

# Klasifikasi Model Percakapan Twitter Mengenai Ujian Nasional

Emily Uly Artha<sup>1\*</sup>, Ahmad Dahlan<sup>2</sup>

<sup>1</sup>Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Magelang

<sup>2</sup>Jurusan Sistem Informasi, Universitas AMIKOM Yogyakarta

<sup>1</sup>Jln. Mayjend Bambang Soegeng, Mertoyudan, Magelang 56172, Indonesia

<sup>2</sup>Jln. Ring Road Utara, Condong Catur, Sleman, Indonesia

email: <sup>1</sup>ully@ummgl.ac.id, <sup>2</sup>alland@amikom.ac.id

**Abstract** – The rise of internet usage especially in social media such as Twitter make a possibility to analyze an existing conversations and derived it to various conversation theme. The data was obtained from the Twitter conversation using text mining that is a "national exam" using the National #ujian, #unas, and #ujianan hastags. Twitter API was used to extract tweet data. This research use the naive bayesian access method to view communication forms based on users and applications. Based on calculation this research obtain a 0.278 result that most of Twitter users (users) are being from @kabarDiknas. The data that used in this research are taken from is a twitter conversation about the "ujian nasional" in 2016.

**Abstrak** – Berkembangnya pengguna internet mengakibatkan data Analisa percakapan yang terdapat pada media sosial seperti Twitter dapat digunakan untuk melihat pola percakapan dengan menggunakan tema atau bentuk percakapan khusus. Data yang didapatkan berasal dari data percakapan Twitter menggunakan teks mining yaitu "ujian nasional" dengan menggunakan *hashtag* #ujianNasional, #unas, dan #ujian nasional. Untuk mengekstrak data tweet menggunakan Twitter API. Metode klasifikasi *naive bayesian* digunakan untuk melihat bentuk percakapan berdasarkan pengguna dan aplikasi yang digunakan. Dari hasil perhitungan didapatkan untuk pengguna (*user*) yang terbanyak adalah dari @kabarDiknas, dan aplikasi twitter yang terbanyak adalah *twitter for android* sebanyak 0.278. Data yang diambil adalah percakapan twitter tentang ujian nasional pada tahun 2016.

**Kata Kunci** – Sosial Media, twitter, *naïve bayes classification*, teks mining, pola data

## I. PENDAHULUAN

Hasil penelitian yang dilakukan oleh Facebook dalam laporan *state of connectivity* tahun 2015 yang berjudul *a report global internet access* yang dibuat oleh Facebook, hingga akhir tahun 2015 jumlah pengguna internet mencapai 3,2 milyar orang, ada kenaikan 200 juta pengguna dari tahun sebelumnya, tetapi masih ada sekitar 4,1 milyar yang belum menggunakan internet, jadi dapat diprediksi pengguna internet ke depannya semakin bertambah. Di Indonesia sendiri dengan penetrasi penggunaannya sendiri sebesar 34 persen, kini pengguna internet mencapai 88,1 juta pengguna dan 79 juta diantaranya merupakan pengguna media sosial

aktif, dan merupakan salah satu negara teraktif di media sosial.

Setiap layanan media sosial pasti terdapat media penyimpanannya. Media penyimpanan atau *database* kaya akan informasi tersembunyi yang dapat digunakan sebagai dasar dalam pengambilan keputusan atau untuk melakukan prediksi tren data dimasa yang akan datang. Data-data yang tersimpan di dalam *database* hanya berisi kumpulan *string*. Untuk melakukan pencarian informasi di dalam database tersebut tentunya harus menggunakan pola pencarian data yang relevan. Berdasarkan data yang dikeluarkan pada April 2014, diketahui bahwa pada kuartal kedua tahun 2014 Twitter memiliki pengguna aktif sebanyak 271 juta jiwa dengan jumlah rata-rata tweet yang dihasilkan sebanyak 58 juta tweet perhari. sebagai sebuah informasi, Twitter dapat menyediakan data dengan jumlah sangat besar dan mudah didapat. Informasi yang disediakan oleh Twitter terdiri dari berbagai ragam topik, menjadikan Twitter sebagai salah satu pusat informasi sehingga pengetahuan dapat diambil di dalamnya. Informasi yang diperoleh di Twitter dapat digunakan untuk berbagai macam keperluan, dalam penelitian ini kicauan (*tweet*) dari Twitter dapat digunakan untuk mengetahui bentuk percakapan yang terjadi pada saat Ujian Nasional (UNAS) tahun 2015. Untuk mendapatkan informasi-informasi yang diinginkan tersebut, perlu dilakukan pemrosesan awal terhadap tweet yang beragam, dimana banyak ditemukan hal-hal yang sangat jarang digunakan seperti kata tidak baku, pengulangan kata yang berlebihan, penggunaan karakter-karakter yang tidak memiliki arti.

Tahapan ekstraksi data/teks mempunyai tujuan adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang dimiliki. *Datamining* sering juga disebut sebagai *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan pola atau hubungan dalam *set* data berukuran besar [1]. Fungsionalitas dari *datamining* adalah untuk menemukan pola /tingkah laku dari suatu *dataset* yang diamati. Ada beberapa metode yang dipakai untuk melakukan *datamining* yaitu asosiasi, klasifikasi dan analisa cluster. Ketika menggunakan beberapa macam algoritma klasifikasi dengan jumlah data yang besar seperti dari Twitter maka dipastikan terdapat data yang rusak yang mengharuskan kita melakukan proses *preprocessing* [2].

Penelitian ini akan melakukan uji pengambilan data Twitter dengan menggunakan teks *mining* yaitu "ujian nasional" dan menggunakan *hashtag* (#) tentang "#ujian

\* ) penulis korespondensi (Emily Uly Artha)  
Email: ully@ummgl.ac.id

nasional”, “unas”, “#ujiannasional” dan berikutnya akan di klasifikasikan berdasarkan tahun 2015, teks yang berhasil ditangkap kemudian akan dianalisa menggunakan metode *naïve bayessian classification* untuk dilihat seringnya kata yang muncul, penggunaan aplikasi Twitter serta nilai distribusinya pada setiap atributnya.

## II. PENELITIAN YANG TERKAIT

Data di internet yang dapat diambil seperti data percakapan twitter dilakukan oleh [3] untuk penelitian pada sekolah SMA terpusat pada bentuk komunikasi pendidikan *non formal* seperti analisis menggunakan soal ujian online. Dapat pula datamining digunakan untuk melihat visualisasi kepadatan lalu lintas di Kota Bandung dengan melihat pola percakapan di twitter [4].

Teks *mining* merupakan salah satu metode analisis data teks untuk memperoleh informasi yang relevan. Teks mining menarik para peneliti karena peningkatan layanan digital seperti layanan web yang disediakan oleh google (*google books*) dan layanan *media social* seperti Twitter. Twitter mempunyai format teks tersendiri dari aplikasi sosial media lainnya yang hanya maksimal jumlah kata yang dapat di tweetkan berjumlah 140 karakter. Contoh pesan yang bisa kita dapatkan adalah semisal kita mengikuti (*following*) Twitter atau biasa disebut dengan (“tweet”): RT @toni has a coll #job. Ini menunjukkan bahwa pengguna tersebut membalas pesan ke orang lain dengan indikasi RT (*retweet*) dan subjek bisa kita lihat dari karakter @, dan tanda *hashtag* (#) digunakan untuk tema percakapan. Kumpulan dari tanda *hashtag* ini yang bisa digunakan untuk melihat percakapan yang sedang menjadi trending topik di Twitter. Dalam konteks teks *mining* di Twitter, ada dua fungsi utama yang berhubungan dengan Twitter: (a) grafik yang berasal dari link antar pesan, dan (b) teks *mining* yang berasal dari analisis dari kumpulan pesan.

## III. METODE PENELITIAN

### A. Twitter Streaming API

API merupakan singkatan dari *Application Programming Interface* yang merupakan sebuah perangkat lunak untuk membuat intruksi komputer dan layanan web dengan mudah. API menggunakan basic otentikasi dari HTTP (*hypertext transfer protocol*) dan menggunakan *account* Twitter. Data dapat diambil dalam format XML atau JSON. Dengan menggunakan API Twitter dapat diambil data dalam format csv yang masih dalam bentuk seadanya. Penelitian ini diawali dengan melakukan *download Tweet* secara *real time* yang mengandung hashtag #unas atau #UjianNasional. Data Tweet yang dikumpulkan antara lain “created-at”, “from-user”, “from-user-id”, “to-user”, “language”, “source”, “text”, “geo-location-lattitude”, “geo-location-longitude”, “retweet-count”, “id”. Tahap berikutnya adalah melakukan pelabelan sesuai dengan kelas yang ditentukan sebelumnya (Rodiyansyah and Winarko, 2012), dan berikutnya adalah melakukan *text preprocessing* dari data tersebut.

### B. Sosial Media Analysis

Analisis pada *social media* mempunyai banyak tujuan. Salah satunya adalah untuk mengklasifikasikan bentuk topik percakapan dan mencari *trending topic* yang menjadi topik

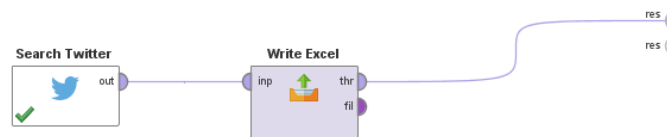
utama. Tujuan lain adalah untuk menganalisa sentiment seseorang terhadap suatu topik. Subjek dari analisa ini bisa seseorang, organisasi, kelompok masyarakat atau produk tertentu [5]. Jutaan pesan yang lalu lalang di internet merubah bentuk komunikasi tradisional yang mengharuskan bertatap muka. Twitter dapat menunjukkan *sentiment* dan opini dari beberapa teks dengan beberapa alasan:

- Platform *microblogging* digunakan oleh orang yang berbeda untuk mengekspresikan pendapat mereka tentang topik yang sama atau berbeda.
- Twitter mengandung sejumlah besar posting dalam bentuk teks
- Pengguna Twitter bervariasi dari selebriti, politikus, bahkan presiden dari suatu negara sehingga lebih beragam dan mempunyai kepentingan yang berbeda.
- Di Twitter tidak menggunakan *emoticon*, sehingga analisa lebih mudah.

### C. Data Tweet

Tweets adalah teks tulisan hingga 140 karakter yang ditampilkan pada halaman profil pengguna. Tweets bisa dilihat secara publik, namun terbatas pengiriman pesan ke daftar pertemanan (*follower*). Data Tweet dalam penelitian ini diperoleh dengan memanfaatkan API (*Application Programming Interface*) yang disediakan oleh Twitter. Dengan memanfaatkan API tersebut maka kita dapat mengambil data sesuai dengan yang kita inginkan, walaupun dalam kenyataannya tidak semua data yang dapat diambil.

Pada saat pengambilan data, sebelumnya sudah mendaftarkan untuk mendapatkan *Access token* dan *consumer key* (API key) pada alamat dev.twitter.com. Setelah mendapatkan kedua token tersebut, langkah berikutnya adalah menggunakan *software Rapidminer* untuk mendapatkan data Twitter, karena di *Rapidminer* mempunyai aplikasi untuk mengambil data dari *rapidminer* yang bisa dilihat pada Gbr.1.



Gbr.1 Desain pengambilan data Twitter

Dari Gbr 1 dapat dijelaskan bahwa hasil pencarian data di Twitter akan di simpan ke dalam sebuah *file* dalam format xls, yang berikutnya akan membentuk sebuah tabel *training* yang digunakan untuk proses training. Tweet yang dikumpulkan adalah Tweet yang mengandung hasta #Unas atau #UjianNasional. Pengambilan data Tweet sebanyak 539 record

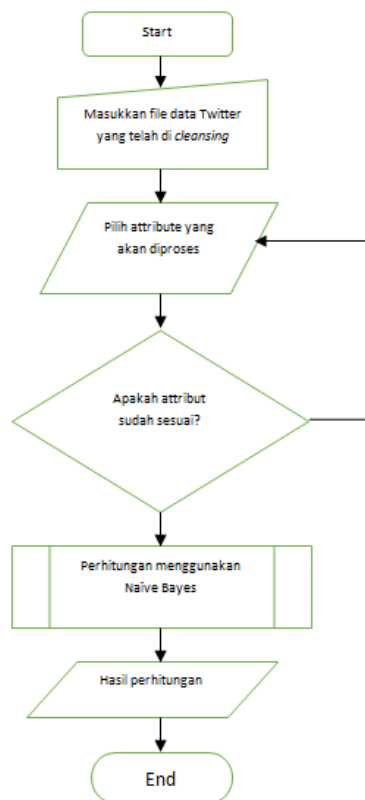
## IV. HASIL DAN PEMBAHASAN

Proses analisis data khususnya data yang didapatkan dari percakapan di Twitter dapat dilakukan menggunakan ekstraksi dari hasil GTI (*Get Tweet Information*). Proses penambangan data sendiri bukan merupakan sebuah proses yang tunggal namun merupakan proses yang berkelanjutan [2], dimulai dari adanya data mentah yang dilakukan di awal diikuti dengan proses penambangan data dan menghasilkan keluaran yang diharapkan [6].

A. Pemrosesan Awal

Tahapan dalam melakukan proses pengambilan data adalah masuk menggunakan akun yang telah terdaftar di Twitter dilanjutkan dengan pembuatan model diawali membuat *flowchart* untuk melihat prosedur secara umum, dan melihat urutan dari prosedur yang ada dalam system. Data yang didapatkan dari Twitter adalah data dalam format teks dan dalam format csv, dimana masih terdapat data yang tidak sesuai sehingga harus dilakukan pembersihan dan pemilihan data terlebih dahulu sebelum dilakukan proses analisa kemudian baru dilakukan validasi dalam tahapan selanjutnya

Dari *flowchart* pada gambar 2 dapat dilihat bahwa langkah pertama adalah melakukan pengambilan data Twitter yang telah dilakukan *cleansing* sebelumnya. Berikutnya adalah memilih atribut yang akan diproses dengan menggunakan algoritma *naïve bayes classification* untuk kemudian dimunculkan hasilnya. Data Twitter yang diambil masih dalam format csv, yang mana format ini harus dirubah terlebih dahulu kedalam bentuk table dalam format excel (.xls). Data *training*, data *testing* disimpan dalam format excel untuk kemudian dilakukan validasi.



Gbr.2 Tahapan proses analisa data Twitter

B. Pembersihan Data

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau tidak relevan. Pembersihan data (*data cleansing*) juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlahnya, setelah itu adalah integrasi data ke dalam satu database *warehouse*. Dalam operasi yang lebih kompleks, pembersihan data dapat dilakukan oleh komputer dengan menggunakan perangkat lunak tertentu, seperti menghapus sebuah baris yang tidak lengkap, memperbaiki

kata-kata yang terdapat salah ejaan, dan menghapus duplikat salinan. Dalam penelitian ini banyak sekali baris yang kosong sehingga harus dihilangkan, selain itu juga terdapat kata yang tidak berpola seperti menggunakan simbol-simbol tertentu.

C. Seleksi dan Transformasi Data

Data yang ada dalam *database*/tabel kemudian direduksi untuk mendapatkan hasil yang akurat. Beberapa cara seleksi antara lain:

- *Sampling* adalah seleksi subset/reprentatif dari populasi data yang besar
- *Denoising* adalah proses menghilangkan noise dari data yang akan ditransformasikan
- *Feature extraction* adalah proses membuka spesifikasi data yang signifikan dalam konteks tertentu.

Data yang telah ditransformasi, kemudian ditambah (*mining*) dengan berbagai teknik. Proses seleksi dan transformasi dilakukan agar data menjadi berguna dan dapat ditelusuri. Proses seleksi dan transformasi data ini mengambil data dalam bentuk *file* excel dengan pemilihan pada kata yang akan dicari.

D. Evaluasi pola dan presentasi pengetahuan

Tahap ini merupakan bagian dari proses pencarian pengetahuan yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan dengan fakta atau hipotesa yang ada sebelumnya dan direpresentasikan dalam bentuk sebuah pengetahuan yang mudah dipahami. Tahapan ini menggunakan desain pada *Rapidminer* untuk memudahkan dalam membaca data yang telah dilakakukan sebelumnya.

E. Algoritma Naïve Bayes Classification

Algoritma *naïve bayes classification* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi yang ditemukan oleh ilmuwan Inggris Thomas Bayes dan dikombinasikan dengan teori bayes. Algoritma ini digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang tepat. Pada algoritma *naïve bayes* sejumlah petunjuk yang disebut dengan atribut diperlukan untuk membantu dalam membetuk kelas yang sesuai bagi sampel yang dianalisa.

Dalam algoritma *naive bayes* setiap dokumen direpresentasikan dengan pasangan atribut  $x_1, x_2, x_3, \dots, x_n$  dimana  $x_1$  adalah kata pertama,  $x_2$  adalah kata kedua dan seterusnya. Sedangkan  $V$  adalah jumlah himpunan kategori yang dipilih. Pada saat klasifikasi algoritma *naïve bayes* akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan ( $V_{map}$ ) dimana persamaannya adalah sebagai berikut

$$V_{map} = \arg \max_{V_j \in V} \left( \frac{P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)} \right) \quad (1)$$

Untuk  $P(x_1, x_2, x_3, \dots, x_n)$  nilainya konstan untuk semua kategori ( $V_j$ ) sehingga persamaannya dapat ditulis sebagai berikut :

$$V_{map} = \arg \max_{V_j \in V} (P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)) \quad (2)$$

Persamaan diatas (2) dapat disederhanakan menjadi :

$$V_{map} = \arg \max_{V_j \in V} \prod_{i=1}^n (P(x_i|V_j)P(V_j)) \tag{3}$$

Keterangan :

- V<sub>j</sub>: Kategori/Atribut yang digunakan
- P(x<sub>i</sub>|V<sub>j</sub>): probabilitas xi pada kategori V<sub>j</sub>
- P(V<sub>j</sub>): probabilitas dari V<sub>j</sub>

Untuk P(V<sub>j</sub>) dan P(x<sub>i</sub>|V<sub>j</sub>) dihitung saat training , dimana persamaannya adalah sebagai berikut :

$$P(V_j) = \frac{|docs\ j|}{|contoh} \tag{4}$$

$$P(x_i|V_j) = \frac{nk+1}{n+|kosakata} \tag{5}$$

Keterangan :

- |docs j|: jumlah dokumen pada kategori j
- |contoh|: jumlah dokumen dari semua kategori
- N<sub>k</sub>: jumlah kemunculan kata xi pada kategori V<sub>j</sub>
- N: jumlah kata dalam setiap kategori
- |kosakata|: jumlah semua kata dari semua kategori

F. Analisa Teks Mining

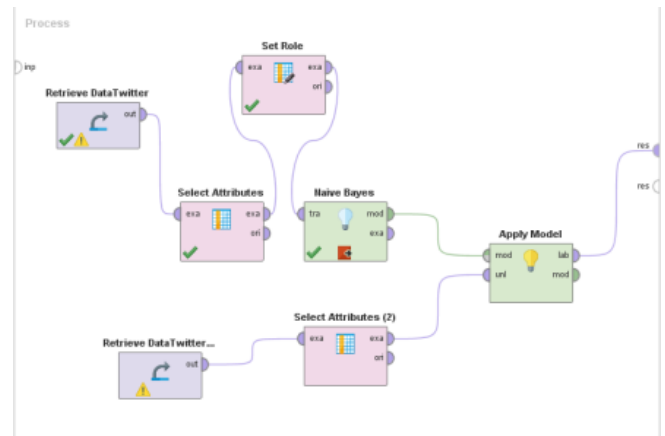
Data yang digunakan untuk penelitian ini adalah data hasil Tweet yang didapatkan pada tahun 2015 dengan memasukkan kata kunci menggunakan tanda *hashtag* (#) yaitu #ujiannasional, #unas, #un. Dari hasil yang sudah diambil terdapat sebanyak 539 baris. Dari kesemua data tersebut ada beberapa yang tidak lengkap dan kosong. Penelitian ini mencari nilai *support* dan *confidence* dari hubungan tingkat kelulusan dengan pengguna (*user*) twitter. Tidak semua data induk akan digunakan, hanya data tertentu yang akan digunakan. Adapun yang akan diproses ke dalam teks mining meliputi :

- 1) *Penggunaan aplikasi Twitter*: Hasil dari proses mining akan melihat aplikasi yang digunakan untuk penggunaan Twitter
- 2) *Hubungan pengguna (user) dengan aplikasi Twitter*: Menganalisa pengguna yang sering menggunakan Twitter dengan tema #unas atau #UjianNasional
- 3) *Atribut yang digunakan*: Atribut berfungsi untuk menggambarkan objek seperti user, id-pengguna, nama, alamat dan sebagainya
- 4) *Tabel Distribusi*: Tabel distribusi digunakan untuk menyederhanakan jumlah data sehingga penyajiannya dapat lebih mudah dipahami.

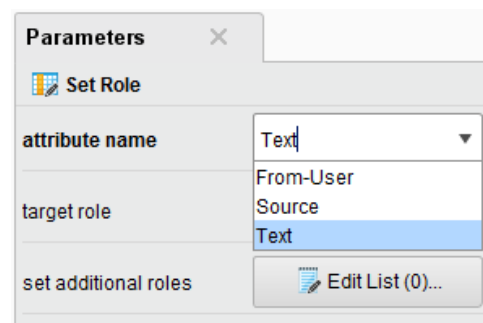
Dari hasil analisa teks mining tersebut maka dapat dimunculkan bentuk dari percakapan yang telah dimasukkan ke *Rapidminer* menjadi bentuk tabel distribusi.

G. Validasi Proses Data Training dan Data Testing

Tahapan ini bertujuan untuk melakukan analisis berbagai model dengan kinerja prediksi yang baik. Proses data *training* yaitu melakukan proses pelatihan pada data model, sedangkan proses *testing* yaitu melakukan pengujian data yang akan menghasilkan grafik atau pola



Gbr.3 Pemodelan data training dan data testing



Gbr.4 Pemilihan atribut.

Dari Gbr.3 dapat dijelaskan bahwa proses data training menggunakan model *naïve Bayesian* yang sebelumnya dihubungkan dengan pemilihan data Twitter, lalu dilanjutkan dengan pemilihan atribut “user”, “text” dan “from-user” pada bagian parameter. keterhubungan antar dokumen atau kalimat, diperlihatkan pada Gbr.4.

H. Hasil Perhitungan Menggunakan Rapidminer

Dari proses perhitungan data mining menggunakan algoritma *naïve Bayesian* dihasilkan suatu informasi baru yaitu perhitungan *data mining* berdasarkan 3 data training yaitu “from user”, “user” dan “text” didapatkan beberapa nilai distribusi. Dari hasil olahan data dapat diejelaskan bahwa penggunaan aplikasi Twitter terbanyak adalah menggunakan aplikasi *twitter for android* dengan jumlah 104, posisi berikutnya adalah *twitter web client* dengan jumlah 59 dan posisi ketiga adalah Instagram sebanyak 29. Untuk lebih jelasnya dapat dilihat pada tabel I, tabel II, dan tabel III. Data tersebut merupakan data dengan nilai tertinggi untuk keseluruhan data yang sudah dimunculkan hasil perhitungan dan analisisnya.

Dari ketiga tabel dapat dijelaskan bahwa dengan menggunakan atribut “user” pada API Twitter maka dapat ditarik sebuah bentuk/model.

TABEL I  
HASIL DISTRIBUSI ATRIBUT USER

Jumlah Distribusi	User	nilai
2	@KabarDiknas	0.027
2	@KabarDiknas	0.021
2	@KompasMuda	0.013

TABEL II  
HASIL DISTRIBUSI ATRIBUT SOURCE

Jumlah Distribusi	Source	nilai
2	Twitter for android	0.278
2	Twitter web client	0.158
2	Instagram	0.078
2	Mobile web	0.067

TABEL III  
HASIL DISTRIBUSI ATRIBUT TEXT

Text	nilai
Menteri pendidikan Anis Baswedan..#UjianNasional	0.359
Selesai UN, Pelajar SMP ...#edukasi #nasional #ujiannasional	0.003
Soal UN SMP terlalu sulit, siswa ...#ujiannasional @unsms2016	0.001

## V. KESIMPULAN

Berdasarkan hasil pengujian dan analisa bahwa algoritma *Naïve Bayes Classification* memiliki kinerja yang cepat dan akurat untuk klasifikasi dokumen teks. Penggunaan hastag #unas atau #UjianNasional dapat digunakan untuk memperjelas tema teks yang akan dianalisa. Proses kerja sistem diawali dengan menggunakan API Twitter yang

disimpan dalam *file* data excel kemudian diolah menggunakan preprosesing. *Cleansing data* dilakukan pada saat preprosesing kemudian diolah menggunakan *naïve Bayesian classification*. Jumlah atribut yang digunakan dalam penelitian ini berjumlah 3 yaitu “from-user”, “source”, dan “text” tentunya harus dicoba lagi dengan menambahkan atribut dan menggunakan pola klasifikasi yang lain seperti *decision tree* untuk memunculkan ilmu pengetahuan yang baru dari proses teks *mining* dengan menggunakan data dari Twitter. Untuk wilayah tidak dapat dimunculkan karena banyaknya data tidak sesuai pada saat preprosesing sehingga menggunakan percakapan Twitter secara umum tidak spesifik di wilayah Yogyakarta.

## UCAPAN TERIMA KASIH

Makalah ini merupakan luaran dari kegiatan Penelitian Dosen Pemula yang dibiayai oleh RISTEKDIKTI tahun 2016

## DAFTAR PUSTAKA

- [1] Ridwan, M., Suyono, H. and Sarosa, M., 2013. Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*.
- [2] Duncan, B. and Zhang, Y., 2015. Neural Networks for Sentiment Analysis on Twitter.
- [3] Karaoglan, B., Candemir, C. and Haytaoglu, E., 2014. Using Twitter as a Diagnostic Teaching and Learning Assessment Tool. pp. 1–4.
- [4] Rodiyansyah, S. and Winarko, E., 2012. Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. *Indonesian Journal of Computing and Cybernetics Systems*, 6(1), pp. 91–100.
- [5] Pingkan, Widyawan, Warsun, N. and Aji, T. B., 2015. An Evaluation of Twitter River and Logstash Performances as Elasticsearch Inputs for Social Media Analysis of Twitter. *IEEE*, 978-1-5090, pp. 181–186.
- [6] Raharjo, S. and Winarko, E., 2014. Klasterisasi, klasifikasi dan peringkasan teks berbahasa indonesia. *Kommit 2014*, 8(Kommit), pp. 391–401.