



ANALISIS SENTIMEN TERHADAP OPERATOR SELULER TELKOMSEL MENGGUNAKAN ALGORITMA NAÏVE BAYES

Candra Naya¹, Eriswandi²

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa
¹candranaya@pelitabangsa.ac.id

Abstraksi

Penggunaan media sosial di era globalisasi sangat diperlukan bagi semua kalangan tidak terkecuali operator seluler telkomsel, telkomsel menggunakan *twitter* sebagai media promosi. Namun digunakannya *twitter* sebagai media promosi telkomsel mendapatkan berbagai pujian masukan maupun kritikan dan dalam hal ini peneliti melakukan penelitian untuk menganalisa sentimen pengguna telkomsel. Dan pada penelitian ini peneliti menggunakan algoritma *Naïve Bayes* untuk melakukan klasifikasi sentimen dan mencari nilai *preference value*. Dari hasil pengujian yang dilakukan menggunakan teknik *cross validation* dan pengukuran akurasi menggunakan *confusion matrix* dengan dilakukan 10 kali pengujian *accuracy* terbaik yang diperoleh adalah 85.33 % dan respon positif yang didapatkan dari hasil penghitungan *preference value* adalah 37.03 %. Dapat disimpulkan bahwa algoritma *Naïve Bayes* dapat digunakan untuk melakukan klasifikasi dengan cukup baik dan dapat mengukur respon pengguna.

Kata kunci : *Twitter, Analisa Sentimen, Naïve Bayes Classifier, Cross Validation, Preference Value*

Abstract

The use of social media in the era of globalization is very necessary for all people including Telkomsel cellular operators, Telkomsel uses Twitter as a promotional media. However, the use of Twitter as a medium for promotion of Telkomsel received various compliments of input and criticism and in this case the authors conducted research to analyze Telkomsel user sentiment. And in this study the authors used the Naïve Bayes algorithm to classify sentiments and look for preference values. From the results of tests conducted using cross validation techniques and accuracy measurements using confusion matrix with 10 times the best accuracy testing obtained was 85.33 % and the positive response obtained from the calculation of preference value was 37.03 %. It can be concluded that the Naïve Bayes algorithm can be used to classify quite well and be able to measure user responses.

Keywords : *Twitter, Sentiment Analysis, Naïve Bayes Classifier, Cross Validation, Preference Value*

1. Pendahuluan

Perkembangan internet di Indonesia cukup pesat, hal ini dibuktikan banyaknya pengguna internet di Indonesia. Menurut Asosiasi Penyelenggara Jasa Internet Indonesia (APJII, 2018) dari total populasi sebanyak 264 juta jiwa penduduk Indonesia, ada sebanyak 171,17 juta jiwa atau sekitar 64,8 persen yang sudah terhubung ke internet. Sementara dengan pesatnya pengguna internet di Indonesia, maka semakin pesatnya juga media sosial. Media sosial adalah sebuah media untuk bersosialisasi. satu sama lain dan dilakukan secara online yang memungkinkan manusia untuk saling berinteraksi tanpa dibatasi ruang dan waktu. Salah satu sosial media yang banyak digunakan di Indonesia adalah Twitter.

Twitter merupakan layanan jejaring sosial dan mikroblog daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter akan tetapi pada tanggal 07 November 2017 bertambah hingga 280 karakter yang dikenal dengan sebutan kicauan (tweet).

Mengacu dari maraknya penggunaan media sosial, Social media marketing menjadi salah

satu kunci strategis dalam kegiatan pemasaran di dunia. Hal ini dibuktikan dengan mayoritas 94% perusahaan di dunia menggunakan media sosial untuk tujuan pemasaran.

Salah satunya operator seluler telkomsel menggunakan twitter sebagai media promosi, telkomsel saat ini sudah mempunyai 1,5 juta followers atau pengikut dengan menggunakan username @Telkomsel. Telkomsel merupakan operator seluler terbesar di Indonesia dengan lebih dari 150 juta pengguna (Kominfo.co.id, 2018). Namun dengan jumlah pengguna terbesar di Indonesia, Telkomsel juga tidak luput dari komentar para penggunanya dan menuai berbagai macam komentar positif dan negatif. Banyaknya pengguna operator seluler telkomsel di twitter yang menyampaikan komentar – komentar tersebut dapat dimanfaatkan untuk mencari sebuah informasi, untuk dapat menganalisis komentar – komentar di twitter tersebut dengan menggunakan analisis sentimen.

2. Tinjauan Studi

2.1 Analisis Sentimen

Analisis sentimen merupakan sebuah proses untuk menentukan sentimen atau opini dari seseorang yang diwujudkan dalam bentuk teks dan bisa dikategorikan sebagai sentimen positif atau negatif (Muchammad Shiddieqy Hadna, Insap Santosa, dan Wahyu Winarno, 2016) Analisis sentimen mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan text mining yang bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau peneliti berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu. Analisis sistem mempunyai tugas untuk mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan dalam dokumen tersebut apakah bersifat positif, negatif atau netral (Manalu, 2014)

2.2 Twitter

Twitter adalah layanan *microblogging* yang dirilis secara resmi pada 13 Juli 2006 (Mostafa, 2013). Aktifitas utama *twitter* adalah *mem-posting* sesuatu (*tweet*) melalui *web* atau *mobile*. Panjang maksimal dari *tweet* adalah 280 karakter. *Twitter* menjadi sumber yang hampir tak terbatas yang digunakan pada *text classification*.

2.3 Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data (Vulandari, 2017).

Menurut Suyanto (2017) *Data mining* adalah kegiatan meliputi pengumpulan, pemakaian *data historis* untuk menemukan keteraturan, pola, atau hubungan dalam data set berukuran besar. Keluaran

dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan.

Sedangkan menurut Muzakir dan Wulandari (2016) *Data Mining* merupakan proses mencari sebuah pola atau informasi yang menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

2.4 Text Mining

Text mining dan analisis teks adalah istilah umum yang dapat menggambarkan berbagai teknologi untuk menganalisa dan memproses data teks yang semi terstruktur dan tidak terstruktur (Miner, 2012).

Sedangkan menurut Priyanto dan Ma'arif (2018) *text mining* adalah bidang interdisipliner yang mengacu kepada pengambilan informasi, *data mining*, pembelajaran mesin, statistic, dan linguistik komputasional.

Ruang lingkup *text mining* adalah sebagai berikut:

1. *Search & Information Retrieval* Adalah kegiatan memperoleh sumber daya sistem informasi yang relevan dengan kebutuhan informasi dari suatu koleksi. Pencarian dapat didasarkan pada teks lengkap atau pengindeksan berbasis konten lainnya. Sistem pengambilan informasi otomatis digunakan untuk mengurangi apa yang disebut informasi yang berlebihan.
2. *Document Clustering* Proses pengelompokan dan pengkategorian istilah, cuplikan, paragraph, atau dokumen menggunakan metode pengelompokan data mining.
3. *Document Classification* Pengelompokan dan pengkategorian cuplikan, paragraph, atau dokumen menggunakan metode klasifikasi *data mining* berdasarkan model yang dilatih pada sampel yang telah diberi label.
4. *Web Mining*
Data mining dan *text mining* pada internet, yang berfokus kepada sekala dan keterkaitan antar web yang dapat di mining.
5. *Information Extraction (IE)*
Adalah tugas mengekstraksi informasi terstruktur secara otomatis dari dokumen yang dapat dibaca mesin yang tidak terstruktur dan / atau semi-terstruktur.
6. *Natural Language Processing (NLP)*
Adalah cabang kecerdasan buatan yang berkaitan dengan menganalisis, memahami, dan menghasilkan bahasa yang digunakan manusia secara alami untuk berinteraksi dengan komputer dalam konteks tertulis dan lisan menggunakan bahasa manusia alami alih-alih bahasa komputer.
7. *Concept Extraction*
Ekstraksi konsep adalah teknik penambangan topik yang paling penting dari suatu dokumen. Dalam konteks *e-commerce*, ekstraksi konsep dapat digunakan untuk mengidentifikasi apa yang dibicarakan oleh halaman Web terkait *product*.

2.5 Text Preprocessing

Menurut Mahmudy dan Widodo (2014) *Text preprocessing* bertujuan untuk mempersiapkan teks menjadi data yang akan mengalami pengolahan pada tahapan berikutnya. Tindakan yang dilakukan meliputi tindakan kompleks dan tindakan sederhana. *Text preprocessing* sendiri terdiri dari:

1. *Case Folding*

Case Folding adalah sebuah proses untuk mengubah huruf dalam dokumen menjadi huruf kecil. Hanya huruf a sampai z yang diterima. Karakter yang tidak sesuai akan dihilangkan dan dianggap delimiter.

2. *Tokenizing*

Tokenizing adalah tahap pemotongan string input berdasarkan setiap kata yang menyusun.

3. *Filtering*

Filtering adalah sebuah tahap untuk mengambil kata-kata penting dari hasil tokenisasi. Terdapat beberapa algoritma dalam *filtering* yaitu *Stop-List* dan *Word-List*. Algoritma *stop-word* merupakan algoritma yang digunakan untuk mengeliminasi kata-kata yang tidak deskriptif. Sedangkan algoritma *Word-List* adalah algoritma yang digunakan untuk menyimpan kata-kata yang memiliki nilai deskriptif.

4. *Stemming*

Stemming adalah proses untuk menggabungkan atau memecahkan setiap varian-varian suatu kata menjadi kata dasar. Proses *Stemming* pada kata Bahasa Indonesia berbeda dengan *stemming* Bahasa Inggris. Proses *stemming* pada kata bahasa Inggris adalah proses untuk mengeliminasi *suffixs* pada kata, sementara proses *stemming* bahasa Indonesia adalah proses untuk mengeliminasi *suffixs*, *prefiks*, dan *konfiks*.

5. *Analyzing*

Analyzing merupakan sebuah tahapan penentuan seberapa jauh kemiripan antar dokumen teks.

2.6 Cross Validation

Cross validation (CV) adalah metode statistik yang dapat digunakan untuk mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua subset yaitu data proses pembelajaran dan data validasi atau evaluasi.

Cross validation adalah membagi dataset menjadi dua bagian dengan satu bagian dijadikan *data training* dan bagian yang lain dijadikan *data testing*. Beberapa penelitian membagi data menjadi 10 bagian, 90% dijadikan *training* dan 10 lainnya digunakan sebagai *testing*. Proses ini dilakukan berulang sampai dengan 10 kali hingga semua record data mendapatkan bagian menjadi data *testing*. Proses ini dikenal juga dengan istilah *10 foldscross validation*. *10 foldscross validation* banyak digunakan peneliti karena terbukti menghasilkan performa algoritma yang lebih stabil.

2.7 Naïve Bayes

Naïve Bayes merupakan sebuah pengklasifikasian probalistik dan statistik sederhana yang bekerja untuk menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang telah diberikan. Algoritma menggunakan teorema Bayes dan mengansumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. *Naïve Bayes* juga dapat didefinisikan sebagai sebuah algoritma pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh seorang ilmuwan Inggris bernama Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Saleh, 2015).

Naïve Bayes mempunyai 2 tahapan dalam proses pengklasifikasiannya, yaitu proses pelatihan dan proses klasifikasi. *Naïve Bayes* adalah metode pengklasifikasian yang membutuhkan jumlah data pelatihan yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naïve Bayes* juga bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan (Saleh, 2015).

2.8 Confusion Matrix

Confusion matrix adalah *tool* yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah *matrix* dari prediksi yang akan dibandingkan dengan kelas yang asli dari inputan atau dengan kata lain berisi informasi nilai *actual* dan prediksi pada klasifikasi (Gorunescu, 2011).

Tabel 1. Confusion Matrix

		Actual Class	
		Class 1	Class 2
Predicted	Class	True	False
	1	Positive	Negative
	Class	False	True
	2	Positive	Negative

Performa yang diperoleh dari *confusion matrix* seperti akurasi, presisi dan *recall* dapat dihitung sesuai dengan persamaan berikut ini:

1. Akurasi

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN}$$

Akurasi merupakan tingkat kedekatan atau seberapa dekat nilai prediksi dengan nilai sebenarnya.

2. Presisi

$$Presisi = \frac{TP}{TP+FP}$$

Presisi merupakan pengukuran seberapa ketepatan informasi yang diberikan oleh sistem.

3. Recall

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

Recall merupakan pengukuran tingkat keberhasilan pengambilan data yang relevan.

2.9 Preference Value

Preference Value merupakan penilaian untuk mengetahui jumlah respon positif pengguna twitter. Semakin tinggi nilai preference value yang dihasilkan maka semakin tinggi pula respon positif yang didapatkan. Berikut rumus dari preference value :

$$PV = \frac{S}{P} = SP + SN * accuracy$$

Keterangan : SP = Sentimen Positif
SN = Sentimen Negatif

2.10 Rapid Miner

Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Teknologi Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan tampilan GUI (Graphical User Interface) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat open source dan dibuat dengan menggunakan program Java di bawah lisensi GNU Public Licence dan Rapid Miner dapat dijalankan di sistem operasi manapun. Dengan menggunakan Rapid Miner, tidak dibutuhkan kemampuan koding khusus, karena semua fasilitas sudah disediakan. Rapid Miner dikhususkan untuk penggunaan data mining. Model yang disediakan juga cukup banyak dan lengkap, seperti Model Bayesian, Modelling, Tree Induction, Neural Network dan lain-lain. Banyak metode yang disediakan oleh Rapid Miner mulai dari klasifikasi, klustering, asosiasi dan lain lain. Jika tidak ada

model atau model algoritma yang tidak ada dalam Weka, pengguna boleh menambahkan modul lain, karena weka bersifat open source, jadi siapapun dapat ikut mengembangkan perangkat lunak ini (Haryati, Sudarsono, dan Suryana, 2015).

2.11 Kerangka Pemikiran

Peneliti perlu membuat gambaran singkat sebagai alur penyusunan laporan ini dengan kerangka pemikiran sebagai berikut:



Gambar 1. Kerangka Pemikiran

Keterangan kerangka pemikiran pada Gambar 2.3 sebagai berikut: Tahap pertama yang dilakukan dalam melakukan penelitian ini adalah melakukan tinjauan penelitian dengan cara mempelajari teori dan pengetahuan dasar mengenai semua yang berkaitan dengan penelitian ini agar dapat memahami dasar teori dan konsep-konsep yang mendukung penelitian.

Tahap kedua yaitu identifikasi masalah, pada tahap ini proses yang dilakukan adalah menggali permasalahan yang ditemukan pada objek yang diteliti serta mengidentifikasi kebutuhan yang dibutuhkan oleh peneliti untuk mencari alternative solusi yang terkait dengan permasalahan tersebut.

Tahap ketiga yaitu pengumpulan data, setelah tahan identifikasi masalah dan semua kebutuhan telah didapatkan maka tahap selanjutnya yang harus dilakukan adalah mengumpulkan data. Pengumpulan data dilakukan dengan menggunakan proses *crawling* data *twitter*. Data yang dikumpulkan merupakan *tweet* yang berbahasa Indonesia yang diambil dari akun *twitter* @Telkomsel yang berisi komentar para pengguna *twitter*.

Tahap keempat yaitu seleksi data, pada tahap ini bertujuan untuk menyeleksi data *tweet* yang akan digunakan nantinya, dikarenakan data *tweet* yang didapatkan saat proses *crawling* data *twitter* belum sepenuhnya bersih dan memiliki data yang tidak diperlukan.

Tahap kelima yaitu pembagian data, untuk melakukan proses uji diperlukan pembagian data yang dibagi menjadi data *Training* dan data *Testing* yang nantinya akan digunakan untuk proses klasifikasi menggunakan *Tools* RapidMiner, data *tweet* yang digunakan dalam penelitian ini sebanyak 1000 data *tweet* yang diambil dari akun @Telkomsel dan nantinya akan dibagi menjadi 80% untuk data *training* dan 20% untuk data *testing* atau masing-masing 800 sebagai data *training* dan 200 sebagai data *testing*.

Tahap keenam yaitu proses dokumen, pada tahapan ini proses *preprocessing* akan menggunakan *tools* RapidMiner, *Preprocessing* ini nantinya akan diterapkan pada data *training* maupun data *testing*, dan komponen yang digunakan dalam proses ini adalah *tokenize*, *transform cases*, *filter stopwords*, *filter token by length* dan *generate n-grams*,

Tahap ketujuh yaitu penerapan algoritma, algoritma yang digunakan adalah *Naïve Bayes Classifier* dan berikut merupakan tampilan penerapan algoritma tersebut didalam sebuah sub proses pada operator *cross validation* dengan menambah kan *Naïve Bayes* pada tab *training* dan pada tab *testing* akan ditambahkan *apply model* dan *performance* yang nantinya akan digunakan sebagai penghitung performa klasifikasi.

Tahap kedelapan yaitu evaluasi, dalam penelitian ini proses evaluasi akan menggunakan *Confusion Matrix* yang mana telah dijelaskan sebelumnya, Evaluasi bertujuan untuk mengetahui kinerja dari algoritma *Naïve Bayes* yang nantinya akan dijadikan sebagai acuan dari perbandingan akurasi pengujian.

3. Metode Penelitian

3.1 Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan proses *crawling* data *twitter* menggunakan perangkat lunak *RapidMiner*, data

yang dikumpulkan merupakan *tweet* yang berbahasa Indonesia yang diambil dari akun *twitter* @Telkomsel yang berisi komentar para pengguna *twitter*.

3.2 Pengolahan Dataset

1. Seleksi Data

Pada tahap ini bertujuan untuk menyeleksi data *tweet* yang akan digunakan, dikarenakan data *tweet* yang didapatkan dari proses *crawling* belum sepenuhnya bersih dan memiliki data yang tidak diperlukan. Pada penelitian ini data yang diambil hanya data yang berhubungan dengan akun @Telkomsel dan proses yang dilakukan pada tahap ini yaitu mengambil *tweet* pada kolom *Text*.

2. Remove Duplicate

Pada proses ini bertujuan untuk memfilter data *tweet* yang sama, hal ini disebabkan karena *twitter* memiliki fitur *retweet* yang menyebabkan banyaknya *tweet* yang sama di *retweet* oleh pengguna lain dan menimbulkan teks berulang dengan topik yang sama.

3. Cleansing

Cleansing merupakan tahapan untuk menghapus kata, simbol dan karakter yang tidak diperlukan dalam penelitian ini, diantaranya adalah sebagaiberikut:

1. Username (@username)
2. Hashtag (#)
3. Emotikon (:0, :(, dll)
4. Karakter HTML(<, >, dll)
5. Url (<http://web>)

3.3 Penentuan Kelas Atribut

Penentuan kelas atribut ini diberikan sesuai subjektifitas peneliti, pembagian atribut pada penelitian ini sendiri hanya akan dibagi menjadi positif dan negatif.

3.4 Pembagian Data

Untuk melakukan proses uji diperlukan pembagian data yang dibagi menjadi data *Training* dan data *Testing* yang nantinya akan digunakan untuk proses klasifikasi menggunakan aplikasi *RapidMiner*, data *tweet* yang digunakan dalam penelitian ini sebanyak 1000 data *tweet* yang diambil dari akun @Telkomsel dan akan dibagi menjadi 90% untuk data *training* dan 10% untuk data *testing* atau masing-masing 900 sebagai data *training* dan 100 sebagai data *testing*.

3.5 Preprocessing

Preprocessing atau pengolahan awal merupakan proses untuk mendapatkan data yang bersih agar dapat dilakukan proses selanjutnya. Proses ini dilakukan secara manual dan pada tahapan ini dilakukan sebagai berikut:

1. Transform cases

Tahapan *transform cases* sendiri dalam penelitian ini bertujuan untuk menyeragamkan bentuk huruf menjadi *lower case* karena pada komentar *Twitter* sendiri memiliki bentuk huruf yang beragam.

2. Tokenize

Tokenize merupakan sebuah proses untuk memisahkan atau pemisahan dokumen menjadi per kata atau beberapa kata dalam suatu *token/unit* kecil.

3. Filter Stopwords

Tahapan *filter stopword (dictionary)* merupakan tahapan proses yang bertujuan untuk menghapus atau mengfilter kata yang tidak memiliki arti dan tidak berhubungan dengan ucapan atau kata *sentiment*.

4. Filter Tokens by Length

Filter Tokens by Length merupakan proses penghapusan kata atau *token* berdasarkan panjang karakter yang ditentukan dan pada penelitian ini minimum panjang karakter yang akan digunakan adalah empat karakter pada setiap kata.

5. Generate N-gram

Proses ini merupakan proses pemotongan *string* atau dokumen (*n-gram*) berdasarkan limit angka tertentu. Dan dalam penelitian ini sendiri *n-gram* yang akan digunakan adalah *bigram*. Tujuan penggunaan *N-gram* sendiri karena dalam Bahasa Indonesia banyak frase yang tidak hanya terdiri dari satu kata.

3.6 Penerapan Algoritma

Dalam penelitian ini algoritma yang digunakan adalah *Naïve Bayes Classifier* dan berikut merupakan tampilan penerapan algoritma tersebut didalam sebuah sub proses pada operator *cross validation* dengan menambahkan *Naïve Bayes* pada tab *training* dan pada tab *testing* akan ditambahkan *apply model* dan *performance* yang nantinya akan digunakan sebagai penghitung performa klasifikasi.

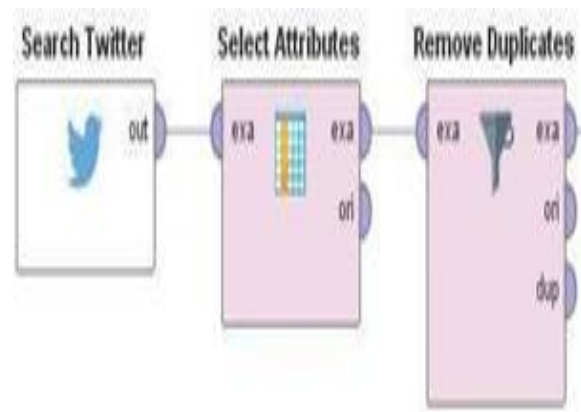
3.7 Evaluasi

Dalam penelitian ini proses evaluasi akan menggunakan *Confusion Matrix* yang mana telah dijelaskan pada bab sebelumnya, Evaluasi bertujuan untuk mengetahui kinerja dari algoritma *Naïve Bayes* yang nantinya akan dijadikan sebagai acuan dari perbandingan akurasi pengujian.

4. Hasil Penelitian dan Pembahasan

4.1 Dataset

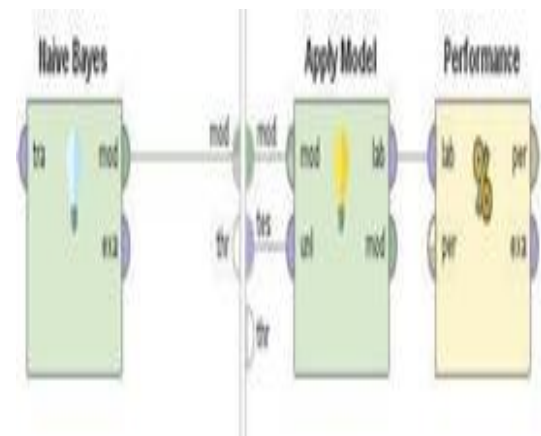
Dataset yang digunakan dalam penelitian ini sebanyak ada 1000 *tweet* yang diambil dari komentar atau *retweet* di akun *twitter @Telkomsel* dan data diambil dari tanggal yang berbeda-beda dikarenakan *twitter* sendiri membatasi pengambilan data atau *crawling* selama 7 hari dari waktu pengambilan data. Berikut adalah tahapan pengumpulan data dan pengolahan data pada perangkat lunak *RapidMiner*:



Gambar 2. Tahapan Pengumpulan Data

4.2 Penerapan Algoritma

Algoritma yang digunakan dalam penelitian ini adalah *Naïve Bayes Classifier* dan berikut merupakan tampilan penerapan algoritma tersebut didalam sebuah sub proses pada operator *cross validation* dengan menambahkan *Naïve Bayes* pada tab *training* dan pada tab *testing* akan ditambahkan *apply model* dan *performance* yang nantinya akan digunakan sebagai penghitung performa klasifikasi.



Gambar 3. Penerapan Algoritma

4.3 Proses Pengujian

Proses Pengujian dilakukan dengan mengikuti alur yang telah ditentukan yaitu dengan metode klasifikasi *Naïve Bayes* dengan *Cross Validation*, proses ini terdiri dari 10 kali proses pengujian yang mana nantinya akan dijadikan tolak ukur dari penelitian ini. Dalam proses ini peneliti menggunakan perangkat lunak *RapidMiner* untuk melakukan pengujian dengan merangkai beberapa proses yaitu *read excel*, *Select Attribute*, *Nominal to text*, *Process Document*, *Cross Validation* dan *Apply Model*.

4.4 Hasil Pengujian

Hasil pengujian berfokus kepada performa klasifikasi yang mana akan dilakukan pengujian *cross validation* sebanyak 10 kali untuk mengetahui tingkat *accuracy*, *recall* dan *precision*.

4.5 Analisa Pengujian

Dari hasil pengujian yang telah dilakukan menggunakan algoritma *Naïve Bayes* dan dilakukannya pengujian menggunakan teknik *cross validation* peneliti mendapatkan hasil pengujian seperti pada tabel berikut:

Tabel 1. Analisa Pengujian

Dataset	Accuracy	Precision	Recall
Dataset 1	82.33	75.79	86.56
Dataset 2	81.44	87.50	77.93
Dataset 3	85.33	90.02	84.19
Dataset 4	84.00	89.51	81.95
Dataset 5	83.00	88.78	79.51
Dataset 6	84.44	88.00	76.41
Dataset 7	83.77	89.42	79.55
Dataset 8	82.11	88.52	78.62
Dataset 9	83.66	89.30	80.67
Dataset 10	85.22	90.37	83.23

4.6 Preference Value

Preference Value bertujuan untuk mengetahui respon positif masyarakat dari data yang telah diambil dan diproses sebelumnya, jumlah data positif pada data *training* dan *testing* berjumlah 434 dan jumlah data negatif pada data *training* dan *testing* berjumlah 566, total dari data tersebut berjumlah 1000 data, dan berikut adalah perhitungan *preference value* dari data tersebut dikalikan dengan nilai *accuracy* terbaik:

$$PV = \frac{434}{434+566} * 85.33\% = 0.3703322 = 37.03 \%$$

Dari perhitungan diatas maka dapat disimpulkan bahwa respon positif pengguna terhadap pelayanan operator seluler telkomsel berdasarkan data yang diambil dari *Twitter* adalah 37.03 %.

5 Penutup

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik kesimpulan bahwa penggunaan algoritma *Naïve Bayes* disertai dengan penggunaan teknik *Cross Validation* menghasilkan hasil yang cukup baik dengan nilai *accuracy* sebesar 77.12 % dan tingkat respon positif pengguna terhadap pelayanan operator seluler telkomsel sebesar 29.84 %, dan nilai respon positif tersebut dapat berubah-ubah dikarenakan respon masyarakat dan data yang diperoleh dapat berubah sewaktu-waktu.

Daftar Pustaka

- [1] Faradhillah, N. Y. A., Kusumawardani, R. P., & Hafidz, I. (2016). Eksperimen Sistem Klasifikasi Analisa Sentimen Twitter pada Akun Resmi Pemerintah Kota Surabaya Berbasis Pembelajaran Mesin (Experiments on Sentiment Classification System for Tweets of the Official Account of the City Government of Surabaya based on Mach. *Prosiding Seminar Nasional Sistem Informasi Indonesia 2016*, 15–24.
- [2] Retrieved from <http://is.its.ac.id/pubs/oajis/index.php/home/detail/1645/EKSPERIMEN-SISTEM-KLASIFIKASI-ANALISA-SENTIMEN-TWITTER-PADA-AKUN-RESMI-PEMERINTAH-KOTA-SURABAYA-BERBASIS-PEMBELAJARAN-MESIN>
- [3] Gorunescu, F. (2011). *Data Mining :Concepts, Models, and Techniques*.
- [4] Gusriani, S., Wardhani, K. D. K., & Zul, M. I. (2016). Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka). *4th Applied Business and Engineering Conference*, 1(1), 1–7.
- [5] Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama*, 11(2), 130–138.
- [6] Mahmudy, W. F., & Widodo, A. W. (2014). Klasifikasi Artikel Berita Menggunakan Naive Bayes Classifier polarity analysis of consumers' airline service tweets. *Social Network Analysis and Mining*, 3(3), 635–649. <https://doi.org/10.1007/s13278-013-0111-2>.
- [7] Muchammad Shiddieqy Hadna, N., Insap Santosa, P., & Wahyu Winarno, W. <https://fti.uajy.ac.id/sentika/publikasi/makalah/2016/95.pdf>.
- [8] Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, 3(1), 19–26. <https://doi.org/10.15294/sji.v3i1.4610>.
- [9] Nugroho, D. G., Chrisnanto, Y. H., & Wahana, A. (2015). Analisis Sentimen Pada Jasa Ojek Online ... (Nugroho dkk.). 156–161.
- [10] Priyanto, A., & Ma'arif, M. R. (2018). Implementasi Web Scrapping dan Text Mining untuk Akuisisi dan Kategorisasi Informasi dari Internet (Studi Kasus: Tutorial Hidroponik). *Indonesian Journal of Information Systems*, 1(1), 25–33. <https://doi.org/10.24002/ijis.v1i1.1664>
- [11] SENTIPOL: Dataset Sentimen Komentar Pada Kampanye PEMILU Presiden Indonesia 2014 dari Facebook Page. *Konferensi Nasional Teknologi Informasi Dan Komunikasi 2017*, (November), 218–228.
- [12] Saleh, A. (2015).

- Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Creative Information Technology Journal*, 2(3), 207–217.
- [13] Suyanto. (2017). Data mining Untuk Klasifikasi dan Klasterisasi Data. SpringerReference, 561–567. https://doi.org/10.1007/SpringerReference_5414.
- [14] Vlandari, R. T. (2017). *DATA MINING (Teori dan Aplikasi Rapidminer)*. 54.