

Klasifikasi Opini Masyarakat Terhadap Jasa ISP MyRepublic dengan *Naïve Bayes*

Hafiz Irsyad¹, Ahmad Farisi², Muhammad Rizky Pribadi³

Abstract— Opinion classification is an analysis that aims to determine the sentiments of the community or a group about a particular entity. Opinion classification can be categorized as positive, negative, and neutral. This research of the classification of public opinion was conducted on the MyRepublic internet service provider. At the moment, MyRepublic has reached seven provinces in Indonesia. MyRepublic has used a lot of media to communicate with its customers, especially Twitter. MyRepublic Twitter account is MyRepublicid with a number of followers of 9,414. This research uses comments or tweets from followers that can be used to see opinions from followers of My Republic, whether positive or negative. The comments or tweets classification on Twitter is using naïve Bayes method. The data used is 1,553. As much as 70% of the data from each category is used as training data and the remaining 30% as testing data. The naïve Bayes method produces positive accuracy value of 0.976%, negative accuracy value of 0.82895%, and neutral accuracy value of 0.8333%, with an average of 0.87949%. Based on the result, it can be concluded that the naïve Bayes method is able to classify the data very well.

Intisari— Klasifikasi opini adalah analisis yang bertujuan untuk menentukan sentimen komunitas atau kelompok tentang entitas tertentu. Klasifikasi opini dapat dikategorikan sebagai positif, negatif, dan netral. Klasifikasi opini publik ini dilakukan pada penyedia layanan internet MyRepublic. Saat ini, MyRepublic telah menjangkau tujuh provinsi di Indonesia. MyRepublic menggunakan banyak media untuk berkomunikasi dengan pelanggannya, terutama menggunakan Twitter. Akun MyRepublic di Twitter adalah MyRepublicid dengan jumlah pengikut 9.414. Makalah ini mengambil komentar atau *tweet* dari pengikut yang dapat digunakan untuk melihat pendapat dari pengikut My Republic, positif atau negatif. Klasifikasi komentar atau *tweet* di Twitter menggunakan metode *naïve Bayes*. Dari sejumlah 1.553 data yang digunakan, diambil 70% dari masing-masing kategori sebagai data latih dan 30% sisanya sebagai data uji. Metode *Naïve Bayes* menghasilkan nilai akurasi positif 0,976%, negatif 0,82895%, dan netral 0,8333%, sehingga memiliki rata-rata 0,87949%. Dengan hasil akurasi ini dapat disimpulkan bahwa metode *Naïve Bayes* mampu mengklasifikasikan data dengan baik.

Kata Kunci— *Naïve Bayes*, Opini, Klasifikasi, MyRepublic.

¹ Dosen, Jurusan Teknik Informatika, STMIK Global Informatika MDP, Jl. Rajawali No 14, Palembang, Sumatera Selatan 30113 Indonesia (tlp: 0711- 376400; e-mail: hafzirsyad@mdp.ac.id)

² Dosen, Jurusan Sistem Informasi, STMIK Global Informatika MDP, Jl. Rajawali No 14, Palembang, Sumatera Selatan 30113 Indonesia (tlp: 0711- 376400; e-mail: ahmadfarisi@mdp.ac.id)

³ Dosen Jurusan Teknik Informatika, STMIK Global Informatika MDP, Jl. Rajawali No 14, Palembang, Sumatera Selatan 30113 Indonesia (tlp: 0711- 376400; e-mail: rizky@mdp.ac.id)

I. PENDAHULUAN

My Republic merupakan sebuah perusahaan yang bergerak di bidang *Internet Service Provider* (ISP) dan *Smart TV*. Berkembangnya perusahaan ISP di Indonesia membuat para penggunanya berlomba-lomba mencari layanan yang paling baik sesuai dengan kebutuhan masing-masing. Sejauh ini MyRepublic telah mempunyai jangkauan area tujuh provinsi di Indonesia. Dalam hal ini, untuk melangsungkan keharmonisan hubungan dengan pelanggan, MyRepublic telah menyediakan media untuk berkomunikasi dalam memberikan pelayanan berupa konsultasi, tanya jawab, dan keluhan seputar jasa yang diberikan oleh MyRepublic menggunakan media sosial.

Salah satu media sosial yang digunakan oleh MyRepublic adalah Twitter. Twitter adalah media yang paling banyak diakses oleh para netizen di Indonesia, sehingga Twitter dapat difungsikan sebagai sarana untuk menampung komunikasi dengan para pelanggan oleh pihak perusahaan atau pebisnis. Twitter memiliki fasilitas untuk menciptakan *statement* atau sering disebut dengan *tweet* (kicauan). *Tweet* tersebut dapat menampung sebanyak 280 karakter, sehingga para pengguna dapat membatasi *tweet*-nya. MyRepublic sendiri mempunyai akun resmi di Twitter yaitu @MyRepublicID sebagai pusat layanan bagi pelanggannya. Akun @MyRepublicID telah bergabung di Twitter sejak bulan April 2015 dan telah mengirimkan *tweet* sebanyak 17.500 serta memiliki pengikut sejumlah 9.414 ribu.

Kehadiran akun MyRepublic di Twitter berdampak pada banyak *tweet* dari pelanggan MyRepublic yang dapat menggambarkan opini-opini masyarakat terhadap pelayanan yang disediakan oleh MyRepublic. Berbagai macam opini dari masyarakat tersebut ditemukan dalam *tweet* terhadap akun MyRepublic, baik yang bersifat positif, negatif, maupun netral. Oleh karena itu, dibutuhkan sebuah sistem yang mampu mengklasifikasikan opini-opini dari masyarakat yang dicurahkan terhadap akun MyRepublic di Twitter. Untuk mengklasifikasikan dan mengelompokkan opini-opini tersebut dapat diterapkan teknik *text mining*.

Penelitian terkait klasifikasi teks sudah banyak dilakukan. Pada [1], berdasarkan hasil analisis dan pengujian yang telah dilakukan dengan seratus data uji, metode *naïve Bayes Classifier* (NBC) dengan *pre-processing* yang digunakan menghasilkan rata-rata akurasi 84%. Dengan menggunakan NBC pada sistem ini, dapat dihasilkan klasifikasi sentimen positif dan negatif [1]. Sementara pada [2] didapatkan bahwa NBC dapat diimplementasikan dengan nilai akurasi 98,40%. Dari 6.338 data uji, 4.049 berhasil terklasifikasi ke dalam sentimen positif dan 3.233 data bersentimen negatif. Komentar terbanyak pada bulan Januari yaitu sebanyak 1.472 data [2]. Maka, berdasarkan rujukan tersebut, makalah ini menggunakan

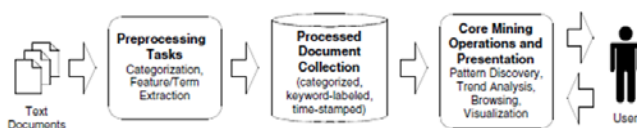
algoritme *naïve Bayes* untuk mengklasifikasikan *tweet* dari pelanggan MyRepublic agar dapat diklasifikasikan menjadi tiga kategori, yakni negatif, positif, atau netral.

II. METODOLOGI

A. Text Mining

Text mining merupakan proses mengekstrak informasi yang berguna dari sekumpulan dokumen dari waktu ke waktu melalui identifikasi pola yang menarik [3]. *Text mining* adalah proses penemuan informasi baru dengan mengekstrak pola secara otomatis dari berbagai sumber teks [4].

Text mining mempunyai beberapa tingkatan fungsional. *Text mining* mengikuti model yang telah disediakan oleh beberapa aplikasi *data mining* sehingga dapat terbagi menjadi empat tingkatan fungsional utama, yakni *pre-processing task*, *processed document collection*, dan *core mining operations* [3]. Tingkatan arsitektur dari *text mining* ditunjukkan pada Gbr. 1.



Gbr. 1 Arsitektur *text mining* [3].

B. Pre-processing

Text pre-processing merupakan awal mula dari *text mining*. Tujuan dari *text pre-processing* adalah membersihkan data dari derau (*noise*) sehingga data menjadi lebih kecil dan lebih terstruktur. Berikut adalah tahapan proses dari *text pre-processing* [3].

1) *Tokenizing*: *Tokenizing* merupakan proses pemecahan teks menjadi kata tunggal dan menghapus tanda baca serta angka, sesuai dengan kamus yang telah ditentukan.

2) *Stopword Removing*: *Stopword removing* merupakan proses menghilangkan kata tidak penting dalam teks. Hal ini dilakukan untuk memperbesar akurasi dari pembobotan. Dalam makalah ini, *stopword removing* digunakan untuk menghilangkan kata-kata seperti dan, atau, mungkin, ini, itu dan sebagainya.

3) *Stemming*: *Stemming* merupakan perubahan berbagai kata imbuhan menjadi sebuah kata dasar. Dalam makalah ini, kata imbuhan yang dihilangkan adalah teks yang berbahasa Inggris, dikarenakan teks bahasa Inggris tersebut memiliki bentuk lampau.

C. Naïve Bayes

Naïve Bayes adalah salah satu algoritme klasifikasi yang ditemukan oleh Tomas Bayes. Klasifikasi adalah proses untuk memperoleh aturan atau model yang dapat mengklasifikasikan data baru yang belum pernah dipelajari dengan mempelajari sekumpulan data yang lama. *Naïve Bayes* adalah algoritme klasifikasi probabilistik yang sederhana berdasarkan teorema Bayes [5]. Prinsip umumnya adalah mengasumsikan bahwa nilai suatu atribut tidak bergantung dan memengaruhi atribut yang lainnya. Model *naïve Bayes*

memungkinkan setiap atribut memiliki kontribusi yang sama terhadap keputusan akhir dan komputasinya lebih efisien bila dibandingkan dengan algoritme pengklasifikasi teks lainnya. Model yang diperoleh dari proses pelatihan berisikan kumpulan konstanta untuk setiap data latih. Model tersebut digunakan pada data pengujian untuk melihat seberapa akurat model yang telah diperoleh [6]. *Naïve Bayes* menjumlahkan frekuensi dan kombinasi nilai dari *dataset* untuk menghitung probabilitas [7]. Probabilitas dihitung dengan menggunakan (1) [5].

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

dengan

X = Data dengan *class* yang belum diketahui

H = Hipotesis data merupakan suatu *class* spesifik

$P(H|X)$ = Probabilitas hipotesis berdasar kondisi

$P(H)$ = Probabilitas hipotesis

$P(X|H)$ = Probabilitas berdasarkan kondisi pada hipotesis

$P(X)$ = Probabilitas H.

D. Analisis Sentimen

Analisis sentimen, juga disebut penambangan opini, adalah bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya. Itu mewakili besar ruang masalah. Ada juga banyak nama dan tugas yang sedikit berbeda, misalnya analisis sentimen, penambangan opini, ekstraksi pendapat, sentimen penambangan, analisis subjektivitas, analisis pengaruh, analisis emosi, peninjauan penambangan. Namun, sekarang semuanya berada di bawah payung analisis sentimen atau penambangan opini. Sementara di industri analisis sentimen jangka lebih umum digunakan, di akademisi baik analisis sentimen maupun penambangan opini sering digunakan.

E. Twitter Rest API

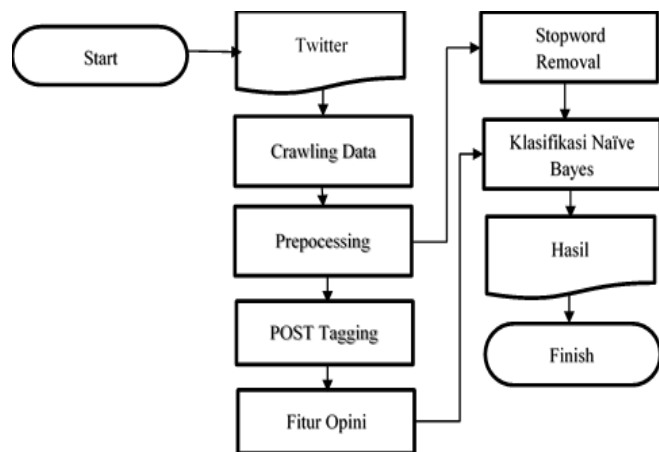
Twitter adalah jejaring sosial populer yang penggunanya dapat berbagi pesan singkat seperti *short message service* (SMS) yang disebut *tweet*. Pengguna berbagi pemikiran, tautan, dan gambar di Twitter; jurnalis mengomentari secara langsung; perusahaan mempromosikan produk dan terlibat dengan pelanggan. Terdapat berbagai macam cara dalam menggunakan Twitter, dan dengan 500 jutaan *tweet* per hari, banyak data untuk dianalisis dan dimainkan. Twitter REST API menyediakan akses terprogram untuk membaca dan menulis data Twitter, seperti membuat *tweet* baru, juga membaca profil penulis dan data pengikut. REST API mengidentifikasi aplikasi Twitter dan pengguna menggunakan OAuth, dengan tanggapan tersedia di JSON [8].

III. HASIL DAN PEMBAHASAN

Gambaran umum dalam proses penelitian diperlihatkan pada Gbr. 2.

Skenario dari penelitian ini adalah sebagai berikut.

1. *Tools* Orange menerima masukan berupa *dataset* Twitter. Data tersebut diambil berdasarkan masukan *hashtag* dan akun dari MyRepublicID.



Gbr. 2 Gambaran umum penelitian.

2. Melakukan *pre-processing* terhadap *dataset* agar dapat memenuhi klasifikasi dan mempermudah dalam memproses data.
3. *POST Tagging* berfungsi sebagai pencarian faktor opini. *POST Tagging* berfungsi juga sebagai ekstraksi fitur opini.
4. Setelah fitur opini didapatkan, dilakukan langkah selanjutnya dengan menggunakan *stopword removal*. *Stopword removal* berfungsi untuk menghapus kata-kata yang sering muncul tetapi tidak memiliki makna dalam penelitian ini.
5. Proses selanjutnya adalah melakukan pengklasifikasian sentimen data dengan menggunakan *naïve Bayes*.

Dataset yang digunakan merupakan opini-opini masyarakat tentang pelayanan yang diberikan oleh *Internet Service Provider (ISP)* MyRepublic pada media sosial Twitter yang mempunyai akun @MyRepublicID. Algoritme yang digunakan untuk klasifikasi adalah *naïve Bayes* dengan menggunakan *tools* Orange.

A. Crawling Data Tweets

Dataset yang digunakan dalam penelitian ini adalah *data mention* dari akun MyRepublicID. *Dataset* diambil menggunakan API Twitter menggunakan Orange.

Kata kunci yang digunakan adalah @MyRepublicID, dengan penarikan data mulai dari tanggal 8 September 2018 sampai 19 September 2018. Dari hasil *crawling* data, diperoleh *tweet* sebanyak 2.911. Adapun hasil dari perolehan *crawling* data disajikan dalam Tabel I.

Dari hasil *crawling* data Twitter, banyak terdapat penulisan kata yang belum benar dan masih berantakan, sehingga membutuhkan proses untuk merapikan kata-kata tersebut dengan *pre-processing* agar kata-kata tersebut menjadi terstruktur.

B. Pre-processing

Setelah dilakukan *Crawling Data* Twitter, dilanjutkan ke proses *pre-processing*, sehingga data-data tersebut dapat terstruktur. Tahapan-tahapan dalam *text preprocessing* dijelaskan dalam subbagian-subbagian selanjutnya.

TABEL I
HASIL CRAWLING DATA TWITTER

No	Komentar Opini	Klasifikasi
1.	@MyRepublicID tiap malem ping kayak gini, tolong dong dibenerin. Bayar full tp yg didapat kayak gini 🙄🙄🙄🙄🙄🙄🙄🙄 https://t.co/EYq5PoYMr6	Negatif
2.	@FarhanGetek @MyRepublicID tergantung amal dan perbuatan tek 🙄	Netral
3.	Speed segini, tapi line loading nya lama banget, instagram juga, OneDrive juga, mesti gimana? Anw pak... https://t.co/tgwGoO5yOR	Negatif
4.	@MyRepublicID jaringan anda paling jelek di tangerang kota khususnya karang tengah di perumahan barata ...Rugi dan nyesel gw	Negatif
5.	Sembari berdoa semoga jaringan @MyRepublicID segera masuk ke daerah sekitar rumah saya.	Netral
6.	@amandawitia @FirstMediaCares Ganti udah man, @MyRepublicID aja stabil	Positif
7.	@MyRepublicID Nama : Ferdinand id: 1044962 gangguan : Up and Down setiap 10 detik, sudah test dengan direct LAN jg... https://t.co/Q3btIa6MoF	Negatif
8.	Setuju. Ditunggu kehadiran @MyRepublicID di Serpong Park. Kalo perlu @GIGbyIndosat tarik kabel dari laverde ke depa... https://t.co/cjkQujR4L8	Netral
9.	@MyRepublicID halo, ada rencana masuk area serpong park tangsel area bvd2 & bvd3? Saat ini yang ngecover area cuma... https://t.co/XpglatxoSd	Netral

C. Tokenizing

Tokenizing merupakan proses pemecahan teks menjadi kata tunggal dan menghapus tanda baca serta angka, sesuai dengan kamus yang telah ditentukan.

Berdasarkan *dataset* yang telah dijelaskan, maka difokuskan pembahasan tentang pelayanan ISP MyRepublic dengan memilih data yang disingkat, dikarenakan Twitter hanya dapat menampung 280 karakter, sehingga para pengguna lebih banyak menyingkat kata serta pengguna banyak menggunakan kata-kata baku. Adapun kata-kata tersebut, yang didapatkan dari *dataset*, adalah berupa “skrng” menjadi “Sekarang”, “Mdh2an” menjadi “Mudah-mudahan”, “kontek” menjadi “Kontak”, dan lain-lain.

D. Stopword Removing

Stopword removing merupakan proses menghilangkan kata tidak penting dalam teks. Hal ini dilakukan untuk memperbesar akurasi dari pembobotan. Dalam makalah ini, *stopword*

recall, sehingga menghasilkan setiap kategori seperti ditunjukkan pada Tabel IV.

TABEL IV
HASIL PENGUJIAN DATA *TWEET*

Kategori	Akurasi	Precision	Recall
Positif	0,97619	0,85366	0,87448
Netral	0,83333	0,87448	0,89581
Negatif	0,82895	0,84917	0,86988
Rata-rata	0,87949	0,85910	0,88005

Dari Tabel IV dapat dilihat bahwa kategori yang memiliki akurasi yang lebih tinggi adalah kategori positif dengan nilai 0,97619 %, kemudian kategori netral dengan nilai 0,8333%, dan terakhir adalah kategori negatif dengan nilai 0,82895%, sehingga didapatkan nilai rata-rata adalah 0,87949%. Dengan nilai rata-rata tersebut terbukti bahwa dari pemodelan dan pengidentifikasi dapat dikenali data yang bukan kategori positif dengan baik.

Pengklasifikasian data netral mempunyai nilai 0,8333%, sehingga dapat diasumsikan pada data dari *tweet* terdapat kekeliruan. Kekeliruan ini diakibatkan *tweet* dari pengikut hampir bisa disebutkan mempunyai kemiripan, sehingga terjadi kesulitan dalam membedakan data netral, negatif, dan positif dalam melakukan pengategorian.

IV. KESIMPULAN

Klasifikasi opini masyarakat terhadap layanan ISP MyRepublicId dengan menggunakan *dataset* pada Twitter dan dilakukan dengan menerapkan metode *naive Bayes* menghasilkan akurasi dengan kategori positif 0,976%, netral 0,833%, dan negatif 0,82895%, dengan nilai rata-rata 0,87949%. Data tersebut membuktikan bahwa komentar yang negatif dan netral dapat diklasifikasikan dengan baik.

Faktor opini yang mendominasi hasil pembahasan ini di antaranya adalah gangguan, tolong, *trying*, *los*, dan merah. Data yang digunakan dalam pembahasan ini adalah *tweet* yang netral

dan negatif. Sedangkan untuk *tweet* yang positif sangat jarang ditemukan, sehingga tidak digunakan. Pada umumnya, *tweet* dari pengikut berkomentar tentang layanan jasa serta kendala yang dihadapi oleh pengikut.

Saran yang dapat disampaikan untuk pembahasan adalah dapat melanjutkan penelitian yang lebih mendalam, sehingga mendapatkan hasil yang lebih bagus lagi.

REFERENSI

- [1] T.A. Lorosae, B.D. Prakoso, Saifudin, dan Kusrin, "Analisis Sentimen Berdasarkan Opini Masyarakat pada Twitter Menggunakan Naive Bayes," *Seminar Nasional Teknologi Informasi dan Multimedia 2018*, 2018, hal. 1.10-25-1.10-30.
- [2] F.N. Zuhri dan A. Alamsyah, "Analisis Sentimen Masyarakat Terhadap Brand Smartfren Menggunakan Naive Bayes Classifier di Forum Kaskus," *e-Proceeding of Management*, Vol. 4, No. 1, hal. 242-251, 2017.
- [3] R. Feldman dan J. Sanger, *The Text Mining Handbook*, New York, USA: Cambridge University Press, 2007.
- [4] M. Kini M., S. Devi H., P.G. Desai, dan N. Chiplunkar, "Text Mining Approach to Classify Technical Research Document Using Naive Bayes," *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 4, No. 7, hal. 386-391, 2015.
- [5] J. Han, M. Kamber, dan J. Pei, *Data Mining Concept and Techniques*, 3rd ed., Waltham, USA: Elsevier Inc., 2012.
- [6] S.L. Ting, W.H. Ip, dan A.H.C. Tsang, "Is Naive Bayes a Good Classifier for Document Classification?" *International Journal of Software Engineering and Its Applications*, Vol. 5, No. 3, hal. 37-46, 2011.
- [7] A. Saleh, "Implementasi Metode Klasifikasi Naive Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Creative Information Technology Journal*, Vol. 2, No. 3, hal. 207-217, 2015.
- [8] M. Rani dan J. Arora, "Twitter Data Predicting Geolocation Using Data Mining Techniques," *International Journal of Innovative Research in Computer*, Vol. 4, No. 6, hal. 10446-10453, 2016.