

Pemanfaatan Big Data Media Sosial Dalam Menganalisa Kemenangan Pilkada

Dewa Ayu Putri Wulandari¹, Made Sudarma², Nyoman Pramaita³

[Submission: 28-01-2019, Accepted: 15-04-2019]

Abstract—The election of the Governor and Deputy Governor of Bali will go through several stages of elections starting from the determination of the Governor and Deputy Governor of Bali to the stages of vote counting. In the election of the Governor and Deputy Governor of Bali the community can be directly involved in the voting stage which will be held on June 27, 2018 (General Commission Election or KPU, 2018). So that it raises many opinions, not only positive and neutral opinions but also negative ones. This research is expected to be able to conduct research on public opinion which contains positive, neutral and negative sentiments. In this research used tokenization preprocessing data N-gram method. N-gram is a token consisting of three words in each one token. In the stemming stages used the *Nzief Adriani* algorithm. For the classification process of this research used the 'Naïve Bays Classifier (NBC) method. In testing the candidate Governor's data the highest accuracy was obtained from the classification of KBS-Ace data on data taken from twitter with 89% accuracy, 91% precision and 94% recall and lowest accuracy when KBS-Ace data calcification process on social media Face book

Intisari— Pemilihan Calon Gubernur dan Wakil Gubernur Bali 2018 akan melalui beberapa tahapan pemilu mulai dari penentuan bakal calon Gubernur dan Wakil Gubernur Bali hingga tahapan penghitungan suara. Dalam pemilihan Gubernur dan Wakil Gubernur Bali masyarakat dapat terlibat langsung dalam tahapan pemungutan suara yang akan dilaksanakan pada tanggal 27 Juni 2018 (KPU, 2018). Sehingga dapat memunculkan banyak komentar atau pendapat, tidak hanya komentar positif dan netral tapi juga komentar yang negatif. Penelitian ini diharapkan mampu untuk melakukan riset atas komentar masyarakat yang mengandung sentimen baik atau positif, sama sekali tidak mengandung senrime atau netral dan mengandung sentimen buruk atau negatif. Dalam penelitian ini metode digunakan untuk *preprocessing* data menggunakan tokenisasi N-gram. N-gram adalah token yang terdiri dari tiga kata setiap satu token. Pada tahap stemming menggunakan algoritma *Nzief Adriani*. Untuk proses klasifikasinya menggunakan metode *Naïve Bayes Classifier (NBC)*. Pada pengujian data calon Gubernur akurasi tertinggi diperoleh dari klasifikasi data KBS-Ace pada data yang diambil dari Twitter dengan nilai akurasi 89%, presisi 91% dan recall 94% dan akurasi terendah pada saat proses kalsifikasi data KBS-Ace pada media sosial Facbook

Kata Kunci— Analisa Sentimen, Calon Gubernur Bali 2018, Naive Bayes Classifier

I. PENDAHULUAN

Pemilihan Gubernur dan Wakil Gubernur Bali akan melalui beberapa tahapan pemilu mulai dari penentuan bakal calon Gubernur dan Wakil Gubernur Bali hingga tahapan penghitungan suara. Dalam pemilihan Gubernur dan Wakil Gubernur Bali masyarakat dapat terlibat langsung dalam tahapan pemungutan suara yang akan dilaksanakan pada tanggal 27 Juni 2018 (KPU, 2018). Sejak tahapan pendaftaran hingga ditetapkan calon Gubernur dan Wakil Gubernur Bali 2018 nama-nama calonnya mulai banyak diperbincangkan, baik didunia nyata maupun dunia maya. Sebelum menentukan pilihan masyarakat akan mengenal lebih jauh figur calon Gubernur dan Wakil Gubernur Bali serta program-programnya dalam tahapan kampanye. Semua orang bebas menyampaikan pendapat maupun berkomentar tentang pasangan calon Gubernur dan Wakil Gubernur Bali 2018 sehingga dapat memunculkan beberapa opini, tidak hanya komentar yang positif dan netral tapi juga komentar yang negatif. Opini-opini tersebut dapat disampaikan oleh masyarakat melalui berbagai media baik dengan cara berkomunikasi secara langsung maupun dengan memanfaatkan teknologi yang ada pada saat ini.

Media sosial menjadi tempat promosi atau kampanye yang efektif dan efisien. Salah satu hal yang menarik pada kampanye di media sosial adalah bagaimana masyarakat pengguna media sosial dapat memberikan opini atau komentar pada suatu topik tertentu. Menelusuri opini pada sosial media menjadi cara yang semakin populer untuk berbagi informasi. Opini pengguna media sosial dapat berisi penilaian atau pandangan terhadap suatu topik tertentu. Melakukan analisis terhadap opini pengguna media sosial tentunya akan bermanfaat untuk mengetahui apakah penilaian yang terdapat pada opini bersifat baik atau buruk bahkan tidak keduanya.

Berdasarkan dari hal-hal yang dijelaskan diatas, penelitian ini bermaksud melakukan analisis sentimen di media sosail Twitter dan Facebook untuk melihat sebuah komentar atau opini seseorang yang ditujukan kepada calon Gubernur dan Wakil Gubernur Bali 2018, komentar atau opini masyarakat bisa dimasukan katagori opini positif, netral atau negatif. Metode yang digunakan dalam penelitian ini, untuk *preprocessing* data menggunakan tokenisasi N-gram. N-gram adalah token yang terdiri dari tiga kata setiap satu token. Pada tahap stemming menggunakan algoritma *Nzief Adriani* [1]. Untuk proses klasifikasinya menggunakan metode *Naïve Bayes Classifier (NBC)*.

II. LANDASAN TEORI

A. Analisa Sentimen

p-ISSN:1693 – 2951; e-ISSN: 2503-2372

¹Mahasiswa Pascasarjana, Program Studi Megister Teknik Elektro Fakultas Teknik Universitas Udayana, Jln. Jalan Kampus Bukit Jimbaran INDONESIA (telp: 0361-703315; fax: 0361-4321982; e-mail: dewaayuputriwulandari@gmail.com)

^{2,3}Dosen, Jurusan Teknik Elektro dan Komputer Fakultas Teknik Universitas Udayana, Jln. Jalan Kampus Bukit Jimbaran 80361 INDONESIA (telp: 0361-703315; fax: 0361-4321; e-mail: msudarma@unud.ac.id npramaita@unud..ac.id)



Sentimen analisa yang juga disebut 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 [2]. Tujuan dari analisa sentimen adalah untuk menentukan perilaku atau opini dari seorang penulis dengan memperhatikan suatu topik tertentu. Perilaku bisa mengindikasikan alasan, opini atau penilaian, kondisi kecenderungan. Sentiment analysis juga dapat menyatakan perasaan emosional sedih, gembira, atau marah [3].

B. Text Mining

Text Mining berpedoman pada proses pengambilan informasi dengan kualitas tinggi dari beberapa text. Informasi dengan kualitas yang tinggi biasanya dapat diperoleh dengan caea peramalan pola dan kecenderungan yang melalui sarana dengan pembelajaran pola statistik [4]. *Text Mining* juga sering melibatkan input dalam proses untuk penataan text (seperti persing dan fitur lingustik turunan dengan proses penambahan secara bersamaan serta proses penghilangan beberapa diantaranya seperti penyisipan *subsequent* ke dalam database), dalam data yang terstruktur harus ditemukan pola, sehingga pada akhirnya dapat mengevaluasi serta menginterpretasi output. Dengan kualitas yang tinggi kebaruan di bidang text akan berpedoman ke beberapa kombinasi relevan, kebaruan dan *interestingness* [5]. Text mining dengan ciri-ciri khas akan meliputi proses seperti kategorisasi teks, text clustering, ekstrasi konsep, produksi taksonomi, granular, analisa sentimen, penyimpulan dokumen dan pemodelan relasi entitas [6]

C. Naïve Bayes Classifier

Naïve bayes classifier merupakan penerapan berdasarkan teorema bayes (dari statistik Bayesian) dan sering disebut *classifier* probabilistik dengan asumsi independen yang kuat. Model fitur independen merupakan sebuah istilah yang digaris bahawi dalam modelprobabilitas. [7]. Kehadiran atau ketidakhadiran suatu fitur tertentu dari kelas yang tidak ada hubungannya dengan kehadiran atau ketidakhadiran fitur lainnya dalam asumsi terminologi sederhana dari sebuah *Naïve Bayes Classifier* [8]. Sebagai contoh sederhana jika ada sesuatu berwarna merah, berbentuk bulat serta berdiameter sekitar 4 inchi maka sesuatu itu mungkin dianggap buah apel. Bahkan fitur ini juga ketergantungan satu sama lain atau atas keberadaan fitur lain. Sebuah [9]. *Naïve Bayes Classifier* menganggap seluruh sifat-sifat berkontribusi mandiri untuk probabilitas bahwa buah ini adalah buah apel [10].

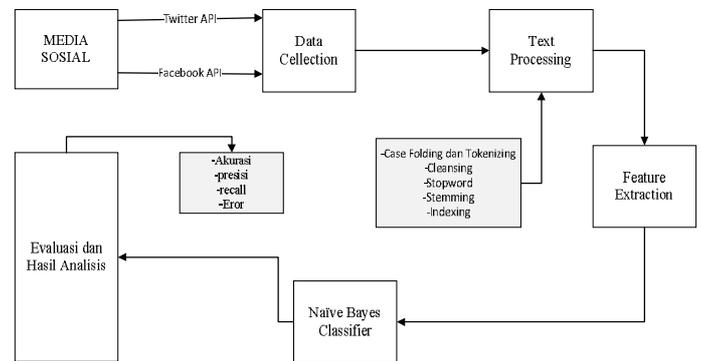
Menggunakan teorema Bayes , kita menulis

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

III. METODELOGI DAN PEMBAHASAN

Sistem yang dirancang ini diharapkan mampu untuk melakukan dan memberikan hasil analisis sentimen pada

calon Gubernur dan Wakil Gubernur Bali 2018 di media sosial Twitter dan Facebook berdasarkan komentar dari masyarakat yang menggunakan dan aktif pada media sosial. Komentar dari masyarakat yang diamati adalah komentar dengan Bahasa Indonesia.



Gambar 1. Contoh Gambaran Umum pada Sistem Analisa Sentimen

Penjelasan dari Gambar 1 gambaran umum pada sistem analisa sentimen yang dilakukan beserta tahapan yang dilakukan dalam penelitian ini.

- Langkah pertama yang harus dilakukan dengan cara mengumpulkan data yang diperlukan dalam penelitian ini. Data yang diperlukan adalah komentar dari masyarakat yang menggunakan media sosial terkait calon Gubernur dan Wakil Gubernur Bali 2018. Sumber data penelitian ini adalah Media Sosial Twitter dan Facebook
- Selanjutnya akan dilakukan tahapan preprocessing text. Untuk text ada yang diperoleh dari komentar masyarakat akan dilakukan proses seleksi kalimat, pembersihan kalimat, *stemming* kalimat dan normalisasi kalimat. Setelah text diolah dan sudah menjadi text yang bersih maka akan siap melakukan tahap selanjutnya
- Setelah melakukan proses *preprocessing* text maka akan dilanjutkan ke tahapan proses klasifikasi. Pada proses klasifikasi dalam penelitian ini menggunakan metode *Naïve Bayes Classifier* untuk menentukan sentimen sebuah kalimat opini.
- Tahapan terakhir yaitu dengan melakukan evaluasi performa Akurasi, Presisi, Recall dan error menggunakan metode *Confusion Matrix* pada eksperimen yang telah dilakukan

Penjelasan dan pembahasan dari hasil penelitian yang dilakukan dalam penelitian ini dengan menggunakan metode *Naïve Bayes Classifier* yang bertujuan untuk dapat menemukan suatu sentimen pada masyarakat pengguna media sosial yang tertuju pada Calon Gubernur dan Wakil Gubernur Bali 2018.

A. Input Pada Sistem

Input Pada Sistem dalam penelitian ini dilakukan mulai dari pencarian data di media sosial Twitter dan Facebook dengan memanfaatkan *Application Programming Interface* (API). Data yang diambil adalah data komentar yang berkaitan dengan Calon Gubernur dan Wakil Gubernur Bali 2018 di

media sosial Twitter dan Facebook. Data yang diperoleh dalam bentuk file CSV. Setelah data diperoleh maka data perlu disiapkan ke tahap *preprocessing* data terlebih dahulu.

B. Preprocessing Data

Tahap *preprocessing* data digunakan untuk melakukan proses perubahan semua karakter untuk menghindari data yang kurang sempurna, gangguan pada data dan data-data yang tidak konsisten seperti merubah karakter huruf menjadi huruf kecil. Kemudian dilakukan penghapusan URL seperti *@mention* dan *@hashtag* yang ada pada *tweet* tersebut. Dalam proses *preprocessing* dokumen atau kalimat dalam penelitian ini, fungsi yang dilakukan antara lain: melakukan *cleansing*, *case folding*, *tokenizing*, *stopword removal* dan *indexing* dan *stemming*

C. Hasil Analisis Sentimen Metode Naive Bayes Classifier

Data yang digunakan sebanyak 1000 tweet, 500 data diambil dari media sosial Twitter dan 500 data diambil dari media sosial Facebook. Data dibagi secara seimbang setiap kelasnya. Data dibagi menjadi KBS-Ace 250 Tweet, Mantra-Kerta 250 Tweet pada media sosial Facebook dan KBS-Ace 250 Tweet, Mantra-Kerta 250 Tweet pada media sosial Twitter

Berikut adalah hasil analisa sentimen pasangan calon Gubernur dan Wakil Gubernur Bali pasangan calon KBS-Ace dan pasangan calon Mantra-Kerta pada media sosial Facebook dan Twitter dalam bentuk tabel

TABEL I
CONTOH HASIL ANALISA SENTIMEN PADA MEDIA SOSIAL

Sentimen	KBS-Ace	Mantra-Kerta
Positif	347	348
Negatif	135	126
Netral	18	28
Total	500	500

TABEL II
CONTOH HASIL ANALISA SENTIMEN PADA MEDIA SOSIAL FACEBOOK

Sentimen	KBS-Ace	Mantra-Kerta
Positif	172	179
Negatif	67	61
Netral	11	10
Total	250	250

TABEL III
CONTOH HASIL ANALISA SENTIMEN PADA MEDIA SOSIAL TWITTER

Sentimen	KBS-Ace	Mantra-Kerta
Positif	175	169
Negatif	68	63
Netral	7	18
Total	250	250

D. Pengujian Menggunakan Confusion Matrix

Pengujian menggunakan *Confusion Matrix* yaitu melakukan evaluasi performa Akurasi, Presisi dan Recall dari eksperimen yang telah dilakukan. Hasil pengujian *Confusion Matrix* dapat dilihat pada Tabel berikut

Dewa Ayu Putri Wulandari : Manfaat Bid Data...

TABEL IV
CONTOH PENGUJIAN CONFUSION MATRIX PADA TOTAL MEDIA SOSIAL

Media Sosial	Akurasi %	Presisi %	Recall %
Facebook	70	77	74
Twitter	90	92	93
Fb dan Tw	80	85	86

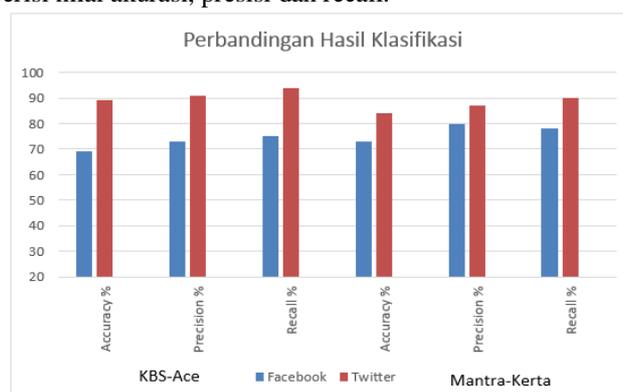
TABEL V
CONTOH PENGUJIAN CONFUSION MATRIX CALON GUBERNUR PADA MEDIA SOSIAL FACEBOOK

Calon Gubernur	Akurasi %	Presisi %	Recall %
KBS-Ace	69	73	75
Mantra-Kerta	90	92	93

TABEL VI
CONTOH PENGUJIAN CONFUSION MATRIX CALON GUBERNUR PADA MEDIA SOSIAL TWITTER

Calon Gubernur	Akurasi %	Presisi %	Recall %
KBS-Ace	89	91	94
Mantra-Kerta	84	87	90

Tabel diatas berisi informasi mengenai nilai akurasi, presisi dan recall dari masing-masing uji coba telah dilakukan. Pada Tabel 4 bagian kolom berisi informasi mengenai media sosial yang digunakan dalam penelitian ini sedangkan bagian baris berisi nilai akurasi, presisi dan recall. Pada Tabel 5 dan Tabel 6 bagian kolom berisi informasi mengenai calon Gubernur Bali 2018 sedangkan bagian baris berisi nilai akurasi, presisi dan recall.



Gambar 2. Contoh Grafik Perbandingan Klasifikasi

Dari Gambar 2 dapat dilihat bahwa nilai akurasi, presisi dan recall pada dua media sosial yaitu Facebook dan Twitter menghasilkan nilai yang berbeda dengan metode klasifikasi yang sama yaitu *Naive Bayes Classifier*. Nilai akurasi tertinggi terdapat pada media sosial Twitter untuk klasifikasi data KBS-Ace, dengan nilai akurasi mencapai 89%, nilai presisi 91% dan recall 94%. Sedangkan akurasi terendah metode *Naive Bayes Classifier* digunakan pada klasifikasi data

p-ISSN:1693 – 2951; e-ISSN: 2503-2372



KBS-Ace di media sosial Facebook, dengan nilai akurasi 69%, presisi 73% dan recall 75%. Nilai akurasi adalah salah satu parameter penilaian dari metode yang telah digunakan, nilai akurasi didapat dari jumlah banyaknya data yang berhasil diklasifikasi dengan benar sesuai kelas sentimennya dari seluruh jumlah data yang diklasifikasi.

IV. KESIMPULAN

Analisa sentimen dapat digunakan untuk mengetahui sentimen masyarakat khususnya netizen Twitter dan Facebook terhadap calon Gubernur Bali 2018. Tujuannya membantu masyarakat menentukan sentimen yang terdapat pada komentar di media sosial Twitter dan Facebook.

Nilai akurasi metode *Naïve Bayes Classifier* saat proses klasifikasi data keseluruhan berupa penggabungan data dari media sosial Twitter dan Facebook dengan nilai akurasi mencapai 80%, presisi 85% dan recall 88%. Sedangkan saat dilakukan pengujian terpisah di masing-masing media sosial hasil akurasi tertinggi diperoleh dari klasifikasi data pada media sosial Twitter dengan nilai akurasi mencapai 90%, presisi 92% dan recall 93%. Pada media sosial Facebook nilai akurasi 70%, presisi 77% dan recall 74%. Pada pengujian data calon Gubernur akurasi tertinggi diperoleh dari klasifikasi data KBS-Ace pada data yang diambil dari Twitter dengan nilai akurasi 89% presisi 91% dan recall 94% dan akurasi terendah pada saat proses klasifikasi data KBS-Ace pada media sosial Facebook

REFERENSI

- [1] Tristiyanti, Diana Ikasari, Elisa Pratiwi dan Nur Syahri, "Analisa Sentimen Terhadap *Review Restoran Fish Streat* Pada Aplikasi Zomato

- Menggunakan *Stemming* Nazief Adriani dan *Naïve Bayes Classifier*," Vol.3.2017
- [2] Murnawan dan Ardiles Sinaga, "Pemanfaatan Analisis Sentimen Untuk Pemingkatan Populalitas Tujuan Wisata," *Jurnal Penelitian Pos dan Informatika*, vol.7, no.2, 109-120, Desember 2017
- [3] Ahmad Fathan, H dan Azhari SN, "Analisis Sentimen Dan Klasifikasi Katagori Terhadap Tokoh Publik Pada Twitter," *Seminar Nasional Informatika*, ISSN: 1979-2328, Agustus 2014
- [4] Bridge, C. 2011. *Unstructured Data and the 80 Percent Rule*. [Online]. Tersediadi: <http://www.clarabridge.com/default.aspx?tabid=137&ModuleID=635&ArticleID=551> [diunduh : 6 Mei 2018].
- [5] Ida Bagus Gede Widnyana Putra, Made Sudarma dan I Nyoman Satya Kumara "Klasifikasi Text Bahasa Bali Dengan Informatin Gain dan *Naïve Bayes Classifier*," *Majalah Ilmiah Teknologi Elektro* 2016
- [6] Mohammad Arif and Gaurav Bathla, "Lexcion Based Sementiment Analysis on Facebook Page," *Journal of Engineering and Applied Sciences*, vol.12, no.2, 2017
- [7] R. Sudrajat, Rudi Rosadi and Harits Muhammad, "Implementation of Data Mining in Analyzing Social Media Users Personality with *Naïve Bayes Classifier*: A Case Study of Instagram Social Media," *International Journal of Computer Science Issues*, Volume 13, Issue 4, July 2016
- [8] Vishal A. Kharde dan S.S. Sona "Analisis Sentimen dari Data Twitter: Survei Teknik," *International Journal of Computer Applications*, (0975 – 8887), Volume 139 – No.11, April 2016
- [9] Konstantinas, Paulius dan Gintautas "SVM and Naive Bayes Classification Ensemble Method for Sentiment Analysis" *Baltic J. Modern Computing*, vol.5. no.4, pp. 398–409, 2017
- [10] Syahmia Gusriani, Kartina Diah K.W DAN Muhammad I.Z, "Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi *Naïve Bayes* (Studi Kasus: Facebook Page BerryBenka) " Conference Paper September 2016
- [11] Agnes Rossi.T, Rizal S. dan M. Ali Fauzi, " Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan *Naïve Bayes* dan Pembobotan *Emoji*" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 1, No. 12, Desember 2017