Matrix* dari Naive Bayes

		Aktual	
		Positif	Negatif
Prediksi	Positif	59	3
	Negatif	5	53

Dari *confusion matrix* pada tabel 1, didapatkan nilai rata-rata dari *accuracy*, *precision*, *recall*, dan *f-measure* seperti terlihat di tabel 2. Perhitungannya menggunakan rumus (4), (5), (6) dan (7). Selain *precision* dan *recall* juga didapatkan *f-measure* sebesar 93.65% dan *Accuracy* sebesar 93.33%.

Tabel 2. *Overall Precision dan Recall* Tiap Kelas

Kelas/Label	Precision (%)	Recall (%)
Postingan Twitter		
Positif	95.16	92.19
Negatif	91.38	94.64

## 4. Simpulan

Paper ini mempresentasikan analisis sentimen terhadap data opini di twitter di beberapa situs *online marketplace* di Indonesia. Metode untuk melakukan analisis sentimen menggunakan algoritma

pengklasifikasi Naive Bayes, yang terdiri dari dua tahap yang dilakukan, yaitu tahap *training* dan tahap *testing*. Untuk menentukan kinerja dari mesin pengklasifikasi Naive Bayes dalam proses klasifikasi dilakukan eksperimen menggunakan 10 kali validasi silang (*10-fold cross validation*). Dari hasil eksperimen tersebut didapatkan rata-rata akurasi sebesar 93.33%.

#### **Daftar Pustaka**

- [1] A. Ishak, "Analisis Kepuasan Pelanggan Dalam Belanja Online : Sebuah Studi Tentang Penyebab ( Antecedents ) dan Konsekuensi ( Consequents )," *Jurnal Siasat Bisnis*, Vol. 16 No. 2, , Hal : 141-154, Juli 2012.
- [2] D. G. Nugroho, Y. H. Chrisnanto, and A. Wahana, "Analisis Sentimen pada Jasa Ojek Online Menggunakan Metode Naive Bayes," *Jur. Inform. Fak. Mat. dan Ilmu Pengetah. Alam Univ. Jenderal Achmad Yani*, pp. 156–161, 2015.
- [3] B. S. D and P. D. Gore, "Sentiment Analysis on Twitter Data Using Support Vector Machine," *International Journal of Computer Science Trends and Technology (IJCST) – Volume 4 Issue 3*, pp. 365–370, May - Jun 2016.
- [4] F. Nurhuda and S. W. Sihwi, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *JURNAL ITSMART Vol 2. No 2. ISSN : 2301–7201*, Desember 2013
- [5] Y. Wan, "An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis," *IEEE 15th International Conference on Data Mining Workshops*, 2015.
- [6] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment Analysis of Review Datasets Using Naive Bayes and K-NN Classifier," *Int. J. Inf. Eng. Electron. Bus.*, vol. 8, no. 4, pp. 54–62, 2016.
- [7] A. Hamzah, "Sentiment Analisis Untuk Memanfaatkan Saran Kuesioner dalam Evaluasi Pembelajaran dengan Menggunakan Naive Bayes Classifier (NBC)," *Pros. Semin. Nas. Apl. Sains Teknol.*, pp. 17–24, November, 2014.
- [8] D. Setyawan and E. Winarko, "Analisis Opini Terhadap Fitur Smartphone Pada Ulasan Website Berbahasa Indonesia," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 10, no. 2, p. 183, 2016.
- [9] B. Liu, E. Blasch, Y. Chen, D. Shen, and G. Chen, "Scalable Sentiment Classification for Big Data Analysis Using Naive Bayes Classifier," *IEEE International Conference on Big Data*, pp. 99–104, 2013.
- [10] A. P. Wijaya and H. A. Santoso, "Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government," *Journal of Applied Intelligent System*, Vol.1, No. 1, pp. 48–55, Februari 2016.
- [11] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," in *Proceedings of the 10th European Conference on Machine Learning*, 1997.
- [12] I. H. Witten, E. Frank, M. A. Hall, "Data mining: practical machine learning tools and techniques. 3rd Edition," San Francisco , MA: Morgan Kaufmann, 2011.