ANALISIS SENTIMEN MASYARAKAT TERHADAP *BRAND* SMARTFREN MENGGUNAKAN *NAIVE BAYES CLASSIFIER* DI FORUM KASKUS

PUBLIC SENTIMENT ANALYSIS OF SMARTFREN BRAND USING NAIVE BAYES CLASSIFIER ON KASKUS FORUM

Faishal Nuruz Zuhri[1], Andry Alamsyah, S.Si., M.Sc.[2]

[1,2]Prodi S1 Manajemen Bisnis Telekomunikasi dan Informatika, Fakultas Ekonomi dan Bisnis, Universitas Telkom

[1]faishalnuruz@gmail.com, [2]andryalamsyah@gmail.com

Abstrak

Analisis sentimen atau opinion mining merupakan analisis yang bertujuan untuk melihat sentimen masyarakat atau kelompok mengenai entitas tertentu. Sentimen yang diekspresikan masyarakat bisa dalam bentuk positif atau negatif. Penelitian ini dilakukan untuk mengklasifikasikan sentimen masyarakat terhadap brand Smartfren. Smartfren merupakan salah satu perusahaan operator telekomunikasi seluler di Indonesia yang masih bertahan dengan jenis jaringan komunikasi CDMA (*Code Division Multiple Access*). Pandangan masyarakat terhadap *brand* Smartfren, tertuang dalam sebuah sentimen, baik itu positif maupun negatif. Dengan semakin berkembangnya teknologi informasi, opini atau sentimen masyarakat banyak dibagikan melalui media sosial. Penelitian ini mengambil komentar-komentar *user* terhadap *brand* Smartfren di forum Kaskus. Pengklasifikasian komentar berupa positif atau negatif menggunakan metode *naive bayes classifier* (NBC). Memvisualisasikan kata-kata menggunakan *wordcloud*. Dari hasil pengujian untuk kasus pada penelitian ini didapatkan bahwa NBC dapat diimplementasikan dengan nilai akurasi 98.40%. Dari 6338 data uji, 4049 berhasil terklasifikasi kedalam sentimen positif dan 3233 data bersentimen negatif. Komentar terbanyak pada bulan januari yaitu sebanyak 1472 data.Dari proses *wordcloud*, dapat disimpulkan bahwa kata-kata dalam komentar user yang mendominasi dari hasil analisis sentimen brand Smartfren, yang bersentimen positif, antara lain smartfren, paket, true unlimited, speed, lancar dan kencang dan bersentimen negatif, antara lain paket, smartfren, beli, true unlimited, lamban, masalah dan gangguan.

Kata Kunci : analisis sentimen; smartfren; *word cloud*; *naive bayes classifier*.

Abstract

Sentiment analysis or opinion mining is an analysis that aims to see public sentiment or group about a particular entity. The sentiment expressed in the form of society could be positive or negative. This research conducted to classify public sentiment on brand of Smartfren. Smartfren is one of the mobile telecommunications operator in Indonesia that still persist with the type of communications network of CDMA (Code Division Multiple Access). Public perceptions of the brand Smartfren, contained in a sentiment, either positive or negative. With the development of information technology, public opinion or sentiment was widely shared through social media. Classifying of comment either positive or negative using the Naive Bayes classifier (NBC). Visualizing the words using a word cloud. The results of the study will show that the sentiment of many in the form of positive or negative and will indicate what words are dominant. From the test results for cases in this study found that NBC can be implemented with a value of 98.40% accuracy. Test data from the 6338, 4049 successfully classified into positive sentiment and the 3233 data is negative grudges. Most commented in January that as many as 1472 data. From wordcloud process, it can be concluded that the words in a user comment which dominated from the analysis Smartfren brand sentiment, the positive grudges, among others smartfren, package, true unlimited, speed, smooth and toned and negative grudges, among other packages, smartfren, buy, true unlimited, slow, problems and disorders.

Keywords : sentiment analysis; smartfren; word cloud; naive bayes classifier.

1. Pendahuluan

Zaman yang terus bergerak maju dan berkembang dengan segala inovasi dan penemuan-penemuan yang ada. Teknologi informasi berperan sangat penting dalam pesatnya perkembangan. Jika dilihat lebih mendalam, perkembangan yang ada tidak luput dari peran internet. Kegunaan internet cukup beragam dan bisa dikatakan multifungsi, seperti komunikasi, jual-beli, bisnis, bertukar data dan informasi, dll. Internet sudah menjadi sumber informasi aktual dan universal yang bisa digunakan siapa saja.

Data terbaru dari We Are Social per maret 2015, menyebutkan, pengguna internet aktif di seluruh dunia kini telah mencapai angka 3,03 miliar dengan angka pertumbuhan dari tahun 2014 bertumbuh hingga 7,6 persen.[1] Pertumbuhan pengguna internet ini juga berpengaruh terhadap pertumbuhan pengguna media sosial dan *smartphone*. Sekali *online*, seseorang di negara berkembang sangat menginginkan untuk interaksi sosial. Mayoritas pengguna internet sering menggunakan media sosial, seperti facebook, twitter, dan instagram.

Untuk menunjang masyarakat melek teknologi di seluruh Indonesia, khususnya bidang telekomunikasi, Pemerintah melalui Kementerian Komunikasi dan Informatika (Kominfo), salah satu langkahnya yaitu dengan menyediakan perusahaan Operator Seluler atau ISP (*Internet Service Provider*) *Mobile Broadband*.[2] Pelayanan internet yang disediakan perusahaan operator seluler sudah menjangkau wilayah terpencil di sudut-sudut Indonesia.

Telkomsel, Indosat Ooredoo, XL, Hutchison 3 Indonesia dan Smartfren merupakan perusahaan operator telekomunikasi seluler yang ada di Indonesia. Dari kelima perusahaan tersebut, keempat diantaranya bergerak pada jenis jaringan komunikasi GSM (*Global System for Mobile*), hanya Smartfren saja yang sampai saat ini masih bertahan dengan jenis jaringan komunikasi CDMA (*Code Division Multiple Access*). Pada tahun 2014, Crowd Voice mencatat bahwa empat operator CDMA, yaitu Flexi, StarOne, Hepi, dan Esia telah memilih untuk menutup layanan bisnis.[3] Selain menjadi perusahaan operator telekomunikasi seluler, Smartfren juga menyediakan produk berupa smartphone, modem, dan router.

Laporan di tahun 2015, maka bukan hal yang tidak mungkin Smartfren bisa bersaing dengan operator GSM yang merajalela di Indonesia. Inovasi dan peningkatan pelayanan dari Smartfren akan meraih banyak lagi pelanggan dan pendapatan. Tetapi, ketika melihat dari sisi pelanggan, bukan tidak mungkin jika banyak pelanggan belum puas dengan pelayanan Smartfren, mengingat jumlah pelanggan yang berhasil diraih di tahun 2015 lebih rendah dari tahun 2014. Dengan meningkatnya pengguna media sosial, para pelanggan dapat memberi kesan, memberi saran, mengulas, mengeluh, mengkritik, bahkan menyumpah ketika sedang *online*. Sifat media sosial yang terbuka membuat para pelanggan bebas mengekspresikan pendapatnya. Dan ketika ada pelanggan yang komplain, maka perusahaan harus meresponnya dengan cepat. Saat ini, ketika konsumen kecewa, mereka akan langsung komplain dan bercerita di media sosial. Hal seperti ini seringkali disebut dengan sentimen.

Media sosial seringkali menjadi sumber data oleh peneliti untuk dianalisa. Kaskus menjadi media sosial yang sering digunakan oleh masyarakat Indonesia. Kaskus juga memiliki forum untuk para *user* berdiskusi terhadap suatu hal. Forum juga menjadi wadah untuk menyampaikan informasi dan opini *user* untuk para pembaca Kaskus. Forum Kaskus memiliki *thread-thread* untuk topik tertentu yang ingin diperbincangkan oleh *user*. *User* juga dapat membuat *thread* tentang suatu perusahaan atau *brand*. *Thread* tentang Smartfren sudah mencapai lebih dari 4.000 komentar atau lebih dari 200 halaman.[4] Banyak komentar yang berisi opini atau sentimen oleh *user*, sehingga perusahaan dapat secara *real-time* mengetahui karakteristik konsumen, persepsi kualitas *brand*, tingkat kepuasan serta memanfaatkannya untuk strategi pemasaran.

Metode klasifikasi yang sering digunakan dan efektif untuk klasifikasi teks dan sudah diadaptasi untuk analisis sentimen yaitu *Naive Bayes Classifier* dan S*upport Vector Machine* (SVM), akan tetapi proses *Naive Bayes* lebih cepat daripada SVM. Hasil dari analisis sentimen dan visualisasi kata menggunakan *word cloud* akan menjadi acuan bagi perusahaan untuk meningkatkan kualitas produk atau jasa agar memberikan layanan terbaik bagi konsumen, dan pada akhirnya mendapatkan persepsi positif dari konsumen.

Tujuan dari penelitian ini yaitu untuk mengetahui sentimen masyarakat terhadap *brand* Smartfren menggunakan metode *Naive Bayes Classifier* pada komentar di forum Kaskus dan untuk mengetahui kata-kata apa yang sering muncul terhadap *brand* Smartfren pada kolom komentar di forum Kaskus.

2. Tinjauan Pustaka

2.1 Dasar Teori

2.1.1 Merek

Merek (*Brand*) merupakan nama, istilah, tanda, simbol, atau rancangan, atau kombinasi dari semuanya, yang dimaksudkan untuk mengidentifikasi sebuah barang atau jasa dari layanan salah satu penjual atau kelompok penjual dan untuk membedakan dari barang atau jasa pesaing.[5] Sebuah merek adalah produk atau jasa yang memiliki perbedaan dimensi dan dengan cara tertentu dari produk atau jasa lain yang dirancang untuk memuaskan pelanggan dengan kebutuhan yang sama.

2.1.2 *Big Data*

Big Data bukan merupakan teknologi tunggal tetapi kombinasi dari teknologi lama dan baru, yang membantu perusahaan mendapatkan informasi untuk ditindaklanjuti. Oleh karena itu, Big Data adalah kemampuan untuk

mengelola banyak data yang berbeda, dengan kecepatan yang tepat, dan dalam kerangka waktu yang tepat untuk memungkinkan analisis *real-time* dan membuat data menjadi informasi sebagai dasar untuk tindakan selanjutnya.[6] Big Data dibagi menjadi tiga karakteristik, yaitu *volume* (berapa banyak data), *velocity* (seberapa cepat data yang diolah, dan v*ariety* (berbagai jenis data).

2.1.3 *Data Mining*

Proses ekstrasi data menjadi sebuah informasi ini disebut sebagai data mining.[7] Data mining merupakan cara yang tepat untuk mengolah data menjadi sebuah informasi, karena dapat menangani data dengan jumlah yang sangat besar.

2.1.4 Analisis Sentimen

Analisis sentimen atau bisa disebut juga dengan *opinion mining*, adalah bidang studi dari data mining yang mempunyai tujuan untuk menganalisis, memahami, mengolah, dan mengekstrak data tekstual yang dapat berupa opini, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas seperti produk, servis, organisasi, individu, dan topik tertentu.[8]

2.1.5 Klasifikasi

Klasifikasi adalah suatu bentuk analisis data yang mengekstrak model untuk menggambarkan kelas data yang penting, seperti model yang disebut klasifikasi, tujuannya untuk memprediksi kategori (diskrit, unordered) label kelas.[7]

2.1.6 *Word Cloud*

Word cloud merupakan perkembangan dari situs jejaring sosial berbasis web. *Word cloud* juga disebut *tag cloud* atau *weighted list*, yaitu gambaran visual dari tabulasi frekuensi kata-kata dalam setiap bahan tertulis yang dipilih, seperti catatan kuliah, teks dalam buku atau sebuah situs internet.[9]

2.1.7 *Preprocessing*

Prepocessing peru dilakukan supaya menjadi data yang berkualitas sebelum masuk ke proses klasifikasi. Ada beberapa faktor untuk menjadi data yang berkualitas, yaitu akurasi, kelengkapan, konsisten, aktual, terpercaya, dan terdefinisi.

Di bawah ini tahapan yang dilakukan pada *preprocessing* komentar:[10]

a) *Tokenization* adalah proses pemecahan kalimat menjadi kata-kata, frase, atau simbol. Daftar token kemudian digunakan sebagai *input data* untuk proses selanjutnya.
b) *Filtering Stopwords* diterapkan untuk membersihkan data dari sebuah karakter-karakter atau kata yang tidak berguna, seperti, tanda baca dan preposisi.
c) *Stemming* adalah proses pembersihan sebuah prefiks, sufiks, infiks, dan konfiks untuk menggabungkan kata yang berasal dari akar yang sama, seperti, jadi, menjadi, menjadikan.

2.1.8 Pembobotan

Pembobotan dengan penggabungan *term frequency* dan *inverse document frequency*, untuk menghasilkan bobot yang komposit untuk setiap *term* kata dalam setiap dokumen, bisa disebut juga dengan *tf-idf*.[11]

2.1.9 Naive Bayes Classifier

Naive bayes merupakan metode pembelajaran secara probabilitas statistik. Asumsi yang harus dipenuhi yaitu adanya independensi dalam variable bebas atau kemunculan sebuah kata tidak mempengaruhi kemunculan kata lainnya, ataupun sebaliknya.[7] Adapun kelebihan dan kekurangan yang dimiliki oleh *naive bayes*, antara lain;

Metode NBC menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap pengujian. Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan *term*, yaitu kata yang mungkin muncul dalam setiap dokumen sampel yang sedapat mungkin dapat menjadi representasi dokumen. Kemudian penentuan probabilitas *prior* bagi setiap kategori berdasarkan sampel dokumen. Pada tahap pengujian, ditentukan nilai kategori dari suatu dokumen berdasarkan *term* yang muncul dalam dokumen yang diklasifikasi.

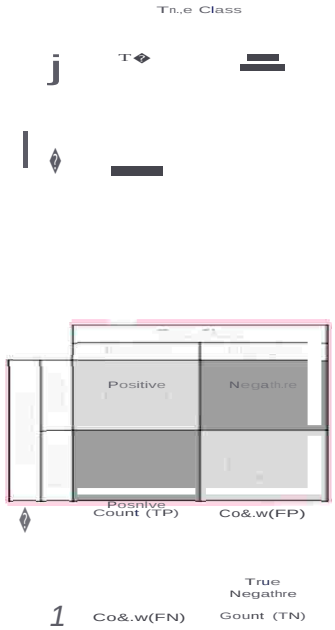
2.1.10 Validasi dan Evaluasi

Dalam rangka meminimalkan bias yang terkait dengan pengambilan sampel acak dari sampel data training dalam membandingkan akurasi prediksi dari dua atau lebih metode menggunakan metodologi yang disebut *k-fold cross validation*. Dalam *k-fold cross validation*, data awal dibagi secara acak menjadi k subset (*fold*), yaitu, ,

Pada iterasi i, data subset ke-i digunakan sebagai dataset pengujian, dan partisi yang tersisa secara kolektif digunakan untuk melatih model agar mendapatkan suatu model klasifikasi yang nantinya digunakan dalam proses pengujian. Model klasifikasi dilatih dan diuji k kali. Setiap kali dilatih pada semua kecuali satu *fold* dan diuji pada *fold* tunggal yang tersisa.

K-fold cross validation juga disebut 10 *k-fold cross validation*, karena k mengambil nilai 10 yang telah menjadi praktek paling umum. Bahkan, studi empiris menunjukkan bahwa sepuluh tampaknya menjadi jumlah yang optimal dari *fold*, karena bias dan varians relatif rendah.[12]

Evaluasi performansi dilakukan untuk menguji hasil dari klasifikasi dengan mengukur nilai performansi dari sistem yang telah dibuat. Parameter pengujian yang digunakan untuk evaluasi yaitu akurasi. Metode pengukuran yang digunakan adalah *confusion matrix* (*classification matrix or a contingency table*).[12]



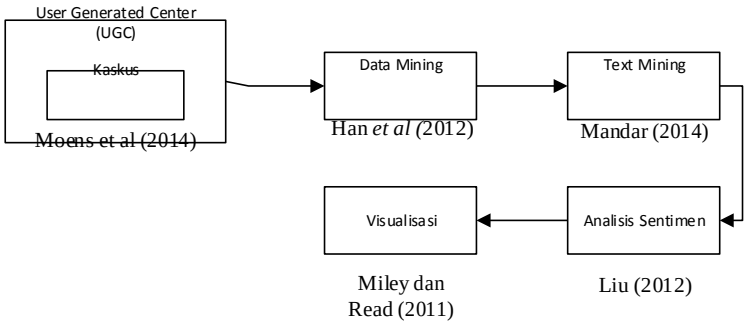
Gambar 1. *Confusion Matrix*.[12]

a) TP (*True Positives*): kelas yang diprediksi positif dan diprediksi oleh sistem klasifikasi kelas positif
b) TN (*True Negatives*): kelas yang diprediksi negatif dan diprediksi oleh sistem klasifikasi kelas negatif
c) FP (*False Positives*): kelas negatif tetapi diprediksi oleh sistem klasifikasi kelas positif
d) FN (*False Negatives*): kelas positif tetapi diprediksi oleh sistem klasifikasi kelas negatif

Precision dan Recall juga banyak digunakan dalam klasifikasi.[7] Tujuannya untuk menentukan apakah data yang dipakai sudah memenuhi keinginan atau belum. *Precision* adalah tingkat ketepatan informasi yang dianggap relevan oleh pengguna dengan jawaban yang diberikan oleh sistem. *Recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah data yang dianggap relevan atau sesuai dengan pengguna.

2.1 Kerangka Pemikiran

Berdasarkan latar belakang, perumusan masalah, tujan penelitian, dan tinjauan pustaka yang telah dijelaskan sebelumnya, maka kerangka pemikiran dalam penelitian ini akan digambarkan sebagai berikut:



Gambar 2. Kerangka Pemikiran

3 Penelitian dan Pembahasan

3.1 Hasil Penelitian

Dataset yang digunakan dalam penelitian ini merupakan opini berbahasa Indonesia mengenai *brand* Smartfren di forum Kaskus. Dataset berupa komentar yang merupakan ungkapan masyarakat mengenai suatu hal melalui media sosial. Proses *crawling* dilakukan secara otomatis melalui website import.io. Dataset yang didapat oleh peneliti yaitu sebanyak 7386 kalimat opini.

Skenario pertama untuk mengukur akurasi dari *corpus* yaitu dengan komposisi data latih tidak seimbang atau acak.

Tabel 1. Data Acak

Data Training	Positif	Negatif	Akurasi
1000	628	372	89.00 %
2000	1295	705	85.65 %
3000	1845	1155	87.30 %
4000	2454	1546	87.28 %
5000	3068	1932	85.90 %
6000	3736	2264	84.50 %
7000	4331	2669	83.14 %

Tingi rendahnya nilai akurasi dipengaruhi oleh banyaknya data yang dievaluasi. Seperti yang terlihat pada tabel 1, semakin banyak data latih yang dievaluasi tingkat akurasi yang dihasilkan cenderung menurun. Dengan 1000 data latih, tingkat akurasi yang berhasil didapatkan merupakan yang tertinggi yaitu sebesar 89.00 % dan yang kedua adalah dengan 3000 data latih berhasil mendapatkan akurasi sebesar 87.30 %. Hal ini dikarenakan jika semakin banyak *dataset* yang digunakan untuk data latih, sehingga keberagaman data semakin tinggi yang sangat berpengaruh pada hasil evaluasi.

Skenario kedua untuk mengukur akurasi dari *corpus* yaitu dengan komposisi data latih seimbang atau jumlah sentimen positif dan negatif sama.

Tabel 2. Data Seimbang

Data Training	Akurasi
1000	98.40 %
2000	92.45 %
3000	91.00 %
4000	90.67 %
5000	88.34 %

Pada tabel 2, jumlah data latih yang dievaluasi memiliki kecenderungan yaitu, jika semakin banyak data latih maka akurasi yang dihasilkan juga menurun. Data latih sejumlah 1000 data menghasilkan akurasi yang tertinggi yaitu sebesar 98.40 %.

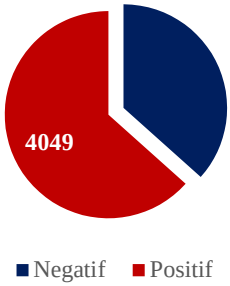
Untuk pembuatan model *machine learning*, peneliti memilih data latih dengan komposisi seimbang sebanyak 1000 data, karena memiliki akurasi tertinggi diantara proses pengevaluasian *corpus*, baik data latih dengan komposisi acak maupun seimbang. Setelah proses pengevaluasian, juga diperoleh *vector* dengan 118 atribut kata yang dilengkapi pembobotan TF-IDF, yang merupakan fitur dari model untuk menguji data uji. Atribut kata yang diperoleh untuk dijadikan sebagai fitur pembeda antar sentimen.

Gambar 3. *Confusion Matrix*

Data latih dengan komposisi sentimen yang sama dapat dilihat pada gambar 3 yang merupakan *confusion matrix* dari proses evaluasi. Dari 500 data bersentimen positif dan data bersentimen negatif, didapatkan TP (*true* positives) atau sentimen positif dan diprediksi oleh sistem klasifikasi bersentimen positif sebanyak 482 data, TN (*True Negatives*) atau sentimen negatif dan diprediksi oleh sistem klasifikasi bersentimen negatif sebanyak 482 data, FP (*False Positives*) atau sentimen negatif tetapi diprediksi oleh sistem klasifikasi bersentimen positif atau dapat dikatakan salah prediksi sebanyak 8 data, dan FN (*False Negatives*) atau sentimen positif tetapi diprediksi oleh sistem klasifikasi bersentimen negatif atau dapat dikatakan salah prediksi sebanyak 8 data.

Dari hasil perhitungan dari proses evaluasi, didapatkan *precision* sebesar 98.42 %, *recall* sebesar 98.40 %, tingkat akurasi sebesar 98.40 %, dan nilai koefisien kappa sebesar 0.948.

Setelah pembuatan model *machine learning* dari data latih, maka proses selanjutnya yaitu menguji data uji untuk mengetahui sentimen apa yang paling banyak didalam *dataset* yang telah diambil. Jumlah total data dalam *dataset* adalah sebanyak 7388 data dan jumlah data latih sebanyak 1000 data, maka data uji sebanyak 6388.



Gambar 4. Hasil klasifikasi data uji

Hasil pengklasifikasian dapat dilihat pada gambar 4, dari 6388 data yang diuji, yang terklasifikasi kedalam sentimen positif sebanyak 4049 data dan sentimen negatif sebanyak 2339 data.

Wordcloud dilakukan setelah hasil klasifikasi telah diketahui. Peneliti membuat dua *wordcloud*, yaitu sentimen positif dan sentimen negatif menurut hasil dari proses pengklasifikasian.

Gambar 5. *Wordcloud* sentimen negatif dan positif

Dari hasil *wordcloud* tersebut, didapatkan 15 kata dominan dari masing-masing sentimen yang sudah melalui proses klasifikasi.

3.2 Pembahasan Hasil Penelitian

3.2.1 Hasil Klasifikasi

Proses pengklasifikasian data uji menjadi 2 kelas, yaitu, sentimen positif dan negatif. Hasil klasifikasi data uji sangat dipengaruhi oleh model yang telah dibuat dari data latih. Dari proses evaluasi dan validasi data latih, didapatkan precision sebesar 98.42 %, recall sebesar 98.40 %, dan tingkat akurasi sebesar 98.40 %. Hal ini menunjukkan bahwa sistem dapat memisahkan data bersentimen positif dan negatif dengan sangat baik. Dan juga memperoleh nilai koefisien kappa sebesar 0.948, maka menunjukkan kesepakatan yang sangat kuat antar penguji karena mendekati nilai 1, maka instrumen data latih dapat dikatakan reliabel.

Dari skala 0% - 100%, nilai precision yang didapatkan dari model klasifikasi sebesar 98.42% dan nilai recall sebesar 98.42%, sehingga dapat diketahui bahwa nilai precision dan recall sama. Tingkat keefektifan dari nilai precision dan recall dari sistem temu kembali informasi, dapat dikatakan efektif. Dari 500 data sentimen positif, hasilnya 492 data yang dideteksi sebagai sentimen positif dan 8 data dideteksi sebagai sentimen negatif. Begitu juga hasil dari 500 data sentimen negatif. Hal ini dikarenakan ada komentar yang hampir sama dengan komentar yang lain di sentimen yang berbeda, jadi komentar tersebut dapat menimbulkan bias.

Model yang telah dibuat memperoleh vector dengan 118 atribut kata yang dilengkapi pembobotan TF-IDF, yang merupakan fitur dari model untuk menguji data uji. Atribut kata yang diperoleh untuk dijadikan sebagai fitur pembeda antar sentimen. Selanjutnya model tersebut, dijadikan machine learning untuk pengklasifikasian data uji atau data baru.

Data uji yang telah diuji di dalam machine learning, berhasil mengklasifikasikan data uji yang telah dipilah. Kelas sentimen positif mendapatkan angka yang cukup signifikan, dari 6388 data uji, 4049 data masuk ke dalam sentimen positif dan hanya 2339 data masuk ke dalam sentimen negatif.

Tabel 3. Hasil klasifikasi oleh *machine learning*

Komentar	Sentimen
Sejawa, ane bawa keliling dr pacitan ke telaga sarangan, trus ke pantura jateng. Sinyal sangat lancar...	Positif
udah 2 hari setiap jam 12 siang sampe jam 6 pagi cma dpt 0.02mbps ini kenapa ya? padahal sinyal Full.... Ane d daerah Cirebon... parah bgt buka google aja kaga sanggup, oh iya dulu sebelum gangguan gini setiap isi paket Booster yg 10rb itu speed bisa sampai 11Mbps tp koq skrng normal lagi jd 500kbps ya	Negatif

4.2.2 Ringkasan kata menggunakan *word cloud*

Tabel 4. Jumlah kata dominan (5 terbanyak)

Sentimen	Kata	Jumlah
Negatif	Paket	1833
	Smartfren	1652
	Beli	971
	True Unlimited	885
	Ribu	798
Positif	Smartfren	1559
	Paket	1164
	True Unlimited	929
	Speed	682
	Gigabyte	662

Berdasarkan gambar diatas menunjukkan kata yang paling sering digunakan oleh user adalah paket yang bersentimen negatif, padahal total data sentimen negatif hanya 2339 komentar. Hal ini menunjukkan bahwa kata "paket" yang notabennya sebuah produk dari smartfren dapat menurunkan nilai brand dari smartfren sendiri. Smartfren yang bergerak dalam bidang telekomunikasi menawarkan paket internet, telepon, sms, dll., tetapi juga sering mendapatkan respon negatif dari pengugunanya.

Sedangkan pada data sentimen positif, kata "paket" berada pada urutan kedua tertinggi sebanyak 1164 kata dengan total keseluruhan data bersentimen positif sebesar 4049 komentar. Jumlah perbandingan antara kata "paket" di sentimen positif dengan negatif sangat jauh berbeda. Padahal data bersentimen positif jauh lebih banyak daripada data bersentimen negatif. Angka yang cukup signifikan untuk masyakarat mempersepsikan "paket" brand Smartfren kurang bagus.

Dalam jangka panjang, rasa percaya diri pengguna dalam mengambil keputusan pembelian akan menurun, karena memiliki pengalaman masa lalu yang kurang baik. Jadi, perusahaan harus dapat meningkatkan kualitas produk dan layanannya, agar kata-kata negatif yang dominan menjadi berkurang.

4 Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil pengujian dan analisis sentimen menggunakan naive bayes classifier dalam bahasa indonesia terhadap brand Smartfren, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Analisis sentimen terhadap data Kaskus mengenai *brand* Smartfren dilakukan dengan metode *Naive Bayes Classifier*, dengan menggunakan 1000 data latih, didapatkan *precision* sebesar 98.42 %, *recall* sebesar 98.40 %, dan tingkat akurasi sebesar 98.40 %. Hal ini menunjukkan bahwa sistem dapat memisahkan data bersentimen positif dan negatif dengan sangat baik. Dan juga memperoleh nilai

koefisien kappa sebesar 0.948, maka menunjukkan kesepakatan yang sangat kuat antar penguji karena mendekati nilai 1, maka instrumen data latih dapat dikatakan reliabel.

2. Respon atau sentimen masyarakat terhadap *brand* Smartfren di Kaskus menggunakan data uji sebanyak 6388 data. Hasil klasifikasi menunjukkan 4049 data masuk ke dalam sentimen positif dan hanya 2339 data masuk ke dalam sentimen negatif. Komentar terbanyak pada bulan januari yaitu sebanyak 1472 data.
3. Berdasarkan *wordcloud*, dapat disimpulkan bahwa kata-kata dalam komentar *user* yang mendominasi dari hasil analisis sentimen *brand* Smartfren, yaitu:
   - Sentimen positif : smartfren, paket, true unlimited, speed, lancar dan kencang
   - Sentimen negatif : paket, smartfren, beli, true unlimited, lamban, masalah dan gangguan.

5.2 Saran

5.2.1 Aspek Teoritis Untuk Penelitian Selanjutnya

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, peneliti selanjutnya diharapkan untuk menggunakan metode klasifikasi yang lain yaitu Support Vector Machine.

5.2.2 Aspek Praktis Untuk Perusahaan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan:

1. Sentimen negatif masyarakat terhadap *brand* smartfren masih cukup tinggi, diharapkan pihak smartfren mempebaiki kualitas, baik produk atau layanannya, karena persaingan operator seluler di Indonesia cukup tinggi dengan 4 kompetitor.
2. Pihak Smartfren diharapkan meningkatkan kualiatas produk dan layanan dari segi "paket" yang ditawarkan, agar mendapatkan respon positif dari pengguna, yang dapat mempengaruhi nilai dari *Brand* Smartfren sendiri.

[1] We Are Social. (2016). *Digital, Social & Mobile in APAC 2015*. [Online]. http://wearesocial.com/sg/special-reports/digital-social-mobile-in-apac-in-2015 [2 Mei 2016]

[2] Kominfo. (2016). *Kemkominfo Terus Berdayakan Masyarakat Semakin Melek TIK*. [Online]. https://kominfo.go.id/index.php/content/detail/5890/Kemkominfo+Terus+Berdayakan+Masyarakat+Semakin+Melek+TIK/0/berita_satker [2 Mei 2016]

[3] CrowdVoice. (2014). *4 dari 5 Operator CDMA Akan Ditutup*. [Online]. https://id.crowdvoice.com/posts/4-dari-5-operator-cdma-akan-ditutup-2uY2. [2 Mei 2016]

[4] Kaskus. (2016). *About KASKUS*. [Online]. https://help.kaskus.co.id/sejarah_kaskus.php [2 Mei 2016]

[5] Kotler, Philip. (2012). *PRINSIP-PRINSIP PEMASARAN Jilid 1 Edisi 12*.Bandung: Erlangga.

[6] Hurwitz, Judith., Nugent, Alan., Halper, Fern., dan Kaufman, Marcia. (2013). *Big Data for Dummies*. John Wiley & Sons, Inc.

[7] Han, J., Kamber, M., dan Pei, J., (2012). *Data Mining: Concepts and Techniques (3rd ed.).* Morgan Kaufmann.

[8] Liu, Bing. (2012*). Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publisher.

[9] Miley, Frances., dan Andrew, Read,. (2011). *Using word clouds to develop proactive learners. Journal of the Scholarship of Teaching and Learning*. 11(2), 91-110. Retrrieved from University of New South Wales at the Australian Defence Force Academy and University of Canberra.

[10] İkiz, Aysun Kapucugil., dan Özdağoğlu, Güzin. (2015). *Text Mining As A Supporting Process For VoC Clarification*. Alphanumeric Journal, 3(1), 025–040. Retrieved from The Journal of Operations Research, Statistics, Econometrics and Management Information Systems.

[11] Manning, Christopher D., Raghavan, Prabhakar., dan Schütze, Hinrich. (2009). *An Introduction to Information Retrieval*. Cambridge University Press

[12] Olson, David L., dan Delen, Dursun. (2008). *Advanced Data Mining Techniques*. Heidelberg: Springer.