PENGGUNAAN N-GRAM PADA ANALISA SENTIMEN PEMILIHAN KEPALA DAERAH JAKARTA MENGGUNAKAN ALGORITMA NAÏVE BAYES

Disusun sebagai salah satu syarat menyelesaikan Program Studi Strata I pada Jurusan Informatika Fakultas Komunikasi dan Informatika

Oleh:

WAHYU CANDRA INDHIARTA

L 200 130 023

PROGRAM STUDI INFORMATIKA

FAKULTAS KOMUNIKASI DAN INFORMATIKA

UNIVERSITAS MUHAMMADIYAH SURAKARTA

2017

HALAMAN PERSETUJUAN

PENGGUNAAN N-GRAM PADA ANALISA SENTIMEN PEMILIHAN KEPALA DAERAH JAKARTA MENGGUNAKAN ALGORITMA NAÏVE BAYES

PUBLIKASI ILMIAH

oleh:

WAHYU CANDRA INDHIARTA
L 200 130 023

Telah diperiksa dan disetujui untuk diuji oleh:

Dosen Pembimbing

Endang Wahyu Pamungkas, S.Kom, M.Kom.

NIK. 100.1704

HALAMAN PENGESAHAN

PENGGUNAAN N-GRAM PADA ANALISA SENTIMEN PEMILIHAN KEPALA DAERAH JAKARTA MENGGUNAKAN ALGORITMA NAÏVE BAYES

OLEH

WAHYU CANDRA INDHIARTA

L 200 130 023

Telah dipertahankan di depan Dewan Penguji
Fakultas Komunikasi dan Informatika
Universitas Muhammadiyah Surakarta
Pada hari Jumat, 4 Agustus 2017
dan dinyatakan telah memenuhi syarat

Dewan Penguji:

1. **Endang Wahyu P., S.Kom, M.Kom.** (.....)
   (Ketua Dewan Penguji)
2. **Helman Muhammad, S.T., M.T.** (.....)
   (Anggota I Dewan Penguji)
3. **Nurgiyatna, M.Sc., Ph.D.** (.....)
   (Anggota II Dewan Penguji)

Publikasi ilmiah ini telah diterima sebagai salah satu persyaratan
Untuk memperoleh gelar sarjana
Tanggal 4 Agustus 2017
Mengetahui,

Dekan
Fakultas Komunikasi dan Informatika

Nurgiyatna, M.Sc.,PhD
NIK. 881

Ketua Program Studi
Informatika

Dr. Heru Supriyono, M.Sc.
NIK. 970

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam naskah publikasi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak terbukti ada ketidakbenaran dalam pernyataan saya di atas, maka akan saya pertanggungjawabkan sepenuhnya.

Surakarta, 4 Agustus 2017

Penulis

WAHYU CANDRA INDHIARTA

L 200 130 023

UNIVERSITAS MUHAMMADIYAH SURAKARTA

FAKULTAS KOMUNIKASI DAN INFORMATIKA

PROGRAM STUDI INFORMATIKA

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448

Surakarta 57102 Indonesia. Web: http://informatika.ums.ac.id. Email: informatika@ums.ac.id

SURAT KETERANGAN LULUS PLAGIASI

247/A.3-II.3/INF-FKI/VIII/2017

Assalamu'alaikum Wr. Wb

Biro Tugas Akhir Program Studi Informatika menerangkan bahwa :

Nama	:	WAHYU CANDRA INDHIARTA
NIM	:	L200130023
Judul	:	PENGGUNAAN N-GRAM PADA ANALISA SENTIMEN PEMILIHAN KEPALA DAERAH JAKARTA MENGGUNAKAN ALGORITMA NAÏVE BAYES
Program Studi	:	Informatika
Status	:	**Lulus**

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Tugas Akhir, dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

Surakarta, 8 Agustus 2017

Biro Tugas Akhir Informatika

Endang Wahyu Pamungkas, S.Kom., M.Kom.

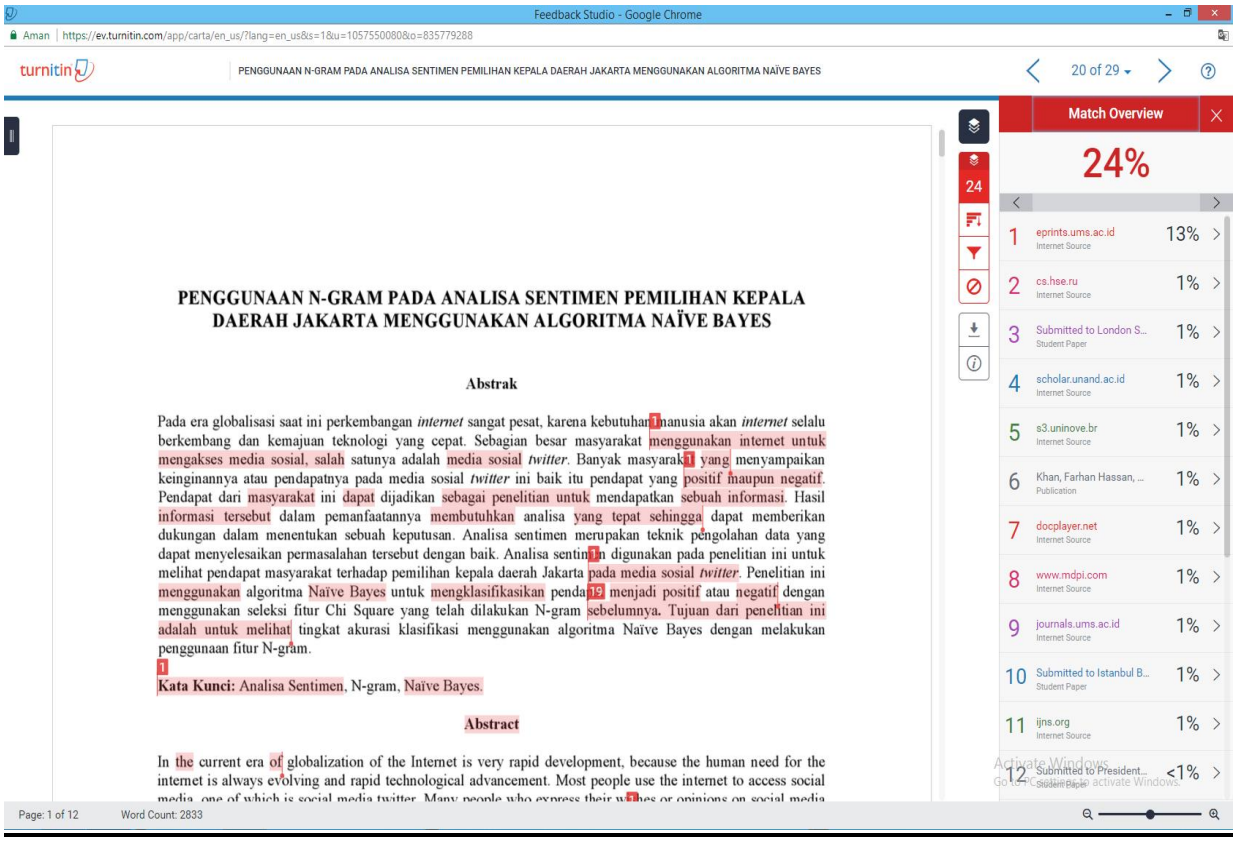
UNIVERSITAS MUHAMMADIYAH SURAKARTA

FAKULTAS KOMUNIKASI DAN INFORMATIKA

PROGRAM STUDI INFORMATIKA

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448

Surakarta 57102 Indonesia. Web: http://informatika.ums.ac.id. Email: informatika@ums.ac.id



PENGGUNAAN N-GRAM PADA ANALISA SENTIMEN PEMILIHAN KEPALA DAERAH JAKARTA MENGGUNAKAN ALGORITMA NAÏVE BAYES

Abstrak

Pada era globalisasi saat ini perkembangan *internet* sangat pesat, karena kebutuhan manusia akan *internet* selalu berkembang dan kemajuan teknologi yang cepat. Sebagian besar masyarakat menggunakan internet untuk mengakses media sosial, salah satunya adalah media sosial *twitter*. Banyak masyarakat yang menyampaikan keinginannya atau pendapatnya pada media sosial *twitter* ini baik itu pendapat yang positif maupun negatif. Pendapat dari masyarakat ini dapat dijadikan sebagai penelitian untuk mendapatkan sebuah informasi. Hasil informasi tersebut dalam pemanfaatannya membutuhkan analisa yang tepat sehingga dapat memberikan dukungan dalam menentukan sebuah keputusan. Analisa sentimen merupakan teknik pengolahan data yang dapat menyelesaikan permasalahan tersebut dengan baik. Analisa sentimen digunakan pada penelitian ini untuk melihat pendapat masyarakat terhadap pemilihan kepala daerah Jakarta pada media sosial *twitter*. Penelitian ini menggunakan algoritma Naïve Bayes untuk mengklasifikasikan pendapat menjadi positif atau negatif dengan menggunakan seleksi fitur Chi Square yang telah dilakukan N-gram sebelumnya. Tujuan dari penelitian ini adalah untuk melihat tingkat akurasi klasifikasi menggunakan algoritma Naïve Bayes dengan melakukan penggunaan fitur N-gram.

Kata Kunci: Analisa Sentimen, N-gram, Naïve Bayes.

Abstract

In the current era of globalization of the Internet is very rapid development, because the human need for the internet is always evolving and rapid technological advancement. Most people use the internet to access social

PENGGUNAAN N-GRAM PADA ANALISA SENTIMEN PEMILIHAN KEPALA DAERAH JAKARTA MENGGUNAKAN ALGORITMA NAÏVE BAYES

Abstrak

Pada era globalisasi saat ini perkembangan *internet* sangat pesat, karena kebutuhan manusia akan *internet* selalu berkembang dan kemajuan teknologi yang cepat. Sebagian besar masyarakat menggunakan internet untuk mengakses media sosial, salah satunya adalah media sosial *twitter*. Banyak masyarakat yang menyampaikan keinginannya atau pendapatnya pada media sosial *twitter* ini baik itu pendapat yang positif maupun negatif. Pendapat dari masyarakat ini dapat dijadikan sebagai penelitian untuk mendapatkan sebuah informasi. Hasil informasi tersebut dalam pemanfaatannya membutuhkan analisa yang tepat sehingga dapat memberikan dukungan dalam menentukan sebuah keputusan. Analisa sentimen merupakan teknik pengolahan data yang dapat menyelesaikan permasalahan tersebut dengan baik. Analisa sentimen digunakan pada penelitian ini untuk melihat pendapat masyarakat terhadap pemilihan kepala daerah Jakarta pada media sosial *twitter*. Penelitian ini menggunakan algoritma Naïve Bayes untuk mengklasifikasikan pendapat menjadi positif atau negatif dengan menggunakan seleksi fitur Chi Square yang telah dilakukan N-gram sebelumnya**.** Tujuan dari penelitian ini adalah untuk melihat tingkat akurasi klasifikasi menggunakan algoritma Naïve Bayes dengan melakukan penggunaan fitur N-gram.

Kata Kunci: Analisa Sentimen, N-gram, Naïve Bayes.

Abstract

In the current era of globalization of the Internet is very rapid development, because the human need for the internet is always evolving and rapid technological advancement. Most people use the internet to access social media, one of which is social media twitter. Many people who express their wishes or opinions on social media twitter is both positive and negative opinions. Opinions from this community can be used as research to obtain an information. The result of such information in its utilization requires proper analysis so as to provide support in determining a decision. Sentiment analysis is a data processing technique that can solve the problem well. Sentiment analysis was used in this study to see the opinion of the public against the election of the Jakarta regional head on social media twitter. This study used the Naïve Bayes algorithm to classify opinions to be positive or negative by using the selection of features of Chi Square that have been done N-gram before. The purpose of this research is to see the level of classification accuracy using Naïve Bayes algorithm by using N-gram feature.

Keywords: Sentiment Analysis, N-gram, Naïve Bayes.

1. PENDAHULUAN

Perkembangan *internet* di era globalisasi kini semakin bertambah pesat, karena kebutuhan manusia akan *internet* selalu berkembang dan kemajuan teknologi yang cepat. *Internet* adalah jaringan komputer dunia, yang terbentuk dari ribuan komputer yang saling berhubungan dengan memanfaatkan protokol yang sejenis untuk berbagi informasi secara bersama (Luthfi & Riasti, 2013). Sebagian besar masyarakat menggunakan *internet* untuk mengakses media sosial, salah satu contohnya adalah media sosial *twitter*.

Media sosial *twitter* merupakan layanan jejaring sosial dan mikroblog yang memungkinkan penggunanya untuk mengirim dan membaca pesan berbasis teks hingga 140 karakter. Pengguna *twitter* pada 14 Oktober 2013 sebanyak 218,3 juta orang diseluruh dunia, sedangkan di Indonesia 19,5 juta orang pada 18 Desember 2013 (Azeharie & Kusuma, 2015). Banyaknya pengguna media sosial *twitter* dapat dimanfaatkan untuk mengetahui pendapat

masyarakat mengenai kebijakan publik yang dikeluarkan pemerintah (Nurfalah & Ardiyanti, 2017). Pendapat masyarakat yang tertuang pada media sosial *twitter* dapat diolah dan menghasilkan informasi namun pemanfaatannya membutuhkan teknik analisa yang tepat yaitu dengan analisa sentimen.

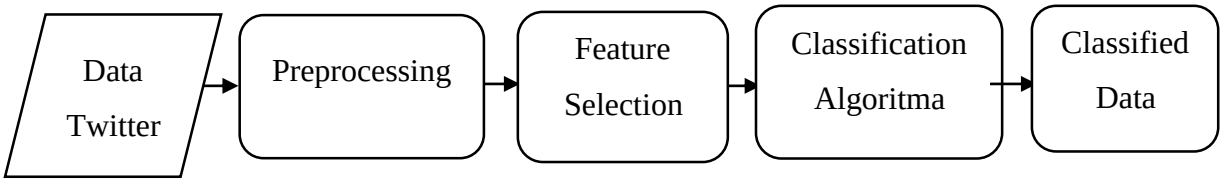
Analisa sentimen adalah cabang ilmu pembelajaran di domain *text mining* yang mempelajari analisa terhadap suatu opini, sentimen, emosi, sikap, evaluasi yang dituangkan ke dalam bentuk tekstual (Liu, 2012). Penggunaan teknik analisa sentimen sering kali digunakan untuk *review* produk, manajemen reputasi, analisa terhadap suatu topik dan lain sebagainya. Pada penelitian Dhande & Patnaik (2014) melakukan analisa sentimen untuk mengetahui *review* terhadap suatu film menggunakan metode Naïve Bayes dengan membagi kategori pendapat menjadi positif dan negatif. Teknik pembelajaran analisa sentimen salah satunya dapat diselesaikan menggunakan Algoritma Naïve Bayes. Algoritma Naïve Bayes merupakan strategi klasifikasi yang sederhana dan intuitif yang kinerjanya mirip dengan pendekatan klasifikasi lainnya tetapi memiliki performa tingkat akurasi yang cukup tinggi (Gamallo, Garcia & Fernández-Lanza, 2013). Algoritma Naïve Bayes merupakan metode klasifikasi populer yang sering digunakan untuk melakukan penelitian, Nugroho (2016) pada penelitiannya yang membandingkan 3 metode klasifikasi yaitu algoritma C4.5, Naïve Bayes dan Algoritma K-means menghasilkan nilai akurasi yang tinggi pada metode Naïve Bayes sehingga metode Naïve Bayes merupakan metode klasifikasi yang lebih baik daripada metode klasifikasi C4.5 dan Algoritma K-Means.

Sentimen analisis pada pemilu Amerika tahun 2012 yang dilakukan oleh Wang et al. (2012) menggunakan pengambilan data langsung dari media sosial *twitter* secara *real-time* yang menggunakan *twitter firehose*, aturan yang akurat dan menggunakan kata kunci untuk mendapatkan gambaran penuh dari pendapat politik yang ada pada *twitter* mengenai topik tersebut. Pada penelitian yang dilakukan oleh Bakliwal et al. (2013) dapat mengklasifikasikan *text twitter* menjadi positif, negatif dan netral yang merujuk pada partai politik tertentu atau pemimpin partai dengan akurasi hampir 59% menggunakan pendekatan dari lexicon. Banyaknya penelitian mengenai sentimen analisis dengan topik mengenai isu politik dapat dijadikan panduan untuk membuat sebuah penelitian yang lain.

Berdasarkan penelitian di atas yang sejenis penelitian ini mencoba melakukan analisa sentimen dengan menggunakan Algoritma Naïve Bayes untuk mengklasifikasikan data *twitter* mengenai topik pemilihan kepala daerah Jakarta dengan membagi pendapat menjadi 2 yaitu positif dan negatif dengan menerapkan penggunaan N-gram. Tujuannya yaitu untuk melihat tingkat akurasi klasifikasi sistem dengan penggunaan N-gram pada Algoritma Naïve Bayes,

sebagaimana pada penelitian Afshoh (2017) penggunaan fitur N-gram sangat berpengaruh dalam perhitungan ketepatan nilai akurasi klasifikasi Algoritma Naïve Bayes.

2. METODOLOGI



Gambar 1. Alur metodologi

Gambar 1 merupakan urutan alur kerja penelitian yang dilakukan. Adapun penjelasan untuk serangkaian urutan alur kerja diatas adalah :

1. Data Twitter

   Tahap pertama yang dilakukan pada penelitian yaitu mengumpulkan data *twitter* yang diambil secara acak tentang pemilihan kepala daerah Jakarta dari 3 calon pasangan, kemudian untuk setiap calon pasangan masing-masing terdiri dari 100 data positif dan 100 data negatif. Data *twitter* tersebut berisi kalimat pendapat dari masyarakat mengenai ketiga calon pasangan gubernur Jakarta yang dicari menggunakan #dukung untuk mencari data positif dan #tolak untuk mencari data negatif. Total data twitter yang berhasil dikumpulkan yaitu 600 data yaitu 100 data positif dan 100 data negatif untuk pasangan calon gubernur pertama, 100 data positif dan 100 data negatif untuk pasangan calon gubernur kedua, serta 100 data positif dan 100 data negatif untuk pasangan calon ketiga.

2. Preprocessing

   Tahap kedua yaitu dengan melakukan *preprocessing* data. Preprocessing merupakan pengolahan awal data dan mempersiapkan data teks untuk dilakukan proses klasifikasi, yaitu dengan melakukan metode:

   1. Normalisasi

      Metode normalisasi merupakan metode untuk menormalisasikan data teks *twitter* menjadi data teks normal. Karena keterbatasan twitter yang membatasi karakternya banyak pengguna menuliskan kata-kata gaul seperti “TDK” jika dinormalisasikan menjadi “TIDAK”.

2. Transfrom Case

   Merupakan metode untuk mengubah data teks *twitter* yang ditulis dengan huruf besar (*upper case*) menjadi huruf kecil semua (*lower case*).

3. Tokenisasi

   Tokenisasi merupakan metode pengambilan data teks pada suatu dokumen untuk dipisahkan menjadi beberapa karakter/token.

4. Generate N-gram

   Setelah data teks *twitter* di normalisasi dan *transform case* berikutnya yaitu dengan tokenisasi menggunakan jenis token unigram, bigram dan trigram sama seperti pada penelitian Nurfalah & Ardiyanti (2017) yang melakukan pembagian N-gram menjadi tiga jenis. N-gram merupakan penggabungan kata sifat yang sering muncul untuk menunjukan suatu sentimen. Pada penelitian menggunakan jenis token unigram yaitu token data teks *twitter* yang hanya terdiri dari satu kata, kemudian bigram yaitu token data teks *twitter* yang terdiri dari dua kata dan trigram yaitu token data teks *twitter* yang terdiri dari tiga kata. Penerapan N-gram dapat dilihat seperti berikut :

   Contoh kalimat : Pemilihan kepala daerah Jakarta tahun ini tidak begitu ramai dibandingkan dengan tahun sebelumnya.

Unigram	Pemilihan, kepala, daerah, Jakarta, tahun, ini, tidak, begitu ramai, dibandingkan, dengan, tahun, sebelumnya.
Bigram	Pemilihan kepala, kepala daerah, daerah Jakarta, Jakarta tahun, tahun ini, ini tidak, tidak begitu, begitu ramai, ramai dibandingkan, dibandingkan dengan, dengan tahun, tahun sebelumnya.
Trigram	Pemilihan kepala daerah, kepala daerah Jakarta, daerah Jakarta tahun, Jakarta tahun ini, tahun ini tidak, ini tidak begitu, tidak begitu ramai, begitu ramai dibandingkan, ramai dibandingkan dengan, dibandingkan dengan tahun, dengan tahun sebelumnya.

   Tujuan pemakaian N-gram dilakukan pada penelitian ini karena dalam bahasa Indonesia banyak frase yang tidak hanya terdiri dari satu kata.

3. *Feature Selection Chi Square*

   Feature Selection merupakan suatu kegiatan yang dilakukan dengan tujuan untuk memilih *feature* yang berpengaruh dan mengesampingkan *feature* yang tidak berpengaruh dalam analisa sistem. Pada penelitian menggunakan *feature selection chi square* untuk memotong fitur-fitur yang tidak penting dalam proses klasifikasi Algoritma Naïve Bayes, sehingga perhitungannya dapat dilakukan denga rumus berikut (Ling, Kencana & Oka, 2014).

$$X^2(D,t,c) = \frac{(N_{00}+N_{11}+N_{10}+N_{01})x(N_{00}N_{11}-N_{10}N_{01})^2}{(N_{11}+N_{01})x(N_{11}+N_{10})x(N_{10}+N_{00})x(N_{01}+N_{00})}$$

4. Algoritma Naïve Bayes

   Algoritma Naïve Bayes yaitu algoritma yang memanfaatkan pencarian nilai probabilitas tertinggi untuk proses klasifikasi pada data uji yang tepat. Pada penelitian menggunakan data uji berupa data teks *twitter* mengenai pemilihan kepala daerah Jakarta dimana setiap pasangan calon memiliki 100 data *twitter* positif dan 100 data *twitter* negatif. Banyak ditemukan penelitian yang menggunakan metode Algoritma Naïve Bayes pada analisa sentimen, hal ini karena Naïve Bayes memiliki kelebihan melakukan proses klasifikasi yang sederhana tetapi cukup tinggi performa akurasinya. Selain itu, metode Naïve Bayes juga memiliki kekurangan yaitu sangat sensitif terhadap pemilihan fitur sehingga jika terlalu banyak jumlah fitur kemungkinan akan mengurangi nilai akurasi klasifikasi. Menurut Markov & Larose (2007) tahapan perhitungan Algoritma Naïve Bayes adalah sebagai berikut:

   1. Mencari nilai probabilitas tertinggi:

      $V_{MAP} = \frac{P(x|C)\,P(C)}{P(x)}$ .....(1)

   2. Jika nilai $P(x)$ adalah konstan maka:

      $V_{MAP} = P(x|C)\,P(C)$ .....(2)

   3. Bentuk dari persamaan (2) dapat disederhanakan lagi menjadi:

      $V_{MAP} = P(x_i|C_j)P(C_j)$ .....(3)

      Keterangan :

      V_{MAP} : semua kategori yang diujikan

      $P(x_i|C_j)$: probabilitas kategori x_i pada kategori *twitter* C_j

      $P(C_j)$: probabilitas kategori twitter C_j,

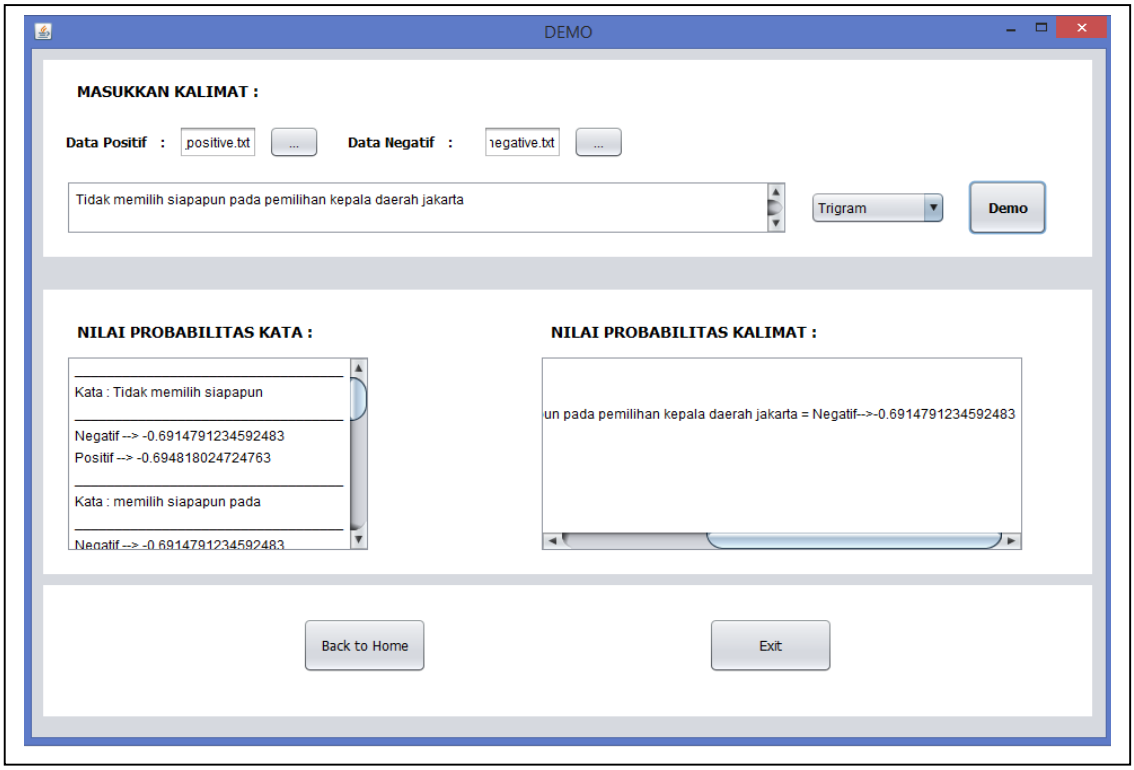
      dengan *j* merupakan kategori sentimen *tweet*.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

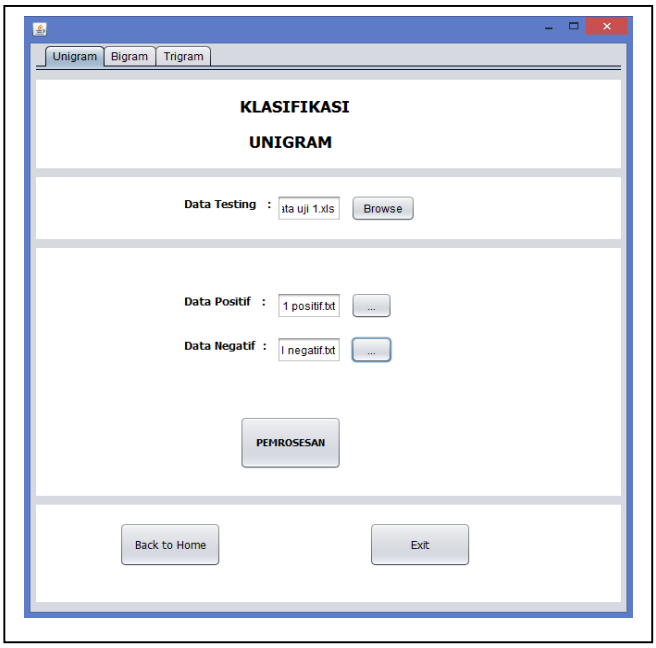
Gambar 2. Halaman utama sistem

Halaman yang ditunjukkan oleh gambar 2 merupakan halaman utama pada sistem analisa sentimen dengan N-gram menggunakan metode Naïve Bayes yang digunakan oleh pengguna untuk menjalankan demo, proses atau *menu* bantuan *help*.



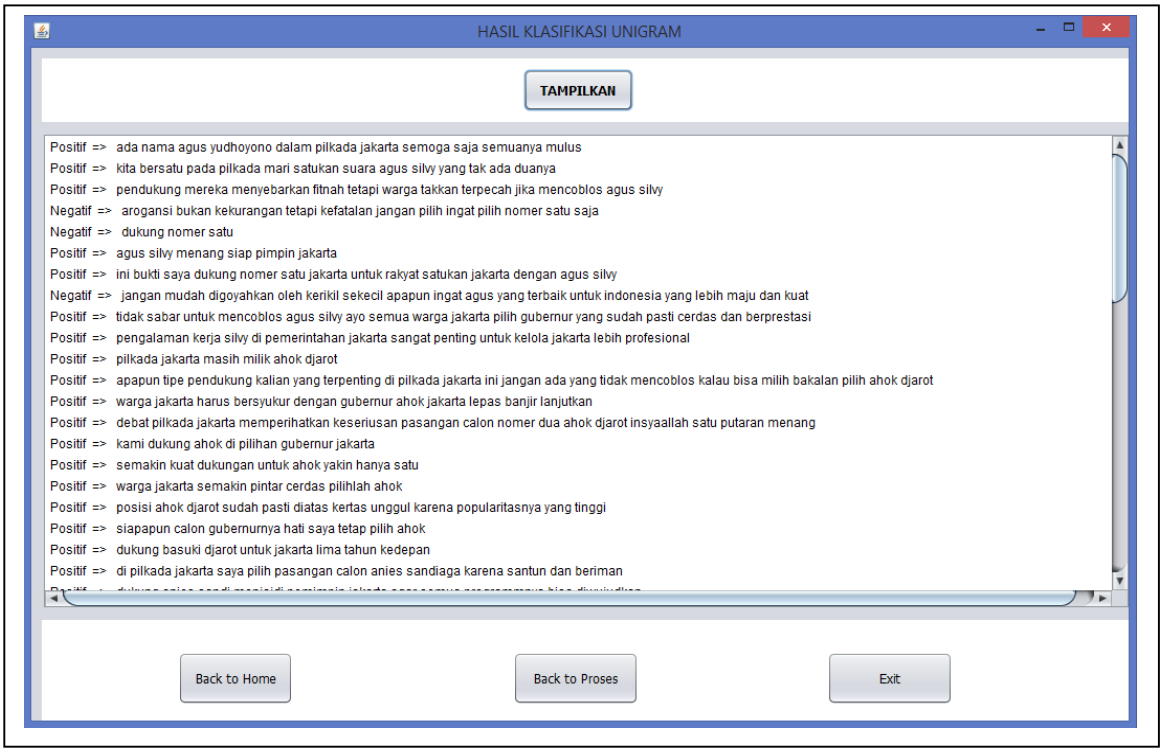
Gambar 3. Halaman *demo*

Halaman *demo* pada gambar 3 digunakan pengguna untuk memasukkan sebuah kalimat dan memilih N-gram yang akan digunakan untuk proses pengujian klasifikasi kalimat oleh sistem.



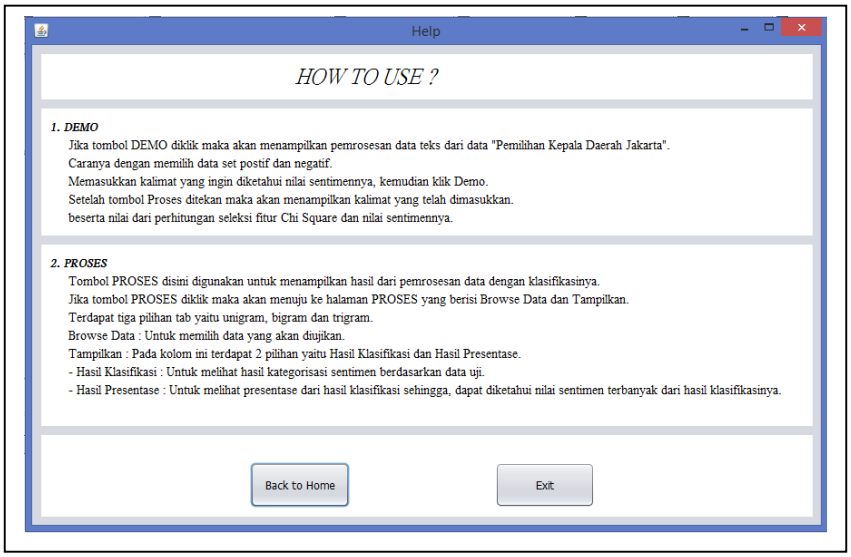
Gambar 4. Proses klasifikasi Unigram

Halaman yang ditujukkan pada gambar 4 untuk proses pengklasifikasian unigram, bigram dan trigram menggunakan data set yang telah disiapkan.



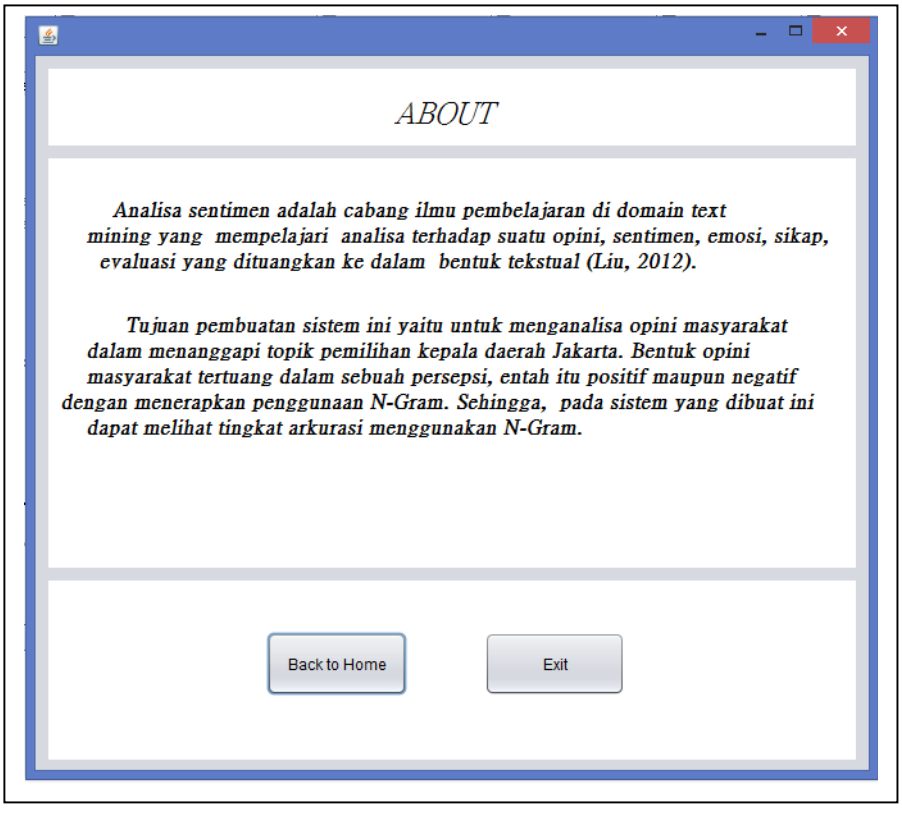
Gambar 5. Hasil pengujian unigram

Halaman yang tertera pada gambar 5 adalah hasil pemrosesan klasifikasi oleh sistem pada gambar 4, menampilkan kategori positif atau negatif dan kalimat yang telah diolah oleh sistem.



Gambar 6. Halaman *help*

Halaman yang ditunjukkan gambar 6 merupakan sebuah panduan singkat untuk mengoperasikan sistem analisa sentimen dengan N-gram menggunakan metode Naïve Bayes yang terdapat dua poin yaitu proses *demo* dan proses.



Gambar 7. Halaman *about*

Halaman yang tertera pada gambar 7 digunakan untuk penjelasan singkat tentang sistem analisa sentimen dengan N-gram menggunakan metode Naïve Bayes dan tujuan dari sistem ini dibuat.

3.2 Hasil Analisis

Penelitian sistem analisis sentimen yang menggunakan pendapat masyarakat berbahasa Indonesia dengan topik pemeilihan kepala daerah Jakarta pada media sosial *twitter* menggunakan spesifikasi perangkat keras dan lunak sebagai berikut :

1. Laptop dengan spesifikasi sebagai berikut :
   - Processor Intel(R) Core(TM) i5-3317U @ 1.70GHz 1.70 GHz, Hardisk 1 TB, OS Windows 8.1.
2. *Software* yang digunakan adalah :
   - NetBeans IDE 8.2, *Java.*
   - Notepad ++.

Hasil dari analisa dan pengujian yang telah dilakukan menggunakan data sebanyak 600 buah dari pendapat masyarakat mengenai pemilihan kepala daerah Jakarta menggunakan N-gram dengan metode Naïve Bayes. Pengujian dilakukan dengan menggunakan teknik pengujian sebanyak sepuluh kali atau *Ten Fold Cross Validation* (10-fold) yaitu memecah data menjadi 10 bagian. Data tersebut berisi 30 data positif dan 30 data negatif dari pemisahan data sebelumnya. Hasil analisa dan pengujian menggunakan unigram diperoleh nilai *precision* terbesar yaitu 0,933 pada data pengujian data ke lima, nilai *recall* terbesar yaitu 0,875 pada pengujian data ke enam, nilai *accuracy* terbesar yaitu 0,883 pada pengujian data ke lima. Pengujian bigram diperoleh nilai *precision* terbesar yaitu 0,9 pada pengujian data ke tujuh, nilai *recall* terbesar yaitu 0,96 pada pengujian data ke lima dan sembilan, nilai *accuracy* terbesar yaitu 0,883 pada pengujian data ke lima dan sembilan. Pengujian trigram diperoleh nilai *precision* terbesar yaitu 0,267 pada pengujian data ke delapan, nilai *recall* terbesar yaitu 1 pada pengujian data ke dua, tiga, empat, lima, enam dan sembilan, nilai *accuracy* terbesar yaitu 0,617 pada pengujian data ke delapan. Dari hasil pengujian yang telah dilakukan kemudian menghitung rata-rata nilai dari *precision, recall* dan *accuracy*. Hasil rata-rata dari data tersebut dapat dilihat pada Tabel 1.

Tabel 1. Hasil Pengujian *Ten Fold Cross Validation*

	Unigram			**Bigram**			**Trigram**		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy
Hasil	0,743	0,773	0,785	0,76	0,889	0,823	0,123	0,898	0,547

Tabel 1 merupakan hasil dari rata-rata nilai *precision, recall* dan *accuracy* menggunakan unigram, bigram dan trigram dengan metode Naïve Bayes. Pada tabel 1 dapat dilihat bahwa nilai *accuracy* paling besar terdapat pada penggunaan bigram yaitu 0,823, ini menunjukkan bahwa dengan menggunakan bigram ketepatan akurasi dari sistem lebih baik dibandingkan menggunakan unigram atau trigram, nilai *precision* tertinggi juga terdapat pada penggunaan bigram dengan 0,76. Namun pada nilai *recall* hasil tertinggi terdapat pada penggunaan trigram yaitu sebesar 0,898. Dapat disimpulkan bahwa penggunaan bigram dalam pengklasifikasian data lebih baik daripada menggunakan unigram atau trigram. Pada token trigram menunjukkan akurasi paling sedikit dengan 0,547, kalimat yang dipisahkan menjadi tiga suku kata akan cenderung menjadi klasifikasi negatif oleh sistem karena hasil dari pemecahan tiga suku kata tersebut kebanyakan sama dengan data set yang ada di negatif daripada data set positif. Untuk token unigram dan bigram hasil dari akurasinya sudah lumayan baik karena pada saat sistem mengklasifikasikannya kalimat yang dipisahkan satu suku kata dan dua suku kata akan mirip dengan data set yang digunakan untuk menentukan positif atau negatif dari hasil pengklasifikasian sistem dan tidak ada kecenderungan menjadi positif atau negatif ketika sistem memproses menggunakan unigram atau bigram. Data set sebaiknya memiliki tiga suku kata atau lebih, jika kurang dari itu maka ketika sistem mencoba menguji menggunakan token trigram maka hasilnya akan *error*.

4. PENUTUP

Penelitian yang telah dilakukan dengan penggunaan fitur N-gram pada analisa sentimen pemilihan Kepala Daerah Jakarta bertujuan untuk melihat peningkatan nilai akurasi klasifikasi sistem. Penggunaan fitur N-gram pada penelitian yaitu dengan melakukan jenis token unigram, bigram dan trigram. Berdasarkan ketiga jenis token N-gram yang digunakan, dapat dilihat bahwa jenis token bigram mampu memberikan hasil akurasi klasifikasi sistem yang lebih baik daripada jenis token unigram dan trigram yaitu dengan menghasilkan nilai akurasi sebesar 0,823. Hal ini karena token bigram memiliki nilai ketepatan klasifikasi antara informasi yang diharapkan oleh pengguna dengan jawaban dari sistem menunjukkan nilai *true positive* atau *correct result* yang lebih besar daripada jenis token yang lainnya. Besarnya tingkat keberhasilan sistem pada pengklasifikasian kalimat dalam menemukan kembali sebuah informasi yang memiliki sedikit kesalahan dalam proses pengklasifikasian atau *missing result.* Data yang berimbang antara ketiga pasangan tersebut dan bigram tetap memiliki jumlah *accuracy* yang paling tinggi menujukkan token jenis bigram yang paling

baik walau nilai dari *true negative* atau *correct absence of result* berimbang antara ketiga pasangan tersebut. N-gram juga dapat digunakan pada metode lain selain Naïve Bayes, penelitian selanjutnya sebaiknya mencoba untuk menggunakan N-gram pada metode lain untuk mengetahui perbedaan akurasinya. Kurangnya data uji juga mempengaruhi akurasi yang dihasilkan, diharapkan penelitian selanjutnya memiliki data uji yang cukup banyak untuk meningkatkan hasil akurasi.

DAFTAR PUSTAKA

Afshoh, F., Pamungkas, E. W., Kom, S., & Kom, M. (2017). Analisa Sentimen Menggunakan Naïve Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter (Skripsi Mahasiswa, Universitas Muhammadiyah Surakarta).

Azeharie, S., & Kusuma, O. (2015). Analisis Penggunaan Twitter Sebagai Media Komunikasi Selebritis Jakarta. *Jurnal Komunikasi, 6*(2), 83-98.

Bakliwal, A., Foster, J., van der Puil, J., O'Brien, R., Tounsi, L., & Hughes, M. (2013). Sentiment Analysis of Political Tweets: Towards an Accurate Classifier. *Association for Computational Linguistics* (pp. 49-58).

Dhande, L. L., & Patnaik, G. K. (2014). Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 3*(4), 313-320.

Gamallo, P., Garcia, M., & Fernández-Lanza, S. (2013). TASS: A Naive-Bayes Strategy for Sentiment Analysis on Spanish Tweets. In *Proceedings on Sentiment Analysis at SEPLN (TASS2013)* (pp. 126-132).

Indrayuni, E., & Wahyudi, M. (2015). Penerapan Character N-gram untuk Sentiment Analysis Review Hotel Menggunakan Algoritma Naive Bayes. *Konferensi Nasional Ilmu Pengetahuan dan Teknologi, 1*(1), 88-93.

Ling, J., N Kencana, I. P. E., & Oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika, 3*(3), 92-99.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures On Human Language Technologies*, 5(1), 1-167.

Luthfi, H. W., & Riasti, B. K. (2013). Sistem Informasi Perawatan dan Inventaris Laboratorium pada SMK Negeri 1 Rembang Berbasis Web. *Speed-Sentra Penelitian Engineering dan Edukasi, 3*(3), 69-77.

Markov, Z., & Larose, D. T. (2007). *Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage*. John Wiley & Sons.

Nugroho, Y. S. (2016). Klasifikasi dan Klastering Penjurusan Siswa SMA Negeri 3 Boyolali. Khazanah Informatika: *Jurnal Ilmu Komputer dan Informatika*, (1), 1-6.

Nurfalah, A., & Adiwijaya, A. A. S. (2017). Analisis Sentimen Berbahasa Indonesia dengan Pendekatan Lexicon-Based pada Media Sosial. *Jurnal Masyarakat Informatika Indonesia, 2*(1), 1-8.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A System for Real-Time Twitter Sentiment Analysis of 2012 us Presidential Election Cycle. In *Proceedings of the ACL 2012 System Demonstrations* (pp. 115-120). Association for Computational Linguistics.