

ANALISIS SENTIMEN PADA SOSIAL MEDIA TWITTER TERHADAP POLITIK DI INDONESIA MENGUNAKAN TEXT MINING DENGAN METODE NAIVE BAYES CLASSIFIER

Rezki Syaputra¹, Rachmansyah²

^{1,2}STMIK GI MDP; Jl. Rajawali No.14, +62(711)376400/376360

³Program Studi Teknik Informatika, STMIK GI MDP, Palembang

email: ¹riski.putra_46@mhs.mdp.ac.id, ²rachmansyah@mdp.ac.id

Abstrak

Kemajuan teknologi dan informasi yang semakin meningkat memberi dampak pada pertukaran informasi dan komunikasi yang semakin mudah. Ini ditandai dengan dengan munculnya media sosial seperti Twitter, Facebook, Github, Gitlab, Linkdln, WhatsApp, Google, Youtube, Instagram, dan Path. Pertumbuhan media sosial ini juga mendorong adanya informasi tekstual yang besar sehingga muncul kebutuhan penyajian data yang memudahkan pengguna mendapatkan informasi yang akurat. Pengguna *twitter* yang semakin meningkat ini terlihat dari jutaan *tweets* yang di *posting* setiap harinya dengan berbagai topik yang berbeda. Data *tweets* ini dapat berupa persepsi publik baik politik, ekonomi, perilaku sosial, fenomena alam, perdagangan yang terjadi diseluruh dunia. Penelitian ini mencoba melakukan analisis sentimen data dengan mengklasifikasi data twitter berbahasa Indonesia terhadap politik di Indonesia. Data tersebut akan diproses dengan *text mining* untuk menghindari data yang kurang sempurna kemudian mengklasifikasi data *tweet* ke dalam tiga klasifikasi yaitu klasifikasi positif, negatif, netral. Klasifikasi ini menggunakan algoritma *Naïve Bayes Classifier*.

Kata Kunci: *Twitter, Naive Bayes Classifier, Text Mining, Analisis Sentimen.*

Abstract

Increased technology and information progress has an impact on easier information and communication exchanges. This is marked by the rise of social media like Twitter, Facebook, Github, Gitlab, Linkdln, WhatsApp, Google, Youtube, Instagram, and Path. The growth of social media also mendorong the existence of large textual information so that the emerging needs of data presentation that allows users to get accurate information. Increasingly twitter users are seen from the millions of tweets that are posted every day with different topics. Data tweets can be a public perception of both political, economic, social behavior, natural phenomena, trade that takes place around the world. This research is trying to analyze data sentiment by classifying Indonesian language twitter data on politics in Indonesia. The data will be processed with text mining to avoid less than perfect data and then classify data tweet into three classifications that is positive, negative, neutral classification. This classification uses the *Naïve Bayes Classifier* algorithm.

Keywords: *Twitter, Naive Bayes Classifier, Text Mining, Sentiment Analysis.*

1. PENDAHULUAN

Perkembangan di bidang Teknologi Informasi Komputer dari masa ke masa sangat pesat dan peranannya dalam kehidupan manusia dapat dirasakan dalam berbagai kegiatan kehidupan manusia, baik secara individu ataupun. Secara umum terdapat dua tipe informasi tekstual di *web* yaitu fakta dan opini. Fakta adalah pernyataan objektif mengenai entitas dan kejadian di dunia sedangkan opini adalah pernyataan subjektif yang merefleksikan sentiment atau persepsi orang mengenai entitas ataupun kejadian di dunia.

Ketika suatu organisasi/perusahaan/perorangan ingin memperoleh opini publik mengenai produk, citra dan layanannya maka mereka tidak perlu melakukan survey konvensional dan fokus group yang mahal biayanya. *Web* melalui situs *review online*, *blog* pribadi, situs jejaring sosial menyediakan sumber-sumber opini yang besar jumlahnya bagi kebutuhan individu maupun organisasi.

Melalui *web* orang dapat mengekspresikan apa saja, termasuk pendapatnya akan suatu hal tanpa adanya keterpaksaan Analisis sentimen atau *opinion mining* merupakan 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. Salah satu contoh penggunaan analisis sentiment dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentiment berkembang pesat. Bahkan di Amerika terdapat sekitar 20- 30 perusahaan yang memfokuskan pada layanan analisis sentimen [1].

Text mining adalah bidang interdisipliner yang mengacu pada pencarian informasi, pertambahan data, pembelajaran mesin, statistik, dan komputasi linguistik. Dikarenakan kebanyakan informasi (perkiraan umum mengatakan lebih dari 80%) saat ini disimpan sebagai teks, text mining diyakini memiliki potensi nilai komersial tinggi. Algoritme pengklasifikasian yang digunakan adalah Naïve Bayes, karena dari penelitian yang dilakukan oleh Sandi dan Edi (2013) tentang klasifikasi *tweet* kemacetan lalu lintas di Bandung dengan hasil akurasi tertinggi mencapai 93,58% (Rodiyansyah & Winarko, 2013), oleh Ahmad dan Azhari (2014) tentang klasifikasi *tweet* yang berisi sentimen masyarakat pada tokoh publik tertentu didapatkan hasil berupa prosentase akurasi pengujian klasifikasi fitur *term frequency* mencapai 79,91% sedangkan fitur TF-IDF mencapai 79,68% dengan metode Naïve Bayes (Hidayatullah & Azhari, 2014) terbukti bahwa metode Naïve Bayes memiliki performansi yang cukup baik untuk melakukan klasifikasi

tweet sehingga memiliki tingkat keberhasilan yang tinggi. Sebagai upaya pembaharuan ditambahkan proses pembobotan pada non-tekstual (*emoji*) karena pada penelitian yang dilakukan Chin (2016) emoji sangat populer digunakan oleh masyarakat untuk mengungkapkan gagasan atau ide yang tidak dapat dituliskan dengan kata-kata (Chin, Zappone, & Zhao, 2016). Pada penelitian Sreeja (2016) terdapat opini sindiran, dimana sentimen yang terkandung dalam non-tekstual lebih mendominasi dari sentimen yang terkandung pada tekstual (kalimat) sehingga menyebabkan keambiguan opini. (Jayanthi & Sreeja, 2016). Dengan adanya pembobotan emoticon, diharapkan hasil sentimen menjadi benar – benar memiliki batas yang jelas antara sentimen positif dan

negatif (Dayalani & Patil, 2014) sehingga akurasi yang dihasilkan oleh sistem ini mengalami peningkatan.

2. METODE PENELITIAN

Alur proses pada penelitian ini diawali dengan memasukkan dokumen latih dan uji, dokumen diolah melalui pembobotan non-tekstual (*emoji*) maupun pembobotan tekstual (teks). Proses awal pada pembobotan non-tekstual ialah mendeteksi adanya *emoji* berdasarkan kamus *emoji* yang telah tersimpan dalam *file csv*, jika ditemukan maka langkah selanjutnya ialah memboboti *emoji* dengan menjumlahkan *emoji* yang ada dalam tiap kategori, kemudian dilakukan normalisasi *min-max*. Pada pembobotan tekstual proses yang dilakukan ialah praproses teks, kemudian dilakukan perhitungan dengan Naïve Bayes dan dilanjutkan dengan normalisasi tekstual sehingga menghasilkan nilai probabilitas positif maupun negatif. Hasil dari perhitungan pembobotan tekstual (Naïve Bayes) dan pembobotan non-tekstual (*emoji*) digabungkan disertai nilai konstanta pengali tertentu sehingga mampu mengklasifikasikan dokumen tersebut dalam kategori sentimen positif maupun negatif.

2.1 Twitter API

Twitter diciptakan oleh Jack Dorsey di tahun 2006 dan pertama meluncur di dunia maya saat Juli 2006 dengan alamat <http://www.twitter.com> yang masih digunakan hingga saat ini. Twitter memiliki *Application Programming Interface* sedemikian hingga developer dapat mengembangkan aplikasi sesuai dengan kebutuhannya masing-masing[1].

1. *Twitter REST API*: Terdiri dari *Twitter REST* dan *Twitter Search*. *Twitter REST* memberikan core data dan core twitter objects. *Twitter search* berfungsi untuk melakukan pencarian mengenai suatu instance objek Twitter maupun mencari trend
2. *Twitter Streaming API*: API ini biasa digunakan untuk penggalan data karena melalui API ini informasi bisa didapatkan secara *realtime* dengan volume yang sangat tinggi. Pada penelitian ini sendiri, Twitter Streaming API akan digunakan untuk mengumpulkan data penelitian.

2.2 Naïve Bayes Classifier

Salah satu tugas *Data Mining* adalah klasifikasi data, yaitu memetakan (mengklasifikasikan) data ke dalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya. Salah satu metoda dalam klasifikasi data adalah *Naïve Bayes Classifier* (NBC). *Naïve Bayes Classifier* merupakan salah satu metoda *machine learning* yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya[2]. Dasar dari *Naïve Bayes* yang dipakai dalam pemrograman adalah rumus Bayes:

$$P(A|B) = (P(B|A) * P(A)) / P(B) \dots\dots\dots (1)$$

Peluang kejadian A sebagai B ditentukan dari peluang B saat A, peluang A, dan peluang B. Pada pengaplikasiannya nanti rumus ini berubah menjadi :

$$P(C_i|D) = (P(D|C_i) * P(C_i)) / P(D) \dots\dots\dots (2)$$

Naïve Bayes Classifier atau bisa disebut sebagai *Multinomial Naïve Bayes* merupakan model penyederhanaan dari Metoda *Bayes* yang cocok dalam pengklasifikasian teks atau dokumen. Persamaannya adalah:

$$V_{MAP} = \arg \max_{V_j \in V} P(V_j | a_1, a_2, \dots, a_n) \dots\dots\dots (3)$$

Menurut persamaan (3), maka persamaan (1) dapat ditulis:

$$V_{MAP} = \arg \max_{V_j \in V} \frac{P(a_1, a_2, \dots, a_n | P(V_j))}{P(a_1, a_2, \dots, a_n)}$$

2.3 Analisis Sentimen

Analisis Sentimen atau *opinion mining* 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 penulis berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu.[3]

2.4 Media Sosial

Media sosial adalah sebuah media *online*, dengan para penggunanya bisa dengan mudah berpartisipasi, berbagi, dan menciptakan isi meliputi blog, jejaring sosial, wiki, forum dan dunia virtual. Blog, jejaring sosial dan wiki merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat di seluruh dunia.[4]

2.5 Text Mining

Text mining adalah bidang interdisipliner yang mengacu pada pencarian informasi, pertambangan data, pembelajaran mesin, statistik, dan komputasi linguistik. Dikarenakan kebanyakan informasi (perkiraan umum mengatakan lebih dari 80%) saat ini disimpan sebagai teks, *text mining* diyakini memiliki potensi nilai komersial tinggi (Bridge, 2011). *Text mining* merupakan teknik yang digunakan untuk menangani masalah klasifikasi, *clustering*, *information extraction* dan *information retrieval* (Berry & Kogan, 2010).[5]

2.6 Stemming

Stemming adalah salah satu proses dari *Information Retrieval* (IR) untuk mencari kata dasar.[6]

2.7 Stopword

Stopword adalah proses penghilangan kata-kata yang tidak berkontribusi banyak pada isi dokumen. Kata-kata yang termasuk ke dalam *stopword* dihilangkan karena memberikan pengaruh yang tidak baik dalam proses *text mining* seperti kata-kata "dan", "saya", "kamu", "dengan", "dia" dan lain-lain.[7]

2.8 Casefolding

Casefolding merupakan tahapan yang mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf a sampai dengan z yang diterima. Karakter selain huruf dihilangkan dan dianggap pembatas.[8]

2.9 Flowchart

Flowchart program adalah suatu bagan yang digunakan untuk menggambarkan suatu algoritma atau prosedur dalam menyelesaikan masalah. *Flowchart* terbagi menjadi bagian yaitu *flowchart system* dan *flowchart program*. *Flowchart system* yaitu bagan yang menggambarkan suatu prosedur dan proses suatu file dalam suatu media menjadi file dalam media lain dalam

suatu sistem. *Flowchart* program adalah bagan yang menggambarkan urutan logika dari suatu prosedur pemecahan masalah.[9]

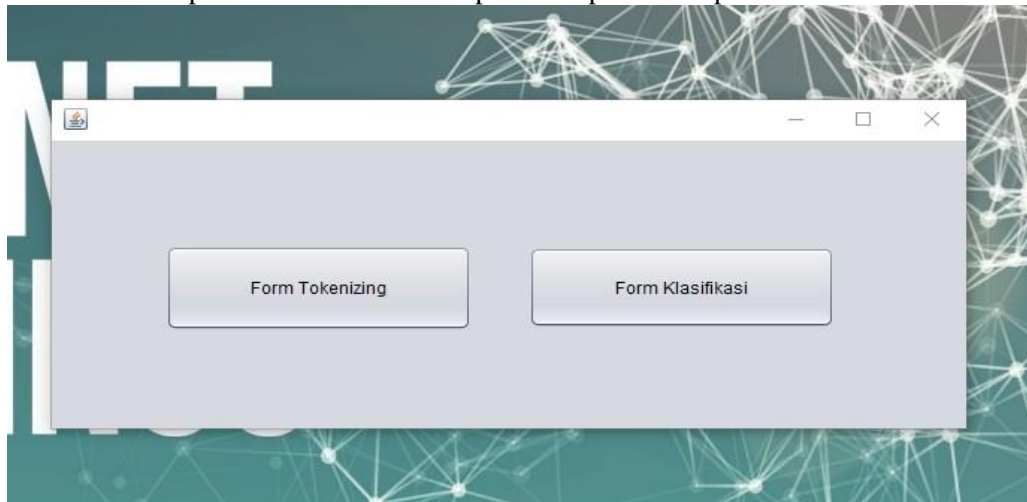
3. HASIL DAN PEMBAHASAN

3.1 Hasil

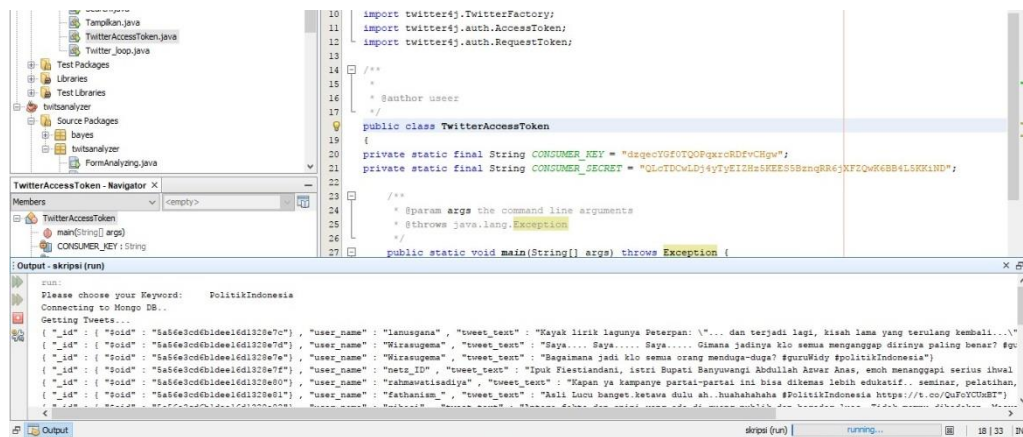
3.1.1 Tampilan beranda

Halaman ini menampilkan dua *button* untuk pengelolahn data. Saat user pertama kali menjalankan aplikasi, user masuk pada tampilan tersebut.

Bentuk dan tampilan halaman beranda aplikasi dapat dilihat pada Gambar 1.



Gambar 1 Tampilan Halaman Beranda



Gambar 2 Tampilan Halaman Hasil Penarikan Data

tweet_text

Partai Politik di Indonesia khususnya PKS: Mendukung Fahri Hamzah Untuk Mencalonkan diri sebaga...â€¦ https://t.co/eNtuGnfqKP

RT @kawanFH: @Fahrihamzah: Klean belajar aja yang bener \\ Politik dah ada yang jagain \\ Tenang aja \\ Tawuran kita kalo ada yang macem-maceâ€¦

RT @ferrykoto: Pilkada kali ini kita makin melihat, ternyata Parpol umumnya telah gagal menjalankan fungsinya sebagai pencetak kader. Tak sâ€¦

RT @pangeransihaan: [new video] Prediksi Pertempuran Politik Indonesia tahun 2018 dari @asumsico \nhttps://t.co/CLmRoHpmmme

Tahukan siapa yang harus dipilih? \n\n https://t.co/7EdCbHclF1

RT @ferrykoto: Pilkada kali ini kita makin melihat, ternyata Parpol umumnya telah gagal menjalankan fungsinya sebagai pencetak kader. Tak sâ€¦

[new video] Prediksi Pertempuran Politik Indonesia tahun 2018 dari @asumsico \nhttps://t.co/CLmRoHpmmme

Ini Surat Gambala yang menurut saya sangat baik&relevan. Ketika discourse seputar agama di negeri ini mjd "sangat mâ€¦ https://t.co/Zp4ajijkh5

@AbuAhmad_411 @InfoPDIPJatim @PDI_Perjuangan @nu_online @savicalli @aananshori Cuma yg tjd saat ini kita dibuat utkâ€¦ https://t.co/CfwTxCsAKb

RT @ferrykoto: Pilkada kali ini kita makin melihat, ternyata Parpol umumnya telah gagal menjalankan fungsinya sebagai pencetak kader. Tak sâ€¦

@ardi_riau #Indonesia ingin di pimpin pria berdasi abri ato pria sipil ? @Takviri @yusejahtera #politik #batakNetâ€¦ https://t.co/DFYRqD71J8

RT @ferrykoto: Pilkada kali ini kita makin melihat, ternyata Parpol umumnya telah gagal menjalankan fungsinya sebagai pencetak kader. Tak sâ€¦

RT @ferrykoto: Pilkada kali ini kita makin melihat, ternyata Parpol umumnya telah gagal menjalankan fungsinya sebagai pencetak kader. Tak sâ€¦

RT @sociotalkr: pada kenyataannya ideologi nasional dan sejarah politik Indonesia adalah dinamika antara gagasan2 liberalisme, sosialisme,â€¦

RT @kerjapembebasan: Keterlibatan Jerman dalam Genosida 1965 dan Peran Agen Rahasia Rudolf Oebser-RÄ¶der (Mantan Perwira Nazi) di Indonesiaâ€¦

RT @TetapAhokDjarot: Gus Yusuf:Abdul somad tdk cocok utk Indonesia.dimanfaatkan utk kepentingan politik di goreng HTI https://t.co/owBjaCNâ€¦

RT @ferrykoto: Pilkada kali ini kita makin melihat, ternyata Parpol umumnya telah gagal menjalankan fungsinya sebagai pencetak kader. Tak sâ€¦

RT @untukmantri: PDIP takut dgn Moreno Anang. Sehingga mereka turunkan paslon terkut: Ipol-Risma. \n\nPDIP ga kapok bahwa politik Indonesiaâ€¦

RT @ferrykoto: Pilkada kali ini kita makin melihat, ternyata Parpol umumnya telah gagal menjalankan fungsinya sebagai pencetak kader. Tak sâ€¦

@GoMobile93458 Abdul K Safei (akas) 089519328314 Selamat Menyambut Tahun Baru 2018. Semoga Indonesia makin damai,â€¦ https://t.co/oVu6sFto7D

2018 menjadi tahun yg penuh tantangan bg politik Indonesia. Tdk hny diselenggarakannya Pilkada serentak di 171 daerâ€¦ https://t.co/csepnvtf4

RT @dulatips: Yuk berjuang dr nol, gaess. Menang itu cm bonus.. tujuan utamanya buat berdiri tegak memegang prinsip. Yuk tanda tangani, sapâ€¦

RT @yusuf_dumdum: Daripada berkuat pada politik identitas, alangkah baiknya jika ada statemen yg negarawan, contoh: "Saya asli putra bangsâ€¦

RT @ruhuhtompul: Presiden RI ke 7 Bpk JOKOWI tegas mengatakan "Jadikan Hukum Panglima" karena itu yg ingin Politik menjadi Panglima, INDOâ€¦

RT @sociotalkr: pada kenyataannya ideologi nasional dan sejarah politik Indonesia adalah dinamika antara gagasan2 liberalisme, sosialisme,â€¦

RT @ferrykoto: Pilkada kali ini kita makin melihat, ternyata Parpol umumnya telah gagal menjalankan fungsinya sebagai pencetak kader. Tak sâ€¦

RT @ferrykoto: Pilkada kali ini kita makin melihat, ternyata Parpol umumnya telah gagal menjalankan fungsinya sebagai pencetak kader. Tak sâ€¦

Adakah politikus Indonesia kayak Gajah Mada & Hayam Wuruk, yakni Politikus Jawa Timur, Politik lokal tapi pemikiran Nasional, kayak Jokowi.

Ini 4 Pasangan Cagub-Cawagub yang akan Bertarung di Pilkada Jabar https://t.co/G0di8ta1IQ

RT @swulll: [#FahR1 voice]\n\nBagi bang @Fahrihamzah Politik & Agama tdk bisa dipisahkan dlm suatu Negara.\n\nKhususnya di Indonesia. Indonesiaâ€¦

RT @ferrykoto: Pilkada kali ini kita makin melihat, ternyata Parpol umumnya telah gagal menjalankan fungsinya sebagai pencetak kader. Tak sâ€¦

Gambar 3 Tampilan Data yang Sudah Ditarik

Form Tokenizing : Memperoleh Token

Browse.. ☒ Positive ☐ Negative

bduKharisA: Mjd narasumber dialog politik @PKS_Sukoharjo dg tema pentingnya masyarakat dlm politik Indonesia @PKSejahtera https://t.co/5a52411b6b1d2ff477fcb2c,maz_doels,5,986,"Twitter for iPhone","RT @swulll: [#FahR1 voice]\n\nBagi bang @Fahrihamzah Politik & Agama tdk bisa dipisahkan dlm suatu Negara.\n\nKhususnya di Indonesia. Indonesia...

Clearing:

Folding:

Tokenizing:

No	Token	Tipe	Total
1	_id,user_name,...	Positif	1
2	rel="nofollow">...	Positif	111
3	for	Positif	96
4	android","p...	Positif	1
5	politik	Positif	63

Statistic

Token Positive 492
Token Negative 0

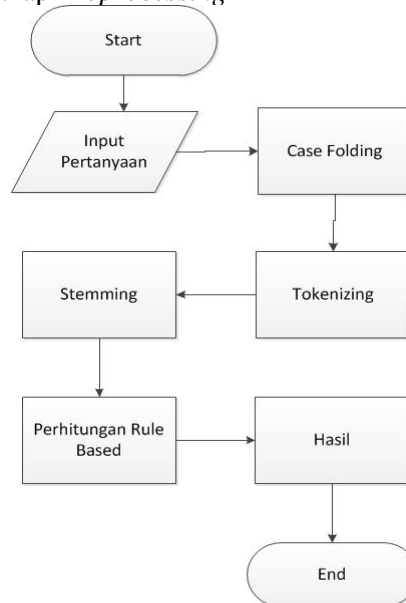
Total Token : **492**

Pra Proses Simpan Tutup

Gambar 4 Tampilan Hasil Data

3.2 Pembahasan

3.2.1 Diagram tahap *Preprocessing*



Gambar 5 Diagram *Preprocessing*

Pertanyaan yang diinput adalah seputar politik di Indonesia. Setelah input pertanyaan maka akan dilakukan proses *preprocessing* yaitu *casefolding*, *tokenizing*, *stopword* lalu akan dilakukan proses perhitungan hasil *Rule Based*. Proses *casefolding* yaitu mengubah huruf besar yang terdapat di dalam pertanyaan yang diinput menjadi huruf kecil. *Tokenizing* merupakan proses menghilangkan spesial karakter yang terdapat di dalam pertanyaan yang diinput dan *stopword* merupakan proses penghilangan kata umum yang sering digunakan. Setelah dilakukan proses *stopword*, setiap kata akan dihitung bobotnya sesuai aturan (*rule based*) yang telah ditetapkan.

4. KESIMPULAN

Dari hasil penelitian dan pembahasan yang dilakukan, maka dapat disimpulkan bahwa:

1. Penarikan data dengan menggunakan media sosial *twitter* telah berhasil dilakukan, dengan melakukan *Authentication* terlebih dahulu untuk mengakses *Twitter API*. Proses *Crawling* data bersifat *real time*, data yang ditarik hari ini tidak akan sama pada saat penarikan data esok harinya.
2. Dari hasil uji analisis sentimen pada data *twitter* didapat *tweet* dengan tiga kategori yaitu kategori positif, negatif dan netral.
3. Analisis sentiment dengan menggunakan *text mining* yang didalamnya terdapat tahapan *preprocessing* dan ekstraksi fitur untuk menghasilkan *tweet* kotor menjadi *tweet* bersih.
4. Algoritma *Naïve Bayes* sebagai algoritma untuk mencari nilai probabilitas dari setiap *tweet* yang di analisa.

5. SARAN

Dari kesimpulan yang telah diuraikan maka terdapat saran yang perlu disampaikan untuk penelitian selanjutnya yaitu penelitian selanjutnya dapat menggunakan metode lain untuk mengklasifikasi dan menggunakan banyak data latih sehingga hasil yang didapat semakin akurat.

DAFTAR PUSTAKA

- [1] Ahmad Fathan Hidayatullah, Azhari SN, (2014). Analisis Sentimen dan Klasifikasi Kategori terhadap Tokoh Publik pada Twitter.
 - [2] Akshay Amolik, Niketan Jivane, Mahavir Bhandari, Dr.M.Venkatesan, (2016). Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques.
 - [3] Lusiani, T, dan Cahyono, A. (2006). *Sistem Berbasis Aturan untuk Mendiagnosa Penyakit Flu Burung Secara Online*. Surabaya: Seminar Nasional Sistem dan Informatika
 - [4] Suarga. (2006). *Algoritma Pemrograman*. Yogyakarta: Andi Offset.
 - [5] Triawati, C. (2009). *Metode Pembobotan Statistical Concept Based untuk Klustering dan Kategorisasi Dokumen Berbahasa Indonesia*. Bandung: Institut Teknologi Telkom.
 - [6] Sismono, H. (2005). *Pengantar Logika Informatika, Algoritma dan Pemrograman*. Yogyakarta: Andi Offset.
 - [7] Bradji, L., & Boufaida, M. (2011). *A Rule Management System for Knowledge Based Data Cleaning*. , 2011(November), 230–239. doi: 10.4236/iim.2011.36028
-