

PENERAPAN METODE NAIVE BAYES DALAM ANALISIS SENTIMEN APLIKASI SENTUH TANAHKU PADA GOOGLE PLAY

Nurul Faridhotul Hidayah, Kurnia Paranita Kartika R., Saiful Nur Budiman

Program Studi Teknik Informatika S1, Fakultas Teknologi Informasi
Universitas Islam Balitar, Jalan Raya Karanglo km 2 Blitar, Indonesia
novahidayah05@gmail.com

ABSTRAK

Masa modern seperti saat ini, sudah hampir semua bidang menggunakan teknologi, salah satu contoh Badan Pertanahan Nasional yang telah membuat Aplikasi Sentuh Tanahku, aplikasi tersebut dibuat untuk kepengurusan hal – hal yang berkaitan dengan pertanahan seperti surat tanah, ptsl, dan lain-lain. Aplikasi ini sangat membantu masyarakat, banyak masyarakat yang merasa puas atau sebaliknya. Komentar masyarakat terhadap aplikasi sentuh tanahku ini mereka sampaikan lewat komentar di google play, dengan keadaan tersebut maka diperlukan adanya analisis sentimen untuk mengklasifikasikan komentar yang positif atau komentar negatif, klasifikasi ini menggunakan metode naive bayes karena metode naive bayes ini merupakan metode klasifikasi yang memanfaatkan probabilitas atau peluang, dengan begitu tujuan penelitian ini membahas tentang penerapan metode naive bayes dalam analisis sentimen aplikasi Sentuh Tanahku pada google play. Hasil yang diperoleh adalah 24% sentimen negatif dan 76% sentimen positif, dan didapatkan akurasi 76%, dengan hasil akurasi tersebut metode naive bayes ini cukup layak untuk analisis sentimen aplikasi Sentuh Tanahku ini.

Kata kunci: *Naive Bayes Classifier (NBC), Aplikasi Sentuh Tanahku, Analisis Sentimen.*

1. PENDAHULUAN

Penggunaan teknologi di era modern ini menjadi bahan utama di kehidupan sehari – hari, penggunaan teknologi ini sudah mencakup di berbagai bidang, salah satunya di pemerintahan, sebagai contoh Badan Pertanahan Nasional, sebelum adanya kecanggihan teknologi [1], lembaga ini dulunya melayani, membuat, dan melakukan kegiatan pertanahan secara manual, tetapi seiring berkembangnya zaman, Badan Pertanahan Nasional ini telah memanfaatkan teknologi dengan membuat Aplikasi Sentuh Tanahku [2][3], Aplikasi ini ditunjukkan untuk masyarakat agar mempermudah dalam mengurus hal – hal yang berkaitan dengan pertanahan seperti pembuatan surat tanah.

Adanya Aplikasi Sentuh Tanahku yang telah dibuat oleh pemerintah tidak luput dari respon masyarakat, Aplikasi Sentuh Tanahku banyak menimbulkan pro dan kontra dari masyarakat terlihat dari kurang lebih 16 Ribu ulasan di Google Play masyarakat memberikan respon negatif ataupun positif. Selain itu, Aplikasi Sentuh Tanahku ini sudah mencapai satu juta downloader, tetapi Aplikasi Sentuh Tanahku ini masih mendapat nilai Rating 3,5, dengan keadaan tersebut maka diperlukan analisis sentimen dengan menerapkan metode Naive Bayes untuk mengklasifikasikan komentar yang positif dan negatif [4], sehingga dapat diketahui komentar yang benar – benar negatif dan komentar yang benar – benar positif, dan untuk mengetahui tingkat akurasi Metode Naive Bayes ini. Data yang dihasilkan dapat digunakan Badan Pertanahan Nasional sebagai bahan acuan untuk pengembangan Aplikasi Sentuh Tanahku.

2. TINJAUAN PUSTAKA

2.1. Metode Naive Bayes

Metode Naive Bayes adalah salah satu metode data mining/machine learning yang bertujuan untuk mengelompokan atau mengklasifikasikan suatu data pada beberapa kelas dengan memanfaatkan probabilitas atau peluang dari data tersebut [5]. Berikut rumus dari metode Naive Bayes:

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

Keterangan

$P(H|X)$ = Peluang hipotesa H berdasarkan kondisi X

X = Data sample dengan label yang tidak diketahui

H = Hipotesa bahwa X adalah data dengan label

$P(H)$ = Peluang hipotesa H

$P(X)$ = Peluang X

$P(X|H)$ = Peluang X berdasarkan kondisi pada hipotesa H

2.2. Analisis Sentimen

Analisis Sentimen adalah suatu proses pengolahan kata atau text yang bersumber dari internet atau media sosial dengan tujuan untuk mendapatkan respon atau opini masyarakat terkait media sosial tersebut [6].

2.3. Aplikasi Sentuh Tanahku

Aplikasi Sentuh Tanahku adalah aplikasi yang dibuat oleh Badan Pertanahan Nasional atau agraria sebagai alat untuk mempermudah masyarakat dalam

mengurus hal – hal yang berkaitan dengan pertanahan seperti pembuatan surat tanah maupun PTSL [7].

2.4. Scraping

Scraping adalah kegiatan untuk mengambil data dari suatu internet atau media sosial. Scraping ini dilakukan dengan memanfaatkan suatu opftware atau aplikasi pembantu seperti aplikasi Web Harvy [8].

2.5. Preprocessing Data

Preprocessing Data adalah proses pengolahan data yang bertujuan untuk mengubah atau membersihkan data yang masih bersifat kasar menjadi data yang bersih [9]. Berikut proses preprocessing ;

- a. Case folding
Case folding adalah proses mengubah kata menjadi huruf kecil semua dan membersihkan atau menghapus tanda – tanda baca.
- b. Tokenizing
Tokenizing adalah proses membuat kalimat menjadi data perkata.
- c. Filtering
Filtering adalah proses menghapus kata – kata yang bersifat umum seperti apa, dan, di, dll.
- d. Stemming
Stemming adalah proses mengubah kata yang erhimbunan menjadi kata dasar.

2.6. Perhitungan Confusion Matrix

Confusion matrik adalah salah satu metode pengujian untuk mengetahui data yang benar positif (TP), data yang salah positif (FP), data yang benar negatif (TN), dan terakhir data yang salah negatif (FN). Confusion matrik ini bertujuan untuk mengetahui tingkat atau seberapa besar akurasi, recall, precision, dan f-measure dari suatu metode [10]. Berikut rumusnya :

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

$$Precision = \frac{TP}{(TP + FP)}$$

$$Recall = \frac{TP}{(TP + FN)}$$

$$f - measure = \frac{2 x (Recall * Precision)}{(Recall + Precision)}$$

- a) True Positives: merupakan jumlah data positive yang diklasifikasikan sebagai nilai positive
- b) False Positive : Merupakan jumlah data negative yang diklasifikasikan sebagai nilai salah positive
- c) False Negatives: Merupakan jumlah data positive yang diklasifikasikan sebagai negative
- d) True Negatives: Merupakan jumlah data negative yang diklasifikasikan sebagai nilai negative

3. METODE PENELITIAN

3.1. Waktu dan Lokasi Penelitian

Waktu yang digunakan peneliti untuk melakukan penelitian ini dilaksanakan sejak Bulan Februari sampai Bulan Agustus tahun 2022 mulai waktu pengumpulan data hingga penyusunan laporan akhir kurang lebih tujuh bulan. Lokasi pada penelitian ini adalah Universitas Islam Balitar.

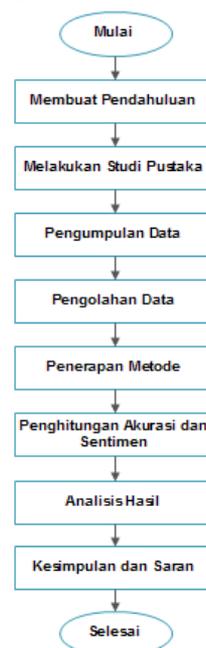
3.2. Jenis Penelitian

Jenis penelitian yang digunakan dalam penelitian ini adalah Jenis penelitian deskriptif kualitatif. Metode deskriptif kualitatif adalah sebuah metode penelitian yang memanfaatkan data kualitatif dan dijabarkan sejarah deskriptif (Sendari, 2019). Artinya dalam penelitian ini mengumpulkan data – data seperti kata – kata atau data kualitatif dan menjabarkan atau mendeskripsikan suatu hasil yang telah diperoleh dari penelitian ini.

3.3. Pengumpulan Data

- a. Jenis Data Studi Dokumentasi
Jenis data yang diperoleh dari pengumpulan data studi dokumentasi berupa dataset, yaitu data yang sudah tersedia dan diambil. Data tersebut adalah data dari ulasan atau komentar dari pengguna Aplikasi Sentuh Tanahku di Google Play yang telah dikumpulkan. Data yang diam
- b. Jenis Data Studi Literatur
Jenis data yang diperoleh dari studi literatur adalah sebagai berikut mulai dari informasi terkait Aplikasi Sentuh Tanahku, langkah-langkah analisis sentimen, cara kerja naive bayes, langkah-langkah preprocessing data, hingga penghitungan nilai akurasi

3.4. Tahap-tahap Penelitian



Gambar 1. Alur Penelitian

Penjelasan sedikit tentang proses penelitian sebagai berikut dalam penelitian ini setelah perumusan masalah dan studi pustaka selanjutnya adalah pengumpulan data ulasan dengan webharving, kemudian data yang disimpan, kemudian masuk dalam pengolahan data, dalam pengolahan data ini software yang digunakan adalah phyton uruta dalam pngolahan data adalah case folding, tokenizing, filtering dan terakhir adalah stemming. Contoh dari setiap proses pengolahan data akan ditampilkan pada poin pengolahan data. Selesai pengolahan langkah selanjutnya adalah penerapan metode naive bayes dengan memilih kata yang di gunakan sebagai probabilitas dibobokan kemudian dihitung untuk menentukan kelas positif atau negatif. Selesai penghitungan maka masuk dalam proses penghitungan akurasi dan sentimen dari Aplikasi Sentuh Tanahku ini. penghitungan akurasi menggunakan confusion setelah itu menghitung sentimen atau besar persentase komentar negatif dan komentar positif dari data ulasan tentang Aplikasi Sentuh Tanahku yang telah diambil. Untuk lebih jelasknya akan dijelaskan per proses penelitian.

4. HASIL DAN PEMBAHASAN

4.1. Pengambilan Data

Hasil yang diperoleh dalam pengambilan data adalah diperoleh 300 data ulasan atau review dari Google Play tentang Aplikasi Sentuh Tanahku. Data yang digunakan dalam penelitian ini adalah ulasan pada Aplikasi Sentuh Tanahku mulai bulan Desember 2021 sampai bulan Juni 2022, selain data ulasan juga diambil waktu ulasan dan rating atau bintang pada ulasan tersebut. Data diambil melalui tahap scraping menggunakan software webharvy

4.2. Case Folding

Pada proses casefolding ini data ulasan diolah dan dirubah yang mulanya terdapat huruf kapital menjadi menjadi huruf kecil semua, berikut hasilnya:

Tabel 1. Contoh Case Folding

Sebelum	Sesudah
Sejak lapor sertifikat menjadi bisa ditampilkan di aplikasi ini berapa lama...?? Tidak ada status proses nya jadi tidak tahu perkembangannya sampai sejauh mana.	sejak lapor sertifikat menjadi bisa ditampilkan di aplikasi ini berapa lama tidak ada status proses nya jadi tidak tahu perkembangannya sampai sejauh mana

4.3. Tokenizing

Pada proses selanjutnya data yang sudah selesai proses casefolding dilanjutkan proses tokenizing, pada proses tokenizing ini data yang awalnya berbentuk kalimat akan diubah menjadi perkata, berikut contoh hasilnya :

Tabel 2. Contoh Tokenizing

Sebelum	Sesudah
sejak lapor sertifikat menjadi bisa ditampilkan di aplikasi ini berapa lama tidak ada status proses nya jadi tidak tahu perkembangannya sampai sejauh mana	sejak,lapor,sertifikat,menjadi,bisa,ditampilkan,di,aplikasi,ini,berapa,lama,tidak,ada,s tatus,proses,nya,jadi,tidak,ta hu,perkembangannya,sampai,i,sejauh,mana

4.4. Filtering

Proses tokenizing selesai masuk ke proses filtering, pada proses filtering ini akan menghapus kata – kata yang bersifat kurang penting atau umum sebagai contoh kata ‘dan’, ‘untuk’ , ‘di’ dll, berikut contoh hasilnya :

Tabel 3. Contoh Filtering

Sebelum	Sesudah
sejak,lapor,sertifikat,menjadi,bisa,ditampilkan,di,aplikasi,ini,berapa,lama,tidak,ada,s tatus,proses,nya,jadi,tidak,ta hu,perkembangannya,sampai,i,sejauh,mana	lapor,sertifikat,ditampilkan, aplikasi,status,proses,nya,pe rkembangannya

4.5. Stemming

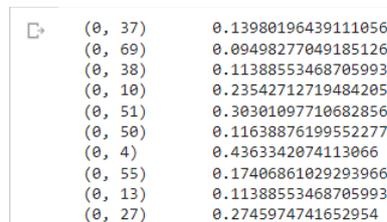
Pada proses stemming kata – kata yang berimbuhan akan diubah menjadi kata dasar dengan menggunakan library Sastrawi yang telah di instal di python, berikut contoh hasilnya :

Tabel 4. Contoh Stemming

Sebelum	Sesudah
lapor,sertifikat,ditampilkan, aplikasi,status,proses,nya,pe rkembangannya	lapor sertifikat tampil aplikasi status proses nya kembang

4.6. Penerapan Metode Naive Bayes

Proses Preprocessing Data selesai langkah selanjutnya penerapan metode, pada proses ini dimulai dengan tahap pembobotan pada setiap ulasan yang telah melewati proses preprocessing data, semakin banyak kata yang berulang maka akan semakin tinggi nilai bobotnya. Berikut hasil yang diperoleh :



Gambar 1. Hasil Pembobotan

Gambar diatas merupakan hasil dari pembobotan kata yang terpilih atau term yang dipilih secara otomatis dari source code yang telah dilakukan sebelumnya. Hasil diatas tertera (0, 37)

yang artinya 0 merupakan ulasan pertama dan 37 merupakan urutan kata di penggabungan, serta 0,13980 merupakan bobot dari kata terpilih atau term tersebut. Selanjutnya adalah pembagian *data training* atau latih dan *data testing* atau uji yang dibagi secara otomatis dan acak, dengan ukuran *data testing* sebesar 0,2 atau 20% dari seluruh data ulasan, sehingga *data training* sebesar 80% dari seluruh data.

Penerapan metode *naive bayes* dilakukan dengan cara pemanggilan library *naive bayes* pada *python*. Penerapan *naive bayes* dibarengi dengan pengujian menggunakan *confusion matrix* yang menghasilkan matrix yang berisi banyak ulasan yang true positive (benar positif), false positive (salah positif), true negatif (Benar negatif) dan false negatif (salah negatif). Berikut hasil *confusion matrix* nya :

```
confusion matrix:
[[32  0]
 [14 14]]
```

Gambar 2. Hasil *Confusion Matrix*

Hasil tersebut akan dilakukan untuk mencari akurasi dari metode *naive bayes*:

Tahap yang selanjutnya adalah perhitungan akurasi dan sentimen, perhitungan akurasi ini untuk pengujian metode *naive bayes* dengan disertai perhitungan *recall*, *precission*, dan *f-measure*. Perhitungan tersebut diterapkan dengan pemanggilan library yang ada pada *python* berikut hasil yang diperoleh :

```
MultinomialNB Accuracy : 0.7666666666666667
MultinomialNB Precision: 0.6956521739130435
MultinomialNB Recall : 1.0
MultinomialNB f1_Score : 0.8205128205128205
```

Gambar 3. Hasil Akurasi

Pada gambar diatas menunjukkan , hasil yang diperoleh pada pengujian ini nilai akurasi nya 76%, *precision* nya 69%, *recall*nya 100% dan *f-measure* nya 82%. Akurasi sebesar76% ini menunjukkan tiga per empat data baik digunakan untuk pembuatan model sudah cukup akurat.

Pada perhitungan sentimen yang dilakukan adalah menghitung berapa persen negatif dan berapa persen positif, hasil yang diperoleh adalah 24% sentimen negatif dan 76% sentimen positif.

4.7. Pembahasan

Penelitian yang dilakukan ini menghitung *accuracy*, *recal*, *precission* dan *f-measure* dari metode *naive bayes*, ada beberapa proses yang dilakukan sebelum masuk dalam tahap perhitungan – perhitungan tersebut yaitu pembobotan dengan *TF-IDF (Term Frequency Inverse Document Frequency)*, dalam proses pembobotan ini menggunakan *function* yang ada di *python*, *function* tersebut akan memilih secara otomatis kata yang digunakan sebagai *term*

karena dalam hasil *running source code* tidak memunculkan kata apa saja yang dijadikan sebagai *term* dan dibobotkan, , langkah berikutnya adalah membagi data *training* dan *data testing*, pembagian ini dilakukan secara otomatis oleh program memanggil *function train_test_split*.

Data *testing* terambil sebanyak 60 data atau 20% dari semua data. Setelah pembagian data *training* dan data *testing* selesai selanjutnya adalah penerapan metode *naive bayes*, dalam penerapannya dilakukan dengan memanggil *function* yang ada dalam *python* dari modul *sklearn* , data yang diklasifikasikan adalah data *testing* yang sudah dibagi secara otomatis sebelumnya dengan menggunakan model dari data *training* yang sudah terbentuk dalam metode *naive bayes* sebelumnya. Langkah selanjutnya adalah pengujian dengan *confusion matrix* yang menghitung *accuracy*, *recall*, *precission*, dan *f-measure*

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan penelitian tentang penerapan metode *naive bayes* dalam analisis sentimen Aplikasi Sentuh Tanahku pada Google Play, diperoleh beberapa kesimpulan yaitu: Metode *naive bayes* dapat diterapkan dalam analisis sentimen Aplikasi Sentuh Tanahku ini, hasil yang diperoleh dari 60 data yang digunakan untuk testing maka diperoleh ulasan positif sebanyak 46 dan 14 ulasan negatif sehingga hasil presentase nya adalah 76% ulasan positif dan 24% ulasan negatif. Pengujian metode *naive bayes* dalam analisis sentimen Aplikasi Sentuh Tanahku ini menggunakan *confusion matrix*, hasil yang diperoleh adalah *accuracy* sebesar 76%, *precission* sebesar 69%, *recall* sebesar 100% dan *f - measure* sebesar 82%, dengan hasil akurasi tersebut metode *naive bayes* ini dikatakan cukup layak diterapkan analisis sentimen aplikasi Sentuh Tanahku.

Dalam pengembangan penerapan metode *naive bayes* untuk analisis sentimen Aplikasi Sentuh Tanahku pada google play ini diperlukan beberapa perbaikan untuk mencapai hasil yang maksimal, yakni: Perlu memilih dataset dari google play yang ratingnya sesuai dengan ulasannya. Pemilihan *function* dalam *tf-idf* perlu diperhatikan dan dikembangkan akan dapat memunculkan term yang terpilih, Salah satu kelemahan dari penelitian ini adalah tidak dapat memunculkan term yang terpilih , dan pembagian data *training* dan *testing* yang tidak ditampilkan, Penggunaan metode *naive bayes* dapat dikembangkan untuk analisis sentimen dengan dataset lain, dan dapat dibuat untuk perbandingan dengan metode lain.

DAFTAR PUSTAKA

[1] Firdaus, R.F. and Gunadhi, E., 2016. ANALISIS TATA KELOLA TEKNOLOGI INFORMASI DI KANTOR BADAN PERTANAHAN NASIONAL KABUPATEN GARUT

- MENGGUNAKAN FRAMEWORK COBIT 4.1. *Jurnal Algoritma*, 13(2), pp.367-373.
- [2] Putri, Y.A., Putera, R.E. and Rahayu, W.K., 2022. Inovasi Pelayanan Informasi melalui Aplikasi Sentuh Tanahku pada Kantor Pertanahan Kota Administrasi Jakarta Utara. *Journal of Social and Policy Issues*, pp.86-94.
- [3] Rachmad, D.S., 2020. EVALUASI USABILITY APLIKASI SENTUH TANAHKU MENGGUNAKAN METODE HEURISTIC. *IJIS-Indonesian Journal On Information System*, 5(1).
- [4] Tanggraeni, A.I. and Sitokdana, M.N., 2022. Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(2), pp.785-795.
- [5] Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*, 207 - 217.
- [6] Annisa, T. (2021, November 1). *Mengenal peran sentiment analysis beserta cara kerjanya : Sentiment analysis dan peran pentingnya*. Dipetik Maret 1, 2022, dari Mengenal peran sentiment analysis beserta cara kerjanya: <https://www.ekrut.com/media/sentiment-analysis-adalah>
- [7] Ruhlesin, M. F. (2021, Desember 1). *Sentuh Tanahku, Aplikasi untuk Atasi Masalah Pertanahan*. Dipetik Maret 1, 2022, dari Aplikasi untuk Atasi Masalah Pertanahan: <https://www.kompas.com>
- [8] Nayoan, A. (2020, Januari 13). *Apa itu Web Scraping? Pengertian, Teknik, dan Manfaatnya : Teknik teknik web scraping*. Dipetik Maret 28, 2022, dari Apa itu Web Scraping? Pengertian, Teknik, dan Manfaatnya: <https://www.niagahoster.co.id>
- [9] Oliver, A. (2021, Oktober 31). *Bikin Data Lebih Mudah Dibaca, yuk, Kenalan dengan Data Preprocessing*. Dipetik April 9, 2022, dari Bikin Data Lebih Mudah Dibaca, yuk, Kenalan dengan Data Preprocessing : <https://glints.com>
- [10] Putra, M. L. (2020, Juli 15). *Menghitung Akurasi Dengan Confusion Matrix : Apa itu Confusion Matrix*. Dipetik April 11, 2022, dari Menghitung Akurasi Dengan Confusion Matrix : <https://laugiwa.medium.com>