

## OPINION MINING PADA KOMENTAR TWITTER E-KTP MENGUNAKAN NAIVE BAYES CLASSIFIER

Mihuandayani<sup>1)</sup>, Eko Feriyanto<sup>2)</sup>, Syarham<sup>3)</sup>, Kusri<sup>4)</sup>

<sup>1), 2), 3), 4)</sup> Magister Teknik Informatika, Universitas AMIKOM Yogyakarta

Jl. Ring Road Utara, Condong Catur, Depok, Sleman, Yogyakarta 55281

Email : mihuandayani20@gmail.com<sup>1)</sup>, feriyanto.eko@gmail.com<sup>2)</sup>, syarham.85@gmail.com<sup>3)</sup>, kusri@amikom.ac.id<sup>4)</sup>

### Abstrak

Twitter menjadi salah satu media sosial yang digunakan untuk mengutarakan opini tentang berbagai isu atau topik yang sedang tren melalui kolom tweet. Isu terkait dengan Electronic Kartu Tanda Penduduk (E-KTP) sempat menjadi topik yang banyak dikomentari oleh pengguna akun twitter. Data yang sangat banyak berupa komentar twitter tentang e-ktip dapat dimanfaatkan dengan melakukan analisis sentimen (Opinion Mining). Opinion mining yang dilakukan bertujuan untuk mengetahui persentase opini positif, negatif dan netral. Hasil opini tersebut juga dapat digunakan sebagai salah satu pertimbangan dalam mendukung suatu keputusan.

Adapun metode yang digunakan untuk melakukan opinion mining yaitu Naive Bayes Classifier. Penelitian ini menjelaskan tahapan yang dilakukan dalam opinion mining yaitu text processing, feature extraction dan klasifikasi menggunakan Naive Bayes. Hasil klasifikasi yang diperoleh melalui data training, dilakukan pengujian dengan data testing.

Pengujian dilakukan untuk mengukur kinerja Naive Bayes Classifier dalam mengklasifikasi data dalam tiga target fitur. Hasil yang diperoleh yaitu tingkat akurasi sebesar 89,67 %.

**Kata kunci:** Opinion Mining, Twitter, E-KTP, Naive Bayes Classifier

### 1. Pendahuluan

Perkembangan internet menjadikan pengguna media sosial terus meningkat. Media sosial seperti Facebook, Twitter, Instagram, Path dan sebagainya bukan lagi menjadi hobi bagi penggunanya, namun juga terkadang menjadi pekerjaan yang mendatangkan penghasilan. Twitter adalah salah satu media sosial yang digunakan untuk melakukan komentar melalui kolom tweet yang dibatasi dengan jumlah karakter tertentu. Menurut berita Country Head Twitter Indonesia, sebagaimana dirilis oleh CNN Indonesia [1], sampai pertengahan tahun 2015 jumlah pengguna Twitter Indonesia telah mencapai 50 juta pengguna. Berdasarkan data yang diperoleh, terdapat sekitar 4,1 juta tweet per hari yang berasal dari Indonesia, sedangkan 77% pengguna Twitter di Indonesia aktif setiap harinya. Dengan data tersebut, maka tidak heran jika kemudian Twitter telah menjadi

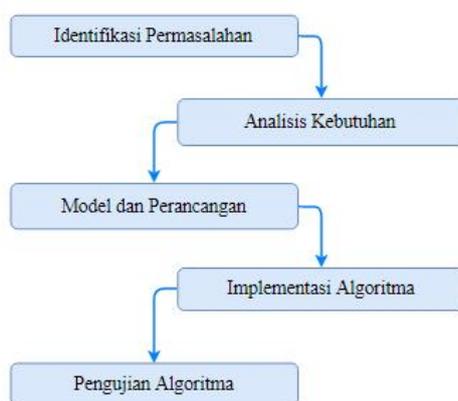
sumber data bagi banyak pihak, baik untuk tujuan yang bersifat ilmiah, komersil, sosial dan lain-lain. Isu-isu yang diangkat dalam berbagai penelitian yang bersumber dari Twitter biasanya beragam, misalnya isu terkait dengan politik, ekonomi, sosial, budaya, korupsi, olahraga, analisis pasar untuk produk tertentu dan lain sebagainya. Salah satu topik terkini yang menjadi tren perbincangan pengguna Twitter adalah mengenai kasus Electronic Kartu Tanda Penduduk (E-KTP) atau Kartu Tanda Penduduk Elektronik (KTP-el). Skandal kasus E-KTP semakin ramai diperbincangkan di Twitter selama beberapa bulan terakhir. Twitter memungkinkan pengguna untuk mengakses data dengan tema tertentu. Dengan demikian maka tweet pengguna Twitter yang bertema kasus E-KTP memungkinkan untuk diakses. Data terkait dengan kasus korupsi E-KTP ini lebih lanjut dapat dianalisa untuk mengetahui sentimen publik melalui text mining.

E-KTP merupakan salah satu produk dari Kementerian Dalam Negeri yang digunakan sebagai tanda pengenal yang berlaku secara nasional [2]. E-KTP mulai hadir untuk mengurangi berbagai kecurangan duplikasi KTP yang terjadi di Indonesia melalui satu database terpadu. Kartu elektronik tersebut memuat dokumen kependudukan dengan sistem keamanan dari sisi administrasi maupun teknologi informasi yang berdasarkan database kependudukan nasional. Undang – Undang Republik Indonesia Nomor 23 Tahun 2006 tentang Administrasi Kependudukan, dijelaskan bahwa: "Penduduk hanya diperbolehkan memiliki 1 (satu) KTP yang tercantum Nomor Induk Kependudukan (NIK). NIK merupakan identitas tunggal setiap penduduk dan berlaku seumur hidup". Saat ini nomor NIK pada KTP dapat dijadikan sebagai dasar untuk menerbitkan paspor, Surat Izin Mengemudi (SIM), Nomor Pokok Wajib Pajak (NPWP), asuransi, sertifikat dan berbagai penerbitan dokumen lainnya. E-KTP dinilai sangat penting dalam mewujudkan terciptanya keakuratan data penduduk untuk mendukung program pembangunan dan mewujudkan sistem administrasi yang sistematis dalam pelayanan publik di Indonesia. E-KTP menggunakan sidik jari yang dikenali melalui chip dan terpasang dalam kartu sebagai biometrik untuk autentikasi kartu identitas dan verifikasi sistem. Data yang terdapat dalam kartu tersebut dienkripsi dengan algoritma kriptografi dan terdiri dari sembilan layer sehingga lebih aman.

Klasifikasi merupakan bagian dari *Machine Learning* yang dikategorikan dalam *Supervised Learning*. Hal ini terlihat pada adanya fitur input dan target fitur yang dapat didefinisikan. Dalam klasifikasi dapat dilakukan dengan berbagai metode yaitu *Support Vector Machine*, *Naive Bayes*, *K-Nearest Neighbor* dan lain-lain. Algoritma *Naive Bayes Classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat [3]. Selain itu, penelitian lain yang membahas tentang analisis sentimen [4][5][6] menunjukkan bahwa algoritma *Naive Bayes* memiliki tingkat akurasi yang lebih baik.

Oleh karena itu, penelitian ini bertujuan untuk membangun sebuah sistem yang dapat mengklasifikasi berbagai *tweet* pada topik E-KTP menggunakan algoritma *Naive Bayes Classifier*. Proses klasifikasi ini menghasilkan kelas sentimen positif, negatif dan netral. Hasil dari penelitian ini dapat digunakan oleh berbagai pihak yang ingin mengetahui sentimen masyarakat terkait dengan topik E-KTP. Misalnya Partai Politik, Organisasi Non Pemerintah yang bergerak di bidang pengawasan korupsi seperti *Indonesian Corruption Watch* dan lain-lain. Dengan hasil penelitian ini, para pihak tersebut tidak hanya dapat mengetahui sentimen masyarakat pengguna Twitter, namun juga dapat mengetahui topik-topik yang dibicarakan pada setiap kelas sentimen yang dihasilkan.

Dalam penelitian ini, untuk melakukan *opinion mining* terdiri dari beberapa tahapan yang diawali dengan melakukan analisis permasalahan, analisis kebutuhan sistem, memodelkan dan membuat rancangan sistem, implementasi algoritma dan pengujian algoritma untuk mengukur tingkat akurasi. Tahapan metode tersebut ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

### Tinjauan Pustaka

Terkait dengan analisis sentimen terhadap topik tertentu, terdapat beberapa penelitian sebelumnya yang pernah dilakukan. Penelitian tentang klasifikasi teks menggunakan *Naive Bayes* dan KNN pada data set

mahasiswa dengan 7 atribut. Tujuan penelitian tersebut yaitu untuk mengukur kinerja metode *Naive Bayes* dan KNN menggunakan *RapidMiner*. Pengukuran dilakukan untuk membandingkan tingkat akurasi penggunaan kedua metode pada dataset mahasiswa. Penelitian tersebut menghasilkan tingkat akurasi penggunaan klasifikasi *Naive Bayes* dan KNN diperoleh sebesar 66,67 % *Naive Bayes* unggul dibandingkan dengan tingkat akurasi KNN yaitu 38,89 % [4]. Selanjutnya penelitian terkait topik yang dibahas tentang penerapan sentimen analisis pada *informative tweets* di India yang menggunakan Twitter API dan data diekstrak dengan teknik NLP *toolkit*. Selain itu menggunakan pengujian dengan *Chi Square* pada *Naive Bayes Classifier* untuk data *training* dan data *testing*. Tujuan penelitian tersebut yaitu untuk mengevaluasi polaritas sentimen menggunakan Python. Berdasarkan pengujian pada 10, 100, 1000 dan 10000 data set, jumlah fitur yang meningkat dapat mempengaruhi tingkat akurasi pada fitur terpilih. Hal ini diperoleh melalui ekstraksi fitur data *tweet* dengan *chi square* [7]. Pengolahan *data mining* terkait klasifikasi dokumen teks juga dilakukan untuk mengevaluasi dokumen multiclass dan mengetahui tingkat keakuratan metode *Naive Bayes* yang diterapkan secara linier dan hierarkis untuk meningkatkan efisiensi model klasifikasi. Dalam kasus tersebut klasifikasi hirarkis dengan *Naive Bayes* menampilkan kinerja yang lebih baik dalam dokumen multi-label [8]. Penelitian yang membahas tentang sentimen ulasan film dan ulasan hotel menggunakan metode statistik untuk mengetahui subjektif dan polaritas kalimat. Selain itu, membahas dua algoritma yaitu *K-Nearest Neighbor* dan *Naive Bayes* untuk membandingkan ketepatan, presisi dan nilai *recall*. Dalam percobaan yang dilakukan menghasilkan pendekatan *Naive Bayes* lebih baik digunakan untuk ulasan film dengan 80 % akurasi daripada pendekatan KNN untuk ulasan hotel [5].

### Landasan Teori

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 [9]. *Opinion Mining* atau yang biasa disebut dengan sentimen analisis merupakan riset komputasional dari suatu opini, sentimen, dan emosi yang diekspresikan secara tekstual. *Opinion mining* bertujuan untuk mengekstrak atribut dan komponen dari objek yang telah dikomentari pada setiap dokumen dan untuk menentukan apakah komentar tersebut positif atau negatif [10]. Berdasarkan sumber datanya, *opinion mining* dapat dibedakan dalam dua kategori yaitu sentimen analisis pada level dokumen atau *Coarse-grained Sentiment Analysis* dan sentimen analisis pada level kalimat atau *Fine-grained Sentiment Analysis* [11]. Terdapat dua pendekatan untuk melakukan analisis sentimen, yaitu *Semantic Orientation Approach* dan *Machine Learning Approach*. *Semantic Orientation Approach* termasuk dalam *Unsupervised Learning* bekerja dengan

menghitung orientasi dokumen dari orientasi semantik kata atau frase dalam dokumen. Orientasi semantik mengacu pada polaritas dan strength pada kata-kata, frasa, atau teks. Cara kerja *Machine Learning Approach* dalam melakukan sentimen yaitu dengan memecah kalimat menjadi kata atau frase, mempresentasikan kalimat *document vector* (*model bag-of word*), dan kemudian mengklasifikasikan kalimat berdasarkan *document vector*. Dalam penelitian ini menggunakan pendekatan dengan *Machine Learning* yang termasuk pada *Supervised Learning*.

*Naïve Bayes Classifier* merupakan teknik yang berdasarkan pada teorema bayes untuk melakukan prediksi berbasis probabilistik sederhana dengan mengasumsikan independensi (ketidaktergantungan) yang kuat (naif) dari setiap kejadian [12]. Menurut penelitian [13] *Naïve Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Proses klasifikasi dokumen dilakukan dengan dua tahap, yang pertama adalah melakukan pelatihan untuk data yang sudah diketahui kelasnya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kelasnya untuk pengujian. Dalam algoritma *Naïve Bayes Classifier* setiap dokumen direpresentasikan dengan pasangan atribut “x1, x2, x3, x4, ... , xn” dimana x1 adalah kata pertama, x2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori *tweet*. Pada klasifikasi algoritma mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (VMAP), dimana persamaan (1) ditunjukkan sebagai berikut:

$$V_{MAP} = \underset{V_j \in V}{\arg \max} \frac{P(x_1, x_2, x_3, x_4, \dots, x_n)P(V_j)}{P(x_1, x_2, x_3, x_4, \dots, x_n)} \dots(1)$$

Untuk persamaan P(x1, x2, x3, x4, . . .xn) nilainya konstan untuk semua kategori (Vj) sehingga persamaan (2) dapat ditulis sebagai berikut:

$$V_{MAP} = \underset{V_j \in V}{\arg \max} P(x_1, x_2, x_3, x_4, \dots, x_n)P(V_j) \dots(2)$$

Persamaan (2) dapat disederhanakan menjadi persamaan (3) seperti berikut:

$$V_{MAP} = \underset{V_j \in V}{\arg \max} \prod_{i=1}^n P(x_i | V_j)P(V_j) \dots(3)$$

Untuk P (xi | Vj) dihitung jumlah frekuensi kemunculan setiap kata tambah satu dibanding dengan jumlah frekuensi kemunculan kata dari setiap kategori ditambah jumlah semua kata dari setiap kategori. Sedangkan P (Vj) dihitung dengan jumlah dokumen setiap kategori j dibandingkan dengan jumlah dokumen latih dari semua kategori.

## 2. Pembahasan

Pembahasan dalam penelitian ini diawali dengan identifikasi tentang topik yang dijadikan objek penelitian yaitu E-KTP. Komentar pengguna Twitter dimanfaatkan sebagai sumber data yang dianalisis. Data tekstual berupa komentar atau *tweets* berbahasa Indonesia diperoleh melalui Twitter API pada proses *crawling* yang secara otomatis mengambil kata yang mengandung *keyword* tertentu. Periode waktu yang digunakan yaitu sejak Juli hingga Desember 2017. *Keyword* yang digunakan dalam pengambilan data komentar Twitter antara lain “e-ktip”, “kasus e-ktip”, “Pansus e-ktip”, “KPK e-ktip”, “kasus korupsi e-ktip” dan “DPR e-ktip”. Kemudian *tweets* tersebut melalui tahapan *text processing* untuk menghasilkan kata yang terpilih yang dapat mewakili tahapan selanjutnya yaitu klasifikasi dalam tiga target fitur, yaitu sentimen positif, negatif dan netral. Data yang dibutuhkan yaitu jenis data latih dan data uji. Data latih diperoleh dari kumpulan *tweets* yang telah diberi label untuk membentuk model analisis sentimen. Sebagian hasil dari *tweets* yang tidak terlabel digunakan sebagai data uji.

### Analisis Kebutuhan Sistem

Sebelum merancang suatu sistem yang dapat digunakan untuk melakukan *opinion mining*, perlu dilakukan analisis kebutuhan fungsional sistem. Hal ini agar sistem yang dihasilkan dapat bekerja menyesuaikan kebutuhan *opinion mining* yang telah dianalisis terlebih dahulu. Berikut merupakan kebutuhan fungsional yang dijelaskan pada Tabel 1.

**Tabel 1.** Tabel Analisis Kebutuhan Fungsional

Tahap	Kebutuhan Fungsional
<i>Crawling</i>	Melakukan pengambilan <i>tweets</i> dari Twitter API
<i>Case Folding</i>	Mengubah kata pada <i>tweets</i> ke dalam huruf kecil
<i>Convert Emoticon</i>	Mengubah <i>emoticon</i> menjadi suatu kata yang dapat diartikan
<i>Tokenizing</i>	Memecah <i>tweets</i> menjadi potongan kata
<i>Stopword Removal</i>	Menghapus setiap kata yang tidak relevan
<i>Stemming</i>	Mereduksi setiap kata untuk mendapatkan kata dasar
<i>Feature Extraction</i>	Menentukan term pada sentimen dengan menghitung frekuensi term pada setiap kategori
Klasifikasi	Mengklasifikasikan <i>tweets</i> dalam kelas positif, negatif dan netral
Visualisasi Diagram	Merepresentasikan hasil sentimen dalam bentuk diagram

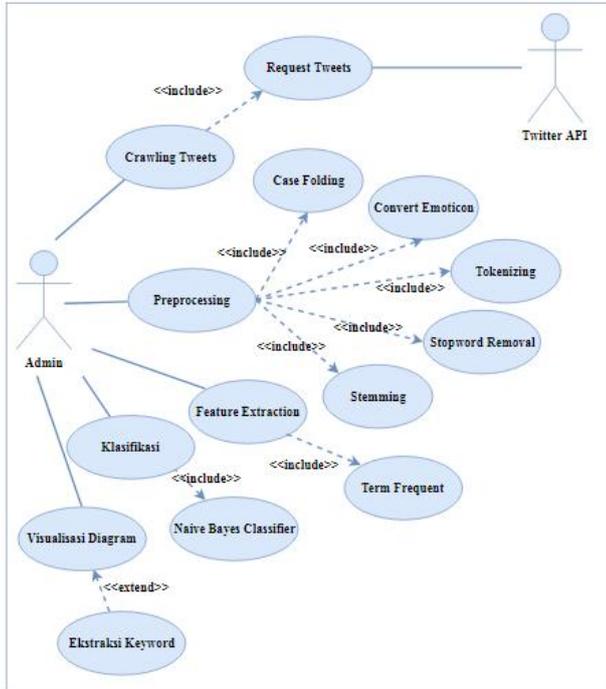
Setelah menentukan kebutuhan fungsional, diperlukan untuk mendeskripsikan kebutuhan non fungsional berupa spesifikasi perangkat keras dan perangkat lunak yang digunakan pada lingkungan implementasi sistem. Kebutuhan non fungsional sistem dapat dijelaskan pada Tabel 2.

**Tabel 2.** Tabel Analisis Kebutuhan Non Fungsional

Spesifikasi Perangkat Keras
Processor Intel Core i3
RAM 4 GB
Spesifikasi Perangkat Lunak
Sistem Operasi Windows 7, 32 Bit
Netbeans 8.2
JDK 1.8
Apache Lucene Core 5.3.1

**Rancangan Pemodelan Sistem**

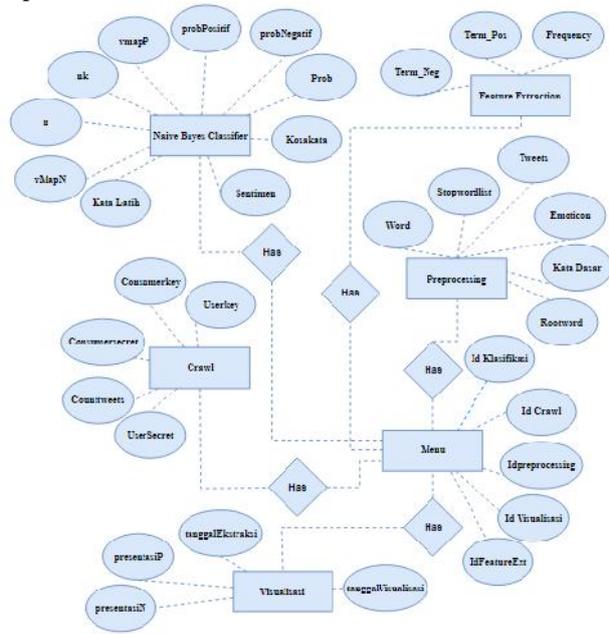
Dalam perancangan *use case*, secara umum terdapat beberapa kejadian yang digunakan yaitu *crawling tweets*, *text processing*, *feature extraction*, klasifikasi dan visualisasi diagram. Masing-masing dalam proses ini terbagi menjadi beberapa kejadian (*case*) lain dan memiliki aktor sebagai admin dan API twitter sebagai perantara sumber data. *Use case* diagram dalam penelitian ini ditunjukkan pada Gambar 2 seperti berikut.



**Gambar 2.** Use Case Diagram

Setelah mengetahui kejadian-kejadian yang terdapat dalam sistem. Diperlukan untuk mengetahui data yang direlaskan. Dalam penelitian ini menggunakan *entity*

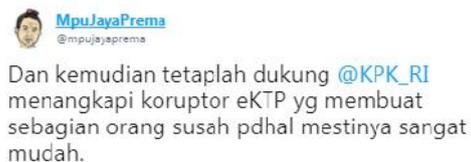
*relational database* yang ditunjukkan pada Gambar 3 seperti berikut.



**Gambar 3.** ER-Diagram

**Implementasi**

Proses yang dilakukan pada tahap pengumpulan data melalui *crawling tweets* menggunakan API Twitter dengan memasukan kata kunci terkait E-KTP. Berdasarkan kata kunci tersebut selanjutnya diinisialisasi sebuah objek baru dengan *tipe Query* yang merupakan *class Query* pada Twitter4J API. Contoh data yang diperoleh dari Twitter yang diolah ditunjukkan pada Gambar 4.



**Gambar 4.** Contoh Data

Tahapan-tahapan *text processing* yang dilakukan yaitu:

1. *Case Folding*: dan kemudian tetaplah dukung menangkapi koruptor ektp yg membuat sebagian orang susah pdhal mestinya sangat mudah.
2. *Convert Emoticon*: “jika terdapat emoticon maka disesuaikan dengan kamus emoticon yang mendefinisikan arti positif atau negatif”. Dalam contoh ini tidak ada maka bentuk kata-kata tetap.
3. *Stopword Removal*: tetaplah dukung menangkapi koruptor ektp membuat susah mestinya.
4. *Tokenizing*: “tetaplah”, “dukung”, “menangkapi”, “koruptor”, “ektp”, “membuat”, “susah”. “mestinya”.
5. *Stemming*: tetap dukung tangkap korup ektp buat susah mesti .

Setelah melakukan *text processing*, selanjutnya melakukan ekstraksi fitur. Dalam prosesnya, user harus menentukan *term* untuk sentimen positif dan negatif. Kemudian sistem menghitung frekuensi untuk *term* pada setiap kategori. *Source Code* dalam perhitungan *term* ditunjukkan pada Gambar 5.

```
private int setFrekuensi(String trm){
    int fre = 0;
    Collections.sort(allTerm);
    for (String t:allTerm){
        if (trm.equals(t.trim())){
            fre++;
        }
    }
    return fre;
}
```

Gambar 5. Source Code Perhitungan Term

Setelah menghitung frekuensi *term*, dilakukan klasifikasi menggunakan algoritma *Naïve Bayes* untuk mengklasifikasikan data pada kelas sentimen positif, negatif dan netral. Klasifikasi yang dilakukan pada tahap ini ditunjukkan pada Gambar 6 sebagai alur proses algoritma *Naïve Bayes* dalam data *training*.



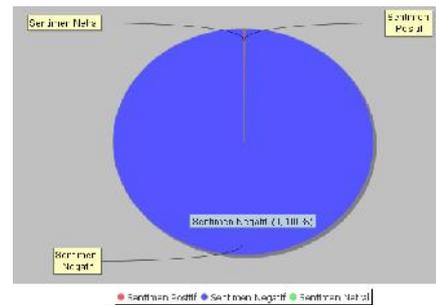
Gambar 6. Alur Klasifikasi Naïve Bayes

Perhitungan yang dilakukan dengan menginputkan data *training* kemudian terdapat pada dokumen kategori *Vj* dilakukan perhitungan nilai  $P(V_j)$  pada setiap kelas dan selanjutnya menghitung nilai  $P(x_i | V_j)$  untuk setiap *xi* pada kosakata dan diperoleh hasil *training* atau model klasifikasi. *Source code* yang digunakan dalam tahapan klasifikasi sentimen positif, negatif dan netral ditunjukkan pada Gambar 7 seperti berikut.

```
private void analisisSentimen(){
    double a,b;
    t = (double) pcs.waet*negtweet;
    readPos(sumber);
    sentiPos = ppcz;
    pwPos = new double[mfcrepos.size()];
    for (int i=0;i<mfcrepos.length;i++){
        a = frePos[i];
        pwPos[i] = a/b;
        sentiPos = sentiPos * pwPos[i];
    }
    sentiPos = sentiPos/posTweet;
    this.wmapPos.setText(String.valueOf(sentiPos));
    pwNeg = new double[mfcreneg.size()];
    for (int i=0;i<mfcreneg.length;i++){
        a = freNeg[i];
        pwNeg[i] = a/b;
    }
    sentiNeg = pncz;
    for (int i=0;i<pwNeg.length;i++){
        sentiNeg=sentiNeg*pwNeg[i];
    }
    pwNet = new double[mfcrenet.size()];
    for (int i=0;i<mfcrenet.length;i++){
        a = freNet[i];
        pwNet[i] = a/b;
    }
    sentiNet = pncz;
    for (int i=0;i<pwNet.length;i++){
        sentiNet=sentiNet*pwNet[i];
    }
}
```

Gambar 7. Source Code Perhitungan Sentimen

Setelah diperoleh hasil perhitungan sentimen, maka hasil klasifikasi divisualisasikan melalui diagram yang terdiri dari sentimen positif, sentimen negatif dan sentimen netral seperti Gambar 8.



Gambar 8. Visualisasi Diagram

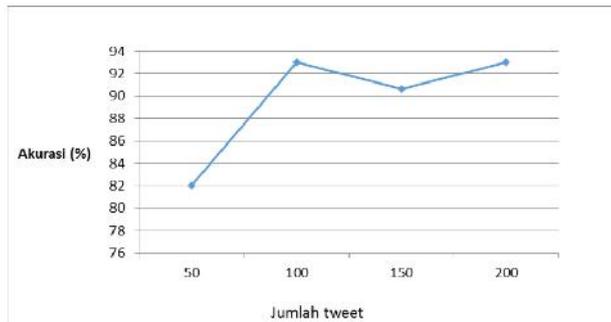
### Pengujian

Tahap pengujian yang dilakukan yaitu mengukur kinerja algoritma *Naïve Bayes*. Dalam proses pengujian akurasi ini, dilakukan pengujian terhadap empat kelompok *tweet* dengan jumlah yang berbeda. Deskripsi untuk kelompok *tweet* yang digunakan sebagai data uji akurasi. Hasil pengujian pada empat kelompok *tweet* ditunjukkan pada Tabel 3.

Tabel 3. Hasil Pengujian Akurasi

Tweet	Jumlah Tweet Hasil Uji			Jumlah Tweet Benar			Akurasi
	Pos	Neg	Net	Pos	Neg	Net	
50	1	33	16	1	24	16	$\frac{1+24+16}{50} \times 100 = 82$
100	10	48	42	10	48	35	$\frac{10+48+35}{100} \times 100 = 93$
150	19	67	64	19	67	50	$\frac{19+67+50}{150} \times 100 = 90.67$
200	22	102	76	22	102	62	$\frac{22+102+62}{200} \times 100 = 93$
<b>Total</b>							358,67

Tabel 3 menjelaskan tingkat akurasi yang dilakukan pada pengujian empat kelompok *tweet*. Selanjutnya pengujian akurasi tersebut ditampilkan dalam bentuk grafik seperti ditunjukkan pada Gambar 9.



Gambar 9. Grafik Hasil Pengujian

Berdasarkan hasil pengujian akurasi pada Tabel 3, maka akurasi rata-rata sistem dalam melakukan analisis sentimen dapat dihitung melalui persamaan (4) seperti berikut:

$$\text{Akurasi} = \frac{\text{Jumlah Total Akurasi}}{\text{Jumlah Kelompok Tweet}} = \frac{358,67}{4} = 89,67\% \quad ..(4)$$

Hasil yang diperoleh untuk tingkat akurasi analisis sentimen yang bertema E-KTP menghasilkan akurasi penggunaan algoritma *Naïve Bayes* sebesar 89,67 %.

### 3. Kesimpulan

Penelitian ini dalam melakukan *opinion mining* menerapkan algoritma *Naïve Bayes Classifier* yang digunakan dalam melakukan klasifikasi untuk topik E-KTP. Klasifikasi dengan *Naïve Bayes* dibagi dalam tiga target fitur yaitu kelas positif, kelas negatif dan kelas netral. Hasil yang diperoleh melalui pengujian sistem pada empat kelompok *tweet* menghasilkan tingkat akurasi sebesar 89,67 %.

### Daftar Pustaka

- [1] Admin. Informasi Pengguna Twitter. Website Online: www.cnnindonesia.com. Diakses pada 26 Oktober 2017.
- [2] Admin. Informasi e-KTP. Website Online: www.e-ktp.com/. Diakses pada 19 November 2017.
- [3] Feldman, R & Sanger, J. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press: New York.
- [4] Rajeswari, R.P, Juliet, K & Aradhana. *Text Classification for Student Data Set using Naïve Bayes Classifier and KNN Classifier*. International Journal of Computer Trends and Technology (IJCTT) Vol. 43 Number 1.
- [5] Lopamudra Dey, et.al. *Sentiment Analysis of Review Datasets Using Naïve Bayes and KNN Classifier*. International Journal Information Engineering and Electronic Business. 2016. DOI: 10.5815/ijieeb.2016.04.07.
- [6] Ayman E. Khedr, S.E.Salama, Nagwa Yaseen, "Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis", International Journal of Intelligent Systems and Applications(IJISA), Vol.9, No.7, pp.22-30, 2017. DOI: 10.5815/ijisa.2017.07.03.
- [7] Vadivukarassi, M. Puviarasan, N. & Aruna, P. *Sentimental Analysis of Tweets Using Naïve Bayes Algorithm*. World Applied Sciences

- Journal 35 (1): 54 – 59, 2017. ISSN 1818-4952. DOI: 10.5829/idosi.wasj.2017.54.59.
- [8] Jadon, E. Sharma, R. *Data Mining: Document Classification using Naïve Bayes Classifier*. International Journal of Computer Applications Vol. 167 – No. 6 June 2017.
- [9] Rozi, I.F. Pramono, S.H. Dahlan, E.A *Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi*. Jurnal EECCIS Vol. 6, No. 1, Juni 2012.
- [10] Liu, B., 2010, *Handbook of Natural Language Processing*, chapter Sentiment Analysis and Analysis, 2nd Edition.
- [11] Clayton R. Fink, et. al. 2011. *Coarse- and Fine-Grained Sentiment Analysis of Social Media Text*. Johns hopkins apl technical digest, volume 30, number 1.
- [12] Eko Prasetyo. *Data Mining Konsep dan Aplikasi menggunakan MATLAB*, 1st ed., Nikodemus WK, Ed. Yogyakarta: Penerbit ANDI, 2012.
- [13] Olson, David L.; & Delen, Dursun. *Advanced Data Mining Techniques*, Springer, 1st edition (February 1, 2008), page 102.

### Biodata Penulis

**Mihuandayani**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika Universitas AMIKOM Yogyakarta, lulus tahun 2016. Saat ini sedang menempuh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknik Informatika Universitas AMIKOM Yogyakarta. Sejak 2016, bekerja sebagai staf Quality Assurance PT. Time Excelindo.

**Eko Feriyanto**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika STMIK AMIKOM Yogyakarta, lulus tahun 2006. Saat ini sedang menempuh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknik Informatika Universitas AMIKOM Yogyakarta.

**Syarham**, memperoleh gelar Sarjana Sistem Informasi (S.Si), Jurusan Sistem Informasi STMIK Bina Bangsa Kendari, lulus tahun 2012. Saat ini sedang menempuh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknik Informatika Universitas AMIKOM Yogyakarta. Saat ini bekerja di Dinas Transmigrasi dan Ketenagakerjaan Bombana.

**Kusrini**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Ilmu Komputer Universitas Gadjah Mada Yogyakarta, lulus tahun 2002. Memperoleh gelar (M.Kom) pada Jurusan Ilmu Komputer Universitas Gadjah Mada Yogyakarta di tahun 2006 dan memperoleh gelar Doktor pada jurusan Ilmu Komputer Universitas Gadjah Mada pada tahun 2010. Sejak 2003, bekerja sebagai Dosen Tetap Universitas Amikom Yogyakarta. Saat ini menjabat sebagai Direktur Program Pascasarjana dan Ketua Program Studi S2 Teknik Informatika.