

ANALISIS SENTIMEN BERINTERNET PADA MEDIA SOSIAL AMIK BSI TEGAL DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES

Ahmad Fauzi¹, Amin Nur Rais², Muhammad Faittullah Akbar³, Windu Gata⁴

¹Jurusan Ilmu Komputer, Fakultas Magister Ilmu Komputer, STMIC NUSA MANDIRI
Jalan Kramat Raya No.18 Jakarta 10420 Telp.(021)31908575
Email: Ahmadfauzi1126@gmail.com

²Jurusan Ilmu Komputer, Fakultas Magister Ilmu Komputer, STMIC NUSA MANDIRI
Jalan Kramat Raya No.18 Jakarta 10420 Telp.(021)31908575
Email: Aminnurraikleca@gmail.com

³Jurusan Ilmu Komputer, Fakultas Magister Ilmu Komputer, STMIC NUSA MANDIRI
Jalan Kramat Raya No.18 Jakarta 10420 Telp.(021)31908575
Email: Faitm76@gmail.com

⁴Jurusan Ilmu Komputer, Fakultas Magister Ilmu Komputer, STMIC NUSA MANDIRI
Jalan Kramat Raya No.18 Jakarta 10420 Telp.(021)31908575
Email: Windu@nusamandiri.ac.id

ABSTRAK

Dalam penelitian analisis sentimen ini proses klasifikasi dokumen di bagi dalam dua kelas, yaitu kelas sentimen positif dan negatif. Media sosial telah menyediakan tempat bagi pengguna *web* untuk mengungkapkan berbagai pemikiran, pendapat dan menyampaikan berita tentang topik yang berbeda dalam sebuah acara. Data diperoleh dari jejaring sosial *Instagram* berdasarkan komentar yang ada. Penelitian ini bertujuan untuk menentukan sentimen publik terhadap AMIK BSI Tegal yang disampaikan di *Instagram*, sehingga membantu usaha untuk melakukan riset *marketing* atas opini publik. Klasifikasi algoritma seperti *NAIVE BAYES* (NB) yang diusulkan oleh banyak peneliti untuk digunakan dalam analisis sentimen teks. Algoritma *NAIVE BAYES* dan metodenya, akan diuji dengan dua masukan dengan menggunakan komentar *Tokenize and Transform Case* yang positif (100 teks komentar) dan negatif (100 komentar teks), akurasi yang didapat algoritma *NAIVE BAYES* 76.50% +/-7.76%(mikro:76.50). Hasil penelitian menunjukkan bahwa *NAIVE BAYES* (NB) mendapatkan hasil terbaik dan akurat.

Kata kunci: Sentiment analysis, NAIVE BAYES, Sosial Media

PENDAHULUAN

latar Belakang

Jumlah pengguna media sosial kini berkembang pesat seiring kemajuan teknologi informasi yang bergerak cepat. Pengguna internet sekarang bebas berbagi informasi hingga membentuk sebuah opini publik yang memiliki berbagai makna. Dari opini tersebut diyakini sebagai data yang akurat sebab opini publik terbentuk dari suatu luapan emosi, pendapat yang muncul, atau cerita yang sedang viral di media sosial. Seperti dalam *Twitter*, *Instagram*, *Facebook* yang kerap menjadi kanal bagi warganet untuk 'curhat'. Selain para *netizen*, partai politik maupun perusahaan dapat menggunakan data tersebut sebagai sumber informasi penting mengenai tanggapan dari masyarakat, mengukur tingkat popularitas dan layanan. Namun mengukur suatu sentimen komentar di suatu media sosial bukan hal yang mudah. Kendala yang dihadapi dalam media sosial yakni menggunakan bahasa formal sehingga dibutuhkan analisis agar opini yang berada di media sosial tersebut dapat diartikan dan digunakan dalam pengambilan keputusan.

Media Sosial yang saat ini ada terdiri dari berbagai media *online* seperti situs media jejaring sosial, aplikasi media jejaring sosial, *game*, dan media *online* lainnya. Macam-macam media sosial yang disediakan: *Path*, *Line*, *Instagram*, *Whatsapp*, *Blackberry Messenger*, *Twitter*, *LinkedIn*, *Facebook*, *Wikipedia*, *Google Plus*, *youtube* dan masih banyak lagi lainnya. Pada umumnya berita yang disampaikan dalam portal tersebut terdiri dari beberapa kategori seperti berita politik, layanan sosial, ekonomi, olahraga, kesehatan, dan lain-



lain. Tentu saja informasi yang akan didapat bukan saja pemberitaan yang bersifat positif saja tetapi ada juga yang bersifat negatif.

Berdasarkan latar belakang diatas, didapatkan rumusan masalah yang dibahas dalam penelitian ini, sebagai berikut:

- a. Bagaimana cara menentukan sentimen publik terhadap AMIK BSI Tegal yang disampaikan di media sosial *instagram*?
- b. Apakah Algoritma *Naive Bayes* dapat membantu menentukan sentimen negatif dan positif publik terhadap AMIK BSI Tegal yang disampaikan di media sosial *instagram*?

Penelitian ini bertujuan untuk menentukan sentimen publik terhadap AMIK BSI Tegal yang disampaikan di *Instagram*, sehingga membantu usaha untuk melakukan riset *marketing* atas opini publik. Klasifikasi algoritma seperti *NAIVE BAYES* (NB) yang diusulkan oleh banyak peneliti untuk digunakan dalam analisis sentimen.

Dalam Surat Edaran (SE) Kapolri soal penanganan ujaran kebencian Nomor SE/06/X/2015 (“SURAT EDARAN KAPOLRI MENGENAI PENANGANAN UJARAN KEBENCIAN.pdf,” 2015). Dalam surat edaran tersebut, disebutkan bahwa persoalan ujaran kebencian semakin mendapatkan perhatian masyarakat baik nasional atau internasional seiring meningkatnya kepedulian terhadap perlindungan hak asasi manusia (HAM). Berikut poin-poin krusial dalam SE tersebut:

Bentuk Ujaran Kebencian disebutkan bahwa "ujaran kebencian dapat berupa tindak pidana yang diatur dalam Kitab Undang-Undang Hukum Pidana (KUHP) dan ketentuan pidana lainnya di luar KUHP, yang berbentuk antara lain: (1) Penghinaan, (2) Pencemaran nama baik, (3) Penistaan, (4) Perbuatan tidak menyenangkan, (5) Memprovokasi, (6) Menghasut, (7) Menyebarkan berita bohong dan semua tindakan di atas memiliki tujuan atau bisa berdampak pada tindak diskriminasi, kekerasan, penghilangan nyawa, dan atau konflik sosial" (“SURAT EDARAN KAPOLRI MENGENAI PENANGANAN UJARAN KEBENCIAN.pdf,” 2015).

Aspek Ujaran Kebencian Selanjutnya, bahwa ujaran kebencian sebagaimana dimaksud di atas bertujuan untuk menghasut dan menyulut kebencian terhadap individu dan atau kelompok masyarakat dalam berbagai komunitas yang dibedakan dari aspek: (1) Suku, (2) Agama, (3) Aliran keagamaan, (4) Keyakinan atau kepercayaan, (5) Ras, (6) Antargolongan, (7) Warna kulit, (8) Etnis, (9) Gender, (10) Kaum difabel, (11) Orientasi seksual (“SURAT EDARAN KAPOLRI MENGENAI PENANGANAN UJARAN KEBENCIAN.pdf,” 2015).

Media Ujaran Kebencian disebutkan bahwa "ujaran kebencian sebagaimana dimaksud di atas dapat dilakukan melalui berbagai media, antara lain: (1) Dalam orasi kegiatan kampanye, (2) Spanduk atau banner, (3) Jejaring media sosial, (4) Penyampaian pendapat di muka umum (demonstrasi), (5) Ceramah keagamaan, (6) Media massa cetak atau elektronik, (7) Pamflet. dengan memperhatikan pengertian ujaran kebencian di atas, perbuatan ujaran kebencian apabila tidak ditangani dengan secara efektif, efisien, dan sesuai dengan ketentuan peraturan perundang-undangan, akan berpotensi menimbulkan tindak kriminal, kekerasan, dan memunculkan konflik sosial yang meluas (“SURAT EDARAN KAPOLRI MENGENAI PENANGANAN UJARAN KEBENCIAN.pdf,” 2015).

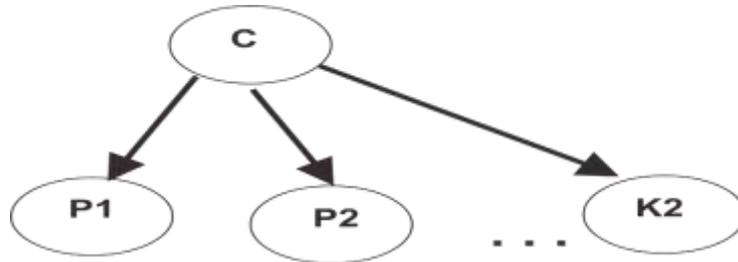
Banyak ide-ide telah muncul selama bertahun-tahun tentang bagaimana mencapai hasil yang berkualitas dari sistem klasifikasi *Web*, sehingga ada pendekatan yang berbeda yang dapat digunakan untuk permasalahan analisis sentimen, seperti *Naive Bayes* (NB) dan *Bayesian Networks*, NNS, DTs, *Support Vector Machines* (SVM) dll. *Mode 1* (*Naive Bayes*) NB yang populer dalam aplikasi *machine learning*, karena kesederhanaan mereka dalam membiarkan setiap atribut berkontribusi terhadap keputusan akhir yang sama dan independen dari atribut lainnya (Xhemali, Hinde and Stone, 2009). Berbagai metode *machine learning* telah digunakan untuk penentuan atribut dengan menggunakan *naive Bayes* (Saleh, El, & Menai, 2014).

Proses klasifikasi teks dapat dibagi menjadi dua *fase klasifikasi* (Darujati et al., 2012) yaitu:

1. *Fase information retrieval* (IR) untuk mendapatkan data numerik dari sebuah dokumen teks. Langkah pertama yang dilakukan pada fase ini adalah *feature extraction*. Merupakan pendekatan yang umum digunakan adalah distribusi frekuensi kata. Nilai angka yang diperoleh dapat berupa berapa kali suatu kata muncul di dalam sebuah dokumen, 1 (satu) jika kata ada didalam dokumen atau 0 (nol) jika tidak ada (biner), atau jumlah kemunculan kata pada awal dokumen. *Feature* yang diperoleh dapat direduksi agar dimensi vektor menjadi lebih kecil. Beberapa pendekatan *feature reduction* dapat diterapkan seperti menghapus *stop-words*, *stemming*, *statistical filtering*. (Darujati et al., 2012).
2. *Fase klasifikasi utama* yaitu ketika suatu algoritma yang memproses sebuah data untuk memutuskan kategori dimana teks itu ditempatkan. Algoritma klasifikasi yang merupakan salah satu kajian di bidang statistika dan *machine learning* yang dapat diterapkan pada fase ini, di antaranya *NAIVE BAYESian*, *Decision Tree*, *Neural Network*, *Nearest Neighbor* dan *Support Vector Machines*. (Darujati et al., 2012).



Klasifikasi–klasifikasi *Bayes* merupakan klasifikasi statistik yang dapat memprediksi atribut kelas suatu anggota probabilitas. Untuk sebuah klasifikasi *Bayes* sederhana yang lebih dikenal sebagai *Naïve Bayesian Classifier* dapat diasumsikan bahwa efek dari suatu nilai atribut sebuah kelas yang diberikan adalah bebas dari atribut-atribut lain. Dalam *Naïve Bayes* diasumsikan prediksi atribut adalah tidak tergantung pada kelas atau tidak dipengaruhi atribut lain (Darujati and Gumelar, 2012), dengan gambaran sebagai berikut:



Gambar 1.Klasifikasi Naïve Bayes sebagai jaringan bayes dengan atribut prediksi (P1, P2,.....Pk) dan kelas (C). (Darujati and Gumelar, 2012)

C adalah anggota kelas dan X adalah variabel acak sebuah vektor sebagai atribut nilai yang diamati. c mewakili nilai label kelas dan x mewakili nilai atribut vektor yang diamati. Jika diberikan sejumlah x tes untuk klasifikasi maka probabilitas tiap kelas untuk atribut prediksi vektor yang diamati adalah:

$$p(C = c|X = x) = \frac{p(C = c)p(X = x|C = c)}{p(C = c)}$$

X = x adalah mewakili kejadian dari Jumlah dari untuk semua kelas adalah 1

$$X_1 = x_1 \wedge X_2 = x_2 \wedge \dots \wedge X_k = x_k$$

$$p(C = c|X = x)$$

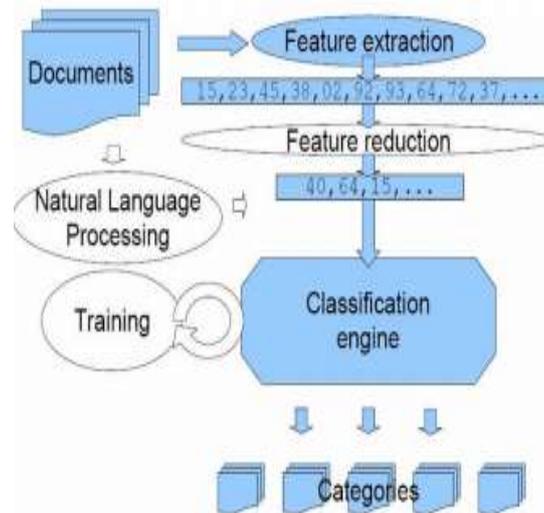
Pada tahap evaluasi sistem, perhitungan menggunakan tabel kontingensi yang diberikan (Ling, Kencana and Oka, 2014):

#	Relevant	Not Relevant
Retrieved	True Positive (TP)	False Positive (FP)
Not Retrieved	False Negatif (FN)	True Negatif (TN)

Tabel 1. Tabel Kontingensi Evaluasi Kinerja Sistem (Ling, Kencana and Oka, 2014)

Proses klasifikasi teks melibatkan banyak teknik IR mulai dari *preprocessing*, pengukuran kemiripan selama operasional klasifikasi sampai dengan evaluasi kinerja algoritma.





Gambar 2. Diagram proses klasifikasi teks (Darujati and Gumelar, 2012)

Pre-Processing pada pengujiannya Analisis Sentimen Menggunakan Metode *Naïve Bayes Classifier* Dengan Seleksi Fitur Chi Square (Ling, Kencana and Oka, 2014) sebagai berikut:

Tokenization adalah tugas pemotongan urutan karakter dan sebuah *set* dokumen yang diberikan menjadi potongan-potongan kata atau karakter yang sesuai dengan kebutuhan sistem. Potongan-potongan tersebut dikenal dengan istilah *token*.

Stemming merupakan salah satu proses dari mengubah *token* yang berimbuhan menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada *token* tersebut. Pentingnya *stemming* dalam proses pembuatan sistem adalah untuk menghilangkan imbuhan pada awalan dan akhiran. Berdasarkan hasil proses tersebut, akan didapatkan sebuah informasi mengenai banyaknya fitur yang muncul dalam sebuah dokumen. *Stopwords* dapat diartikan sebagai menghilangkan karakter, tanda baca, serta kata-kata umum yang tidak memiliki makna atau informasi yang dibutuhkan. *Stopwords* umumnya digunakan dalam Penelitian tentang klasifikasi sentimen terhadap *review* film telah dilakukan oleh Dhande dan Patnaik (2014) dengan menggunakan algoritma *NAIVE BAYES*, *Neural Network*, dan *NAIVE BAYES Neural Classifier*. Dari hasil penelitian akhir yang diuji menggunakan ketiga algoritma tersebut menyebutkan bahwa *NAIVE BAYES* menghasilkan akurasi yang lebih tinggi dibandingkan *Neural Network*. Dan algoritma *NAIVE BAYES Neural Classifier* yang merupakan penggabungan antara metode *NAIVE BAYES* dan *Neural Network* menghasilkan akurasi yang paling tinggi diantara kedua algoritma tersebut. *NAIVE BAYES* memiliki beberapa keunggulan seperti sederhana, cepat dan akurasi yang tinggi.

Banyak peneliti telah melakukan klasifikasi sentimen dengan menggunakan *NAIVE BAYES*. Namun klasifikasi ini memiliki keterbatasan utama yang tidak mungkin selalu memenuhi asumsi independensi antara atribut. Dan ini mempengaruhi tingkat akurasi klasifikasi.

K-fold Cross-validation merupakan teknik validasi dengan membagi data awal secara acak kedalam *k* bagian yang saling terpisah atau "*fold*" [6]. Grafik *Receiver Operating Characteristics (ROC)* adalah teknik untuk memvisualisasikan dan memilih pengklasifikasi berdasarkan kinerja mereka [2]. Kurva ROC digunakan untuk mengukur nilai *Area Under Curve (AUC)*. Kurva ROC memiliki properti yang menarik: mereka tidak sensitif terhadap perubahan distribusi kelas. Jika proporsi positif terhadap kasus negatif berubah dalam satu set tes, kurva ROC tidak akan berubah. Untuk melihat mengapa demikian, dapat dilihat pada *confusion matrix* [13](Indrayuni et al., 2015)

Penelitian lain yang pernah dilakukan Kang, Yoo, dan Han (2012) adalah analisa sentimen pada *review* restoran menggunakan algoritma *NAIVE BAYES* dengan fitur unigrams dan bigrams untuk meningkatkan akurasi *NAIVE BAYES*. Pada penelitian ini dengan menerapkan metode senti leksikon yaitu fitur unigrams dan bigrams, menunjukkan bahwa selisih akurasi antara *class* positif dan negatif sekitar 3,6% dibandingkan dengan penggunaan *NAIVE BAYES* saja.

Pada penelitian ini menggunakan algoritma *NAIVE BAYES* untuk mendapatkan nilai akurasi tertinggi dengan model tersebut:

1. Banyaknya pemberitaan dimedia sosial yang bersifat negatif sehingga perlunya pembatasan atau penyaringan pada pemberitaan
2. Bagaimana tingkat akurasi analisa



3. sentimen yang dihasilkan algoritma *Naïve Bayes*

Dan adapun manfaat dari penelitian ini adalah:

1. Manfaat dari penelitian ini adalah membantu para pembaca atau penerima informasi yang lebih sehat dan bersifat positif
2. Manfaat teoritis penelitian ini adalah memberikan bukti secara empiris untuk teori yang berkaitan sehingga dapat dijadikan sumbangan pemikiran untuk pengembangan teori berikutnya.

Text mining

Text mining adalah proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar sumber data tak terstruktur. Penambangan teks memiliki tujuan dan menggunakan proses yang sama dengan penambangan data, namun memiliki masukan yang berbeda. Masukan untuk penambangan teks adalah data yang tidak (atau kurang) terstruktur, seperti dokumen Word, PDF, kutipan teks, dll., sedangkan masukan untuk

penambangan data adalah data yang terstruktur (Feldman, Goldenberg, & Netzer, 2010).

Text mining adalah untuk mengetahui cakupan atau topik dari permasalahan dalam teks (Maning, et al. [6]). *Text mining* penting dalam analisis sentimen sebagai pengidentifikasi emosional suatu pernyataan, sehingga banyak studi tentang analisis sentimen dilakukan (Zhang, et al. [12]). (Ling & Oka, 2014)

Naïve Bayes

NAIVE BAYES adalah model sederhana untuk klasifikasi. Model ini bekerja dengan baik untuk klasifikasi teks. Model ini merupakan bentuk sederhana dari *Bayesian Network*, dimana semua atribut independen diberi nilai kelas variabel. *Naïve Bayes* memiliki beberapa keunggulan seperti sederhana, cepat dan akurasi yang tinggi. (Indrayuni et al., 2015).

Rapid Miner

Rapidminer (RM) adalah lingkungan untuk pembelajaran mesin dan *data mining* proses. Ini adalah *open source*, proyek gratis diimplementasikan di Jawa. Ini merupakan pendekatan baru untuk merancang masalah yang sangat rumit - konsep operator modular yang memungkinkan disain rantai operator bersarang kompleks untuk sejumlah besar masalah belajar (Keel dkk 2009).

Rapidminer adalah salah satu yang paling populer dan memiliki antarmuka pengguna yang bersih sehingga memudahkan untuk belajar. Selanjutnya *Rapidminer* dapat diperpanjang, sehingga pengembang dapat menyediakan fungsionalitas tambahan pada perangkat lunak dasar (Ak dkk 2011).

Rapidminer memiliki banyak kelebihan. Beberapa di antaranya adalah: disediakan secara gratis sebagai perangkat lunak *open source*, namun disediakan juga berdasarkan lisensi komersial yang sesuai untuk pengembangan aplikasi komersial *closed-source*, komunitas bebas yang berpengalaman juga mendapat dukungan berbayar yang sangat profesional, pengguna dewasa dan yang berkembang. Dan pengembang komunitas, biaya pengembangan dengan *Rapidminer* relatif rendah bila dibandingkan dengan solusi *data mining* lainnya, karena ini adalah perangkat lunak *open source*, sangat sesuai untuk tujuan penelitian, antarmuka pengguna grafis yang intuitif, terorganisir, dan *user friendly*, akses ke banyak *database* dan kemungkinan untuk membaca banyak format *file*, kode modular dan tersusun dengan baik, sejumlah besar operator *built-in* untuk *data mining* yang berisi lebih dari 250 algoritma pembelajaran dan banyak operator pemrosesan dan *preprocessing*, *ostprocessing* dan *meta-operator*, tampilan data berlapis-lapis konsep (berbeda pandangan pada data yang sama), *Rapidminer* data core mirip dengan DBMS standar (*Data Base Management System*), *powerful* bahasa *cripting* (proses DM dirancang sebagai pohon operator dengan struktur yang didefinisikan dalam *file XML* - alur data biasanya mengikuti Penelusuran Kedalaman Pertama) (Burget, Karasek, Smekal, Uher, & Dostal, 2010).

Sentiment Analysis

Analisis sentimen, adalah proses bertujuan untuk menentukan apakah polaritas korpus tekstual (dokumen, kalimat, paragraf dll) cenderung ke arah positif, negatif atau netral (Kontopoulos, Berberidis, Dergiades, & Bassiliades, 2013). Sentimen terutama mengacu pada perasaan, emosi, pendapat atau sikap. Dengan meningkatnya *World Wide Web*, orang sering mengungkapkan perasaan mereka di melalui internet melalui media sosial, *blog*, *rating* dan *reviews*. Karena ini meningkatkan data tekstual, ada kebutuhan untuk menganalisa konsep mengekspresikan sentimen dan menghitung wawasan untuk menjelajahi bisnis (Tripathy, Agrawal, & Rath, 2015).

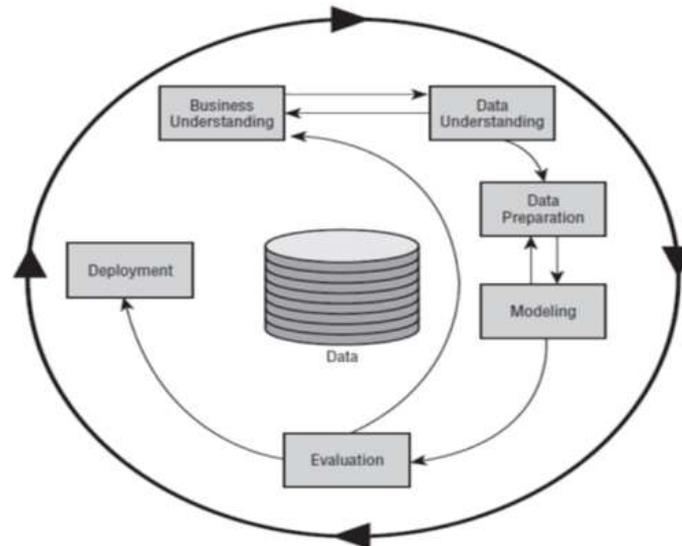
Analisis sentimen merupakan studi komputasi dari opini-opini, sentimen, serta emosi yang diekspresikan di dalam teks. Tugas dasar dalam suatu analisis sentimen yaitu mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau pendapat. Polaritas mempunyai arti apakah teks yang ada dalam dokumen,



kalimat, atau pendapat memiliki aspek positif atau negatif. Salah satu teknik pembelajaran mesin untuk analisis sentimen adalah *Naïve Bayes classifier* (NBC). *Naïve Bayes classifier* merupakan teknik pembelajaran mesin yang berbasis probabilistik. *Naïve Bayes classifier* adalah metode sederhana tetapi memiliki akurasi serta performansi yang tinggi dalam pengklasifikasian pengklasifikasian teks. (Ling & Oka, 2014)

Metode CRISP-DM

Berikut ini adalah penjelasan mengenai enam tahap siklus hidup pengembangan *data mining* *Cross-Industry Standard Process for Data mining* (CRISP-DM) yang telah dimodifikasi dan digunakan dalam penelitian ini sebagaimana pada gambar 1.



Sumber: jurnal Anisa Paramitha Fadillah

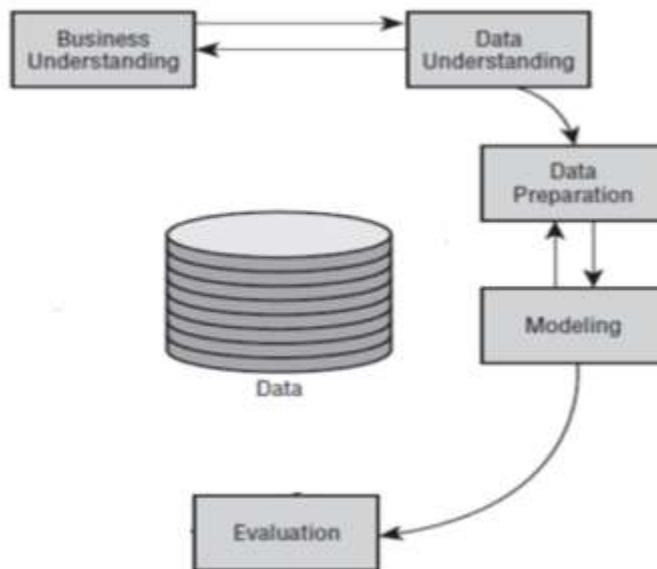
Gambar 3. Model CRISP-DM

Langkah-langkah yang terdapat pada model CRISP-DM untuk diterapkan pada sebuah penelitian adalah sebagai berikut:

- a. *Business Understanding*
Business Understanding atau pemahaman domain (penelitian). Pada fase ini dibutuhkan pemahaman tentang substansi dari kegiatan *data mining* yang akan dilakukan, kebutuhan dari perspektif bisnis. Kegiatan yang dilakukan antara lain: menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan *data mining* dan membuat perencanaan strategi serta jadwal penelitian.
- b. *Data Understanding*
Data Understanding atau pemahaman data adalah fase mengumpulkan data awal, mempelajari data untuk bisa mengenal data yang akan dipakai. Fase ini mencoba mengidentifikasi masalah yang berkaitan dengan kualitas data, mendeteksi subset yang menarik dari data untuk membuat hipotesa awal.
- c. *Data Preparation*
Data preparation atau persiapan data. Fase ini sering disebut sebagai fase yang padat karya. Kegiatan yang dilakukan yaitu memilih sebuah *table* dan *field* yang akan ditransformasikan ke dalam sebuah *database* baru untuk bahan *data mining* (set data mentah).
- d. *Modeling*
Modeling adalah fase menentukan teknik *data mining* yang digunakan, menentukan *tools data mining*, teknik *data mining*, algoritma *data mining*, menentukan parameter dengan nilai yang optimal.
- e. *Evaluation*
Evaluation merupakan suatu fase interpretasi terhadap hasil *data mining* yang ditunjukkan dalam proses pemodelan pada fase yang sebelumnya. Evaluasi dilakukan secara mendalam dengan tujuan untuk menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase yang pertama..
- f. *Deployment*
Deployment atau penyebaran yaitu suatu fase penyusunan laporan atau presentasi dari pengetahuan yang didapat dari evaluasi yang dilakukan pada proses *data mining*. (Proporsi & Tridharma, 2012)



Metode Penelitian



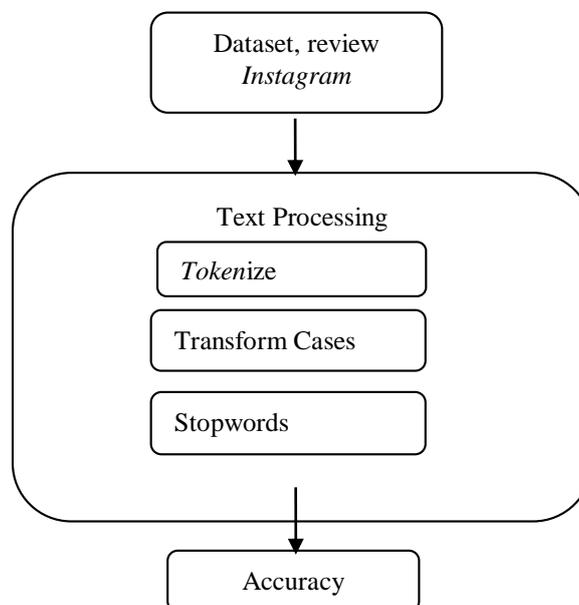
Gambar 4. Model modifikasi CRISP-DM pada penelitian

Metode yang dilakukan pada penelitian ini hanya menggunakan tahap *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation* dan *Deployment*

HASIL DAN PEMBAHASAN

Penelitian yang dilakukan pada analisa sentimen dengan menggunakan Algoritma *NAIVE BAYES*(NB) pada pemberitaan “Surat Edaran (SE) Kapolri soal penanganan ujaran kebencian atau hate speech Nomor SE/06/X/2015” dengan mengukur akurasi data eksperimen yang dilakukan dalam penelitian dengan menggunakan aplikasi *Rapidminer Studio 6.5.002*.

Pada gambar 2 menggambarkan kerangka pemikiran yang diusulkan adalah sebagai berikut :



Gambar 3. Kerangka Pemikiran

Untuk melakukan eksperimen penelitian ini menggunakan data komentar review <https://www.instagram.com/amikbsitegal/>. Maka didapat 200 data review yang terdiri dari 100 review positif dan 100 review negatif. Data review positif disatukan dalam folder dengan nama Positif. Sedangkan data review negatif disatukan penyimpanannya dalam folder dengan nama Negatif. Tiap dokumen berekstensi.txt yang dapat dibuka menggunakan aplikasi *Notepad*.

Kemudian setelah mendapatkan data positif dan negatif dalam bentuk.txt dengan jumlah data yang sama, yaitu 100 review data positif maka yang bersifat negatif harus didapatkan 100 review data negatif, data akan diproses menggunakan Aplikasi *Rapid Miner*. Pada aplikasi *Rapid Miner* untuk *text processing* yang peneliti gunakan *Tokenize, Transform Cases, Stopwords(Dictionar)*. Pengujian yang dilakukan dengan memilih *feature selection NAIVE BAYES(NB)*.

Penelitian yang dilakukan menggunakan Aplikasi *Rapidminer Studio 6.5.002* dengan menggunakan spesifikasi komputer CPU *Intel Core i3 1.90GHz*, RAM 4GB, dan *system operasi Microsoft Windows 7 Professional 64-bit*. Hasil pengujian review Pemberitaan “*Awas, Penebar Kebencian di Media Sosial Bisa Diancam Pidana*” dengan menggunakan teks bahasa Indonesia.

Hasil dari penelitian menggunakan *Text Processing* pada table 1 yang dilalui terdiri dari *Tokenize, Transform Cases, Stopwords(Dictionar)*. Hasil pengolahan sebagai berikut :

Tabel 2. Text Processing menggunakan Tokenize, Transform Cases, Stopwords(Dictionar)

<i>Text Review</i>	<i>Tokenize</i>	<i>Transform Cases</i>	<i>Stopwords(Dictionar)</i>
YG BEGINI DAPET SOSIAL MEDIA AWARD? AMPUN MALU MALUIN	YG BEGINI DAPET SOSIAL MEDIA AWARD AMPUN MALU MALUIN	begini dapet Sosial media award ampun malu maluin	begini dapet sosial media award
wanjir ada aku wkwk, MIN katanya pihak penyelenggara wajib pelajar semua min? .	wanjir ada aku wkwk MIN katanya pihak penyelenggara wajib pelajar semua min	wanjir ada aku wkwk, katanya pihak penyelenggara wajib pelajar semua .	ada aku wkwk, katanya pihak penyelenggara wajib pelajar semua
BSI keren coy!! Nambah banyak teman baru, salam dari tim kuning	BSI keren coy Nambah banyak teman baru salam dari tim kuning	bsi keren nambah banyak teman baru salam dari tim kuning	bsi keren nambah teman baru salam tim kuning

Pada tabel 2 terlihat *Text Processing* dari *text review* yang bersifat negatif akan berubah menjadi *text review* yang bersifat positif. Pada *text review* melalui *Tokenizes* eluruh tanda baca akan dihilangkan, kemudian *Transform Cases* seluruh *text review* yang menggunakan huruf besar akan berubah menjadi huruf kecil dan yang terakhir *Stopwords(Dictionar)* kosa kata atau kata yang bersifat negatif akan dihilangkan.



Pengujian pertama penelitian dilakukan menggunakan algoritma *NAIVE BAYES* untuk mendapatkan *accuracy* dapat kita lihat pada tabel 3 dengan nilai *accuracy* 76.50%, sebagai berikut:

	true Negatif	true Positif	Class precision
pred. Negatif	75	22	77.32%
pred. Positif	25	78	75.73%
Class recall	75.00%	78.00%	

Tabel 3. Accuracy Algoritma NAIVE BAYES

KESIMPULAN

Pengujian model dengan menggunakan *NAIVE BAYES*(NB) dengan menggunakan data *review* analisis sentimen Etika berinternet dimedia social dari 200 data *review* dalam teks bahasa indonesia yang terdiri dari 100 *review* positif dan 100 *review* negatif. Model yang dihasilkan mendapatkan hasil *accuracy* pengujian pertama dilakukan dengan menggunakan *NAIVE BAYES*(NB) yaitu mendapatkan nilai *accuracy* 76.50%+/- 7.76(mikro: 76.50%) maka dapat disimpulkan pengujian analisa sentimen dengan menggunakan *NAIVE BAYES*(NB) sudah memenuhi syarat batas minimum akurasi sebuah penelitian. sehingga kita dapat memecahkan masalah analisa sentimen dengan memilih pemberitaan atau informasi dimedia sosial khususnya lebih akurat.

DAFTAR PUSTAKA

- [1] "SURAT EDARAN KAPOLRI MENGENAI PENANGANAN UJARAN KEBENCIAN.pdf." pp. 1–3, 2015.
- [2] A. Saleh, M. El, and B. Menai, "Naï ve Bayes classifiers for authorship attribution of Arabic texts," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 26, no. 4, pp. 473–484, 2014.
- [3] C. Darujati *et al.*, "PEMANFAATAN TEKNIK SUPERVISED UNTUK," vol. 16, no. 1, pp. 1–8, 2012.
- [4] E. Indrayuni, M. Wahyudi, S. Informasi, J. Selatan, I. Komputer, and J. Selatan, "PENERAPAN CHARACTER N-GRAM UNTUK SENTIMENT ANALYSIS REVIEW HOTEL MENGGUNAKAN ALGORITMA NAIVE BAYES," pp. 88–93, 2015.
- [5] R. Feldman, J. Goldenberg, and O. Netzer, "Mine your own business: Market structure surveillance through *text mining*," *Mark. Sci. Institute, Spec. Rep.*, vol. 10, no. 3, pp. 10–202, 2010.
- [6] J. Ling and T. B. Oka, "ANALISIS SENTIMEN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DENGAN SELEKSI FITUR CHI SQUARE," vol. 3, no. 3, pp. 92–99, 2014.
- [7] U. Keel, M. Graczyk, T. Lasota, and B. Trawiński, "Comparative Analysis of Premises Valuation Models," *N.T. Nguyen, R. Kowalczyk, S.-M. Chen ICCCI 2009, LNAI 5796, pp. 800–812, 2009. © Springer-Verlag Berlin Heidelb. 2009, pp. 800–812, 2009.*
- [8] Z. P. Ak, G. Makrai, T. Henk, and C. Gar-Papanek, "Radoop : Analyzing Big Data with *Rapidminer* and Hadoop," *Proc. 2nd Rapidminer Community Meet. Conf.*, pp. 1–12, 2011.
- [9] R. Burget, J. Karasek, Z. Smekal, V. Uher, and O. Dostal, "*Rapidminer* Image Processing Extension:A Platform for Collaborative Research," *Int. Conf. Telecommun. SIGNAL Process.*, no. November 2015, pp. 1–5, 2010.
- [10] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, "Expert Systems with Applications Ontology-based sentiment analysis of twitter posts," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4065–4074, 2013.
- [11] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques," *Procedia - Procedia Comput. Sci.*, vol. 57, pp. 821–829, 2015.
- [12] P. Proporsi and P. Tridharma, "28. 29. 1," pp. 1–56, 2012.

