

ANALISIS SENTIMEN DI MEDIA SOSIAL TWITTER TERHADAP BRAND XIAOMI DENGAN ALGORITMA NAÏVE BAYES CLASSIFIER

Zakaria Mursid¹⁾, Anggit Dwi Hartanto²⁾,

¹⁾ Informatika Universitas AMIKOM Yogyakarta

²⁾ Informatika Universitas AMIKOM Yogyakarta

Jl Ringroad Utara, Condongcatur, Depok, Sleman, Yogyakarta Indonesia 55283

Email : zakaria.mursid@students.amikom.ac.id¹⁾, anggit@amikom.ac.id¹⁾

Abstract - *Xiaomi is a Chinese electronics company. Chief Financial Officer (CFO) Xiaomi Chew Shouzi during his visit to Indonesia in November 2018 stated that Indonesia had an important role for Xiaomi's business growth. Xiaomi must be competitive and differentiate itself from the competition as new challengers continue to enter the market. Sentiment analysis is a tool to produce a collection of results that want to get the attributes (quality, features, etc.) and the process of obtaining the results of their opinions. Social media users are free to express their opinions. Twitter is a social media that is favored by the people of Indonesia. This research tries to utilize what users write or post on Twitter social media or better known as tweets. Tweets will be processed with text mining and processed using the Naïve Bayes Classifier algorithm.*

Keywords - Sentiment analysis, Social media, Naïve Bayes Classifier

1. Pendahuluan

Xiaomi adalah perusahaan yang bergerak di bidang elektronik dan software yang berasal dari Beijing, Cina. Xiaomi mendesain, mengembangkan, dan menjual smartphone, aplikasi ponsel, laptop dan produk lain yang berhubungan dengan elektronik. Xiaomi didirikan pada tahun 2010, perusahaan ini dengan cepat telah menjadi salah satu perusahaan teknologi terkemuka di Cina. Perusahaan ini kini bernilai lebih dari 10 miliar USD dan memiliki lebih dari 3000 karyawan [1]

Chief Financial Officer (CFO) Xiaomi Chew Shouzi dalam kunjungannya ke Indonesia pada bulan November 2018 menyatakan bahwa Indonesia punya peran penting bagi pertumbuhan bisnis Xiaomi. Menurut CFO Xiaomi tersebut, Indonesia akan menjadi pasar perekonomian terpenting di dunia di masa yang akan datang, dan karenanya Indonesia adalah salah satu fokus ekspansi internasional yang paling penting [2]

Dalam upaya lebih kompetitif dan membedakan dirinya dengan pesaing di pasar ponsel pintar. Xiaomi perlu memperhatikan / mendengarkan konsumennya di Indonesia. Dengan mengetahui sentimen masyarakat Indonesia dan istilah yang seringkali muncul / dibicarakan di media sosial terhadap merk / brand. Xiaomi dapat

mengetahui sentiment, pendapat, keinginan masyarakat Indonesia terhadap brand, hal itu dapat digunakan sebagai acuan bagi perusahaan untuk berbenah diri dan meningkatkan kualitas produk atau jasa agar memberikan layanan terbaik bagi konsumen, atau menyusun strategi baru apabila dibutuhkan untuk memperkuat pasarnya di Indonesia. Oleh karena itu penulis merasa penting untuk mengajukan penelitian dengan judul Analisis Sentimen di Media Sosial Twitter Terhadap Brand Xiaomi dengan Algoritma Naïve Bayes Classifier.

2. Dasar Teori

2.1 Brand

Brand atau merk adalah tanda yang dapat ditampilkan secara grafis berupa gambar, logo, nama, kata, huruf, angka, susunan warna, dalam bentuk 2 (dua) dimensi dan/atau 3 (tiga) dimensi, suara, hologram, atau kombinasi dari 2 (dua) atau lebih unsur tersebut untuk membedakan barang dan atau jasa yang diproduksi oleh orang atau badan hukum dalam kegiatan perdagangan barang dan/atau jasa. [3]

2.2 Text mining

Text mining juga dikenal sebagai data mining teks atau penemuan pengetahuan dari database tekstual. Sesuai dengan buku The Text Mining Handbook, text mining dapat didefinisikan sebagai suatu proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponen-komponen dalam data mining. Tujuan dari text mining adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan dalam text mining adalah sekumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Adapun tugas khusus dari text mining antara lain yaitu pengkategorisasian teks dan pengelompokan teks. [4]

2.3 Analisis Sentimen

Analisis sentimen adalah proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan

untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung beropini negatif atau positif. [5]

Naïve Bayes Classifier adalah salah satu metode yang populer digunakan untuk keperluan data mining karena kemudahan penggunaannya [6] serta waktu pemrosesannya yang cepat, mudah diimplementasikan dengan strukturnya yang cukup sederhana dan tingkat efektifitas yang tinggi. [7]

2.4 Evaluasi

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 yang perhitungannya dari tabel confusion matrix (matriks klasifikasi atau tabelkontigensi). [8] Tabel confusion matrix Positif, Netral & Negatif dapat dilihat pada Tabel 1 sebagai berikut :

Tabel 1 Confusion Matrix Positif, Netral & Negatif

		Predicted		
		Positive	Neutral	Negative
Actual	Positive	True	False	False
		Positive	Positif	Positive
		(TP)	(FP)	(FP)
	Neutral	False	True	False
		Neutral	Neutral	Neutral
		(FNeu)	(TNeu)	(FNeu)
	Negative	False	False	True
		Negative	Negative	Negative
		(FNeg)	(FNeg)	(TNeg)

Matriks tersebut memiliki enam nilai yang dijadikan acuan dalam perhitungan, yaitu :

- A. True Positive (TP) adalah kelas yang diprediksi positif dan faktanya positive.
- B. True Neutral (TNeu) adalah kelas yang diprediksi netral dan faktanya netral.
- C. True Negative (TNeg), adalah kelas yang diprediksi negatif dan faktanya negatif.
- D. False Positif (FP) adalah kelas yang diprediksi positif dan faktanya tidak positif
- E. False Neutral (FNeu) adalah kelas yang diprediksi netral dn faktanya tidak netral

F. False Negative (FNeg) adalah kelas yang diprediksi negatif dan faktanya tidak negatif

Alternatif dalam menilai sebuah sistem adalah dengan akurasi. Akurasi adalah ketepatan suatu sistem melakukan klasifikasi yang benar. Untuk menghitung nilai akurasi prediksi kelas positif, netral dan negatif menggunakan Persamaan 1 sebagai berikut

$$Akurasi = \frac{TP+TNeu+TNeg}{(TP+TNeu+TNeg+FP+FNeu+FNeg)} \times 100\%$$

(1)

3. Metode Penelitian

3.1 Metode Pengumpulan Data

Dataset yang digunakan merupakan teks berbahasa Indonesia yang diambil dari <https://twitter.com/> dengan kata kunci Xiaomi. Proses pengumpulan data / *crawling* dibedakan menjadi dua yaitu : data latih dan data uji.

3.1.1 Data Latih

Data latih diambil dalam kurun waktu 3 bulan terhitung mulai Maret – Juni 2019. . Sekumpulan tweet yang sudah dikumpulkan kemudian diberi label sentiment / diklasifikasikan secara manual menjadi tiga kategori yaitu positif, netral dan negatif yang berjumlah 1060 tweet. Dengan komposisi jumlah tweet masing-masing kategori sentiment yang dapat dilihat pada Tabel 2

Tabel 2. Komposisi Data Uji

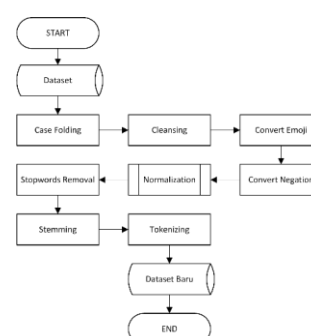
Positif	Netral	Negatif
425	318	317

3.1.2 Data Uji

Data uji pada penelitian ini merupakan data yang diambil 7 hari kebelakang semenjak fungsi *crawling* dijalankan.

3.2 Preprocessing

Sebelum data diolah lebih lanjut, perlu dilakukan preprocessing. Tahapan preprocessing adalah proses penyeleksian data yang akan diproses, dari semula merupakan data mentah yang masih kotor menjadi data bersih berkualitas yang diharapkan akan menghasilkan keluaran yang akurat. Diagram alur tahapan-tahapan proses preprocessing dapat dilihat pada Gambar 1



Gambar 1. Diagram Alur Preprocessing

3.3 Pembobotan TF-IDF

Metode TF-IDF merupakan metode pembobotan term yang banyak digunakan sebagai metode pembandingan terhadap metode pembobotan baru. Pada metode ini, perhitungan bobot term t dalam sebuah dokumen dilakukan dengan mengalikan nilai *Term Frequency* dengan *Inverse Document Frequency*.

$$w_{ij} = tf_{ij} \times idf_j \quad (2)$$

$$w_{ij} = tf_{ij} \times \log \frac{D}{df_j} \quad (3)$$

Keterangan :

w_{ij} : adalah bobot term t_j terhadap dokumen d_i

tf_{ij} : adalah jumlah kemunculan term t_j dalam dokumen d_i

3.4 Naïve Bayes Classifier

Bergantung pada model probabilitasnya, Naïve Bayes Classifier dapat dilatih untuk melakukan supervised learning dengan sangat efektif. Dalam berbagai macam penerapannya, estimasi parameter untuk model Naïve Bayes menggunakan metode maximum likelihood, yang artinya pengguna dapat menggunakan model Naïve Bayes tanpa perlu mempercayai probabilitas Bayesian atau tanpa menggunakan metode Bayesian. [9]

$$P(V_j) = \frac{docs_j}{training} \quad (5)$$

$$P(Va_i|v_j) = \frac{n_i + 1}{n + kosakata}$$

Keterangan :

$docs_j$ = jumlah dokumen pada kategori j

$training$ = jumlah dokumen yang digunakan dalam

proses *training*

n_i = jumlah kemunculan kata a_i pada kategori v_j

n = jumlah kosakata yang ada pada kategori v_j

$kosakata$ = jumlah kata unik pada semua data *training*.

[10]

4. Pembahasan

Pada tahap pengujian peneliti melakukan ujicoba pengaruh seleksi pembobotan terms berdasarkan nilai bobot TF-IDF. Peneliti menemukan bahwa penyeleksian

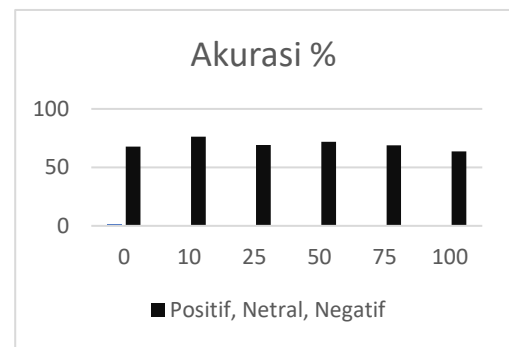
terms akan mempengaruhi tingkat akurasi. Pada tahap pengujian peneliti menggunakan data testing yang di crawling pada pertengahan bulan September 2019. Pengujian dilakukan dengan berbagai macam batas nilai minimum terms antara lain 0, 10, 25, 50, 75, 100. Hasil prediksi klasifikasi, dihitung akurasinya secara manual dengan mengimplementasikan persamaan 1.

Pada ujicoba ini peneliti menemukan bahwa penyeleksian *terms* dengan menentukan batas minimum terms berdasarkan nilai bobot TF-IDF dapat mempengaruhi nilai akurasi. Perbandingan nilai akurasi masing-masing batas minimum *terms* dapat dilihat pada Tabel 3 sebagai berikut :

Tabel 3. Hasil Perhitungan Akurasi

0	10	25	50	75	100
67.89	76.22	69.04	71.78	68.80	63.77

Atau disajikan dalam bentuk diagram lingkaran seperti pada Gambar 2 sebagai berikut :



Gambar 2. 10 Terms Yang Paling Sering Muncul

Berdasarkan data tersebut, terlihat bahwa batas minimum 10 meningkatkan akurasi sedangkan batas minimum 100 menurunkan nilai akurasi.

Halaman bobot terms dapat digunakan untuk mengetahui terms yang paling sering keluar pada koleksi data. 10 Terms yang paling sering muncul pada teks tweet berbahasa Indonesia mengenai brand Xiaomi dengan terms 'Xiaomi' sebagai pengecualian dapat dilihat pada gambar 3 berikut :

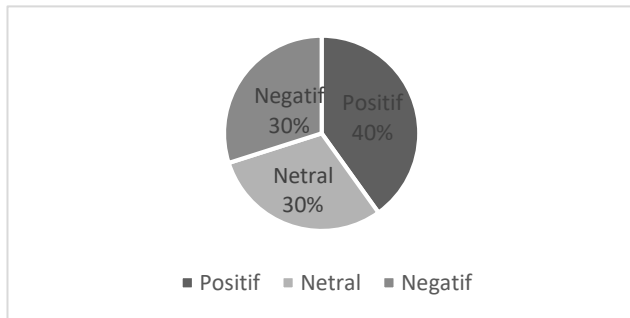
NO	Kata	Kecenderungan	TF	IDF	TF-IDF
644	handphone	NEGATIF	298	0.594	177.012
1383	redmi	POSITIF	138	0.965	133.17
1252	pake	POSITIF	122	0.972	118.584
800	kalo	POSITIF	101	1.081	109.181
18	aja	POSITIF	88	1.091	96.008
652	harga	POSITIF	78	1.144	89.232
1046	mi	POSITIF	78	1.15	89.7
1188	note	POSITIF	73	1.233	90.009
1449	samsung	NEGATIF	69	1.226	84.594
146	bell	POSITIF	67	1.247	83.549

Gambar 3. 10 Terms Yang Paling Sering Muncul

5. Penutup

5.1 Kesimpulan

Berdasarkan penelitian yang sudah dilakukan, dapat disimpulkan bahwa berdasarkan koleksi data training sentimen yang diambil antara bulan April, Mei dan Juni 2019, sentimen positif lebih mendominasi. Dengan komposisi sebagai berikut : tweet yang mempunyai sentiment positif sebanyak 425 tweet, netral sebanyak 318, dan negatif sejumlah 317. Komposisi sentiment tweet berbahasa Indonesia mengenai brand Xiaomi dapat dilihat pada Gambar 4



Gambar 4. Diagram Lingkaran Komposisi Sentimen Tweet Berbahasa Indonesia Mengenai Brand Xiaomi

10 Terms yang paling sering muncul pada teks tweet berbahasa Indonesia mengenai brand Xiaomi dengan terms 'xiaomi' sebagai pengecualian adalah : 'handphone', 'redmi', 'pake', 'kalo', 'aja', 'harga', 'mi', 'note', dan 'samsung'.

Berdasarkan data yang ada terms 'redmi', 'mi', 'note', 'pocophone' yang merupakan jajaran produk ponsel pintar Xiaomi mempunyai kecenderungan sentimen positif. Kemudian hal-hal yang berhubungan dengan ponsel pintar seperti 'spesifikasi', 'kamera', 'baterai' juga sama. Terms 'harga', 'murah' juga tercatat lebih sering digunakan pada tweet yang mempunyai sentimen positif.

Meskipun demikian, tercatat 'handphone' adalah terms yang paling sering muncul pada teks tweet berbahasa Indonesia mengenai brand Xiaomi dengan terms 'xiaomi' sebagai pengecualian. Akan tetapi terms 'handphone' yang mempunyai frekuensi kemunculan sebanyak 298 kali tersebut lebih banyak digunakan pada tweet yang mempunyai sentimen negatif dibanding positif. Kemudian tweet dengan terms 'samsung', 'iphone', 'vivo', 'oppo', 'huawei', 'nokia', 'realme', 'apple' yang notabene adalah pesaing brand Xiaomi di pasar handphone juga mempunyai kecenderungan bersentimen negatif. Selain itu terms 'layar', 'video', 'suara', 'iklan', 'resmi', 'miui', 'browser', 'flagship', juga lebih sering digunakan dalam tweet bersentimen negatif.

Pada penelitian ini juga terbukti bahwa penyeleksian terms dengan menentukan batas minimum terms berdasarkan nilai bobot TF-IDF dapat mempengaruhi nilai akurasi. Hasil uji coba pada penelitian ini batas

minimum 10 meningkatkan akurasi sedangkan batas minimum 100 menurunkan nilai akurasi.

5.2 Saran

Untuk brand Xiaomi :

Beberapa produk Xiaomi dalam sektor ponsel pintar bisa dibilang lebih banyak diperbincangkan oleh pengguna media sosial Twitter ketimbang produk lainnya. Selain itu tercatat ada beberapa produk lain seperti earphone, dan masker, akan tetapi apabila kita melihat dengan perspektif yang lebih luas tentu seharusnya banyak produk lain dari brand Xiaomi yang diperbincangkan di media sosial Twitter oleh masyarakat Indonesia. Peneliti menyarankan agar Xiaomi lebih gencar memperkenalkan produk-produk mereka di sektor lain seperti bidang smart home.

Meskipun beberapa nama produk ponsel pintar Xiaomi mayoritas lebih banyak digunakan dalam tweet yang mempunyai sentiment positif. Akan tetapi yang menarik adalah terms 'handphone' lebih banyak digunakan pada tweet bersentimen negatif. Selain itu, ketika pengguna media sosial Twitter membicarakan tentang brand Xiaomi, dan disaat bersamaan menggunakan terms nama merek pesaing Xiaomi di bidang ponsel pintar juga mayoritas mempunyai sentimen negatif.

Peneliti menyarankan agar pihak Xiaomi mempertahankan beberapa aspek berikut 'spesifikasi', 'kamera', 'baterai', 'harga', 'murah' yang sudah lebih banyak diperbincangkan dengan membawa sentimen positif. Dan juga meninjau ulang beberapa aspek seperti 'iklan', 'resmi' yang merupakan kebijakan pemasaran, 'miui' yang merupakan sistem operasi besutan Xiaomi, 'layar', 'suara' yang berkaitan dengan perangkat keras, 'browser', 'video' yang berkaitan dengan perangkat lunak, dan 'flagship' yang merupakan istilah untuk sebuah produk inti atau produk unggulan dari suatu perusahaan.

Untuk peneliti selanjutnya :

1. Pada penelitian ini, peneliti menemukan beberapa masalah yang disebabkan oleh Sastrawi Stemmer dan TwitterOAuth diharapkan akan ada perbaikan dimasa yang akan datang atau penggunaan library lain yang lebih baik.
2. Pada penelitian selanjutnya peneliti menyarankan menambahkan metode Information Gain dan Adaboost untuk meningkatkan akurasi klasifikasi.
3. Penambahan koleksi kamus kesalahan tipografi (typo).
4. Peningkatan akurasi dengan mengganti metode pembobotan TF-IDF dengan metode TF-IDF-CF.

Daftar Pustaka

- [1] "Sekilas tentang Perusahaan",
<https://www.mi.co.id/id/about/>, (diakses pada 4 maret 2019)

- [2] Fidinillah, Ramdoni. (2018), "Ini Peran Penting Indonesia bagi Bisnis Xiaomi", <https://20.detik.com/e-flash/20181106-181106019/ini-peran-pentingindonesia-bagi-bisnis-xiaomi>, (diakses pada 4 maret 2019)
- [3] Republik Indonesia, (2016), Undang-undang No 20 Tahun 2016 Tentang Merk dan Indikasi Geografis, LEMBARAN NEGARA REPUBUK INDONESIA TAHUN 2016 NOMOR 252, Kementrian Hukum dan HAM Republik Indonesia: Jakarta
- [4] Nurhuda, F., Sihwi, S.W., Doewes, A., (2013), Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia (2014) berdasarkan Opini dari Twitter menggunakan Metode Naïve Bayes Classifier, Jurnal ITSMART
- [5] Rozi, I. F., Pranomo, S. H., & Dahlan, E. A. (2012), Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Electrics, Electronics, Communications, Controls, Informatics, Systems (EECCIS)* .
- [6] Hall, M. (2006). A Decision Tree-Based Attribute Weighting Filter for Naïve Bayes. *KnowledgeBased Systems*, hal 120–126. <http://www.cs.waikato.ac.nz/pubs/wp/2006/uow-cs-wp-2006-05.pdf>.
- [7] Taheri, S., Mammadov, M. (2013). Learning The Naive Bayes Classifier With Optimization Models. In *International Journal of Applied Mathematics and Computer Science*, hal 787–795
- [8] Novantirani, A., Sabariah, M. K., Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. Program Studi Teknik Informatika, Fakultas Informatika, Universitas Telkom, Bandung .
- [9] Hadna, N. M., Santosa, P. I., Winarno, W. W. (2016). Studi Literatur tentang Perbandingan Metode untuk Proses Analisis Sentimen di Twitter, Seminar Teknologi Informasi dan komunikasi (SENTIKA)
- [10] Rodiyansyah, S. F., Winarko, E. (2013). Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. *Indonesian Journal of Computing and Cybernetics Systems (IJCCS)*

AMIKOM Yogyakarta, lulus tahun 2009. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknik Informatika STMIK AMIKOM Yogyakarta, Lulus tahun 2011. Saat ini menjadi Dosen di Universitas AMIKOM Yogyakarta, pada Program Studi Informatika..

Biodata Penulis

Zakaria Mursid, memperoleh gelar Sarjana Komputer (S.Kom), Program Studi Informatika Universitas AMIKOM Yogyakarta, lulus tahun 2019.

Anggit Dwi Hartanto, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika STMIK