

Penerapan Algoritma Naïve Bayes Untuk Analisis Kepuasan Penggunaan Aplikasi Bank

Domi Sepri¹

Fakultas Sains & Teknologi, Sistem Informasi, UIN Imam Bonjol, Padang, Indonesia

Email: domisepri@uinib.ac.id

Submitted: 20/11/2020; Accepted: 27/11/2020; Published: 27/11/2020

Abstrak—Dunia perbankan telah bertransformasi dalam penggunaan aplikasi berbasis mobile dalam pelayanan dan memberikan kemudahan kepada konsumen. Aplikasi berbasis android seperti mobile banking belum sepenuhnya diminati pengguna karena nilai keamanannya. Pada bidang text mining terdapat metode menganalisa kepuasan pengguna melalui konten sentimen analisis dengan pendekatan yang digunakan penulis dalam penelitian ini adalah menggunakan Algoritma Naïve Bayes. Penelitian ini diolah dengan bahasa pemrograman python dimana terdapat modul textblob untuk proses penganalisaan keefektifan penggunaan Algoritma Naïve Bayes. Dari hasil pengolahan penelitian ini dihasilkan akurasi penggunaan Algoritma Naïve Bayes sebesar 0,8714 (87%).

Kata Kunci: Algoritma Naive Bayes, Sentimen Analisis, Teks Mining

Abstract—The world of banking has been transformed by using mobile-based applications in services and providing convenience to consumers. Android-based applications such as mobile banking are not yet fully attractive to users because of their security value. In the field of text mining, there is a method of analyzing user satisfaction through content sentiment analysis. The approach used by the author in this study is to use the Naïve Bayes Algorithm. This research was processed with the python programming language where there is a textblob module for the process of analyzing the effectiveness of using the Naïve Bayes Algorithm. From the results of the processing of this research, the accuracy of using the Naïve Bayes Algorithm is 0.8714 (87%).

Keywords: Naïve Bayes Algorithm, Sentiment Analysis, Text Mining

1. PENDAHULUAN

Aplikasi berbasis android telah banyak dikembangkan oleh korporat dalam menunjang proses pelayanan bisnisnya. Sektor perbankan menjadi sektor yang harus memperhatikan aspek keamanan data dalam pelayanan berbasis online dalam transaksi keuangan yang dilakukan.

Teks mining adalah proses penambangan data (data mining) yang mengambil informasi dari data yang berupa teks. teks mining mengekstrak data-data berupa teks seperti email, dokumen teks, file HTML serta data dari social media [1].

Tahapan proses pada teks mining adalah dimulai dari proses tokenisasi, yaitu pemisahan per kata dari keseluruhan sumber teks asli yang akan diolah. Tokenisasi mengubah ke satu tipe huruf (tidak ada lagi huruf besar dan kecil), dan tanda baca sudah hilang. Berikutnya setelah tokenisasi adalah proses *stemming* yaitu metode meningkatkan kinerja token potensial, dimana dilakukan pencarian varian token potensial dengan proses menghilangkan imbuhan kata. Setelah terbetuk kata dasar maka proses berikutnya adalah pengelompokan teks yaitu mengelompokkan kata per kata berdasarkan pengkategorian data. Setelah pengelompokan ini maka data bisa diolah menggunakan algoritma data mining dalam hal ini menggunakan algoritma klasifikasi naïve bayes untuk penemuan pola dari data training yang digunakan. Setelah didapatkan keberhasilan data training maka proses berikutnya adalah pengaplikasian kedalam data latih untuk proses evaluasi dari keseluruhan data.

Kepuasan pengguna pada aplikasi berbasis android bisa dilihat dari ulasan maupun komentar dari pengguna yang bisa dilihat pada play store android. Ulasan inilah yang akan menjadi sumber penulis dalam melihat dan menganalisa sentimen analisis dari aplikasi Bank. Sentimen analisis adalah proses menganalisa opini dan ekspresi masyarakat di media sosial dengan berbagai pendekatan. Ini merupakan gabungan antara ilmu teks mining dan proses pemrosesan bahasa alami (*Natural Language Processing*) [2]. Sentimen analisis didapatkan dari proses *crowling* menggunakan *tool* dan disimpan ke dalam *database*. Dengan sentimen analisis ini maka kecenderungan positif maupun negatif perilaku sosial media di masyarakat bisa dinilai tingkat akurasi dengan baik.

Menurut Liu [3], sentimen analisis sering disebut *opinion mining* yang mengidentifikasi dan mengungkapkan pendapat, emosi, evaluasi, sikap, emosi, subjektivitas, Penilaian atau opini yang terkandung dalam teks. Dalam hal ini teks banyak sumber dalam pengungkapan emosi seseorang termasuk di media sosial. Fenomena yang ada di dunia nyata sekarang ini banyak diekspresikan dalam bentuk tulisan opini di sosial media oleh manusia. Di dalam media sosial berbagai macam sumber opini untuk melihat sentimen analisisnya seperti *hashtag* maupun komentar-komentar dari grup maupun status dari media sosial.

Sentimen Analisis menganalisa opini dan ekspresi masyarakat di media sosial dengan berbagai pendekatan. Tugas dasar analisis sentimen adalah Polaritas teks dalam grup dokumen, apakah akan mengungkapkan pendapat Dokumen tersebut positif, negatif atau netral [4]. Penelitian tentang analisis sentimen telah berkembang Sejak 2003 yang umum digunakan adalah menganalisa dari twitter, facebook, instagram sebagai wadah pengekspresian

diri pengguna social media. Pendekatan yang akan penulis lakukan adalah dengan menganalisa menggunakan algoritma Naïve Bayes pada review *playstore* aplikasi Mobile Banking Bank Muamalat.

Pada Penelitian ini sentimen analisis dikembangkan menggunakan algoritma naïve bayes. Sistem kerjanya ialah dengan pengelompokan statistik dengan pendekatan yang digunakan adalah menggunakan metode hitung dari probabilitas suatu kejadian [5]. Algoritma Naive Bayes adalah sebuah metode klasifikasi probabilitas sederhana kedalam proses hipotesis dimana terdapat independensi yang tinggi dalam penerapannya. Di samping itu algoritma ini dalam proses klasifikasi dilakukan dengan cepat dan metode presisi tinggi [6]. Keuntungan menggunakan pengklasifikasi Naive Bayes adalah Metode ini hanya membutuhkan data training kecil untuk menentukan nilai taksiran parameter yang dibutuhkan dalam proses klasifikasi [7] .

Ilmu sentimen analisis di sosial media telah banyak digunakan dalam penelitian sebelumnya diantaranya penelitian dari Winda EN, dkk tentang analisis sentimen terhadap tayangan televisi[8] . Pada penelitian di atas sentimen analisis mengambil sumber data dari twitter dan mengklasifikasi sentimen menjadi positif dan negatif menggunakan algoritma K-Nearest Neighbor. Di samping itu penelitian Ratnawati Fajar tentang sentimen opini film pada twitter menggunakan algoritma naïve bayes menghasilkan akurasi algoritma sebesar 90 % . Ini menghipotesa penulis bahwa ilmu teks mining menggunakan algoritma naïve bayes cocok digunakan dalam penganalisaan data sentimen analisis [9].

Tujuan Penelitian ini adalah untuk menganalisa perilaku sentimen analisa masyarakat sehingga didapatkan tingkat akurasi menggunakan algoritma naïve bayes dalam melihat kecenderungan masyarakat pada komentar pengguna aplikasi bank berbasis android. Diharapkan kecenderungan masyarakat ini menjadi bahan evaluasi juga bagi pihak bank dalam menentukan kebijakan yang berbasis teknologi informasi.

2. METODOLOGI PENELITIAN

Tahapan penelitian pada tulisan penelitian ini adalah :

2.1 Pengumpulan data

Proses ini dengan mengambil ulasan dari aplikasi bank pada *playstore*. Dalam hal ini ulasan yang akan diambil adalah ulasan pada aplikasi Mobile Banking Muamalat yang dimasukkan ke dalam format csv. Jumlah data ulasan pada play store aplikasi mobile banking bank muamalat berjumlah 1985 data yang terdiri dari komentar positif, negatif dan netral.

2.2 Pemrosesan Teks Mining

a. Proses Data

Tahapan ini dilakukan dengan proses normalisasi kata sesuai dengan kaidah bahasa Indonesia. Proses yang dilakukan seperti tokenisasi, *casefolding* (pengubahan ke huruf kecil), penghilangan stop word (kata-kata yang tidak penting, dan *remove punctuation* (penghilangan tanda baca). Proses tokenisasi menggunakan *Natural Language Toolkit (NLTK)* dimana tokenisasi yang dipakai adalah *regextokenizer* dan pada penghilangan stop word modul yang digunakan adalah *nlk corpus*.

b. Stemming

Pada tahapan ini data yang sudah dilakukan proses penghilangan imbuhan dengan teknik stem. Proses penghilangan imbuhan ini menggunakan modul *sastrawi stemmer* yang sudah tersedia pada modul pemrograman python.

c. Pengelompokan Kata

Tahapan ini adalah mengelompokkan kata berdasarkan kategorinya. Dari 1890 data total dilakukan pengelompokan menjadi positif dan negatif dimana pengkategorian ini berdasarkan:

- 1 Bintang 4 dan 5 adalah review positif
- 2 Bintang 1 dan 2 adalah review Negatif
- 3 Bintang 3 adalah review netral

Setelah dilakukan pengkelompokkan data ini didapatkan rincian sebagai berikut :

- 1 Review Positif : 1506 Data
- 2 Review Negatif : 277 Data
- 3 Review Netral : 107 Data

Pada penelitian ini review netral tidak masuk kedalam kategori pengelompokan sehingga total data tanpa kategori netral adalah 1783 data.

d. Data Mining

Data mining dalam penelitian ini menggunakan algoritma naïve bayes, dimana pada pengaplikasiannya penulis memakai library *textblob* dalam algoritma klasifikasi naïve bayes yang sudah tersedia pada pemrograman python. Dari jumlah review data positif dan negatif terdapat ketidakseimbangan jumlah data, maka diperlukan penyeragaman jumlah data yang akan diolah. Untuk data training digunakan 200 data positif dan 200 negatif.

e. Algoritma Naïve Bayes

Secara umum cara penghitungan Naïve Bayes terdapat pada rumus :

$$P(Y | X) = \frac{P(X | Y)PY}{PX} \quad (1)$$

Keterangan :

- X : Data dari kelas yang belum diketahui
- Y : Hipotesis data yang merupakan spesifik kelas tertentu
- P(Y|X) : Probabilitas hipotesis pada kondisi tertentu X
- P(Y) : Probabilitas hipotesis Y
- P(X|Y) : Probabilitas X saat kondisi hipotesis Y
- P (X) : Probabilitas X

Tahapan dari proses algoritma naive bayes terdapat dua langkah pada proses klasifikasi teks, tahap pelatihan dan tahap klasifikasi. Selama fase pelatihan, proses akan dilakukan menggunakan data training 400 data. Pada tahap klasifikasi dilakukan dengan menentukan nilai kategori data sesuai dengan kondisi muncul dalam data klasifikasinya [10]. Pada tahap klasifikasi ini 400 data training dibagi kedalam 2 klasifikasi data yaitu 200 data positif dan 200 data negative.

f. Evaluasi

Pada tahapan ini dilakukan proses evaluasi dan pengecekan akurasi algoritma naïve bayes dari data training yang digunakan dengan melihat akurasi pada data testing sebanyak 70 data review positif dan 70 data review negatif.

3. HASIL DAN PEMBAHASAN

3.1 Data

Gambar 1 di bawah adalah 5 sampel teratas dari review pengguna dari 1980 total review aplikasi pada playstore android. Data belumlah dilakukan proses normalisasi kata, stemming dan pengelompokan data untuk diolah menggunakan algoritma data mining. Dapat dilihat gambar di atas masih terdapat data review yang menunjukkan sentimen netral (bintang 3) yang dalam hal ini akan dilakukan proses preprocessing dan pengolahan data mining.

	Komentar	Bintang
0	Force close di Zenfone 5f	2
1	Kecewa Aneh,ap saya mau log in malah diminta kode challenge. Sungguh merepotkan ketika saya berada di luar negri seperti sekarang	1
2	Keren	5
3	Perbaiki lagi Mhn kualitas aplikasi di perbaiki lagi	3
4	Thankyou so much. Mempermudah transaksi dan gak ribet harus ke ATM kalo mau transfer.. Bagi yg ada masalah coba langsung ke cs nya aja. Selamat mencoba 🍀	5

Gambar 1. Data Sampel Review Aplikasi Bank

3.2 Normalisasi

Berikut hasil dari beberapa proses tokenisasi, penghilangan stopwords, menjadikan huruf kecil dan tanda baca. Berikut sampel daftar normalisasi kata dari *regextokenizer*.

1	aktif,aktif
2	aktifitas,aktivitas
3	Apotik,Apotek
4	apotik,Apotek
5	analisa,analisis
6	azas,asas

Gambar 2. Sampel Normalisasi Kata

Berikut tampilan pembersihan kata yang tidak penting (stopword) dengan sampel stopwords dari *nlTK corpus*:

1	ada
2	adanya
3	adalah
4	adapun
5	agak
6	agakny
7	agak
8	akan
9	akankah
10	akhirnya

Gambar 3. Sampel Stopword

Setelah dilakukan proses normalisasi di atas menghasilkan hasil pada gambar di bawah :

1	,Komentar,Bintang						
2	0,force close zenfone 5,2						
3	1,"kecewa aneh,tiap log diminta kode challenge sungguh merepotkan berada luar negri",1						
4	2,keren,5						
5	3,perbaiki mohon kualitas aplikasi perbaiki,3						

Gambar 4. Hasil Normalisasi

3.3 Stemming

Setelah dilakukan proses normalisasi maka dilakukan stemming untuk penghilangan imbuhan dan tanda baca, sehingga bisa dilihat hasil pada gambar 5 di bawah telah terjadi penghilangan imbuhan dan tanda baca dari hasil proses normalisasi pada Gambar 5. Berikut hasil yang didapatkan setelah dilakukan proses stemming.

Unnamed: 0		Komentar	Bintang
0	0	force close zenfone 5	2
1	1	kecewa aneh tiap log minta kode challenge sungguh repot ada luar negri	1
2	2	keren	5
3	3	baik mohon kualitas aplikasi baik	3
4	4	thankyou much mudah transaksi ribet atm transfer masalah coba langsung cs sajaselamat coba	5

Gambar 5. Hasil Stemming

Dari proses stemming ini telah selesai tahapan preprocessing normalisasi Teks Mining dimana telah terjadi perubahan bentuk kata sampel pada gambar 1 menjadi gambar 5.

3.4 Pengelompokan kata

Pada gambar 6 dan 7 di bawah diperlihatkan telah terjadi pengelompokan sampel data menjadi positif dan negatif.

	Komentar	Bintang
0	keren	pos
1	thankyou much mudah transaksi ribet atm transfer masalah coba langsung cs sajaselamat coba	pos
2	good good	pos
3	muamalat mobile top dahh	pos
4	mantap percaya	pos

Gambar 6. Sampel Komentar Positif

	Komentar	Bintang
0	keren	pos
1	thankyou much mudah transaksi ribet atm transfer masalah coba langsung cs sajaselamat coba	pos
2	good good	pos
3	muamalat mobile top dahh	pos
4	mantap percaya	pos

Gambar 7. Sampel Komentar Negatif

3.5 Data Mining

Hasil percobaan data dari 200 review positif dan 200 review negatif berhasil diterapkan menggunakan modul textblob untuk penerapan algoritma naïve bayes.

```
In [23]: c1.accuracy(test.values)
```

```
Out[23]: 0.8714285714285714
```

Gambar 8. Akurasi Data Training Naïve Bayes

3.6 Evaluasi

Hasil pengujian menggunakan data testing sebanyak 70 review positif dan 70 review negatif didapatkan hasil akurasi sebesar 0,87 atau (87 %).

4. KESIMPULAN

Pengolahan Sentimen analisis penggunaan aplikasi bank menggunakan algoritma naïve bayes menghasilkan keakuratan data sebesar 0.8714 (87 %) dimana pada data training yang digunakan adalah 200 komentar positif dan 200 komentar negatif serta data testing yang digunakan adalah 70 komentar positif dan 70 komentar negatif.

REFERENCES

- [1] "PENERAPAN TEXT MINING PADA SISTEM KLASIFIKASI EMAIL SPAM MENGGUNAKAN NAIVE BAYES," *PENERAPAN TEXT Min. PADA Sist. KLASIFIKASI EMAIL SPAM MENGGUNAKAN NAIVE BAYES*, 2014.
- [2] M. W. A. Putra, Susanti, Erlin, and Herwin, "Analisis Sentimen Dompot Elektronik Pada Twitter Menggunakan Metode Naïve Bayes Classifier," *IT J. Res. Dev.*, 2020.
- [3] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, 2012.
- [4] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in *Proceedings of the 12th International Conference on World Wide Web, WWW 2003*, 2003.
- [5] Y. Lukito and A. R. Chrismanto, "Perbandingan Metode-Metode Klasifikasi untuk Indoor Positioning System," *J. Tek. Inform. dan Sist. Inf.*, 2015.
- [6] D. T. Larose, "Naïve Bayes Estimation and Bayesian Networks," in *Data Mining Methods and Models*, 2006.
- [7] A. Kao and S. R. Poteet, *Natural language processing and text mining*. 2007.
- [8] W. E. Nurjanah, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, 2017.
- [9] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Inform.*, 2018.
- [10] F. Nurhuda, S. Widya Sihwi, and A. Doewes, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *J. Teknol. Inf. ITSmart*, 2016.