

Klasifikasi Ulasan Fasilitas Publik Menggunakan Metode Naïve Bayes dengan Seleksi Fitur *Chi-Square*

Adhitya Prayoga Permana ⁽¹⁾, Totok Chamidy ^{(2)*}, Cahyo Crysdian ⁽³⁾
Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim, Malang
e-mail : {adhityaprayoga48,totokchamidy}@gmail.com, cahyo@ti.uin-malang.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 26 Desember 2022, direvisi 4 April 2023, diterima 5 April 2023, dan dipublikasikan 26 Mei 2023.

Abstract

Government builds public facilities to support the needs of the community. The use of these public facilities needs to be re-evaluated, and one way to do it is through community response. Google Maps is one platform that receives the most responses from the community about location. Google Maps Reviews allow us to see how the public reacts to a location. Naïve Bayes method is used for classification in this study because it is one of the simple methods in machine learning that can be easily applied to several experiments conducted by the author. In the classification process, reviews produce many features that will be calculated based on their class. More features generated, more features processed too in the system. Chi-Square feature selection will be used to reduce features that have low dependence on the system. In this study, performance values will be calculated based on the experimental use of feature ratios of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100%. The results show that the use of 10% Chi-Square features produces the best performance, with an accuracy rate of 86.94%, precision of 80.42%, recall of 80.42%, and f-measure of 80.42%.

Keywords: Sentiment Analysis, Public Facility, Google Maps Reviews, Naïve Bayes, Chi-Square Feature Selection

Abstrak

Pemerintah membangun fasilitas publik untuk mendukung kebutuhan masyarakat. Penggunaan fasilitas publik ini perlu ditinjau kembali di mana salah satu caranya adalah melalui respon masyarakat. Google Maps merupakan salah satu platform yang paling banyak menerima respon dari masyarakat tentang suatu tempat atau lokasi. Ulasan Google Maps memungkinkan untuk melihat bagaimana reaksi masyarakat terhadap suatu lokasi. Metode Naïve Bayes digunakan untuk melakukan klasifikasi dalam penelitian karena termasuk salah satu metode dalam *machine learning* yang sederhana sehingga mudah diterapkan pada beberapa percobaan yang dilakukan penulis. Dalam proses klasifikasi, ulasan-ulasan tersebut menghasilkan banyak fitur dan akan dihitung berdasarkan kelasnya. Semakin banyak fitur yang dihasilkan maka semakin banyak pula fitur yang diproses dalam sistem. Seleksi fitur *Chi-Square* akan digunakan untuk mereduksi fitur-fitur yang memiliki ketergantungan rendah dalam sistem. Dalam penelitian ini akan menghitung nilai performa berdasarkan uji coba penggunaan rasio fitur sebanyak 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, dan 100%. Hasilnya penggunaan 10% fitur *chi-square* menghasilkan performa terbaik yaitu dengan nilai akurasi 86,94%, presisi 80,42%, *recall* 80,42% dan *f-measure* 80,42%.

Kata Kunci: Analisis Sentimen, Fasilitas Publik, Ulasan Google Maps, Naïve Bayes, Seleksi Fitur *Chi-Square*

1. PENDAHULUAN

Fasilitas publik atau fasilitas umum adalah sarana dan prasarana yang disediakan oleh pemerintah yang digunakan untuk kepentingan bersama dalam melaksanakan kegiatan sehari-hari. Dengan adanya fasilitas-fasilitas publik yang telah dibangun oleh pemerintah daerah, diharapkan masyarakat pada suatu daerah tersebut dapat memanfaatkan fasilitas tersebut dengan semestinya.



Fasilitas publik ini perlu ditinjau kembali penggunaannya karena fasilitas publik yang pembangunannya tidak sesuai dengan kebutuhan masyarakat tidak akan membantu produktivitas. Untuk menghindari adanya fasilitas-fasilitas yang tidak sesuai dengan tujuan pembuatan ini, maka perlu adanya evaluasi dari pemerintah. Evaluasi penggunaan fasilitas publik dapat dilakukan melalui respon masyarakat. Masyarakat biasa memberikan responnya melalui media sosial, di mana salah satu platform yang banyak mendapat respon masyarakat apabila berkaitan dengan suatu tempat atau lokasi adalah Google Maps. Google Maps sebenarnya merupakan aplikasi peta dunia milik perusahaan Google yang memiliki fitur ulasan sehingga tempat-tempat yang tertera pada peta yang ditampilkan dapat diberikan ulasan berdasarkan pengalaman orang-orang yang pernah mengunjungi tempat tersebut. Dari fitur ulasan Google Maps tersebut maka akan diketahui bagaimana respon masyarakat terhadap suatu tempat tersebut.

Ulasan-ulasan yang diberikan masyarakat melalui platform Google Maps tersebut terdiri dari ulasan yang bersifat mendukung adanya fasilitas dan menentang adanya fasilitas. Jumlah ulasan tersebut semakin lama akan meningkat secara signifikan sehingga untuk mengetahui sebuah ulasan tergolong ulasan positif atau negatif akan membutuhkan banyak waktu apabila dilakukan secara manual. Untuk mengatasi hal tersebut akan dikembangkan sistem otomatis untuk mengklasifikasikan ulasan tersebut menggunakan metode dalam *machine learning*.

Analisis sentimen, juga disebut klasifikasi opini, adalah bidang studi yang menganalisis pendapat orang, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas seperti produk, layanan, organisasi, individu, isu, peristiwa, topik, dan atributnya (B. Liu, 2012).

Linawati et al. (2020) melakukan penelitian tentang perbandingan penggunaan algoritma Naïve Bayes dan Support Vector Machine untuk penentuan pemberian beasiswa PPA. Data yang digunakan berasal dari data pemberian beasiswa PPA berjumlah 122 data dengan 5 variabel untuk menentukan kelas diterima atau ditolak. Variabelnya terdiri dari semester, pekerjaan, orang tua, penghasilan, dan IPK. Pengujian dilakukan menggunakan aplikasi WEKA 3.8 dan hasilnya algoritma Naïve Bayes memiliki tingkat akurasi lebih tinggi dengan nilai 90,90% sedangkan metode SVM memiliki nilai akurasi 89,25%.

Dulhare (2018) dalam penelitiannya membuat sistem prediksi penyakit jantung menggunakan Naïve Bayes dan Particle Swarm Optimization (PSO). Data yang digunakan merupakan data *stat log (heart)* dari UCI *dataset* yang berjumlah 270 data dengan 14 fitur dan dibagi menjadi dua kelas yaitu *presence* dan *absence*. Pengujian 7 dilakukan dengan sepuluh kali iterasi dengan 100 persen sampai dengan 90 persen seleksi fitur PSO dengan akurasi tertinggi 89,71 persen pada iterasi dengan fitur 70 persen dan 100 persen. Harahap et al. (2018) juga melakukan penelitian mengenai implementasi metode Naïve Bayes untuk memprediksi pembelian. Hasilnya 15 dari 20 data berhasil memprediksi dengan benar yang diujikan berdasar data latih dan nilai akurasinya sebesar 75%.

Ruz et al. (2020) dalam penelitiannya menganalisis data *tweet* dari Twitter selama terjadi peristiwa penting seperti bencana alam maupun pergerakan sosial. Data yang digunakan merupakan *dataset* dalam bahasa Spanyol pada saat peristiwa gempa di Chile tahun 2010 yang berjumlah 2187 *tweet* serta peristiwa *Catalan Independence Referendum* tahun 2017 yang berjumlah 60000 *tweet*. Data tersebut dibagi menjadi dua kelas yaitu positif dan negatif. Metode *bayesian network* digunakan untuk menyelesaikan permasalahan ini. Mereka juga menggunakan *Synthetic Minority Over-sampling Technique* pada data *training* untuk menangani masalah ketidakseimbangan kelas. Hasilnya pada *dataset* gempa di Chile didapatkan nilai akurasi sebesar $0,742 \pm 0,027$, presisi $0,895 \pm 0,004$, *recall* $0,790 \pm 0,009$, *F1-score* $0,841 \pm 0,020$, *Nº Edges* 0. Sedangkan pada *dataset* kedua didapatkan nilai akurasi sebesar $0,781 \pm 0,013$, presisi $0,885 \pm 0,005$, *recall* $0,852 \pm 0,004$, *F1-score* $0,868 \pm 0,000$, *Nº Edges* 0.

Banyak metode dalam *machine learning* yang dapat digunakan untuk melakukan analisis sentimen. Berdasar beberapa penelitian yang telah dilakukan sebelumnya, penulis memutuskan



untuk menggunakan metode Naïve Bayes. Metode Naïve Bayes merupakan metode statistik klasifikasi sederhana berdasarkan probabilitas. Klasifikasi Naïve Bayes mengasumsikan independensi kondisi kelas yaitu nilai atribut pada kelas tertentu tidak tergantung pada nilai atribut lainnya di mana hal ini digunakan untuk menyederhanakan perhitungan yang terlibat (Han & Kamber, 2001). Dalam Naïve Bayes, hipotesis langsung dibentuk tanpa proses pencarian, hanya dengan menghitung frekuensi kemunculan suatu kata pada data training, sedangkan pada metode pembelajaran lainnya pencarian hipotesis biasanya dilakukan dari ruang hipotesis (Aziz et al., 2021).

Analisis sentimen dalam proses klasifikasinya akan membagi dokumen ke dalam kelas dengan memisahkan fitur-fitur dari setiap dokumen untuk dihitung nilai fitur berdasarkan kelasnya. Banyaknya fitur yang dihasilkan tentu akan membebani jalannya sistem komputasi. Penggunaan seleksi fitur *chi-square* digunakan untuk mengurangi dimensi data inputan sehingga tidak membebani sistem. Seleksi fitur *chi-square* akan mengurangi fitur yang tidak memiliki ketergantungan terhadap kelas label sehingga diharapkan dapat meningkatkan performa dari analisis sentimen menggunakan metode Naïve Bayes.

Pratama et al. (2018) melakukan analisis sentimen dengan seleksi fitur *chi-square* terhadap ulasan konsumen untuk rekomendasi lokasi makanan tradisional. Metode yang digunakan dalam penelitian ini merupakan metode Naïve Bayes. Data yang diperoleh merupakan data ulasan konsumen makanan tradisional di Kota Malang dari website tripadvisor.co.id dan diklasifikasikan menjadi dua kelas yaitu positif dan negatif. Hasilnya dilakukan seleksi fitur sebesar 25%, 50%, 75%, 100%, di mana seleksi fitur 25% mendapatkan nilai akurasi tertinggi yaitu 81% serta seleksi fitur 75% mendapatkan nilai terendah dengan akurasi 77%.

Irvantoro (2019) melakukan seleksi fitur menggunakan *chi-square* dan *n-gram* untuk menganalisis ulasan produk elektronik. Metode yang digunakan menggunakan metode Naïve Bayes. Data yang digunakan didapat dari *library RStudio* sejumlah 500 data yang dibagi menjadi dua kelas positif dan negatif dengan jumlah yang sama dan diambil 100 data uji secara acak dari data tersebut. Pengujian dilakukan 4 kali dengan 25% fitur, 50% fitur, 75% fitur dan 100% fitur. Prediksi dilakukan menggunakan *N-gram tokenizer* di mana terdapat dua model yaitu *uni-gram* yang menggunakan satu per satu kata dan *bi-gram* yang menggunakan dua kata terkait. Hasilnya pada model *unigram* mendapat nilai tertinggi dengan 89% akurasi, 89% presisi, dan 100% *recall* pada seleksi fitur 75%. Sedangkan model *bi-gram* mendapatkan nilai tertinggi dengan 89% akurasi, 89% presisi, dan 100% *recall* pada seleksi fitur 25%.

Xu et al. (2019) menganalisa data komentar berbasis teks menggunakan metode BiLSTM. Data yang digunakan merupakan data komentar hotel yang didapat dari website ctrip.com yang berjumlah 15.000 teks komentar. Pada penelitian ini juga dilakukan ekstraksi fitur yang mana mengintegrasikan kontribusi informasi sentimen kedalam TF-IDF. Hasil dari integrasi ini didapatkan nilai presisi sebesar 91.50, *recall* 92.87, dan *F1-score* 92.18 yang mana secara keseluruhan lebih baik dari seluruh ekstraksi fitur yang dicoba. Sedangkan hasil akhir metode BiLSTM yang digunakan pada penelitian ini didapatkan nilai terbaik dari keseluruhan metode yang dicoba dengan nilai presisi sebesar 91.54, *recall* 92.82, dan *F1-score* 92.18

Singh et al. (2019) dalam penelitiannya membandingkan penggunaan *multinomial* Naïve Bayes dengan *multivariate* Bernoulli Naïve Bayes. Data yang digunakan berjumlah 312 data polaritas berita. Hasilnya *multinomial* Naïve Bayes memiliki nilai akurasi 73% sedangkan *multivariate* Bernoulli Naïve Bayes memiliki nilai akurasi 69%.

Melihat penelitian-penelitian yang telah dilakukan sebelumnya, penulis memutuskan untuk melakukan klasifikasi ulasan fasilitas publik menggunakan seleksi fitur *chi-square* dengan metode Naïve Bayes. Dalam penelitian ini, akan dilakukan uji coba penggunaan rasio fitur sebanyak 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, dan 100% untuk menghitung nilai performa dalam sistem. Perbedaan rasio fitur ini digunakan untuk mengetahui pengaruh penggunaan seleksi fitur



chi-square pada klasifikasi ulasan fasilitas publik serta menganalisis faktor-faktor apa saja yang memengaruhi nilai performa dalam sistem.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data yang diperoleh dari ulasan pada platform Google Maps. Penulis melakukan pengambilan data menggunakan fitur pada *website* <https://apify.com/compass/google-maps-reviews-scrapert> terhadap data ulasan tempat-tempat fasilitas publik di Kota Malang yang dengan total jumlah data sebanyak 3000 data ulasan yang didapat dari 44 entitas fasilitas publik. Data yang digunakan cukup besar dan diharapkan dapat memberikan hasil yang cukup akurat dan representatif dalam penelitian. Data ulasan yang dipilih hanya data ulasan yang berupa teks sehingga data ulasan yang hanya menampilkan *rating* dan atau foto saja tidak akan digunakan. Ulasan akan diklasifikasikan menjadi 3 kelas yaitu mendukung (D), netral (N), dan tidak mendukung (TD). Ulasan yang termasuk kelas mendukung (D) merupakan ulasan positif dan pro terhadap fasilitas publik. Sebaliknya, ulasan yang termasuk kelas tidak mendukung (TD) merupakan ulasan negatif dan kontra terhadap fasilitas publik. Sedangkan ulasan yang termasuk kelas netral (N) merupakan ulasan yang tidak termasuk kategori dua kelas sebelumnya.

2.2 Desain Sistem

Untuk tercapainya penelitian, peneliti membangun sistem menggunakan bahasa pemrograman Python. Data ulasan fasilitas publik yang telah didapatkan akan diproses melalui beberapa langkah yaitu *preprocessing*, *bag-of-words*, seleksi fitur *chi-square*, Naïve Bayes sehingga hasilnya ulasan-ulasan tersebut akan terklasifikasi berdasarkan kelas dalam label.

2.3 Preprocessing

Ulasan fasilitas publik dari pengumpulan data perlu melalui tahapan *preprocessing* sebelum digunakan dalam proses selanjutnya karena dalam tahapan *preprocessing* dapat memperbaiki struktur inputan data. Proses yang dibutuhkan dalam tahapan *preprocessing* disesuaikan berdasarkan karakteristik dari data inputan. Dalam penelitian ini, proses yang akan dilakukan adalah *cleaning*, *lowercase*, *tokenizing*, *stopword removal*, dan *stemming*.

Sentimen perlu melewati tahapan *preprocessing* di mana pada tahap ini data sentimen akan diolah terlebih dahulu sebelum data diproses dalam klasifikasi atau sebelum diproses dalam algoritma (Faisal et al., 2020).

Cleaning akan menghapus karakter selain huruf yang ada pada data inputan sehingga data inputan hanya akan tersisa data berupa teks. Setelah itu teks perlu diseragamkan untuk menghindari *case sensitive*. Sehingga untuk menyeragamkan data, seluruh teks akan diubah menjadi huruf kecil dalam proses *lowercase*.

Data ulasan yang masih berupa kalimat dipecah menjadi satuan kata atau token, dan dilanjutkan pada proses *stemming*. Token kata diubah kembali ke dalam bentuk kata dasar tanpa imbuhan apapun. Segala jenis prefiks maupun sufiks yang ada dalam token dihilangkan. Terakhir adalah proses *stopword removal*. Pada proses ini kata-kata yang cenderung tidak penting dan tidak ada hubungannya dengan dokumen seperti konjungsi dan preposisi akan dihilangkan.

2.3.1 Bag-of-Words

Bag-of-words adalah representasi kemunculan kata atau *term* terhadap sebuah dokumen. Dalam *bag-of-words* dihitung frekuensi kemunculan kata dalam setiap kategori dokumen. *Bag-of-words* adalah teknik sederhana namun efektif untuk merepresentasikan teks sebagai vektor jumlah kata. Teknik ini melibatkan dua langkah: tokenisasi dan penghitungan. Tokenisasi membagi teks



menjadi kata-kata dan menghapus tanda baca dan karakter non-kata lainnya. Penghitungan menghitung berapa kali setiap kata muncul dalam teks. Vektor yang dihasilkan dapat digunakan sebagai masukan untuk berbagai algoritma pembelajaran mesin seperti Naive Bayes dan regresi logistik (Jurafsky & Martin, 2023). Tahapan ini perlu dilakukan sebab data yang telah diperoleh akan dilakukan transformasi data untuk mentransformasikan atribut-atribut pada variable ke dalam bentuk yang dapat diolah pada tahap selanjutnya (Baskoro et al., 2022).

2.3.2 Seleksi Fitur *Chi-Square*

Proses seleksi fitur diharapkan dapat mengurangi jumlah *noise* dan mengeliminasi fitur yang kurang relevan (Putri, 2018). Fitur yang nilainya rendah dianggap tidak banyak memengaruhi kinerja sistem sehingga bisa mengurangi beban sistem. *Chi-square* dipilih karena dapat digunakan untuk mengetahui tingkat independensi sebuah *term* atau fitur terhadap kelas label. *Chi-square* adalah alat yang berguna dan andal untuk diskritisasi dan pemilihan fitur atribut numerik (Huan Liu & Setiono, 1995). Bobot *chi-square* tiap kelas dapat dihitung menggunakan Pers. (1).

$$x^2(t, c) = \frac{N(AD-CB)^2}{(A+C)(B+D)(A+B)(C+D)} \quad (1)$$

Di mana x^2 merupakan nilai *chi-square* fitur t dalam kelas c , dengan N merupakan jumlah keseluruhan dokumen. Nilai A didapat dari probabilitas jumlah dokumen pada kelas c yang mengandung fitur t , nilai B merupakan probabilitas jumlah dokumen selain pada kelas c yang mengandung fitur t , nilai C didapat dari probabilitas jumlah dokumen pada kelas c yang tidak mengandung fitur t , dan nilai D merupakan probabilitas jumlah dokumen selain pada kelas c yang tidak mengandung fitur t . Kemudian nilai total *chi-square term* dihitung menggunakan Pers. (2).

$$X^2(t) = \sum_{c=1}^k x^2(t, c) \quad (2)$$

Di mana X^2 merupakan nilai *chi-square* fitur t yang didapatkan dari penjumlahan $x^2(t, c)$ *chi-square term* sebanyak c kelas pertama sampai dengan k jumlah keseluruhan kelas pada label.

Setelah nilai *chi-Square* masing-masing *term* diketahui, maka *term* tersebut akan dilakukan pemeringkatan secara *descending* dari nilai *chi-square* dari *term* yang tertinggi. *Term* yang memiliki nilai *chi-square* tinggi merupakan *term* yang memiliki ketergantungan tinggi terhadap kelas label dan berpengaruh dalam sistem,

2.3.3 Naïve Bayes

Klasifikasi Naïve Bayes yang merupakan metode statistik yang dihitung menggunakan konsep probabilitas. Klasifikasi Naïve Bayes mengasumsikan bahwa efek dari nilai atribut pada kelas tertentu tidak tergantung pada kelas lainnya. Nilai atribut tersebut kemudian dicari nilai probabilitas *prior*, probabilitas atribut tiap kelas, dan hasilnya nilai probabilitas yang paling tinggi merupakan hasil dari klasifikasi Naïve Bayes. Nilai probabilitas *prior* dihitung menggunakan Pers. (3).

$$P(c) = \frac{Nc}{N} \quad (3)$$

Probabilitas *prior* $P(c)$ dihitung dari Nc yang merupakan jumlah dokumen pada kelas c dibagi dengan N jumlah keseluruhan dokumen. Selanjutnya adalah menghitung nilai probabilitas *likelihood* kelas menggunakan Pers. (4).

$$P(t|c) = \frac{freq(t,c) + 1}{freq(c) + V} \quad (4)$$

Dengan $P(t|c)$ merupakan probabilitas *term* t pada kelas c , $freq(t,c)$ merupakan jumlah frekuensi *term* t pada kelas c , $freq(c)$ merupakan jumlah *term* yang ada pada kelas c , dan V



merupakan jumlah keseluruhan *term* pada data latih. Kemudian dihitung probabilitas *posterior term* menggunakan Pers. (5).

$$P(c|t) = \frac{P(t|c)P(c)}{P(t)} \quad (5)$$

Dengan $P(c|t)$ merupakan probabilitas kelas c untuk *term* t , $P(t|c)$ merupakan probabilitas *term* t pada kelas c , $P(c)$ merupakan probabilitas *prior* kelas dan $P(t)$ merupakan probabilitas *prior* prediktor atau probabilitas jumlah dokumen yang terdapat *term* t terhadap keseluruhan dokumen.

Tetapi, hasil perhitungan $P(t)$ berbanding terbalik terhadap $P(c|t)$ serta nilainya selalu tetap saat nilai hasil $P(c|t)$ dibandingkan untuk mencari nilai kelas tertinggi. Sehingga $P(t)$ nilainya dapat dianggap 1 dan hanya menyisakan Pers. (6).

$$P(c|t) = P(t|c).P(c) \quad (6)$$

Metode Naïve Bayes yang digunakan dalam penelitian ini merupakan multinomial Naïve Bayes. Dalam multinomial Naïve Bayes digunakan distribusi multinomial untuk menghitung likelihood, di mana hal ini bekerja dengan baik untuk data yang merepresentasikan frekuensi atau jumlah, misalnya jumlah kemunculan kata pada dokumen (Russell & Norvig, 1995). Untuk menghitung probabilitas posterior dokumen, didapatkan dengan Pers. (7).

$$P(c|u) = \prod_{i=1}^n P(t_i|c) \cdot P(c) \quad (7)$$

Berdasar pada Pers. (6), $P(c|t)$ diubah menjadi $P(c|u)$ yang merupakan probabilitas kelas c untuk dokumen u , $P(t|c)$ diubah menjadi $\prod_{i=1}^n P(t_i|c)$ yang merupakan probabilitas *term* t pertama sampai *term* ke- n pada dokumen terhadap kelas c , probabilitas *prior* tetap.

Dalam pengklasifikasian Naïve Bayes kita perlu mendapatkan nilai tertinggi dari nilai probabilitas setiap kelas, yang dinyatakan sebagai *Hypothesis Maximum A Posteriori* (HMAP) dengan Pers. (8).

$$HMAP = \arg \max (P(c|u)) \quad (8)$$

2.4 Skenario Uji Coba

Data yang telah dikumpulkan akan dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Data *training* akan diimplementasikan ke dalam sistem yang telah dibuat untuk membangun model klasifikasi dokumen Naïve Bayes. Sedangkan data *testing* digunakan untuk melakukan pengujian kinerja pada sistem. Ulasan-ulasan yang ada akan dibagi menjadi data *testing* dan data *training* dengan rasio perbandingan 8:2 berdasarkan *Pareto Principle* (Sanders, 1987). *Stratified sampling* digunakan untuk membagi data berdasar jumlah label yang mendukung (D), netral (N), dan tidak mendukung (TD) sehingga perbandingan data *training* dan data *testing* terhadap rasio label tetap sama. Uji coba dilakukan berdasarkan seleksi fitur *chi-square* dengan menggunakan skenario variasi rasio penggunaan fitur sebanyak 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, serta 100% di mana penggunaan 100% fitur juga dianggap sebagai pengujian tanpa seleksi fitur. Hasil klasifikasi akan dibandingkan apakah sama dengan label (*ground truth*). Dan dicari nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN) dan akan dijumlahkan untuk menghitung nilai performa dari sistem menggunakan Pers. (9) sampai (12).

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (9)$$

$$Presisi = \frac{TP}{TP+FP} \quad (10)$$



$$Recall = \frac{TP}{TP+FN} \quad (11)$$

$$F - measure = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \quad (12)$$

3. HASIL DAN PEMBAHASAN

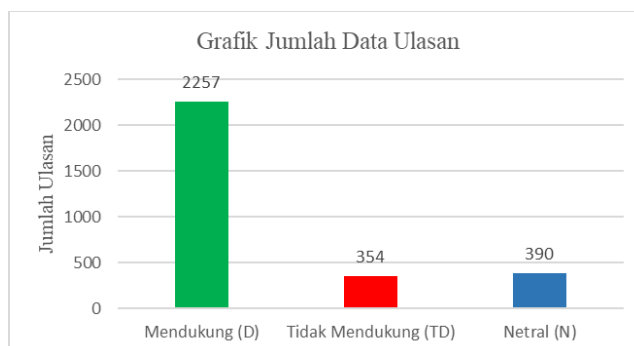
3.1 Data Understanding

Tabel 1 Entitas Penelitian

No.	Entitas	Jumlah Ulasan
1	Taman Merjosari	300
2	Rumah Sakit Saiful Anwar	97
3	Rumah Sakit Lavalette	136
4	RSUD Kota Malang	67
5	Bandara Abdurrahman Saleh	100
6	Terminal Arjosari	93
7	Stasiun Malang Kotabaru	107
8	Vihara Vajra Earth Kertanegara	4
9	Pura Luhur Dwijawarsa	22
10	Klenteng Eng An Kiong	26
11	Gereja Cathedral Ijen	86
12	Masjid Sabilillah	162
13	MAN 1 Kota Malang	13
14	MAN 2 Kota Malang	28
15	SMA N 1 Kota Malang	9
16	SMA N 3 Kota Malang	9
17	SMA N 4 Kota Malang	7
18	SMK N 2 Kota Malang	43
19	SMK N 3 Kota Malang	26
20	SMK N 4 Kota Malang	66
21	SMK Telkom	41
22	SMAK Kolese Santo Yusup	58
23	Lapangan Rampal	131
24	Stadion Gajayana	169
25	Perpustakaan Kota Malang	63
26	Perpustakaan Universitas Negeri Malang	57
27	Perpustakaan Universitas Muhammadiyah Malang	11
28	Perpustakaan UIN Maulana Malik Ibrahim Malang	21
29	Perpustakaan Universitas Brawijaya	26
30	Pasar Buku Wilis	113
31	Toko Buku Togamas	9
32	Pasar Besar Kota Malang	300
33	Mall Olympic Garden	171
34	Malang Town Square	129
35	Indigo Space	20
36	Ruang Perintis	19
37	East Java Super Corridor Malang	92
38	Ngalup Coworking Space	53
39	MOA Coworking Space	5
40	Seikopi Coworking Space	9
41	Malang Digital Core	30
42	EZO Coworking Space	3
43	Orbits Coworking	6
44	Alfath Coworking Space	63



Data ulasan fasilitas publik di Kota Malang yang berjumlah 3000 data ulasan yang berasal dari 44 entitas fasilitas publik. Berikut Tabel 1 merupakan tabel entitas dan Gambar 1 merupakan grafik *ground truth* data ulasan yang telah dilabeli.



Gambar 1 Grafik Jumlah Data Ulasan

Berdasarkan data tersebut didapatkan 2415 data *training* dan 585 data *testing* yang akan digunakan untuk dalam sistem.

3.2 Preprocessing dan Bag-of-Words

Data ulasan yang telah dilabeli dan disimpan dalam format csv diinputkan ke dalam sistem dalam bentuk *dataframe* dan akan diambil kolom yang berisi ulasan untuk masuk ke tahap *preprocessing* di antaranya *cleaning*, *lowercase*, *tokenizing*, *stopword removal*, dan *stemming*. Tabel 2 merupakan contoh hasil dari tahap *preprocessing*.

Tabel 2 Tahap *Preprocessing*

Keterangan	Hasil
Ulasan	Untuk tempatnya strategis sebenarnya, tapi sangat kurang perawatan menurut saya. Seandainya dirawat dengan baik pasti lebih oke. Skybikenya juga mangkrak.
<i>Cleaning</i>	Untuk tempatnya strategis sebenarnya tapi sangat kurang perawatan menurut saya Seandainya dirawat dengan baik pasti lebih oke Skybikenya juga mangkrak
<i>Lowercase</i>	untuk tempatnya strategis sebenarnya tapi sangat kurang perawatan menurut saya seandainya dirawat dengan baik pasti lebih oke skybikenya juga mangkrak
<i>Tokenizing</i>	"untuk", "tempatnya", "strategis", "sebenarnya", "tapi", "sangat", "kurang", "perawatan", "menurut", "saya", "seandainya", "dirawat", "dengan", "baik", "pasti", "lebih", "oke", "skybikenya", "juga", "mangkrak"
<i>Stemming</i>	untuk tempat strategis benar tapi sangat kurang awat turut saya anda rawat dengan baik pasti lebih oke skybikenya juga mangkrak
<i>Stopword Removal</i>	untuk tempat strategis benar tapi kurang awat turut anda rawat baik pasti lebih oke skybikenya mangkrak

Proses selanjutnya adalah *bag-of-word*. Data ulasan yang telah melewati tahap *preprocessing* akan dihitung frekuensi kemunculan kata unik atau *unique term* dalam tiap ulasan. Gambar 2 merupakan hasil dari tahap *bag-of-words*.



Gambar 2 Hasil Tahap *Bag-of-Words*

3.3 Seleksi Fitur *Chi-Square*

Tahapan seleksi fitur digunakan untuk menyeleksi fitur atau *term* berdasarkan bobot fitur. Fitur yang nilainya rendah dianggap tidak banyak memengaruhi kinerja sistem sehingga bisa menambah beban sistem. *Term* yang memiliki nilai *chi-square* tinggi merupakan *term* yang memiliki ketergantungan tinggi terhadap kelas label dan berpengaruh dalam sistem. Sedangkan *term* yang nilainya rendah merupakan *term* yang tidak memiliki ketergantungan pada label dan tidak berpengaruh banyak terhadap kinerja sistem sehingga nantinya akan dilakukan seleksi. Hasil perhitungan nilai *chi-square term* dapat dilihat pada Gambar 3.

	term	D	N	TD	Xtot	rank
0	sempit	5.546670	23.458788	0.814283	29.819740	1.0
1	peduli	4.250257	0.820863	20.452176	25.523296	2.0
2	bumi	2.779750	0.273167	17.257320	20.310237	3.0
3	cangjih	2.779750	0.273167	17.257320	20.310237	4.0
4	customer	2.779750	0.273167	17.257320	20.310237	5.0
...
3395	butuh	0.011430	0.024750	0.000798	0.036978	3396.0
3396	diskusi	0.024728	0.006882	0.004997	0.036606	3397.0
3397	komplit	0.024728	0.006882	0.004997	0.036606	3398.0
3398	naik	0.024728	0.006882	0.004997	0.036606	3399.0
3399	samping	0.024728	0.006882	0.004997	0.036606	3400.0

Gambar 3 Hasil Perhitungan Nilai *Chi-Square Term*

3.4 Uji Coba

Percobaan selanjutnya akan dilakukan berdasarkan seleksi fitur *chi-square* dengan menggunakan skenario variasi rasio penggunaan fitur sebanyak 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, serta 100% di mana penggunaan 100% fitur juga dianggap sebagai pengujian tanpa seleksi fitur. Dari setiap penggunaan variasi rasio fitur ini, kemudian nilai performa akan dihitung sehingga dapat diketahui di mana rasio penggunaan fitur dengan performa terbaik.

Dalam penelitian ini menggunakan 3 kelas label sehingga sistem akan membuat klasifikasi untuk masing-masing ulasan yang diproses kemudian hasil klasifikasi tersebut akan dipadankan dengan label (*ground truth*). Jika hasil klasifikasi sistem dengan *ground truth* sama, nilai TP adalah 1 sebab hasil klasifikasi sistem benar pada satu kelas, dan nilai TN adalah 2 sebab hasil klasifikasi sistem terdapat 2 kelas selain *ground truth* dan diklasifikasikan secara benar, selain



kelas dari ulasan. Kemudian jika hasil klasifikasi sistem dengan *ground truth* berbeda atau salah, nilai TP adalah 0 sebab tidak ada hasil klasifikasi sistem yang sama dengan *ground truth*, nilai FP adalah 1 sebab seharusnya ada hasil klasifikasi sistem benar pada satu kelas, nilai FN adalah 1 sebab seharusnya ada hasil klasifikasi sistem salah pada satu kelas. Berdasarkan pemaparan tersebut didapatkan nilai yang ditunjukkan pada Tabel 3.

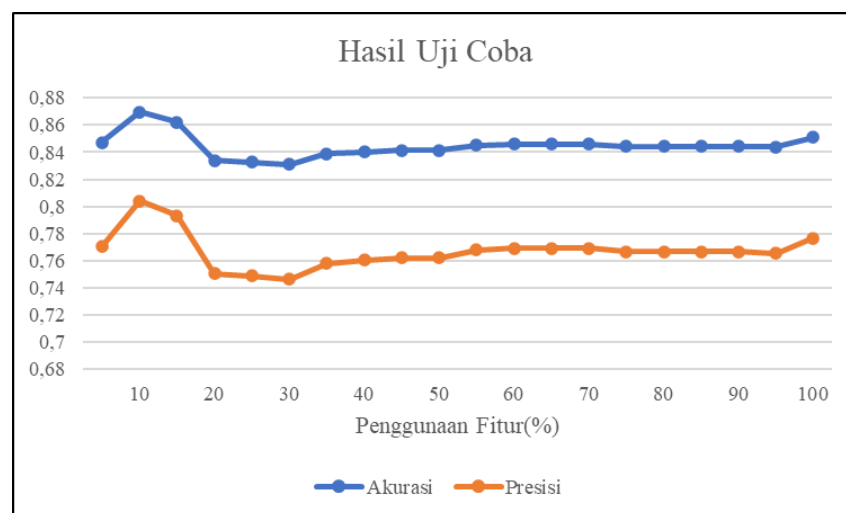
Tabel 3 Hasil Uji Coba

Features (%)	TP	TN	FP	FN
10	152	341	37	37
20	286	667	95	95
30	321	751	109	109
40	381	882	120	120
50	391	904	122	122
60	413	950	124	124
70	413	950	124	124
80	414	954	126	126
90	414	954	126	126
100	445	1018	128	128

Kemudian dihitung nilai performa dalam sistem yang ditunjukkan pada Tabel 4 serta representasinya pada Gambar 4.

Tabel 4 Nilai Performa

Penggunaan Fitur (%)	Akurasi	Presisi	Recall	F-Measure
10	0,869489	0,804233	0,804233	0,804233
20	0,833771	0,750656	0,750656	0,750656
30	0,831008	0,746512	0,746512	0,746512
40	0,840319	0,760479	0,760479	0,760479
50	0,841455	0,762183	0,762183	0,762183
60	0,846058	0,769088	0,769088	0,769088
70	0,846058	0,769088	0,769088	0,769088
80	0,844444	0,766667	0,766667	0,766667
90	0,844444	0,766667	0,766667	0,766667
100	0,851076	0,776614	0,776614	0,776614



Gambar 4 Grafik Nilai Performa



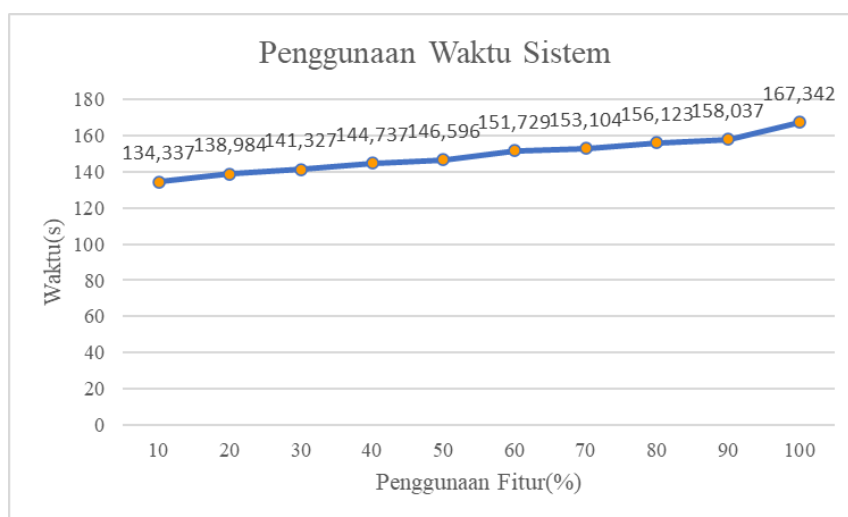
Berdasarkan percobaan yang telah dilakukan, terdapat kesalahan atau kelalaian dalam pengklasifikasian pemberitahuan fasilitas umum. Hal ini dikarenakan kesalahan ketik yang menyebabkan terdapat kekurangan pada tahap *preprocessing* dalam mengubah kata menjadi kata yang tepat seperti penggunaan huruf yang berulang-ulang “mantulll”, “hantuuu”, dan lain-lain. Selain itu, terdapat unsur kesalahan ejaan dan dua kata yang saling menempel seperti “pelajaeen”, “ruangtunggunya” dan lain-lain. Hal ini dikarenakan gramatika yang digunakan pengunjung tempat-tempat fasilitas publik dalam memberikan ulasan menggunakan kata-kata informal. Penggunaan kata pada ulasan yang bercampur dengan bahasa lain seperti “experience”, “nice” juga memengaruhi hasil klasifikasi.

Seleksi fitur *chi-square* yang digunakan dalam penelitian ini juga dapat memengaruhi hasil pengujian performa karena seleksi fitur *chi-square* akan menyeleksi kata yang tidak memiliki ketergantungan terhadap label sistem. Diketahui semakin banyak rasio penggunaan fitur yang digunakan dalam sistem, nilai performa cenderung semakin naik. Sebaliknya, semakin sedikit rasio fitur yang digunakan dalam sistem, nilai performa cenderung turun. Tetapi terdapat anomali pada hasil penggunaan fitur sebesar 10% yang menghasilkan nilai akurasi 86,94%, presisi 80,42%, *recall* 80,42% dan *f-measure* 80,42% di mana hasil ini merupakan nilai performa terbaik.

Faktor yang dapat memengaruhi hasil dan anomali pada hasil uji coba adalah fitur-fitur pada seleksi fitur *chi-square* itu sendiri. Misalnya, dalam penggunaan 10% fitur, fitur-fitur ini merupakan fitur dengan peringkat teratas yang memiliki nilai *chi-square* total tertinggi dari fitur-fitur yang digunakan dalam sistem. Meskipun tiap-tiap fitur tersebut merupakan fitur dengan peringkat teratas, dalam fitur kata tersebut masih banyak terdapat kata yang tidak baku, kata singkatan, serta kata bahasa asing.

Faktor lain yang dapat memengaruhi nilai performa adalah ulasan-ulasan yang tidak dapat terklasifikasi akibat pengurangan penggunaan fitur dalam sistem. Jika seluruh kata yang terdapat dalam sebuah ulasan tereduksi akibat seleksi fitur, maka ulasan tersebut tidak dapat diklasifikasikan dan akan dihilangkan dari data uji coba yang akan digunakan untuk menghitung nilai performa.

Durasi penggunaan waktu sistem dalam melakukan *training data* hingga data terklasifikasi keseluruhan dapat dilihat pada Gambar 5.



Gambar 5 Grafik Penggunaan Waktu Sistem

Diketahui bahwa semakin banyak rasio fitur yang digunakan, maka semakin lama pula durasi waktu yang dibutuhkan untuk melakukan klasifikasi. Di mana hal ini menunjukkan bahwa penggunaan seleksi fitur *chi-square* secara efisien dapat melakukan klasifikasi dengan durasi



waktu relatif lebih cepat. Hal ini dikarenakan fitur-fitur telah direduksi sebelum diproses dalam sistem sehingga tidak memberatkan sistem dalam melakukan perhitungan.

4. KESIMPULAN

Penggunaan 10% fitur *chi-square* menghasilkan performa terbaik yaitu nilai akurasi 86,94%, presisi 80,42%, *recall* 80,42%, dan *f-measure* 80,42%. Faktor yang dapat memengaruhi hal tersebut adalah banyaknya kata tidak baku yang masih terproses dalam sistem karena kurang optimalnya tahap *preprocessing* dalam menangani kata-kata yang memiliki huruf berulang, kesalahan penulisan, hingga kata yang seharusnya ditulis terpisah tetapi tertulis secara tergabung. Penggunaan seleksi fitur *chi-square* juga memengaruhi hasil performa di mana semakin banyak penggunaan fitur semakin tinggi pula nilai performa dalam sistem, akan tetapi terdapat anomali pada penggunaan 10% fitur pada penelitian ini yang menghasilkan nilai performa tertinggi. Tetapi, semakin banyak fitur yang digunakan juga menyebabkan waktu yang dibutuhkan untuk melakukan klasifikasi semakin lama.

Dikarenakan keterbatasan yang ada dalam penelitian ini, penulis menyarankan hal-hal berikut untuk dicoba pada penelitian selanjutnya:

- 1) Mengoptimalkan tahap *preprocessing* terutama untuk mengatasi, memperbaiki, mengeliminasi kesalahan kata, kata tidak baku, maupun kata yang tidak berpengaruh terhadap sistem.
- 2) Melakukan percobaan menggunakan algoritma *oversampling* atau *undersampling* untuk mengatasi *imbalance dataset*.
- 3) Melakukan percobaan menggunakan *N-gram* pada penggunaan fitur untuk mengetahui apakah fitur yang dipasang dapat memengaruhi performa sistem.

DAFTAR PUSTAKA

- Aziz, K. E., Crysdian, C., & Imamudin, M. (2021). Identification of Student Academic Performance in Computer Science Based on Naive Bayes. *MATICS*, 13(1), 33–41. <https://doi.org/10.18860/mat.v13i1.9726>
- Baskoro, S. E., Suhartono, S., Chamidy, T., & Zaman, S. (2022). Pengujian akurasi model regresi logistik multinomial untuk memprediksi keberhasilan mahasiswa di perguruan tinggi menggunakan r. *Fair Value: Jurnal Ilmiah Akuntansi Dan Keuangan*, 5(3), 1551–1565. <https://doi.org/10.32670/fairvalue.v5i3.2472>
- Dulhare, U. N. (2018). Prediction system for heart disease using Naive Bayes and particle swarm optimization. *Biomedical Research (India)*, 29(12), 2646–2649. <https://doi.org/10.4066/BIOMEDICALRESEARCH.29-18-620>
- Faisal, M., Nugroho, F., Sulthan, M. M. El, Amini, F., Hariyadi, M. A., & Sedayu, A. (2020). Plagiarism Detection Using Manber and Winnowing Algorithm. *International Journal of Advanced Science and Technology*, 29(6s), 2130–2136. <http://sersc.org/journals/index.php/IJAST/article/view/10924>
- Han, J., & Kamber, M. (2001). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Harahap, F., Harahap, A. Y. N., Ekadiansyah, E., Sari, R. N., Adawiyah, R., & Harahap, C. B. (2018). Implementation of Naive Bayes Classification Method for Predicting Purchase. *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, 1–5. <https://doi.org/10.1109/CITSM.2018.8674324>
- Irvantoro, D. (2019). *Feature Selection Chi-Square N-Gram Naive Bayes Classifier Review* [Universitas Muhammadiyah Jember]. <http://repository.unmuhjember.ac.id/7128/>
- Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd ed.). Prentice Hall. <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>
- Linawati, S., Nurdiani, S., Handayani, K., & Latifah. (2020). *Prediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma Random Forest dan C4.5*. 8(1), 6–13. <https://doi.org/10.31294/jki.v8i1.7827>
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language*



- Technologies*, 5(1), 1–167. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- Huan Liu, & Setiono, R. (1995). Chi2: feature selection and discretization of numeric attributes. *Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence*, 388–391. <https://doi.org/10.1109/TAI.1995.479783>
- Pratama, N. D., Sari, Y. A., & Adikara, P. P. (2018). Analisis Sentimen pada Review Konsumen Menggunakan Metode Naive Bayes dengan Seleksi Fitur Chi Square untuk Rekomendasi Lokasi Makanan Tradisional. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(9), 2982–2988. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2494>
- Putri, L. A. A. R. (2018). Seleksi Fitur dalam Klasifikasi Genre Musik. *Jurnal Ilmu Komputer*, 10(1), 19–26. <https://ojs.unud.ac.id/index.php/jik/article/view/39772>
- Russell, S., & Norvig, P. (1995). Artificial intelligence—a modern approach. In *The Knowledge Engineering Review* (Issue 1). Pearson Education. https://www.cambridge.org/core/product/identifier/S0269888900007724/type/journal_article
- Ruz, G. A., Henríquez, P. A., & Mascareño, A. (2020). Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Future Generation Computer Systems*, 106, 92–104. <https://doi.org/10.1016/j.future.2020.01.005>
- Sanders, R. (1987). THE PARETO PRINCIPLE: ITS USE AND ABUSE. *Journal of Services Marketing*, 1(2), 37–40. <https://doi.org/10.1108/eb024706>
- Singh, G., Kumar, B., Gaur, L., & Tyagi, A. (2019). Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification. *2019 International Conference on Automation, Computational and Technology Management (ICACTM)*, 593–596. <https://doi.org/10.1109/ICACTM.2019.8776800>
- Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment Analysis of Comment Texts Based on BiLSTM. *IEEE Access*, 7, 51522–51532. <https://doi.org/10.1109/ACCESS.2019.2909919>

