

ANALISIS SENTIMEN KEPUASAN PEMANGKU KEPENTINGAN MENGUNAKAN METODE NAÏVE BAYES CLASSIFIER DAN K-NEAREST NEIGHBOUR

Ni Luh Ratniasih¹, Indi Larasati Nabila Putri²

Institut Teknologi dan Bisnis (ITB) STIKOM Bali¹²

Jalan Raya Puputan No 86, Renon, Denpasar

*e-mail: ratni@stikom-bali.ac.id¹, indilarasati@gmail.com²

ABSTRAK

Pemangku kepentingan merupakan individu atau kelompok yang memiliki kepentingan dan dapat memberikan pengaruh kepada suatu objek. Pengukuran kepuasan pemangku kepentingan (*stakeholder*) sangat penting dilakukan untuk mendapatkan umpan balik dan masukan bagi keperluan pengembangan dan implementasi strategi peningkatan kepuasan pemangku kepentingan, sehingga perlu diketahui opini dari pemangku kepentingan. Metode penelitian terdiri dari beberapa tahap diantaranya tahap pertama dilakukan identifikasi masalah dan studi pustaka, tahap kedua pengumpulan data kepuasan pemangku kepentingan (Mahasiswa), tahap ketiga preprocessing data, tahap keempat adalah ekstraksi fitur agar dapat mempermudah klasifikasi menggunakan metode *Naïve Bayes Classifier (NBC)* dan *K-Nearest Neighbour (KNN)*. Tahap keempat merupakan tahap pengujian dan evaluasi model. Tahap kelima adalah pengujian tingkat akurasi metode. Tujuan dari penelitian ini adalah mendapatkan perbandingan tingkat akurasi antara metode *Naïve Bayes Classifier (NBC)* dan *K-Nearest Neighbour (KNN)* pada analisis sentimen dari komentar hasil pengukuran kepuasan pemangku kepentingan. Hasil tingkat akurasi menggunakan metode *Naïve Bayes Classifier (NBC)* sebesar 91,13% dan *K-Nearest Neighbour (KNN)* sebesar 83,06% sehingga *performance* metode *Naïve Bayes Classifier (NBC)* lebih tinggi dalam analisis sentiment kepuasan pemangku kepentingan.

Kata kunci: Analisis Sentimen, Kuesioner Pemangku Kepentingan, NBC, K-NN

ABSTRACT

Stakeholders are individuals or groups who have interests and can exert influence on an object. It is very important to measure the satisfaction of stakeholders (stakeholders) to get feedback and input for the purposes of developing and implementing strategies to increase stakeholder satisfaction, so it is necessary to know the opinions of stakeholders. The research method consists of several stages including the first stage is problem identification and literature study, the second stage is collecting data on stakeholder satisfaction (Students), the third stage is data preprocessing, the fourth stage is feature extraction in order to facilitate classification using the Naïve Bayes Classifier (NBC) method and K-Nearest Neighbor (KNN). The fourth stage is the testing and evaluation stage of the model. The fifth stage is testing the accuracy of the method. The purpose of this study is to compare the level of accuracy between the Naïve Bayes Classifier (NBC) and K-Nearest Neighbor (KNN) methods on sentiment analysis from comments on the results of measuring stakeholder satisfaction. The results of the accuracy rate using the Naïve Bayes Classifier (NBC) method of 91.13% and K-Nearest Neighbor (KNN) of 83.06% so that the performance of the Naïve Bayes Classifier (NBC) method is higher in the analysis of stakeholder satisfaction sentiment.

Keywords: Sentiment Analysis, Stakeholder Questionnaire, NBC, K-NN

1. PENDAHULUAN

Menjadi perguruan tinggi unggulan dan berkualitas Internasional di bidang ilmu pengetahuan, teknologi dan seni merupakan

Visi dari Institut Teknologi dan Bisnis (ITB) STIKOM Bali. Institut Teknologi dan Bisnis (ITB) STIKOM Bali secara terjadwal melakukan pengukuran terhadap kepuasan pemangku kepentingan. Pemangku kepentingan merupakan individu atau kelompok yang memiliki kepentingan dan dapat memberikan pengaruh kepada suatu objek [1]. Masing – masing pemangku kepentingan memiliki kepentingan dan pengaruh yang berbeda – beda (Bryson, 2004), sehingga tanggung jawab yang dimiliki juga berbeda – beda. Peran pemangku kepentingan dikelompokkan menjadi tiga yaitu pemangku kepentingan utama (primer), pemangku kepentingan pendukung (sekunder) dan pemangku kepentingan kunci [2].

Pengukuran kepuasan pemangku kepentingan (*stakeholder*) sangat penting dilakukan untuk mendapatkan umpan balik dan masukan bagi keperluan pengembangan dan implementasi strategi peningkatan kepuasan pemangku kepentingan, sehingga perlu diketahui opini dari pemangku kepentingan. Opini pemangku kepentingan Institut Teknologi dan Bisnis (ITB) STIKOM Bali dapat dilihat pada komentar kuesioner yang disebarakan Direktorat Penjaminan Mutu dan Pengawasan Internal setiap semester. Saat ini belum pernah dilakukan analisis terhadap hasil komentar kuesioner kepuasan pemangku kepentingan tersebut sehingga para pimpinan tidak mengetahui apakah komentar cenderung positif atau negatif.

Berdasarkan studi literatur terdapat penelitian sebelumnya yang dilakukan oleh Retno Sari tentang Analisis Sentimen dapat diketahui bahwa peneliti menggunakan metode KNN, namun pada penelitian tersebut tidak menjelaskan tools yang digunakan serta menggunakan satu metode saja [3]. Selanjutnya pada penelitian yang dilakukan oleh Ghulam Asrofi Buntoro tentang Analisis Sentimen [4] dapat diketahui bahwa peneliti menggunakan metode SVM, namun tidak menjelaskan tools dalam analisisnya. Penelitian analisis sentimen yang menggunakan algoritma Naïve Bayes sudah banyak dilakukan sebelumnya. Penelitian yang dilakukan oleh Samsir., dkk dalam analisis sentimen opini masyarakat tentang pembelajaran daring di masa pandemi COVID-19 menghasilkan 30% sentiment positif, 69% sentimen negatif, dan 1% netral [5]. Penelitian yang dilakukan oleh Safitri Juanita terkait

dengan persepsi masyarakat terhadap pemilu 2019 pada media sosial twitter menyimpulkan persepsi negatif sebesar 52% jauh lebih besar dari persepsi positif 18% dan persepsi netral 31% lebih tinggi dari persepsi positif [6].

Pada penelitian ini dilakukan analisis sentiment pada komentar para pemangku kepentingan menggunakan dua metode yaitu Naïve Bayes Classifier (NBC) dan K-Nearest Neighbour (KNN), dimana tujuan penelitian ini adalah mengetahui tingkat akurasi dari kedua metode tersebut.

2. METODE PENELITIAN

Metode penelitian terdiri dari beberapa tahap diantaranya tahap pertama dilakukan identifikasi masalah dan studi pustaka, tahap kedua pengumpulan data kepuasan pemangku kepentingan (Mahasiswa), tahap ketiga *preprocessing data*, tahap keempat adalah ekstraksi fitur agar dapat mempermudah klasifikasi menggunakan algoritma *Naïve Bayes Classifier (NBC)* dan *K-Nearest Neighbour (KNN)*. Tahap kelima merupakan tahap pengujian dan evaluasi model. Tahap keenam adalah pengujian tingkat akurasi metode. Pada Gambar 1 menunjukkan tahapan atau alur penelitian ini.

a. Identifikasi Masalah dan Studi Pustaka.

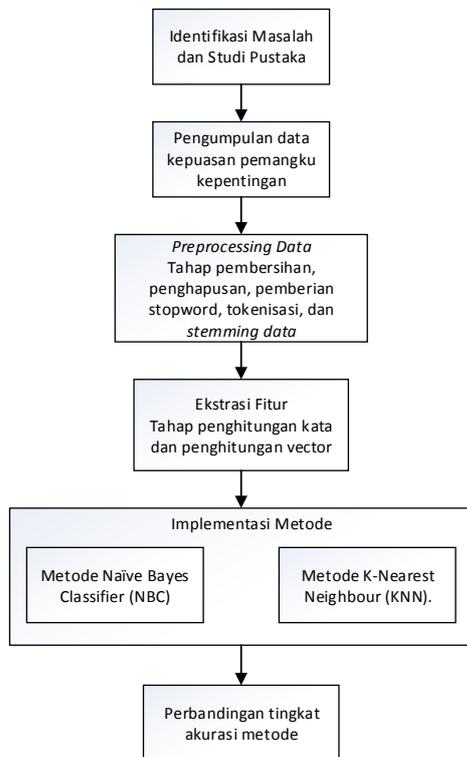
Pada tahap ini dilakukan identifikasi permasalahan tentang kepuasan pemangku kepentingan (Mahasiswa). Hasil pengisian kuesioner kepuasan pemangku kepentingan diperoleh langsung dari Bagian Direktorat Penjaminan Mutu dan Pengawasan Internal ITB STIKOM Bali. Pengukuran kepuasan pemangku kepentingan dilakukan satu kali dalam satu semester sesuai aturan yang tertuang dalam Prosedur Manual.

Kemudian mencari teori–teori terkait *text mining* dan analisis sentimen serta mempelajari pustaka yang relevan terhadap masalah penelitian sejenis yang pernah dilakukan sebelumnya yang digunakan untuk menunjang penelitian ini.

b. Pengumpulan Data Penelitian

Pada tahap ini dilakukan pengumpulan data penelitian, data diperoleh dari Direktorat penjaminan Mutu dan Pengawasan internal. Data berupa komentar hasil pengukuran

kepuasan pemangku kepentingan (Mahasiswa) semester Ganjil 2020/2021 yang diisi melalui kuesioner. Jumlah Mahasiswa yang aktif di Semester Ganjil 2020/2021 adalah sebanyak 4338 orang, sedangkan jumlah responden pada semester tersebut adalah 3856 orang (88,89%).



Gambar 1. Alur Penelitian

c. Preprocessing Data

Tahap *preprocessing data* merupakan tahap seleksi data dan mengubahnya terstruktur. Terdapat beberapa proses dalam *preprocessing data* diantaranya tahap *Cleansing*. Kalimat yang didapat biasanya masih terdapat noise, yaitu kesalahan acak atau varian dalam *variable* terukur kita harus menghilangkan noise tersebut. Tahap berikutnya *parsing* yaitu proses memecah dokumen menjadi sebuah kata dengan melakukan analisa terhadap kumpulan kata dengan memisahkan kata tersebut dan menentukan struktur sintaksis dari tiap kata tersebut. Selanjutnya tahap normalisasi kalimat yang bertujuan untuk menormalkan kalimat sehingga kalimat gaul menjadi normal sehingga bahasa gaul tersebut dapat dikenali sebagai bahasa yang sesuai dengan KBBI. Proses

tokenization digunakan dalam identifikasi kata – kata dan memecah kalimat menjadi istilah berdasarkan spasi dan tanda baca. Tahap terakhir dalam *preprocessing* yaitu *stemming* merubah kata imbuhan menjadi kata dasar [7].

Selanjutnya dalam penelitian ini adalah ekstrak fitur untuk mempermudah klasifikasi Naive Bayes. Tahap ini menghasilkan model dan digunakan untuk menunjukkan ketepatan hasil klasifikasi. Pada tahap ini dilakukan pembobotan terminologi kata dengan Algoritma TF/IDF, cara menghitung frekuensi kemunculan kata. Persamaan yang digunakan untuk menghitung bobot (W) dokumen (dalam hal ini kalimat satu enter pada saat chatting) terhadap kata kunci (dalam hal ini kata spam) menggunakan persamaan berikut ini :

$$W_{d,t} = tf_{d,t} * IDF$$

dimana : *W* = bobot dokumen ke – *n*
d = dokumen
t = kata kunci
tf = *terms frequency* (jumlah kemunculan kata)
IDF = *Inverse Document Frequency*

d. Implementasi Metode

Pada tahapan ini dilakukan implementasi kedua metode yaitu *Naive Bayes Classifier (NBC)* dan *K-Nearest Neighbour (KNN)*. Data dibagi menjadi dua dataset yaitu data latih dan data uji. Jumlah data yang digunakan sebanyak 170 data dimana 90% data latih dan 10% data uji. Pembagian persentase ini digunakan berdasarkan hasil pengujian sebelumnya yang menunjukkan pengujian 3 kelas menggunakan dataset 90% data latih dan 10% data uji memiliki akurasi tertinggi sebesar 77.78% [8].

e. Membandingkan Tingkat Akurasi

Pada tahap ini dilakukan analisis terhadap hasil implementasi metode *Naive Bayes Classifier (NBC)* dan *K-Nearest Neighbour (KNN)*. Analisis dilakukan untuk mengetahui tingkat akurasi dari kedua metode sehingga kesimpulan hasil penelitian dapat diperoleh.

3. HASIL DAN PEMBAHASAN

3.1 Preprocessing Data

Data diperoleh dalam format *excel* dimana data bersumber dari Bagian Direktorat Penjaminan Mutu dan Pengawasan Internal. Data hasil pengukuran kepuasan pemangku kepentingan (Mahasiswa) terdiri dari NIM, nilai harapan dan aktual, serta komentar. Pada penelitian ini yang digunakan sebagai data penelitian hanya NIM dan komentar mahasiswa seperti yang terlihat pada Tabel 1.

Tabel 1. Data Komentar Pemangku Kepentingan (Mahasiswa)

NO	NIM	KOMENTAR
1	140010021	Bagus
2	140010045	Selama ini kebetulan belum ada pengalaman mengecewakan yang saya alami. terima kasih
3	140010092	top
4	140010136	Mantap
5	140010139	Sudah Bagus
6	140010211	Tidak ada
7	140010221	Mantap
8	140010229	Tidak ada
9	140010287	semua sudah baik
10	140010306	Pemberian informasi perlu ditingkatkan

Proses *preprocessing* data dilakukan untuk menghilangkan *noise* dan memperjelas fitur serta membersihkan dengan menghapus atribut – atribut yang tidak diperlukan. Proses ini akan mengolah sebuah teks menjadi data yang mudah diterima oleh sistem saat proses utama dilakukan. Tahap *preprocessing* diperlukan sebelum melakukan implementasi algoritma. Proses *preprocessing* data menggunakan Rapid Miner, dimana terdapat beberapa tahapan proses yang dilakukan antara lain :

a. *Import Data* : mengambil data yang akan diproses dari tempat penyimpanan sehingga data dapat dibaca pada *local* Rapid Miner. Pada tahap ini akan terlihat data secara keseluruhan dan *missing data* yang perlu *preprocessing*. Berdasarkan hasil analisis pada data statistics terdapat 153 data missing. Pada proses *import data* dilakukan dengan mengambil data dari file *excel* namun kolom yang digunakan hanya kolom

komentar seperti yang terlihat pada Gambar 2 berikut ini. Jumlah data yang diproses pada tahap ini adalah 4338 data.

Row No.	KOMENTAR
1	Bagus
2	Selama ini kebetulan belum ada pengalaman mengecewakan yang s...
3	top
4	Mantap
5	Sudah Bagus
6	Tidak ada
7	semua sudah baik
8	Pemberian informasi perlu ditingkatkan
9	Ok
10	Kurang adanya aplikasi untuk android atau ios supaya setiap ad pen...
11	baik
12	menurut saya tidak ada sejauh ini
13	Lebih baik lagi
14	Perbaikan absensi/kehadiran mahasiswa dan pendaftaran proposal ...
15	Pelayanan sudah cukup baik

Gambar 2. Proses Import Data

- b. *Replace Data* : merupakan operator ini digunakan untuk menghilangkan beberapa karakter yang ada pada komentar. Karakter atau simbol yang dihilangkan seperti tanda titik (.), tanda seru (!), tanda tanya (?), dan lain – lain. Jumlah data komentar setelah diproses menggunakan operator *replace* menjadi 3855 komentar.
- c. *Filter Examples* : merupakan operator yang digunakan untuk menghilangkan baris – baris data yang kosong yang ditandai dengan simbol tanda tanya (?) (*missing attribute*). Jumlah data setelah dilakukan proses filter menjadi 3688 komentar.
- d. *Remove Duplicates* : merupakan operator yang digunakan untuk menghapus beberapa komentar yang sama untuk mempermudah proses analisis sentimen dan mengurangi waktu pelabelan komentar. Jumlah data komentar setelah diproses dengan operator ini adalah 2333 komentar. Tahap ini menghasilkan data yang sudah siap untuk diproses.
- e. *Write CSV* : merupakan operator untuk menyimpan data hasil terakhir dengan format data *.CSV. Data tersebut merupakan data set yang akan diberikan label sentimen secara manual.

Tahapan *preprocessing* data menggunakan Rapid Miner dapat dilihat pada Gambar 3 sedangkan hasil (*result*) *preprocessing* data dapat dilihat pada Gambar 4 berikut ini :

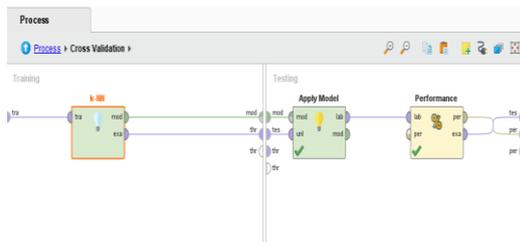
$$Accuracy = \frac{(120 + 106)}{18 + 120 + 106 + 4}$$

$$Accuracy = \frac{226}{248} = 91,13$$

Gambar 8. Hasil Prediksi dari Metode NBC

3.3 Implementasi Metode K-Nearest Neighbour (KNN)

Sentiment analisis menggunakan metode *K-Nearest Neighbour (KNN)* dilakukan dengan tahap yang sama pada implementasi metode *Naive Bayes Classifier (NBC)*. Tahapan yang sama dilakukan dari tahap pertama sampai tahap kelima, dimana pada tahap kelima ini akan digunakan metode *K-Nearest Neighbour (KNN)* sebagai metode dalam pemodelan seperti yang terlihat pada Gambar 9.



Gambar 9. Mengukur Performance Metode KNN

Pada Gambar 9 terlihat implementasi model tersebut dengan operator *Apply Model* yang merupakan tahap keenam. Tahap ketujuh adalah melihat *performance* metode dengan operator *Performance Classification* seperti yang terlihat pada Gambar 10. Tabel 2 merupakan hasil eksperimen nilai K dimana tingkat akurasi tertinggi yaitu 83,03% dihasilkan pada saat nilai K ditentukan 10.

Gambar 10. Tingkat Akurasi Metode K-NN

Tabel 2. Eksperimen Nilai K

Nilai K	Accuracy
1	81.48%
2	81.49%
3	68.58%
4	68.60%
5	60.52%
6	66.17%
7	67.33%
8	77.80%
9	81.82%
10	83.06%
11	78.98%
12	80.20%
13	77.38%
14	77.38%
15	76.18%

Tingkat akurasinya pada Gambar 10 adalah sebesar 83,06%, jika diturunkan ke dalam *Confusion Matrix* hasilnya adalah sebagai berikut.

$$Accuracy = \frac{(TN + TP)}{TN + FN + TP + FP}$$

$$Accuracy = \frac{(120 + 86)}{38 + 120 + 86 + 4}$$

$$Accuracy = \frac{206}{248} = 83,06$$

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan sebagai berikut : analisis sentiment menggunakan metode *Naive Bayes Classifier (NBC)* menghasilkan tingkat kurasi akurasi sebesar 91,13% sedangkan dengan

metode *K-Nearest Neighbour (KNN)* menghasilkan tingkat akurasi sebesar 83,06% dengan nilai $K=10$ sehingga dapat disimpulkan bahwa tingkat akurasi metode *Naive Bayes Classifier (NBC)* lebih tinggi.

DAFTAR PUSTAKA

- [1] Hidayah, N., Hutagalung, S. S., & Hermawan, D. 2019. Analisis peran stakeholder dalam pengembangan wisata talang air peninggalan kolonial Belanda di Kelurahan Pajaresuk Kabupaten Pringsewu. *Jurnal Ilmu Administrasi Publik*. Vol. 7. No. 1. Hal. 55–71. doi:10.31289/publika.v7i1.2179.
- [2] Bryson, J. M. 2004. What to do when stakeholders matter: stakeholder identification and analysis techniques. *Public Management Review*, Vol. 6. No. 1. Hal. 21–53. doi:10.1080/14719030410001675722.
- [3] Sari, R. (2020). Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma *KNearest Neighbor (K-Nn)*. *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(1), 10–17. <https://doi.org/10.31294/evolusi.v8i1.7371>.
- [4] Buntoro, Ghulam Asrofi. 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Integer Journal*. Vol. 2. No. 1. Hal. 32-41.
- [5] Samsir, A., Verawardina, U., Edi, F., Watrianthos, R. 2021. Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode *Naive Bayes*. *Jurnal Media Informatika Budidarma*. Vol. 5, No. 1. Page 157-163.
- [6] Juanita, S. 2020. Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan *Naive Bayes*. *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 3, p. 552.
- [7] Rasenda, R., H. Lubis, and R. Ridwan 2020. “Implementasi *K-NN* Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter,” *J. MEDIA Inform. BUDIDARMA*, doi: 10.30865/mib.v4i2.2051.
- [8] Gunawan, Billy, Helen Sasty Pratiwi, Enda Esyudha Pratama. 2018. Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode *Naive Bayes*. *Jurnal Edukasi dan Penelitian Informatika*. Vol. 4. No. 2. Hal. 113-118.
- [9] Akbari, M. I. H. A. D., Astri Novianty S.T., M. & Casi Setianingsih S.T., M., 2012. Analisis Sentimen Menggunakan Metode *Learning Vector Quantization*. Telkom University.
- [10] Han, J., Kamber, M. & Pei, J. 2012. *Data Mining Concepts and Techniques*. 3rd ed. Waltham: Morgan Kaufmann Publishers.
- [11] Kalpit G. Soni and Dr. Atul Patel. 2017. *Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data*. *International Journal of Computational Intelligence Research* ISSN 0973-1873 Vol. 13, No. 5, pp. 899-906.