p- ISSN : 2407 – 1846
e-ISSN : 2460 – 8416

Website : jurnal.umj.ac.id/index.php/semnastek

PENGARUH TOKOH AHOK PADA MEDIA SOSIAL MENJADI TRENDING TOPIC MENGGUNAKAN METODE CLASSIFICATION

Yudi Permana Wiyadi[1*], Taufiqurrochman[2]
[*12]Ilmu Komputer, STMIK Nusa Mandiri, Jakarta
Jl.Kramat Raya No.18 Jakarta Pusat, 10420
[*]E-mail : yudi.wiyadi@gmail.com

ABSTRAK

Tingginya pemberitaan seorang tokoh dalam sebuah pemberitaan baik dimedia elektronik maupun media cetak, menjadikan seorang tokoh tersebut buah bibir pembicaraan khalayak ramai. Terlebih pada era digital saat ini memberikan kemudahan bagi masyarakat banyak untuk mengakses situs pemberitaan secara online, dan kadang berita-berita yang dimuatnya itu dibuat secara real time tentang issue-issue yang sedang hot pada saat ini dan bahkan bisa menjadi trending topik pembicaraan dalam situs pemberitaan tersebut. Penelitian disini akan mengambil satu media sosial yang menjadi wadah masyarakat banyak membicarakan akan hal-hal berkaitan dengan tokoh yang banyak diberitakan yaitu Twitter.

Kata kunci: tokoh, issue-issue, berita, twitter

ABSTRACT

The high coverage of a character in a good news electronic media and print media, making a figure is the fruit of the talk of the general public. Especially in the current digital era makes it easy for many people to access news sites online, and sometimes the news that was published was made in real time about the issues that are hot at the moment and can even be trending the topic of conversation in the news site the. The research here will take a social media that becomes a place of society to talk about things related to the much-publicized Twitter.

Keywords : *characters, issues, news, twitter*

PENDAHULUAN

Tingginya pemberitaan seorang tokoh dalam sebuah pemberitaan baik dimedia elektronik maupun media cetak, menjadikan seorang tokoh tersebut buah bibir pembicaraan khalayak ramai. Terlebih pada era digital saat ini memberikan kemudahan bagi masyarakat banyak untuk mengakses situs pemberitaan secara online, dan kadang berita-berita yang dimuatnya itu dibuat secara real time tentang issue-issue yang sedang hot pada saat ini dan bahkan bisa menjadi trending topik pembicaraan dalam situs pemberitaan tersebut.

Keingintahuan masyarakat akan pemberitaan yang sedang hot menjadikan tokoh tersebut menjadi perbincangan dibanyak media sosial hingga nama seorang tokoh tersebut menjadi trending topic karena banyaknya yang mempostingkan tokoh yang sedang diberitakan itu. Akan hal tersebut pada paper ini akan melakukan penelitian seberapa berpengaruhnya tokoh yang sedang di beritakan dengan kasus nya itu menjadi trending topic. Penelitian disini akan mengambil satu media sosial yang menjadi wadah masyarakat banyak membicarakan akan hal-hal berkaitan dengan tokoh yang banyak diberitakan yaitu Twitter.

Twitter adalah adalah layanan jejaring sosial dan mikroblog daring yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter, yang dikenal dengan sebutan kicauan (tweet). Twitter didirikan pada bulan Maret 2006 oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan Juli. Sejak diluncurkan, Twitter telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di Internet, dan dijuluki dengan pesan singkat dari Internet. Di Twitter, pengguna tak terdaftar hanya bisa membaca kicauan, sedangkan pengguna terdaftar bisa menulis kicauan melalui antarmuka situs web, pesan singkat (SMS), atau melalui berbagai aplikasi untuk perangkat seluler.

Tipe Artikel

p- ISSN : 2407 – 1846
e-ISSN : 2460 – 8416
Website : jurnal.umj.ac.id/index.php/semnastek

Dalam paper ini akan menerapkan satu algoritma klasifikasi untuk menguji seberapa akuratkah informasi, issue atau pemberitaan yang menjadi banyak pembicaraan orang melalui media sosial twitter. Pada paper ini akan menerapkan algoritma C4.5 dalam pengolahan text mining data yang diambil dari cuitan di twitter, yang nantinya akan menghasilkan tingkat akurasi benarkah seorang tokoh sedang sedang di beritakan bisa menjadi pembicaraan trending topic disebuah sosial media dalam kasus ini di twitter.

Algoritma C4.5 diperkenalkan oleh Quinlan (1996) sebagai versi perbaikan ID3. Dalam ID3, induksi decision tree hanya bisa dilakukan pada fitur bertipe kategorikal (nominal atau ordinal), sedangkan tipe numerik (interval tau rasio) tidak dapat digunakan. Perbaikan yang membedakan algoritma C4.5 dari ID3 adalah dapat menangani fitur dengan tipe numerik, melakukan pemotongan (prunning) decision tree, dan penurunan (deriving) rule set. Algoritma C4.5 juga menggunakan kriteria gain dalam menentukan fitur yang menjadi pemecah node pada pohon yang diinduksi (Eko Prasetyo : 2014)

Model algoritma pengklasifikasian ini sebenarnya bukan hanya algoritma C4.5, ada jenis-jenis lain algoritma yang tergolong kedalam proses algoritma pengklasifikasian. Dalam penelitian yang sudah pernah dilakukan kebanyakan peneliti menggunakan algoritma Naive Bayes Classifier untuk menghasilkan tingkat akurasi.

Rumusan Masalah

Apakah algoritma C4.5 akan memberikan hasil yang akurat untuk pengolahan text mining di twitter menjadi sebuah trending topic, sehingga akan diperoleh sebuah pola atau aturan-aturan yang dapat dijadikan sebagai acuan dalam memprediksi sebuah cuitan akan menjadi trending topic atau tidak?

Apakah algoritma Naive Baiyes akan memberikan hasil yang akurat untuk pengolahan text mining di twitter menjadi sebuah trending topic, sehingga akan diperoleh sebuah pola atau aturan-aturan yang dapat dijadikan sebagai acuan dalam memprediksi sebuah cuitan akan menjadi trending topic atau tidak?

Apakah algoritma Naive Baiyes lawan C4.5 akan memberikan hasil yang akurat untuk pengolahan text mining di twitter menjadi sebuah trending topic, sehingga akan diperoleh sebuah pola atau aturan-aturan yang dapat dijadikan sebagai acuan dalam memprediksi sebuah cuitan akan menjadi trending topic atau tidak?

METODE

Agar paper ini tetap didalam alurnya dan tidak menyimpang dari apa yang telah diterapkan serta pembahasannya tidak terlalu luas dan sesuai dengan yang direncanakan penulis, maka paper ini penulis batasi permasalahannya yaitu masalah yang terkait dengan algoritma C4.5 dan Naive Baiyes menggunakan klasifikasi data mining dengan cara menganalisis sejumlah atribut yang menjadi parameter dalam prediksi text dari twitter yang menggunakan hastag (#) dan nama tokoh yang dimaksud.

Dalam paper ini penulis memakai algoritma C4.5 dan Naive Baiyes juga untuk pembanding menggunakan metode algoritma klasifikasi yang lainnya. Serta didukung dengan perangkat lunak RapidMiner yang merupakan aplikasi data mining berbasis open source (GPL) dan berengine Java, dengan Graphical User Interface (GUI) menggunakan java dan Microsoft Excel sebagai Databasenya.

Proses Data Mining

Secara sistematis, ada tiga langkah utama dalam data mining (Gonunescu, 2011):

1. Eksplorasi/pemrosesan awal data

Eksplorasi/pemrosesan awal data terdiri dari ‘pembersihan’ data, normalisasi data. Transformasi data, penanganan data yang salah. Reduksi dimensi, pemilihan subset fitur, dan sebagainya.

2. Membangun model dan melakukan validasi terhadapnya

Membangun model dan melakukan validasi terhadapnya berarti melakukan analisis berbagai model dan meilih model dengan kinerja presdiksi yang terbaik. Dalam langkah ini digunakan metode-metode seperti klasifikasi, regresi, analisis cluster, deteksi anomali, analisis asosiasi, analisis pola sekuensial, dan sebagainya. Dalam beberapa referensi, deteksi anomali juga masuk dalam langkah eksplorasi. Akan tetapi, deteksi anomali juga dapat digunakan sebagai

p- ISSN : 2407 – 1846
e-ISSN : 2460 – 8416
Website : jurnal.umj.ac.id/index.php/semnastek

algoritma utama, terutama utuk mencari data-data yang spesial.

3. Penerapan

Penerapan berarti menerapkan model pada data yang baru untuk menghasilkan perkiraan/ prediksi masalah yang diinvestigasi.

Klasifikasi

Klasifikasi adalah sebuah proses analisa data yang menghasilkan modelmodel untuk menggambarkan kelas-kelas yang terkandung di dalam data (Han, Kamber, & Pei, 2012). Model - model tersebut disebut classifier. Jadi, classifier inilah yang akan digunakan untuk menyusun kelas-kelas yang terkandung didalam data. Ada banyak jenis algoritma klasifikasi, dua diantaranya adalah Decision Tree dan K Nearest Neighbour (k-NN).

Decision Tree

Decision Tree atau pohon keputusan adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawabah dari masalah yang dimasukkan. Pohon yang dibentuk tidak selalu berupa pohon biner. Jika semua fitur dalam data set menggunakan 2 macam nilai kategorikal maka bentuk pohon yang didapatkan berupa pohon biner. Jika didalam fitur berisi lebih dari dua macam nilai kategorikal atau menggunakan tipe numerikal maka bentuk pohon yang didapatkan biasanya tidak berupa pohon biner (Eko Prasetyo : 2014).

Text Mining

Text Mining (TM) dapat didefinisikan sebagai sebuah proses ilmiah dimana seorang peneliti berinteraksi dengan sekumpulan dokumen menggunakan berbagai perangkat untuk menganalisa teks yang terkandung dalam kumpulan dokumen tersebut (Feldman R. & Sanger James, 2007). Tujuan utama dari TM adalah menganalisa informasi untuk menemukan pola-pola (Aggarwal & Zhai,2012).

Hal ini sejalan dengan DM yang bertujuan untuk mengekstrak pola-pola menarik dari data, bedanya untuk TM data ini berbentuk teks sedangkan untuk DM data ini berbentuk angka (Weiss Sholom M., Indurkhya Nitin, & Zhang Tong, 2010). DM berasumsi bahwa data sudah dalam bentuk terstruktur sedangkan pada TM data dalam bentuk tidak terstruktur dari koleksi dokumen (corpus) harus diolah terlebih dahulu (preprocessing) menjadi bentuk terstruktur (Feldman R. & Sanger James, 2007).

Teknologi *Text Categorization*

Pendekatan teknologi yang digunakan untuk Text Categorization ada dua jenis, yaitu (Feldman R. & Sanger James, 2007):

1. Knowledge Engineering (KE)

KE menggunakan pengetahuan seorang ahli dalam menentukan kategorikategori yang akan ditanamkan ke dalam sistem, baik dalam bentuk deklaratif maupun dalam bentuk aturan klasifikasi. Kelemahan utama dari KE untuk membuat sistem TC adalah diperlukannya banyak tenaga ahli yang terlatih untuk merawat aturan klasifikasi yang dimiliki sistem.

2. Machine Learning (ML)

ML menentukan kategori-kategori melalui sebuah proses pembelajaran dari satu set contoh data yang sudah dikategorisasi sebelumnya. Kelebihan utama dari ML adalah tidak diperlukannya banyak tenaga ahli yang terlatih sebagaimana pada KE karena ML hanya memerlukan satu set contoh data saja yang sudah diklasifikasi.

Teknologi *Machine Learning* untuk *Text Categorization*

Ada beberapa metode Machine Learning untuk Text Categorization, dua diantaranya adalah sebagai berikut (Aggarwal & Zhai, 2012):

1. Metode Decision Tree

Decision Tree pada hakikatnya adalah mengurai ruang data secara hirarki menggunakan kondisi nilai yang dimiliki oleh attribut. Dalam konteks data berbentuk teks maka kondisi attributnya adalah ada atau tidak adanya satu atau lebih kata pada dokumen.

2. Metode Proximity Based

Proximity Based pada hakikatnya adalah menggunakan ukuran jarak untuk melaksanakan klasifikasi. Asumsi utamanya adalah dokumendokumen dalam kelas yang sama kemungkinan besar memiliki jarak yang dekat satu sama lain berdasarkan ukuran kemiripannya. Ukuran kemiripan ini misalnya diukur dengan mempresentasikan dokumen kedalam bentuk document vector.

Tahapan *Text Categorization*

TC memiliki empat tahapan proses utama yang satu sama lain salin terkait, yaitu:

1. Pengumpulan Dokumen

p- ISSN : 2407 – 1846
e-ISSN : 2460 – 8416
Website : jurnal.umj.ac.id/index.php/semnastek

Pada tahapan pertama ini, dokumen teks dikumpulkan menggunakan berbagai perangkat yang sesuai dengan sumber dokumen teks. Misalnya, untuk teks dari situs web maka digunakan web crawler (Weiss Sholom M., Indurkhya Nitin, & Zhang Tong, 2010).

2. Pre-processing

Pre-processing mengkonversi setiap dokumen dari koleksi dokumen ke dalam bentuk canonical-nya (Feldman R. & Sanger James, 2007). Misalnya dalam bentuk satuan kata yang terkandung dalam dokumen tersebut.

3. Document Representation

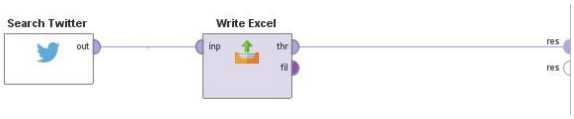
Setelah dokumen di-preprocessing selanjutnya dokumen harus direpresentasikan dalam bentuk terstruktur yang sesuai dengan bentuk masukan algoritma klasifikasi yang akan digunakan. Representasi yang umum dari dokumen adalah dalam bentuk bag-of-words (BoW). Selanjutnya BoW ini akan di-weighting (Feldman R. & Sanger James, 2007).

4. Classification

Dokumen yang telah direpresentasikan dalam bentuk yang terstruktur merupakan dataset untuk algoritma klasifikasi.

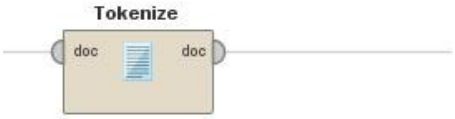
HASIL DAN PEMBAHASAN

Data yang digunakan adalah hasil aplikasi RapidMiner dengan tools yang digunakan adalah Search Twitter dengan langkah dan cara nya sebagai berikut dalam gambar 1 di bawah ini:



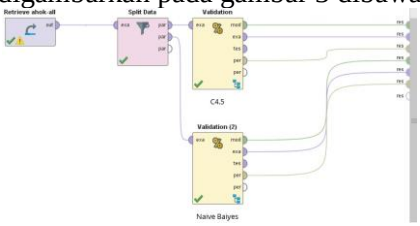
Gambar 1. Proses pengambilan data

Setelah didapat hasil cuitan di twitter menggunakan aplikasi RapidMiner tersebut barulah data tersebut kita olah menggunakan aplikasi yang sama juga. Langkah pertama adalah Tokenisasi kata-kata hasil dari pengambilan data tersebut, seperti yang terlihat pada gambar 2 dibawah ini:



Gambar 2. Proses Tokenisasi Data

Langkah selanjutnya adalah memasukan dua algoritma yang digunakan pada penelitian paper ini seperti yang tertulis pada Metode penelitian sebelumnya, langkah-langkah tersbut dapat digambarkan pada gambar 3 dibawah ini:



Gambar 3. Proses perbandingan dua metode algoritma

Hasil dari pengolahan di atas akan menampilkan tingkat akurasi dan juga auc antara dua metode algortima yang kita gunakan pada paper ini. Berikut dibawah ini ada table tingkat accuracy dari hasil prose perbandingan algoritma diatas:

accuracy: 99.41% +/- 1.76% (mikro: 99.43%)

	true in	true en	class precision
pred. in	83	1	98.81%
pred. en	0	91	100.00%
class recall	100.00%	98.91%	

Tabel 1. Tingkat Accuracy

Dari hasil proses di atas juga akan mendapatkan hasil recall dari perbandingan dua metode algortima yang digukankan, berikut adalah hasil recall yang keluar:

recall: 98.89% +/- 3.33% (mikro: 98.91%) (positive class: en)

	true in	true en	class precision
pred. in	83	1	98.81%
pred. en	0	91	100.00%
class recall	100.00%	98.91%	

Tabel 2. Tingkat Recall

Setelah semua proses diatas dijalankan maka dari proses tersebut akan menampilkan hasil juga berupa kurva AUC atau bisa disebutkan juga denga Area Under Curve seperti yang terlihat pada gambar dibawah ini:

Gambar 4. Hasil AUC

p- ISSN : 2407 – 1846
e-ISSN : 2460 – 8416
Website : jurnal.umj.ac.id/index.php/semnastek

SIMPULAN DAN SARAN

Dari hasil penelitian pada paper ini menunjukkan bahwa hasil yang di dapat dari penggunaan metode algoritma yaitu C4.5 (*Decision Tree*) dan juga Naive Baiyes itu sangat akurat. Terlihat dari hasil accuracynya yaitu berda di angka 99,41%, bahwa benar adanya tingkt pemberitaan suatu tokoh tersebut bisa membuat topoc dengan mencantumkan nama tokoh tersebut bisa menyebabkan satu *Hastag* (#) tersebut menjadi sebuah *Trending Topic* di sebuah media sosial Twitter. Saran untuk penelitian selanjutnya mungkin agar bisa membandingkan lagi metode-metode algoritma yang lain mungkin hasilnya akan lebih baik dari paper penelitian kali ini.

UCAPAN TERIMAKASIH

Saya ucapkan banyak terimakasih kepada pendamping saya dalam menuliskan paper penelitian ini yaitu saudara Taufiqurrochman atas dorongannyalah penelitian ini selesai.

DAFTAR PUSTAKA

Aprilla, D, Baskoro, Donny Aji, Ambarwati, Lia, & Wicaksana, IWayan Simri. (2013). Belajar Data Mining dengan Rapid Miner. Jakarta.

Aggarwal, Zhai (2012). Mining Text Data. Springer Sciences+Business Media, LLC.

Feldman R., Sanger James (2007). THE TEXT MINING HANDBOOK: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press.

Fakhrurroja, Hanif. (2009).Twitter Ngoceh Dapet Duit. Yogyakarta: Great

Gorunescu, F. (2011) Data Mining Comcept, Models and Techniques, Springer-Verlag.

Han, Kamber, Pei (2012) Data Mining Concept and Technique. Morgan Kaufman Pubisher.

Hermawati, Melayu S.P (2002). Manajemen Sumber Daya Manusia. Jakarta: Bumi Aksara.

https://books.google.co.id/books?id=_X57DQAAQBAJ&printsec=frontcover&dq=ahok&hl=jv&sa=X&redir_esc=y#v=onepage&q=ahok&f=false

Kothari, C. R. (2004). Research Methodology Methods and Techniques, Second Revised Edition. New Delhi: New Age International Publishers.

North, Matthew (2012). Data Mining for The Masses. A Global Text Project Book.

Publisher. https://books.google.co.id/books?id=I6lzCLqlyDQC&pg=PA1&dq=twitter+hanif&hl=en&sa=X&ved=0ahUKEwj7sIT_vIfVAhXDPI8KHUaZBeoQ6AEIJzAA#v=onepage&q=twitter%20hanif&f=false.

Prasetyo, Eko. (2014). Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta: Andi Offset.

Weiss Sholom M., Indurkhya Nitin, Zhang Tong (2010). Fundamentals of Predictive Text Mining. Springer-Verlag London Limited.