



TESIS - TE2599

**ANALISA *MICROBLOGGING* UNTUK
MENENTUKAN PRIORITAS KEBIJAKAN PUBLIK
BERDASARKAN OPINI MASYARAKAT
MENGUNAKAN ALGORITMA *NAIVE BAYES* DAN
*ANALYTICAL HIERARCHY PROCESS***

MOHAMMAD KHOIRON
2214206708

DOSEN PEMBIMBING
Dr. Surya Sumpeno, S.T., M.Sc.
Dr. Adhi Dharma Wibawa, S.T., M.T.

PROGRAM MAGISTER
BIDANG KEAHLIAN TELEMATIKA-CIO
JURUSAN TEKNIK ELEKTRO
FAKULTAS TEKNIK INDUSTRI
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2016



THESIS - TE2599

**MICROBLOGGING ANALYSIS FOR DETERMINING
THE PRIORITY OF PUBLIC POLICY BASED ON
PUBLIC OPINION USING NAIVE BAYES AND
ANALYTICAL HIERARCHY PROCESS ALGORITHM**

MOHAMMAD KHOIRON
2214206708

SUPERVISOR
Dr. Surya Sumpeno, S.T., M.Sc.
Dr. Adhi Dharma Wibawa, S.T., M.T.

MASTER PROGRAM
AREAS OF EXPERTISE TELEMATIKA-CIO
ELECTRICAL ENGINEERING DEPARTMENT
FACULTY OF INDUSTRIAL TECHNOLOGY
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2016

LEMBAR PENGESAHAN TESIS

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Teknik (M.T)

di
Institut Teknologi Sepuluh Nopember
oleh:

Mohammad Khoiron
Nrp. 2214206708

Tanggal Ujian : 7 Januari 2016
Periode Wisuda : Maret 2016

Disetujui oleh:

1. 
Dr. Surya Sumpeno, S.T., M.Sc. (Pembimbing I)
NIP. 196906131997021003
2. 
Dr. Adhi Dharma Wibawa, S.T., M.T. (Pembimbing II)
NIP. 197605052008121003
3. 
Prof. Dr. Ir. Mauidhi Hery P, M. Eng. (Penguji)
NIP. 195809161986011001
4. 
Dr. Ir. Yoyon Kusnendar Suprpto, M.Sc. (Penguji)
NIP. 195409251978031001
5. 
Dr. Ir. Wirawan, DEA. (Penguji)
NIP. 196311091989031011
6. 
Eko Setijadi, S.T, M.T., Ph.D. (Penguji)
NIP. 197210012003121007



Direktur Program Pascasarjana,

Prof. Fauhar Manfaat, M.Sc, Ph.D
NIP. 195409251978031001

KATA PENGANTAR

Alhamdulillah, segala puji syukur penulis panjatkan kehadirat Allah SWT, atas segala karunia dan ridho-NYA, sehingga tesis dengan judul “Analisa *Microblogging* untuk Menentukan Prioritas Kebijakan Publik Berdasarkan Opini Masyarakat Menggunakan Algoritma *Naïve Bayes* dan *Analytical Hierarchy Process*” ini dapat diselesaikan.

Tesis ini disusun untuk memenuhi salah satu persyaratan memperoleh gelar Magister Teknik (M.T.) dalam bidang keahlian Telematika-CIO, Jurusan Teknik Elektro, Institut Teknologi Sepuluh Nopember Surabaya. Pada kesempatan ini penulis ingin menyampaikan rasa hormat dan ucapan terima kasih kepada:

1. Dr. Surya Sumpeno, S.T., M.Sc. atas bimbingan, arahan dan waktu yang telah diluahkan kepada penulis untuk berdiskusi selama menjadi dosen wali, dosen pembimbing dan perkuliahan.
2. Dr. Adhi Dharma Wibawa, S.T., M.T. atas bimbingan, arahan dan waktu yang telah diluahkan kepada penulis untuk berdiskusi selama menjadi dosen pembimbing dan perkuliahan.
3. Prof. Dr. Ir. Mauridhi Hery P, M. Eng. atas masukan dan arahnya selama menjadi penguji dalam sidang tesis, sehingga memberikan pemahaman baru dan menyempurnakan pengerjaan tesis ini.
4. Dr. Ir. Yoyon Kusnendar Suprpto, M.Sc , Dr. Ir. Wirawan, DEA , Eko Setijadi, S.T, M.T., Ph.D, atas masukan dan arahnya selama menjadi penguji dalam sidang tesis maupun sebagai dosen pengajar yang menginspirasi penulis untuk selalu berjuang sampai akhir dan memahami apa arti sebuah kesabaran.
5. Dr. Ir. Djoko Purwanto, M.Eng. selaku Ketua Jurusan Teknik Elektro, yang memberikan ruang kepada penulis untuk mengembangkan diri dan berkarya.
6. Prof. Ir. Djauhar Manfaat, M.Sc, Ph.D. selaku Direktur Pasca Sarjana, yang memberikan kesempatan kepada penulis untuk mengembangkan diri dan berkarya di Institut Teknologi Sepuluh Nopember

7. Semua Dosen-dosen Telematika-CIO ITS yang telah memberikan ilmu yang bermanfaat bagi penulis, mudah-mudahan dianggap sebagai amal jariyah oleh Allah SWT.
8. Semua Civitas Akademika Institut Teknologi Sepuluh Nopember, atas semua kebersamaan dan dukungan yang selama ini diberikan kepada penulis.
9. Ayahanda Moh. Maksum, Ibunda Islicha, Bapak mertua Achmad Roesmadji dan Ibu mertua Lilik Istiningsih dan saudara-saudaraku atas semua dukungan, bantuan dan doa-doa yang tidak pernah putus selama penulis belajar di S-2.
10. Istriku tercinta Endah Trihapsari atas semua pengorbanan jiwa raga maupun kesabaran yang luar biasa selama mendampingi penulis menyelesaikan studi S-2, yang tidak akan pernah tergantikan oleh apapun. Mudah-mudahan Allah menjadikan pahala atas semua yang telah dilakukannya kepada penulis.
11. Anakku Khalida Nazeeya An Nasira dan Muhammad Baqir Al Zahravi atas semua rasa cinta dan sayang yang kalian hadirkan dalam kehidupan ini. Mudah-mudahan menjadi anak yang sholeha dan sholeh.
12. Rekan-rekan S2 Telematika-CIO dan S2 Telematika seangkatan maupun beda angkatan yang selalu memberikan keceriaan, motivasi maupun bantuan lainnya. Terima kasih, kalian adalah teman yang terbaik.
13. Rekan-rekan SMKN 1 Mojokerto atas semua dukungan dan doanya, sehingga penulis dapat menyelesaikan studi S-2.
14. Kepada semua pihak yang telah membantu, mendoakan, memberikan motivasi dan dorongan yang tidak dapat saya sebutkan satu persatu.

Penulis menyadari bahwa penulisan tesis ini masih sangat jauh dari sempurna. Oleh sebab itu, penulis sangat mengharapkan kritik dan saran agar tesis ini menjadi lebih baik. Akhir kata, penulis berharap tesis ini memberikan manfaat bagi kita semua terutama untuk pengembangan ilmu pengetahuan.

Surabaya, Januari 2016

Mohammad Khoiron

ANALISA *MICROBLOGGING* UNTUK MENENTUKAN PRIORITAS KEBIJAKAN PUBLIK BERDASARKAN OPINI MASYARAKAT MENGGUNAKAN ALGORITMA *NAIVE* *BAYES* DAN *ANALYTICAL HIERARCHY PROCESS*

Nama Mahasiswa : Mohammad Khoiron
NRP : 2214206708
Pembimbing 1 : Dr. Surya Sumpeno, S.T., M.Sc.
Pembimbing 2 : Dr. Adhi Dharma Wibawa, S.T., M.T.

ABSTRAK

Kondisi ketidakpuasan masyarakat terhadap kebijakan publik yang dibuat pemerintah akan menimbulkan sikap apatis, kontraproduktif, bahkan destruktif. Melihat permasalahan tersebut, perlu dicari solusi yang cepat dan murah untuk mendapatkan data mengenai harapan masyarakat terhadap sebuah kebijakan publik, melalui analisa *microblogging* dalam kurun waktu tertentu.

Kebijakan publik yang dijadikan simulasi untuk mengambil prioritas kebijakan dalam penelitian ini adalah MDGs (*Millenium Development Goals*). *Tweet harvesting* dilakukan selama periode bulan Juni-Juli 2015 menggunakan enambelas kata kunci yang berkaitan dengan MDGs, dengan hasil 282124 *tweet*. Sebagai data latih diambil sebanyak 3200 *tweet* berdasarkan prosentase perolehan *tweet* dari masing-masing kata kunci dan dilabeli secara manual sebagai *tweet* negatif, *tweet* positif, *tweet* netral dan *tweet* tanya.

Analisa terhadap data latih memanfaatkan perangkat lunak Weka, dengan menggunakan algoritma *Naïve Bayes* menghasilkan akurasi klasifikasi terbaik sebesar 72,1 % dengan kondisi pembobotan yang digunakan adalah TF-IDF , minimal *term frequency*=3, tanpa melakukan proses *stopword*, tokenizer yang dipakai adalah *word tokenizer* dengan *procentage split* sebesar 66% dan *stemmer* yang digunakan adalah Lovins Stemmer. Hasil dari proses pembelajaran digunakan untuk melakukan prediksi terhadap 278924 data *tweet* yang tidak menjadi data latih.

Hasil klasifikasi digunakan untuk menentukan prioritas kebijakan publik menggunakan algoritma *Analytical Hierarchy Process* (AHP) dengan tahapan pembentukan hirarki dilanjutkan pembentukan matrik perbandingan berpasangan yang mempertimbangkan tingkat kepentingan *tweet* negatif, *tweet* positif, *tweet* netral, *tweet* tanya, *direct tweet* dan *retweet*. Berikutnya dilakukan proses normalisasi matrik perbandingan berpasangan, pemeriksaan rasio konsistensi dan evaluasi bobot keseluruhan. Analisa tersebut menempatkan memberantas kemiskinan dan kelaparan ekstrem menjadi prioritas kebijakan publik yang utama.

Kata Kunci : Kebijakan publik, Analisa sentimen, Klasifikasi, *Naïve bayes*, *Analytical Hierarchy Process*

MICROBLOGGING ANALYSIS FOR DETERMINING THE PRIORITY OF PUBLIC POLICY BASED ON PUBLIC OPINION USING NAIVE BAYES AND ANALYTICAL HIERARCHY PROCESS ALGORITHM

By : Mohammad Khoiron
Student Identity Number : 2214206708
Supervisor : Dr. Surya Sumpeno, S.T., M.Sc.
Co-Supervisor : Dr. Adhi Dharma Wibawa, S.T., M.T.

ABSTRACT

Dissatisfaction conditions against public policy made by the government would lead to apathy, counterproductive, even destructive. Seeing these problems, need to find the quick and cheap solution to obtain data about people's expectations towards a public policy, through the microblogging analysis within a certain time

Public policy that is used as a simulation to take policy priorities in this research are the MDGs (Millennium Development Goals). Tweet harvesting is done during the period June - July 2015 using sixteen keywords related to the MDGs, with the result 282124 tweets. For the training data 3200 tweet data is taken, based on the percentage of each keyword and manually labeled in a negative tweet, tweet positive tweet, neutral tweet and a question tweet..

Analysis of training data utilizing Weka software, using Naïve Bayes algorithm produces the best classification accuracy of 72.1 % with the condition of the weighting used is the TF - IDF, the minimum term frequency = 3 , without the stopword , tokenizer used is = word procentage tokenizer with a split of 66% and is used Lovins stemmer stemmer . Results of the learning process is used to make predictions on the data 278924 tweets that do not become training data.

Classification results are used to determine the priority of public policy using Analytical Hierarchy Process (AHP) algorithms with a sequence of steps, the establishment of the hierarchy, the formation of pairwise comparison matrix taking into consideration the interests of a negative tweet, positive tweet, neutral tweet, question tweet, direct tweet, and retweet. Next is pairwise comparison matrix normalization process, consistency ratio checking and overall weight evaluation. The analysis puts eradicate extreme poverty and hunger become a major public policy priority.

Keywords: Public Policy, Sentiment Analysis, Classification, Naïve Bayes, Analytical Hierarchy Process

DAFTAR ISI

| | |
|---|----------|
| Halaman Sampul | i |
| Pernyataan Keaslian Tesis | ii |
| Lembar Pengesahan | iii |
| Abstrak | iv |
| Abstract | v |
| Kata Pengantar | vi |
| Daftar Isi | vii |
| Daftar Gambar | x |
| Daftar Tabel | xi |
| Daftar Notasi | xii |
| Bab I. Pendahuluan | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Masalah | 2 |
| 1.3 Tujuan | 2 |
| 1.4 Manfaat | 3 |
| 1.5 Metode | 3 |
| Bab II. Kajian Pustaka | 5 |
| 2.1 Kebijakan Publik | 5 |
| 2.2 Text Mining | 5 |
| 2.3 Text Preprocessing | 6 |
| 2.4 Feature Selection | 6 |
| 2.5 Stop Word | 7 |
| 2.6 Stemming | 9 |
| 2.7 Stemming Bahasa Indonesia | 10 |
| 2.8 N-Gram | 19 |
| 2.9 Pembobotan Kata | 20 |
| 2.10 Analisa Sentimen | 23 |
| 2.11 Microblogging Twitter | 24 |
| 2.12 Algoritma Naïve Bayes | 26 |
| 2.13 Analytical Hierarchy Process (AHP) | 28 |

| | |
|---|----|
| 2.14 Perangkat Lunak Weka | 32 |
| 2.15 Penelitian Sebelumnya | 34 |
| Bab III. Metodologi | 37 |
| 3.1 Diagram Alir Metodologi Penelitian..... | 37 |
| 3.2 Penentuan Kebijakan Publik yang Akan Dianalisa..... | 37 |
| 3.3 Penentuan Kata Kunci..... | 38 |
| 3.4 Penentuan Kata Negatif dan Positif | 39 |
| 3.5 Pengumpulan Data | 40 |
| 3.6 Pra Proses | 40 |
| 3.7 Seleksi Fitur | 41 |
| 3.8 Pembelajaran dan Klasifikasi..... | 41 |
| 3.9 Validasi dan Evaluasi..... | 41 |
| 3.10 Penentuan Prioritas Kebijakan..... | 42 |
| Bab IV. Hasil dan Pembahasan | 43 |
| 4.1 Pengumpulan Data | 43 |
| 4.2 Data Training | 45 |
| 4.3 Pra Proses..... | 47 |
| 4.4 Seleksi Fitur | 48 |
| 4.5 Pembelajaran dan Klasifikasi | 50 |
| 4.6 Penentuan Prioritas Kebijakan Publik..... | 60 |
| Bab V. Kesimpulan | 67 |
| Daftar Pustaka | |

DAFTAR TABEL

| | |
|--|----|
| Tabel 2.1 Kombinasi Awalan Akhiran yang Tidak Dijinkan | 8 |
| Tabel 2.2 Cara Menentukan Tipe Awalan untuk Awalam “te-“ | 14 |
| Tabel 2.3 Jenis Awalan Berdasarkan Tipe Awalannya | 14 |
| Tabel 2.4 Contoh Pemotongan N-Gram Berbasis Karakter | 19 |
| Tabel 2.5 Contoh Pemotongan N-Gram Berbasis Kata | 20 |
| Tabel 2.6 Skala Perbandingan AHP | 30 |
| Tabel 2.7 Indeks Konsistensi | 31 |
| Tabel 2.8 Penelitian Sebelumnya | 35 |
| Tabel 3.1 Kata Kunci | 35 |
| Tabel 3.2 Kata Positif, Negatif dan <i>Emoticon</i> | 40 |
| Tabel 4.1 Perolehan Hasil <i>Tweet Harvesting</i> | 45 |
| Tabel 4.2 Contoh Pelabelan <i>Tweet</i> Secara Manual | 46 |
| Tabel 4.3 Hasil Tahap Pra Proses Data Training | 47 |
| Tabel 4.4 Hasil <i>Stemming</i> dengan Menggunakan PHP <i>Library</i> Sastrawi | 48 |
| Tabel 4.5 Hasil Pembelajaran dengan Pembobotan TF | 53 |
| Tabel 4.6 Hasil Pembelajaran dengan Pembobotan IDF | 54 |
| Tabel 4.7 Hasil Pembelajaran dengan Pembobotan TF-IDF | 54 |
| Tabel 4.8 Hasil Klasifikasi dan <i>Filterisasi</i> | 54 |
| Tabel 4.9 Matrik Perbandingan Berpasangan | 63 |
| Tabel 4.10 Normalisasi Matrik Perbandingan Berpasangan | 64 |
| Tabel 4.11 Tabel Evaluasi Bobot Keseluruhan | 65 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2.1 Contoh <i>Stoptlist</i> dalam bahasa Indonesia..... | 8 |
| Gambar 2.2 Algoritma Nazief dan Andriani..... | 11 |
| Gambar 2.3 Pembentuk Hirarki | 11 |
| Gambar 2.4 Weka <i>GUI Chooser</i> | 32 |
| Gambar 2.5 Weka <i>GUI Explorer</i> | 32 |
| Gambar 2.6 Weka <i>GUI Explorer (Clasify)</i> | 34 |
| Gambar 3.1 Metodologi Penelitian | 37 |
| Gambar 4.1 Perangkat Lunak Tags V.6.0..... | 43 |
| Gambar 4.2 Contoh Hasil <i>Tweet Harvesting</i> | 44 |
| Gambar 4.3 <i>Script</i> PHP untuk Melakukan Tahap Praproses dan Seleksi Fitur ..49 | |
| Gambar 4.4 Hasil Praproses dan Seleksi Fitur..... | 49 |
| Gambar 4.5 Format data ARFF..... | 50 |
| Gambar 4.6 Konversi Data <i>Sring</i> ke <i>Word Vector</i> | 51 |
| Gambar 4.7 Beberapa Kombinasi pada Proses <i>Filtering</i> | 51 |
| Gambar 4.8 Contoh Hasil Klasifikasi pada Weka | 52 |
| Gambar 4.9 Format data ARFF untuk Proses Prediksi..... | 57 |
| Gambar 4.10 Proses Pengambilan <i>Data Set</i> pada Weka..... | 57 |
| Gambar 4.11 Proses Visualisasi Kesalahan Klasifikasi pada Weka..... | 58 |
| Gambar 4.12 Proses Penyimpanan Hasil Prediksi pada Weka | 58 |
| Gambar 4.13 Contoh Hasil Prediksi pada Weka..... | 59 |
| Gambar 4.14 Pembentuk hirarki | 61 |

DAFTAR NOTASI

| | | |
|--------------|---|---|
| W_{ij} | = | bobot term t_j terhadap dokumen d_i |
| T_{fij} | = | jumlah kemunculan term t_j di dalam dokumen |
| D | = | jumlah semua dokumen yang ada dalam database |
| df_j | = | jumlah dokumen yang mengandung term t_j (minimal ada satu kata |
| | | yaitu term t) |
| V_j | = | Kategori <i>tweet</i> $j=1,2,3,\dots,n$. Dimana dalam penelitian ini |
| | | j_1 = kategori <i>tweet</i> sentimen negatif, |
| | | j_2 = kategori <i>tweet</i> sentimen positif, |
| | | j_3 = kategori <i>tweet</i> sentiment netral dan |
| | | j_4 = kategori <i>tweet</i> sentiment tanya |
| $P(x_i V_j)$ | = | Probabilitas x_i pada kategori V_j |
| $P(V_j)$ | = | Probabilitas dari V_j |
| $ docs_j $ | = | jumlah dokumen setiap kategori j |
| $ contoh $ | = | jumlah dokumen dari semua kategori |
| n_k | = | jumlah frekuensi kemunculan setiap kata |
| n | = | jumlah frekuensi kemunculan kata dari setiap kategori |
| $ kosakata $ | = | jumlah semua kata dari semua kategori |

BAB I PENDAHULUAN

1.1 Latar Belakang

Salah satu ciri masyarakat modern yang demokratis adalah dilibatkannya masyarakat dalam mengambil suatu kebijakan publik. Keterlibatan tersebut dimulai sejak akan dibuat suatu kebijakan publik oleh pemerintah sampai kepada implementasinya. Keterlibatan masyarakat diperlukan karena kebijakan publik akan mempengaruhi kehidupan mereka sehari-hari. Oleh karena itu, sudah seharusnya pemerintah yang demokratis selalu melibatkan masyarakat dalam menentukan sebuah kebijakan publik.

Saat ini di Indonesia, masyarakat terlihat semakin bergairah dalam membahas sebuah kebijakan publik yang dihasilkan oleh pemerintah. Antusiasme semacam ini sangat positif sejauh memberikan perspektif lain bagi kepentingan masyarakat. Debat publik menandakan dinamika suatu masyarakat. Besarnya keterlibatan masyarakat tidak bisa dilepaskan dari era reformasi yang sampai saat ini masih terus bergulir dengan berbagai dinamika dan resikonya.

Salah satu media yang paling sering digunakan masyarakat untuk menyampaikan opini maupun pendapatnya adalah dengan menggunakan media sosial *microblogging*. Pada saat ini situs *microblogging* telah menjadi alat komunikasi yang sangat populer di kalangan pengguna internet. Dimana jutaan pesan yang muncul setiap hari di situs web populer yang menyediakan layanan *microblogging* seperti Twitter, Tumblr, dan Facebook.

Penulis pesan tersebut menulis tentang kehidupan mereka, berbagi opini tentang berbagai topik dan membahas isu-isu yang terjadi pada saat ini. Format pesan yang bebas dan aksesibilitas dari berbagai platform yang mudah, pengguna internet cenderung untuk beralih dari blog atau milis ke layanan *microblogging*. Hal tersebut menyebabkan semakin banyak pengguna yang melakukan posting tentang suatu produk dan layanan yang mereka gunakan, mengekspresikan pandangan mereka tentang politik dan agama maupun mengkritisi sebuah kebijakan publik.

Twitter sebagai salah satu situs *microblogging* dengan pengguna lebih dari 500 juta dan 400 juta tweet perhari, memungkinkan pengguna untuk berbagi pesan menggunakan teks pendek disebut *tweet*. Twitter dapat menjadi sumber data pendapat dan sentimen masyarakat. Data tersebut dapat digunakan secara efisien untuk pemasaran atau studi sosial.

Salah satu hal yang bisa dilakukan pada data *tweet* adalah melakukan *sentiment analysis* atau *opinion mining* yaitu studi komputasional dari opini-orang, sentimen dan emosi melalui entitas dan atribut yang dimiliki yang diekspresikan dalam bentuk teks. Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral.

Dalam tesis ini akan dibahas mengenai analisa sentimen *microblogging* twitter dengan menggunakan algoritma *naïve bayes* yang hasilnya dapat dimanfaatkan sebagai bahan pertimbangan untuk menentukan prioritas kebijakan publik dengan menggunakan algoritma *Analytical Hierarchy Process*, sehingga kualitas kebijakan yang dibuat mampu memenuhi harapan dan keinginan masyarakat.

1.2 Masalah

Permasalahan yang akan diangkat dalam tesis ini adalah, adanya kesulitan untuk memperoleh prioritas kebijakan publik yang mendekati aspirasi dan harapan masyarakat dengan memanfaatkan analisa terhadap *microblogging*.

1.3 Tujuan

Tujuan yang akan dicapai dalam tesis ini adalah, menghasilkan sebuah analisa dari *microblogging*, yang layak untuk menjadi bahan pertimbangan dan kajian untuk menentukan prioritas sebuah kebijakan publik berdasarkan aspirasi dan keinginan masyarakat.

1.4 Manfaat

Manfaat yang akan diperoleh dalam tesis ini adalah, meningkatnya kualitas kebijakan publik yang diambil oleh Pemerintah, sehingga dapat memenuhi ekspektasi dan harapan masyarakat yang berujung pada pelayanan prima.

1.5 Metode

Secara garis besar metode yang akan digunakan untuk menyelesaikan tesis ini dibagi menjadi beberapa tahap seperti di bawah ini:

1. Menentukan kebijakan publik yang akan dibuat prioritasnya.
2. Menentukan *keyword* yang bisa mewakili masing-masing kebijakan publik.
3. Proses pengumpulan data (*harvesting*) Twitter dengan menggunakan *streaming APIs* berdasarkan *keyword* yang telah ditentukan sebelumnya dalam kurun waktu tertentu dan menyimpannya dalam database.
4. Melakukan *preprocessing* terhadap data Twitter yang telah tersimpan di database, meliputi *cleansing*, *case folding* dan *parsing*
5. Melakukan pemilihan fitur (*feature selection*) terhadap data Twitter yang telah mengalami pra proses dengan melakukan *stopword removal* dan melakukan proses *stemming*.
6. Melakukan pembelajaran dan klasifikasi sentimen negatif, positif, netral maupun pertanyaan terhadap data opini masyarakat yang telah melalui tahap pra proses dan pemilihan fitur dengan menggunakan algoritma naïve bayes.
7. Melakukan validasi dan evaluasi terhadap hasil pembelajaran yang telah dilakukan.
8. Menentukan prioritas kebijakan publik berdasarkan sentimen opini publik yang telah dihasilkan dengan menggunakan algoritma *Analytical Hierarchy Process*.

BAB II **KAJIAN PUSTAKA**

2.1 Kebijakan Publik

Kebijakan Publik merupakan suatu aturan-aturan yang dibuat oleh pemerintah dan merupakan bagian dari keputusan politik untuk mengatasi berbagai persoalan dan isu-isu yang ada dan berkembang di masyarakat.

Kebijakan publik juga merupakan keputusan yang dibuat oleh pemerintah untuk melakukan pilihan tindakan tertentu untuk tidak melakukan sesuatu maupun untuk melakukan tindakan tertentu.

Dalam kehidupan masyarakat yang ada di wilayah hukum suatu negara sering terjadi berbagai permasalahan. Negara yang memegang penuh tanggung jawab pada kehidupan rakyatnya harus mampu menyelesaikan permasalahan-permasalahan tersebut. Kebijakan publik yang dibuat dan dikeluarkan oleh negara diharapkan dapat menjadi solusi akan permasalahan-permasalahan tersebut. Kebijakan Publik adalah suatu keputusan yang dimaksudkan untuk tujuan mengatasi permasalahan yang muncul dalam suatu kegiatan tertentu yang dilakukan oleh instansi pemerintah dalam rangka penyelenggaraan pemerintahan (Mustopadidjaja, 2002).

2.2 Text Mining

Text mining (penambangan teks) adalah penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui sebelumnya atau menemukan kembali informasi yang tersirat secara implisit, yang berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda-beda (Feldman & Sanger, 2007). Text mining merupakan teknik yang digunakan untuk menangani masalah klasifikasi, clustering, information extraction dan information retrieval (Berry & Kogan, 2010).

Pada dasarnya proses kerja dari text mining banyak mengadopsi dari penelitian Data Mining namun yang menjadi perbedaan adalah pola

yang digunakan oleh text mining diambil dari sekumpulan bahasa alami yang tidak terstruktur sedangkan dalam Data Mining pola yang diambil dari database yang terstruktur (Han & Kamber, 2006). Tahap-tahap text mining secara umum adalah text preprocessing dan feature selection (Feldman & Sanger 2007, Berry & Kogan 2010) .

2.3 Text Preprocessing

Tahap text preprocessing adalah tahap awal dari text mining. Tahap ini mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi knowledge discovery sistem text mining (Feldman & Sanger, 2007). Tindakan yang dilakukan pada tahap ini adalah *toLowerCase*, yaitu mengubah semua karakter huruf menjadi huruf kecil dan *Tokenizing* yaitu proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan delimiter- delimiter seperti tanda titik (.), koma (,), spasi dan karakter angka yang ada pada kata tersebut (Weiss et al, 2005).

2.4 Feature Selection

Tahap seleksi fitur (*feature selection*) bertujuan untuk mengurangi dimensi dari suatu kumpulan teks, atau dengan kata lain menghapus kata-kata yang dianggap tidak penting atau tidak menggambarkan isi dokumen sehingga proses pengklasifikasian lebih efektif dan akurat (Do et al, 2006., Feldman & Sanger, 2007., Berry & Kogan 2010). Pada tahap ini tindakan yang dilakukan adalah menghilangkan *stopword* (*stopword removal*) dan *stemming* terhadap kata yang berimbuhan (Berry & Kogan 2010., Feldman & Sanger 2007).

Stopword adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen (Dragut et al. 2009). Misalnya “di”, “oleh”, “pada”, “sebuah”, “karena” dan lain sebagainya. Sebelum proses *stopword removal* dilakukan, harus dibuat daftar *stopword* (stoplist). Jika termasuk di dalam *stoplist* maka kata-kata tersebut akan dihapus dari deskripsi sehingga kata-kata yang tersisa di dalam deskripsi dianggap sebagai kata-kata yang

mencirikan isi dari suatu dokumen atau keywords. Daftar kata stopword di penelitian ini bersumber dari Tala (2003).

Setelah melalui proses *stopword removal* tindakan selanjutnya adalah yaitu proses *stemming*. *Stemming* adalah proses pemetaan dan penguraian berbagai bentuk (variants) dari suatu kata menjadi bentuk kata dasarnya (stem) (Tala, 2003). Tujuan dari proses *stemming* adalah menghilangkan imbuhan-imbuhan baik itu berupa prefiks, sufiks, maupun konfiks yang ada pada setiap kata. Jika imbuhan tersebut tidak dihilangkan maka setiap satu kata dasar akan disimpan dengan berbagai macam bentuk yang berbeda sesuai dengan imbuhan yang melekatinya sehingga hal tersebut akan menambah beban database. Hal ini sangat berbeda jika menghilangkan imbuhan-imbuhan yang melekat dari setiap kata dasar, maka satu kata dasar akan disimpan sekali walaupun mungkin kata dasar tersebut pada sumber data sudah berubah dari bentuk aslinya dan mendapatkan berbagai macam imbuhan. Karena bahasa Indonesia mempunyai aturan morfologi maka proses stemming harus berdasarkan aturan morfologi bahasa Indonesia.

Berdasarkan penelitian sebelumnya, ada beberapa algoritma stemming yang bisa digunakan untuk stemming bahasa Indonesia diantaranya algoritma confix- stripping, algoritma Porter stemmer bahasa Indonesia, algoritma Arifin dan Sutiono, dan Algoritma Idris (Tala 2003, Agusta 2009, Asian et al 2005, Adriani et al 2007). Dimana, Algoritma confix-stripping adalah algoritma yang akurat dalam stemming bahasa Indonesia (Tala 2003, Agusta 2009, Asian et al 2005, Adriani et al 2007).

2.5 Stop Word

Kebanyakan bahasa resmi di berbagai negara memiliki kata fungsi dan kata sambung seperti artikel dan preposisi yang hampir selalu muncul pada dokumen teks. Biasanya kata-kata ini tidak memiliki arti yang lebih di dalam memenuhi kebutuhan seorang searcher di dalam mencari informasi. Kata-kata tersebut (misalnya a, an, the on pada bahasa Inggris) disebut sebagai *stopwords*.

Sebuah sistem *text retrieval* biasanya disertai dengan sebuah *stoplist*. *Stoplist* berisi sekumpulan kata yang 'tidak relevan', namun sering sekali muncul dalam sebuah dokumen. Dengan kata lain *Stoplist* berisi sekumpulan *Stopwords*.

Stopwords removal adalah sebuah proses untuk menghilangkan kata yang 'tidak relevan' pada hasil parsing sebuah dokumen teks dengan cara membandingkannya dengan *stoplist* yang ada

| Word | Root | Part of Speech | Word | Root | Part of Speech |
|-------------|---------|----------------|----------------|----------|----------------|
| ada | ada | verb | lah | lah | particle |
| adanya | ada | noun | lain | lain | adjective |
| adalah | adalah | verb | lainnya | lain | adjective |
| adapun | adapun | particle | melainkan | lain | verb |
| agak | agak | adverb | selaku | laku | particle |
| agaknyanya | agak | adverb | lalu | lalu | verb |
| agar | agar | particle | melalui | lalu | verb |
| akan | akan | particle | terlalu | lalu | adverb |
| akankah | akan | particle | lama | lama | adjective |
| akhirnya | akhir | noun | lamanya | lama | noun |
| aku | aku | pronomia | selama | lama | noun |
| akulah | aku | pronomia | selama-lamanya | lama | adjective |
| amat | amat | adverb | selamanya | lama | adjective |
| amatlah | amat | adverb | lebih | lebih | adjective |
| anda | anda | noun | terlebih | lebih | adverb |
| andalah | anda | noun | bermacam | macam | adjective |
| antar | antar | particle | bermacam-macam | macam | adjective |
| diantaranya | antar | verb | macam | macam | noun |
| antara | antara | noun | semacam | macam | adverb |
| antaranya | antara | particle | maka | maka | particle |
| diantara | antara | verb | makanya | maka | particle |
| apa | apa | pronomia | makin | makin | adverb |
| apaan | apa | pronomia | malah | malah | adverb |
| mengapa | apa | pronomia | malahan | malah | adverb |
| apabila | apabila | particle | mampu | mampu | adjective |
| apakah | apakah | pronomia | mampukah | mampu | adjective |
| apalagi | apalagi | pronomia | mana | mana | pronoun |
| apatah | apatah | pronomia | manakala | manakala | particle |
| atau | atau | particle | manalagi | manalagi | particle |
| ataukah | atau | particle | masih | masih | adverb |
| ataupun | atau | particle | masihkah | masih | adverb |
| bagai | bagai | noun | semasih | masih | adverb |
| bagaimana | bagai | particle | masing | masing | pronomia |

continue to next page

Gambar 2.1 Contoh *Stoplist* dalam Bahasa Indonesia (Tala, 2003)

2.6 Stemming

Stemming merupakan suatu proses yang terdapat dalam sistem IR yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (root word) dengan menggunakan aturan-aturan tertentu. Sebagai contoh, kata bersama, kebersamaan, menyamai, akan distem ke root word-nya yaitu “sama”. Proses stemming pada teks berbahasa Indonesia berbeda dengan stemming pada teks berbahasa Inggris. Pada teks berbahasa Inggris, proses yang diperlukan hanya proses menghilangkan sufiks. Sedangkan pada teks berbahasa Indonesia, selain sufiks, prefiks, dan konfiks juga dihilangkan.

Metode stemming memerlukan input berupa term yang terdapat dalam dokumen. Sedangkan outputnya berupa stem. Ada tiga jenis metode stemming, antara lain :

1. **Successor Variety (SV)** : lebih mengutamakan penyusunan huruf dalam kata dibandingkan dengan pertimbangan atas fonem. Contoh untuk kata-kata : *corpus*, *able*, *axle*, *accident*, *ape*, *about* menghasilkan SV untuk kata *apple*:
 - a. Karena huruf pertama dari kata “*apple*” adalah “a”, maka kumpulan kata yang ada substring “a” diikuti “b”, “x”, “c”, “p” disebut SV dari “a” sehingga “a” memiliki 4 SV.
 - b. Karena dua huruf pertama dari kata “*apple*” adalah “ap”, maka kumpulan kata yang ada substring “ap” hanya diikuti “e” disebut SV dari “ap” sehingga “ap” memiliki 1 SV.
2. **N-Gram Conflation** : ide dasarnya adalah pengelompokan kata-kata secara bersama berdasarkan karakter-karakter (substring) yang teridentifikasi sepanjang N karakter.
3. **Affix Removal** (penghilangan imbuhan) : membuang prefix (awalan) dan suffix (akhiran) dari term menjadi suatu stem. Yang paling sering digunakan adalah algoritma Porter Stemmer karena modelnya sederhana dan efisien.

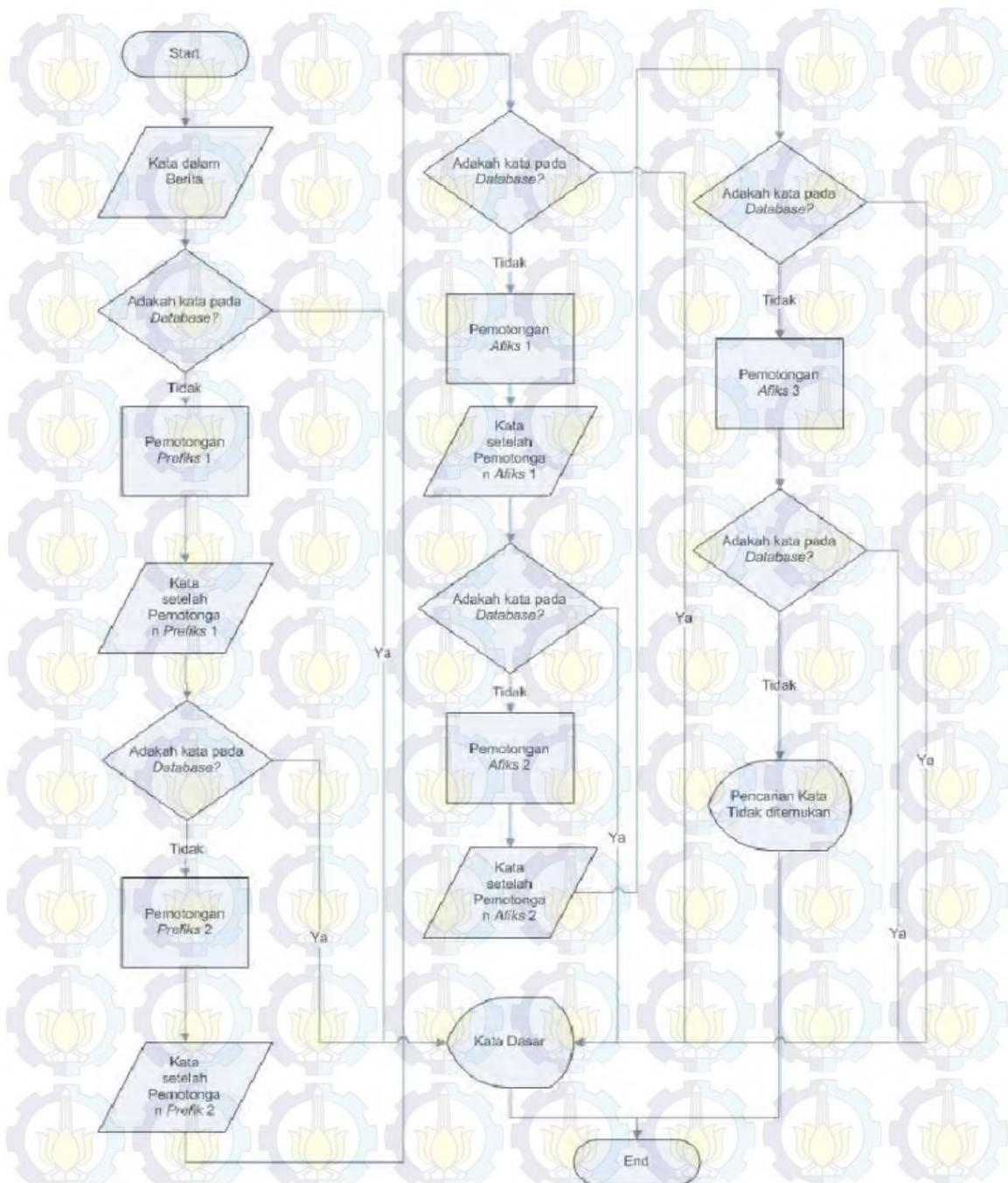
- a. Jika suatu kata diakhiri dengan “ies” tetapi bukan “eies” atau “aies”, maka “ies” di-replace dengan “y”
- b. Jika suatu kata diakhiri dengan “es” tetapi bukan “aes” atau “ees” atau “oes”, maka “es” di-replace dengan “e”
- c. Jika suatu kata diakhiri dengan “s” tetapi bukan “us” atau “ss”, maka “s” di-replace dengan “NULL”

2.7 Stemming Bahasa Indonesia dengan Algoritma Nazief dan Andriani

Stemming adalah salah satu cara yang digunakan untuk meningkatkan performa IR (*information retrieval*) dengan cara mentransformasi kata-kata dalam sebuah dokumen teks ke bentuk kata dasarnya. Algoritma *stemming* untuk bahasa yang satu berbeda dengan algoritma *stemming* untuk bahasa lainnya. Sebagai contoh bahasa Inggris memiliki morfologi yang berbeda dengan bahasa Indonesia sehingga algoritma *stemming* untuk kedua bahasa tersebut juga berbeda. Proses *stemming* pada teks berbahasa Indonesia lebih rumit/kompleks karena terdapat variasi imbuhan yang harus dibuang untuk mendapatkan *root word* (kata dasar) dari sebuah kata. Pada umumnya kata dasar pada bahasa Indonesia terdiri dari kombinasi:

Prefiks 1 + Prefiks 2 + Kata dasar + Sufiks 3 + Sufiks 2 + Sufiks 1

Dari kombinasi tersebut dapat dibuat sebuah flowchart seperti tampak pada gambar 1.



Gambar 2.2 Algoritma Nazief dan Andriani (Nazief dan Andriani, 1996)

Algoritma Nazief & Adriani yang dibuat oleh Bobby Nazief dan Mirna Adriani ini memiliki tahap-tahap sebagai berikut:

1. Pertama cari kata yang akan distem dalam kamus kata dasar. Jika ditemukan maka diasumsikan kata adalah root word. Maka algoritma berhenti.
2. Inflection Suffixes (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa particles (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus Possesive Pronouns (“-ku”, “-mu”, atau “-nya”), jika ada.
3. Hapus Derivation Suffixes (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti. Jika tidak maka ke langkah 3a
 - a) Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b.
 - b) Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
4. Hapus Derivation Prefix. Jika pada langkah 3 ada sufiks yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b.
 - a) Periksa tabel kombinasi awalan-akhirian yang tidak diijinkan. Jika ditemukan maka algoritma berhenti, jika tidak
 - b) pergi ke langkah 4b.
 - c) For $i = 1$ to 3, tentukan tipe awalan kemudian hapus awalan. Jika root word belum juga ditemukan lakukan langkah 5, jika sudah maka algoritma berhenti. Catatan: jika awalan kedua sama dengan awalan pertama algoritma berhenti.
5. Melakukan Recoding.
6. Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai root word. Proses selesai.

Tipe awalan ditentukan melalui langkah-langkah berikut:

1. Jika awalannya adalah “di-”, “ke-”, atau “se-” maka tipe awalannya secara berturut-turut adalah “di-”, “ke-”, atau “se-”.
2. Jika awalannya adalah “te-”, “me-”, “be-”, atau “pe-” maka dibutuhkan sebuah proses tambahan untuk menentukan tipe awalannya.
3. Jika dua karakter pertama bukan “di-”, “ke-”, “se-”, “te-”, “be-”, “me-”, atau “pe-” maka berhenti.
4. Jika tipe awalan adalah “none” maka berhenti. Jika tipe awalan adalah bukan “none” maka awalan dapat dilihat pada Tabel 2. Hapus awalan jika ditemukan.

Tabel 2.1 Kombinasi awalan akhiran yang tidak diijinkan

| Awalan | Akhiran yang tidak diijinkan |
|--------|------------------------------|
| be- | -i |
| di- | -an |
| ke- | -i, -kan |
| me- | -an |
| se- | -i, -kan |

(Sumber : L agusta, 2009)

Tabel 2.2 Cara Menentukan Tipe Awalan Untuk awalan “te-”

| Following Characters | | | | Tipe Awalan |
|-------------------------|-------------------------|------------|-----------|-------------|
| Set 1 | Set 2 | Set 3 | Set 4 | |
| “-r-“ | “-r-“ | - | - | none |
| “-r-“ | “-r-“ | - | - | Ter-luluh |
| “-r-“ | not (vowel or “-r-”) | “-er-“ | vowel | ter |
| “-r-“ | not (vowel or “-r-”) | “-er-“ | not vowel | ter- |
| “-r-“ | not (vowel or “-r-”) | not “-er-“ | - | ter |
| not (vowel or “-r-”) | “-er-“ | vowel | - | none |
| not (vowel or “-r-”) | “-er-“ | not vowel | - | te |

(sumber : L agusta, 2009)

Tabel 2.3. Jenis Awalan Berdasarkan Tipe Awalannya

| Tipe Awalan | Awalan yang harus dihapus |
|-------------|---------------------------|
| di- | di- |
| ke- | ke- |
| se- | se- |
| te- | te- |
| ter- | ter- |
| Ter-luluh | ter |

(sumber : L agusta, 2009)

Untuk mengatasi keterbatasan pada algoritma di atas, maka ditambahkan aturan-aturan dibawah ini:

1. Aturan untuk reduplikasi.
 - a) Jika kedua kata yang dihubungkan oleh kata penghubung adalah kata yang sama maka root word adalah bentuk tunggalnya, contoh : “buku-buku” root word-nya adalah “buku”.
 - b) Kata lain, misalnya “bolak-balik”, “berbalas-balasan, dan ”seolah-olah”. Untuk mendapatkan root word-nya, kedua kata diartikan secara terpisah. Jika keduanya memiliki root word yang sama maka diubah menjadi bentuk tunggal, contoh: kata “berbalas-balasan”, “berbalas” dan “balasan” memiliki root word yang sama yaitu “balas”, maka root word “berbalas-balasan” adalah “balas”. Sebaliknya, pada kata “bolak-balik”, “bolak” dan “balik” memiliki root word yang berbeda, maka root word-nya adalah “bolak-balik”.
2. Tambahan bentuk awalan dan akhiran serta aturannya.
 - a) Untuk tipe awalan “mem-“, kata yang diawali dengan awalan “memp-” memiliki tipe awalan “mem-”.
 - b) Tipe awalan “meng-“, kata yang diawali dengan awalan “mengk-” memiliki tipe awalan “meng-”.

Berikut contoh-contoh aturan yang terdapat pada awalan sebagai pembentuk kata dasar.

1. Awalan SE-

Se + semua konsonan dan vokal tetap tidak berubah

Contoh :

Se + bungkus = sebungkus

Se + nasib = senasib

Se + arah = searah

Se + ekor = seekor

2. Awalan ME-

Me + vokal (a,i,u,e,o) menjadi sengau “meng”

Contoh :

Me + inap = menginap

Me + asuh = mengasuh

Me + ubah = mengubah

Me + ekor = mengekor

Me + oplos = mengoplos

Me + konsonan b menjadi “mem”

Contoh :

Me + beri = member

Me + besuk = membesuk

Me + konsonan c menjadi “men”

Contoh :

Me + cinta = mencinta

Me + cuci = mencuci

Me + konsonan d menjadi “men”

Contoh :

Me + didik = mendidik

Me + dengkur = mendengkur

Me + konsonan g dan h menjadi “meng”

Contoh :

Me + gosok = menggosok

Me + hukum = menghukum

Me + konsonan j menjadi “men”

Contoh :

Me + jepit = menjepit

Me + jemput = menjemput

Me + konsonan k menjadi “meng” (luluh)

Contoh :

Me + kukus = mengukus

Me + kupas = mengupas

Me + konsonan p menjadi “mem” (luluh)

Contoh :

Me + pesona = mempesona

Me + pukul = memukul

Me + konsonan s menjadi “meny” (luluh)

Contoh :

Me + sapu = menyapu

Me + satu = menyatu

Me + konsonan t menjadi “men” (luluh)

Contoh :

Me + tanama = menanam

Me + tukar = menukar

Me + konsonan (l,m,n,r,w) menjadi tetap “me”

Contoh :

Me + lempar = melempar

Me + masak = memasak

Me + naik = menaik

Me + rawat = merawat

Me + warna = mewarna

3. Awalan KE-

Ke + semua konsonan dan vokal tetap tidak berubah

Contoh :

Ke + bawa = kebawa

Ke + atas = keatas

4. Awalan PE-

Pe + konsonan (h,g,k) dan vokal menjadi "per"

Contoh :

Pe + hitung + an = perhitungan

Pe + gelar + an = pergelaran

Pe + kantor + = perkantoran

Pe + konsonan "t" menjadi "pen" (luluh)

Contoh :

Pe + tukar = penukar

Pe + tikam = penikam

Pe + konsonan (j,d,c,z) menjadi "pen"

Contoh :

Pe + jahit = penjahit

Pe + didik = pendidik

Pe + cuci = pencuci

Pe + zina = penzina

Pe + konsonan (b,f,v) menjadi "pem"

Contoh :

Pe + beri = pemberi

Pe + bunuh = pembunuh

Pe + konsonan "p" menjadi "pem" (luluh)

Contoh :

Pe + pikir = pemikir

Pe + potong = pemotong

Pe + konsonan “s” menjadi “peny” (luluh)

Contoh :

Pe + siram = penyiram

Pe + sabar = penyabar

Pe + konsonan (l,m,n,r,w,y) tetap tidak berubah

Contoh :

Pe + lamar = pelamar

Pe + makan = pemakan

Pe + nanti = penanti

Pe + wangi = pewangi

2.8 N-Gram

N-gram adalah potongan n karakter dalam suatu string tertentu atau potongan n kata dalam suatu kalimat tertentu (Cavnar & Trenkle, 1994).

Misalnya dalam kata “Teknik” akan didapatkan n-gram sebagai berikut.

Tabel 2.4. Contoh pemotongan N-gram berbasis karakter

| Nama | n-gram karakter |
|-----------|---------------------------------------|
| Uni-gram | T, E, K, N, I, K |
| Bi-gram | _T, TE, EK, KN, NI, IK, K_ |
| Tri-gram | _TE, TEK, EKN, KNI, NIK, IK_, K__ |
| Quad-gram | _TEK, TEKN, EKNI, KNIK, NIK_, IK__, K |

Karakter blank “_” digunakan untuk merepresentasikan spasi di depan dan diakhir kata. Dan untuk word-based n-gram contohnya adalah sebagai berikut.

Kalimat : “N-gram adalah potongan n karakter dalam suatu string tertentu”

Tabel 2.5. Contoh pemotongan N-gram berbasis kata

| Nama | n-gram kata |
|----------|---|
| Uni-gram | n-gram, adalah, potongan, n, karakter, dalam, suatu, string, tertentu |
| Bi-gram | n-gram adalah, adalah potongan, potongan n, n karakter, karakter dalam, dalam suatu, suatu string, string tertentu |
| Tri-gram | n-gram adalah potongan, adalah potongan n, potongan n karakter, n karakter dalam, karakter dalam suatu, dalam suatu string, suatu string tertentu |
| Dst.. | |

2.9 Pembobotan Kata (*term weighting*)

Sistem Temu Kembali Informasi berhadapan dengan pencarian informasi yang sesuai dengan query pengguna dari koleksi dokumen. Koleksi dokumen tersebut terdiri dari dokumen-dokumen yang beragam panjangnya dengan kandungan term yang berbeda pula. Hal yang perlu diperhatikan dalam pencarian informasi dari koleksi dokumen yang heterogen adalah pembobotan term. Term dapat berupa kata, frase atau unit hasil indexing lainnya dalam suatu dokumen yang dapat digunakan untuk mengetahui konteks dari dokumen tersebut. Karena setiap kata memiliki tingkat kepentingan yang berbeda dalam dokumen, maka untuk setiap kata tersebut diberikan sebuah indikator, yaitu *term weight*.

Term weighting atau pembobotan term sangat dipengaruhi oleh hal-hal berikut ini (Mandala, 2002):

1. *Term Frequency (tf) factor*, yaitu faktor yang menentukan bobot term pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut. Nilai jumlah kemunculan suatu kata (*term frequency*) diperhitungkan dalam pemberian bobot terhadap suatu kata. Semakin besar jumlah kemunculan suatu term (*tf* tinggi) dalam dokumen, semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuaian yang semakin besar.
2. *Inverse Document Frequency (idf) factor*, yaitu pengurangan dominansi term yang sering muncul di berbagai dokumen. Hal ini diperlukan karena term yang banyak muncul di berbagai dokumen, dapat dianggap sebagai term umum (*common term*) sehingga tidak penting nilainya. Sebaliknya faktor kejarangmunculan kata (*term scarcity*) dalam koleksi dokumen harus diperhatikan dalam pemberian bobot. Menurut Mandala 'Kata yang muncul pada sedikit dokumen harus dipandang sebagai kata yang lebih penting (*uncommon terms*) daripada kata yang muncul pada banyak dokumen. Pembobotan akan memperhitungkan faktor kebalikan frekuensi dokumen yang mengandung suatu kata (*inverse document frequency*). Hal ini merupakan usulan dari George Zipf. Zipf mengamati bahwa frekuensi dari sesuatu cenderung kebalikan secara proposional dengan urutannya.

Metode TF-IDF merupakan metode pembobotan term yang banyak digunakan sebagai metode pembandingan terhadap metode pembobotan baru. Pada metode ini, perhitungan bobot term t dalam sebuah dokumen dilakukan dengan mengalikan nilai Term Frequency dengan Inverse Document Frequency.

Pada Term Frequency (*tf*), terdapat beberapa jenis formula yang dapat digunakan yaitu (Mandala, 2002):

1. tf biner (*binary tf*), hanya memperhatikan apakah suatu kata ada atau tidak dalam dokumen, jika ada diberi nilai satu, jika tidak diberi nilai nol
2. tf murni (*raw tf*), nilai tf diberikan berdasarkan jumlah kemunculan suatu kata di dokumen. Contohnya, jika muncul lima kali maka kata tersebut akan bernilai lima.
3. tf logaritmik, hal ini untuk menghindari dominansi dokumen yang mengandung sedikit kata dalam query, namun mempunyai frekuensi yang tinggi.

$$tf = 1 + \log (tf) \quad (2.1)$$

4. tf normalisasi, menggunakan perbandingan antara frekuensi sebuah kata dengan jumlah keseluruhan kata pada dokumen.

$$tf = 0.5 + 0.5 \times [tf / \max tf] \quad (2.2)$$

Inverse Document Frequency (*idf*) dihitung dengan menggunakan formula

$$Idfj = \log (D / dfj) \quad (2.3)$$

dimana

D adalah jumlah semua dokumen dalam koleksi

dfj adalah jumlah dokumen yang mengandung term t_j

Menurut Defeng (dalam Robertson, 2004) jenis formula yang akan digunakan untuk perhitungan term frequency (tf) yaitu tf murni (*raw tf*).

Dengan demikian rumus umum untuk TF-IDF adalah penggabungan dari formula perhitungan raw tf dengan formula idf (rumus b.3) dengan cara mengalikan nilai term frequency (tf) dengan nilai inverse document frequency (idf) :

$$w_{ij} = tf_{ij} \times idf_j$$

$$w_{ij} = tf_{ij} \times \log (D/df_j) \quad (2.4)$$

Keterangan :

w_{ij} adalah bobot term t_j terhadap dokumen d_i

tf_{ij} adalah jumlah kemunculan term t_j di dalam dokumen d_i

D adalah jumlah semua dokumen yang ada dalam database

df_j adalah jumlah dokumen yang mengandung term t_j (minimal ada satu kata yaitu term t_j)

Berdasarkan rumus 2.4, berapapun besarnya nilai tf_{ij} , apabila $D = df_j$ maka akan didapatkan hasil 0 (nol) untuk perhitungan idf_j . Untuk itu dapat ditambahkan nilai 1 pada sisi idf_j , sehingga perhitungan bobotnya menjadi sebagai berikut:

$$w_{ij} = tf_{ij} \times (\log (D/df_j) + 1) \quad (2.5)$$

2.10 Analisis sentimen (sentiment analysis)

Sentiment analysis atau *opinion mining* mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *text mining* yang bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu (Liu, 2011).

Tugas dasar dalam analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral. *Sentiment analysis* juga dapat menyatakan perasaan emosional sedih, gembira, atau marah.

Kita dapat mencari pendapat tentang produk-produk, merek atau orang-orang dan menentukan apakah mereka dilihat positif atau negatif di web (Saraswati, 2011). Hal ini memungkinkan kita untuk mencari informasi tentang:

- a. Deteksi Flame (rants buruk)
- b. Persepsi produk baru.
- c. Persepsi Merek.
- d. Manajemen reputasi.

Ekspresi atau sentiment mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subyek yang berbeda. Oleh karena itu pada beberapa penelitian, terutama pada review produk, pekerjaan didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses opinion mining.

2.11 Microblogging Twitter

Twitter adalah sebuah situs web yang dimiliki dan dioperasikan oleh Twitter Inc, yang menawarkan jaringan sosial berupa mikroblog sehingga memungkinkan penggunanya untuk mengirim dan membaca pesan Tweets. Mikroblog adalah salah satu jenis alat komunikasi online dimana pengguna dapat memperbarui status tentang mereka yang sedang memikirkan dan melakukan sesuatu, apa pendapat mereka tentang suatu objek atau fenomena tertentu. Tweets adalah teks tulisan hingga 140 karakter yang ditampilkan pada halaman profil pengguna. Tweets bisa dilihat secara publik, namun pengirim dapat membatasi pengiriman pesan ke daftar teman-teman mereka saja. Pengguna dapat melihat Tweets pengguna lain yang dikenal dengan sebutan pengikut (*follower*).

Tidak seperti Facebook, LinkedIn, dan MySpace, Twitter merupakan sebuah jejaring sosial yang dapat digambarkan sebagai sebuah graph berarah (Wang, 2010), yang berarti bahwa pengguna dapat mengikuti pengguna lain, namun pengguna kedua tidak diperlukan untuk mengikutinya kembali. Kebanyakan akun berstatus publik dan dapat diikuti tanpa memerlukan persetujuan pemilik..

Semua pengguna dapat mengirim dan menerima Tweets melalui situs Twitter, aplikasi eksternal yang kompatibel (telepon seluler), atau dengan pesan singkat (SMS) yang tersedia di negara-negara tertentu. Pengguna dapat menulis pesan berdasarkan topik dengan menggunakan tanda # (hashtag). Sedangkan untuk menyebutkan atau membalas pesan dari pengguna lain bisa menggunakan tanda @.

Pesan pada awalnya diatur hanya mempunyai batasan sampai 140 karakter disesuaikan dengan kompatibilitas dengan pesan SMS, memperkenalkan singkatan notasi dan slang yang biasa digunakan dalam pesan SMS. Batas karakter 140 juga meningkatkan penggunaan layanan memperpendek URL seperti bit.ly, goo.gl, dan tr.im, dan jasa hosting konten, seperti Twitpic, Tweepphoto, memozu.com dan NotePub untuk mengakomodasi multimedia isi dan teks yang lebih panjang daripada 140 karakter. Twitter menggunakan bit.ly untuk memperpendek otomatis semua URL yang dikirim-tampil. Fitur yang terdapat dalam Twitter, antara lain:

1. Laman Utama (*Home*)

Pada halaman utama kita bisa melihat Tweets yang dikirimkan oleh orang-orang yang menjadi teman kita atau yang kita ikuti (*following*).

2. Profil (*Profile*)

Pada halaman ini yang akan dilihat oleh seluruh orang mengenai profil atau data diri serta Tweets yang sudah pernah kita buat.

3. Followers

Pengikut adalah pengguna lain yang ingin menjadikan kita sebagai teman. Bila pengguna lain menjadi pengikut akun seseorang, maka Tweets seseorang yang ia ikuti tersebut akan masuk ke dalam halaman utama.

4. Following

Kebalikan dari pengikut, following adalah akun seseorang yang mengikuti akun pengguna lain agar Tweets yang dikirim oleh orang yang diikuti tersebut masuk ke dalam halaman utama.

5. Mentions

Biasanya konten ini merupakan balasan dari percakapan agar sesama pengguna bisa langsung menandai orang yang akan diajak bicara.

6. Favorite

Tweets ditandai sebagai favorit agar tidak hilang oleh halaman sebelumnya.

7. Pesan Langsung (Direct Message)

Fungsi pesan langsung lebih bisa disebut SMS karena pengiriman pesan langsung di antara pengguna.

8. Hashtag

Hashtag “#” yang ditulis di depan topik tertentu agar pengguna lain bisa mencari topik yang sejenis yang ditulis oleh orang lain juga

9. List

Pengguna Twitter dapat mengelompokkan ikutan mereka ke dalam satu grup sehingga memudahkan untuk dapat melihat secara keseluruhan para nama pengguna (*username*) yang mereka ikuti (*follow*).

10. Topik Terkini (*Trending Topic*)

Topik yang sedang banyak dibicarakan banyak pengguna dalam suatu waktu yang bersamaan.

2.12 Algoritma Naïve Bayes

Algoritma naive bayes classifier merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman & Sanger 2007). Dalam penelitian ini yang menjadi data uji adalah dokumen Tweets. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

Dalam algoritma naïve bayes classifier setiap dokumen direpresentasikan dengan pasangan atribut “ $x_1, x_2, x_3, \dots, x_n$ ” dimana x_1 adalah kata pertama, x_2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori Tweet.

Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (VMAP), dimana persamaannya adalah sebagai berikut :

$$V_{MAP} = \underset{v_j \in V}{\arg \max} \frac{P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)} \quad (2.6)$$

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

$$V_{MAP} = \underset{v_j \in V}{\arg \max} P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j) \quad (2.7)$$

Persamaan diatas dapat disederhanakan menjadi sebagai berikut :

$$V_{MAP} = \underset{v_j \in V}{\arg \max} \prod_{i=1}^n P(x_i | V_j) P(V_j) \quad (2.8)$$

Keterangan :

V_j = Kategori *tweet* $j = 1, 2, 3, 4, \dots, n$. Dimana dalam penelitian ini

j_1 = kategori *tweet* sentimen negatif,

j_2 = kategori *tweet* sentimen positif

j_3 = kategori *tweet* sentiment netral, dan

j_4 = kategori *tweet* sentiment tanya

$P(x_i | V_j)$ = Probabilitas x_i pada kategori V_j

$P(V_j)$ = Probabilitas dari V_j

Untuk $P(V_j)$ dan $P(x_i|V_j)$ dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut :

$$P(V_j) = \frac{|docs\ j|}{|contoh|} \quad (2.9)$$

$$P(x_i|V_j) = \frac{n_k+1}{n + |kosakata|} \quad (2.10)$$

Keterangan :

$|docs\ j|$ = jumlah dokumen setiap kategori j

$|contoh|$ = jumlah dokumen dari semua kategori

n_k = jumlah frekuensi kemunculan setiap kata

n = jumlah frekuensi kemunculan kata dari setiap kategori

$|kosakata|$ = jumlah semua kata dari semua kategori

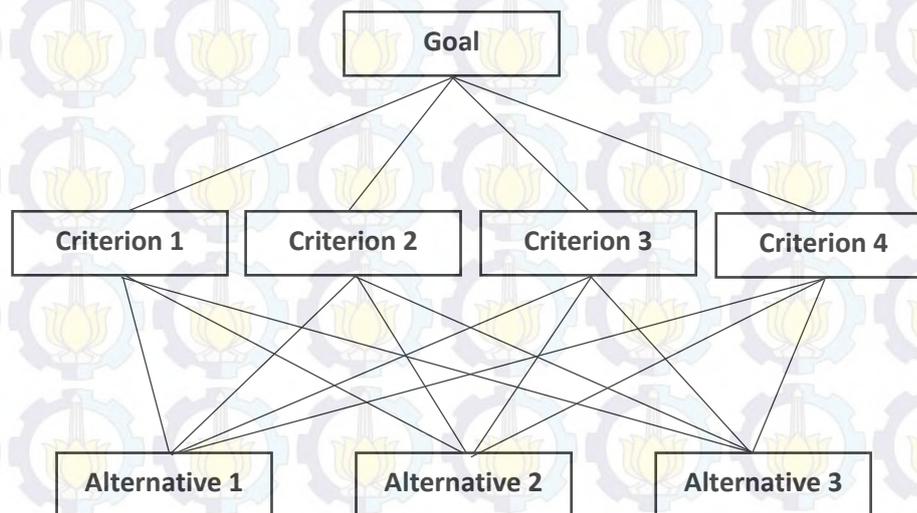
2.13 Analytical Hierarchy Process

AHP (Analytical Hierarchy Process) adalah merupakan suatu sistem pendukung keputusan yang mendekomposisikan suatu problem multi faktor yang kompleks kedalam suatu hirarki, di mana setiap levelnya dibentuk dari elemen-elemen yang spesifik. Peralatan utama AHP adalah sebuah hierarki fungsional dengan input utamanya persepsi manusia. Keberadaan hierarki memungkinkan dipecahnya masalah kompleks atau tidak terstruktur dalam sub-sub masalah, lalu menyusunnya menjadi suatu bentuk hierarki (Kusrini, 2007).

Pengambil keputusan dilibatkan untuk memberikan pertimbangan dalam penentuan kepentingan relatif dari faktor-faktor tersebut. Tujuan umum dari keputusan yang akan diambil terletak pada puncak hirarki, sedangkan kriteria, serta alternatif keputusan secara berurutan masing-masing berada pada level yang lebih rendah. Adapun tahapan AHP adalah sebagai berikut:

1. Pembentukan hirarki

Hirarki merupakan suatu pohon struktur yang dipergunakan untuk merepresentasikan penyebaran pengaruh mulai dari tujuan turun hingga sampai pada struktur yang terletak pada level yang paling dasar.



Gambar 2.3 Pembentuk hirarki (<http://leabhar.ie/literature-review-analytic-hierarchy-process>)

2. Perbandingan berpasangan

Langkah dalam AHP melibatkan estimasi prioritas bobot suatu himpunan kriteria atau alternatif dari suatu matriks bujursangkar yang digunakan dalam perbandingan berpasangan $A = [a_{ij}]$, yang mana nilai bobot ini haruslah positif dan jika kebijakan mengenai perbandingan berpasangan sudah benar-benar konsisten maka dibuat suatu perbandingan terbalik dari nilai tersebut, contoh : $a_{ij} = 1/a_{ji}$ untuk semua $i, j = 1, 2, 3, \dots, n$

Selanjutnya, bobot akhir dari faktor ke- i yang telah dinormalkan, yaitu w_i , adalah sebagai berikut :

$$w_{ij} = a_{ij} / \left(\sum_{i=1}^n a_{ij} \right) \quad \forall i = 1, 2, \dots, n \quad (2.12)$$

Skala perbandingan berpasangan untuk kepentingan relatif yaitu menilai secara perbandingan tingkat kepentingan suatu elemen dengan elemen lain. Skala perbandingan yang digunakan dalam AHP menurut Kusrini adalah:

Tabel 2.6. Skala perbandingan AHP

| Nilai | Keterangan |
|---------|---|
| 1 | Kriteria/alternatif A sama penting dengan Kriteria/Alternatif B |
| 3 | A sedikit lebih penting dari B |
| 5 | A jelas lebih penting dari B |
| 7 | A sangat jelas lebih penting dari B |
| 9 | A mutlak lebih penting dari B |
| 2,4,6,8 | Apabila ragu-ragu antara dua nilai yang berdekatan |

(Sumber: Kusrini, 2007)

Di sini w_i menunjukkan bobot dari elemen C_n sedangkan $a_{ij} = w_i / w_j$ adalah angka yang menunjukkan kekuatan dari C_i ketika diperbandingkan dengan C_j . Matriks dari angka-angka a_{ij} ini ditunjukkan oleh A , yaitu sebagai berikut:

$$A = [a_{ij}] = \begin{bmatrix} 1 & w_1/w_2 & \dots & w_1/w_n \\ w_2/w_1 & 1 & \dots & w_2/w_n \\ \vdots & \vdots & \dots & \vdots \\ w_n/w_1 & w_n/w_2 & \dots & 1 \end{bmatrix} = \begin{bmatrix} A^1 \\ \vdots \\ A^i \\ A^n \end{bmatrix} \quad (2.13)$$

3. Pemeriksaan konsistensi

Memeriksa apakah perbandingan berpasangan yang dilakukan berdasarkan kebijakan pengambil keputusan masih berada dalam batas yang ditentukan atau tidak. Pengukuran konsistensi secara alamiah atau

deviasi dari konsistensi disebut dengan indeks konsistensi (CI = Consistency Index) yang didefinisikan sebagai berikut :

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (2.14)$$

Indeks konsistensi dari suatu matriks perbandingan terbalik skala 1 sampai 9 yang di-generate secara random, dengan hasil perbandingan terbaliknya, untuk tiap ukuran matriks disebut dengan Indeks Random (RI = Random Index) yang ditunjukkan pada tabel berikut ini :

Tabel 2.7 Indeks konsistensi

| Matrix Ordo | 1,2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-------------|------|------|------|------|------|------|------|------|------|
| RI | 0.00 | 0.58 | 0.90 | 1.12 | 1.24 | 1.32 | 1.41 | 1.45 | 1.49 |

(Sumber: Kusrini, 2007)

Sehingga rasio konsistensi (CR = Consistency Ratio) didefinisikan sebagai rasio antara CI dan RI untuk ordo matriks yang sama

$$CR = CI / RI$$

CR < 0.1 maka kebijakan dapat diterima. Jika nilai CR lebih dari 0.01 maka pimpinan perlu meninjau kembali kebijakan yang diambil.

4. Evaluasi bobot keseluruhan

Bobot dari setiap kriteria yang telah didapatkan di kalikan dengan nilai kriteria pada masing-masing alternatif sehingga alternatif terbaik adalah alternatif yang memiliki prioritas tertinggi.

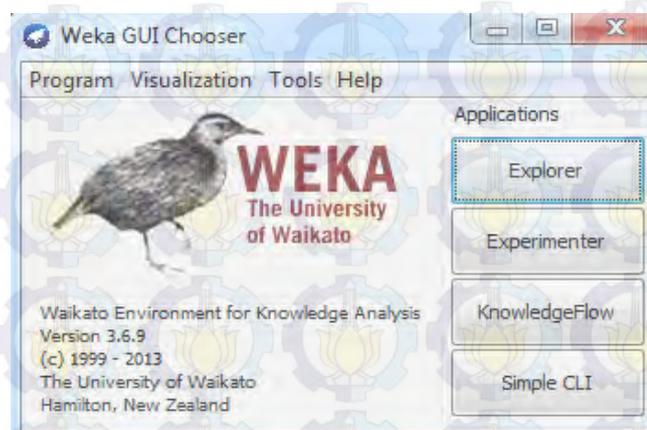
5. Pengambilan keputusan kelompok/penetapan kebijakan

Untuk menurunkan hasil kebijakan kelompok, tiap anggota kelompok membuat kebijakan-kebijakan sendiri pada copy model yang mereka miliki dan kemudian mengkombinasikan hasilnya

2.14 Perangkat Lunak Weka

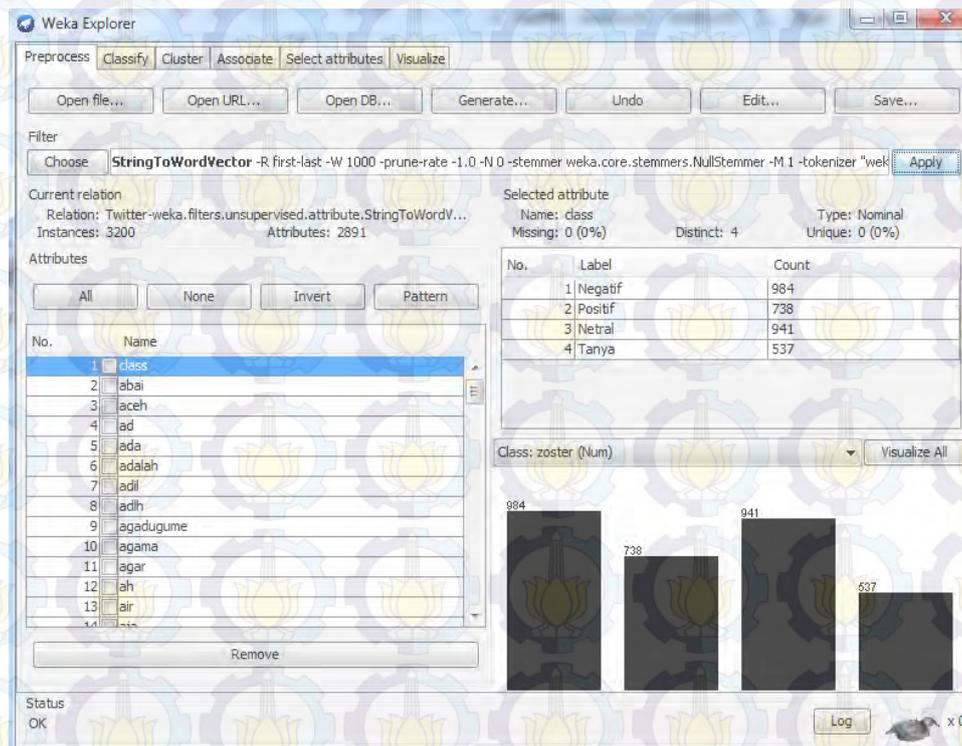
Weka (*Waikato Environment for Knowledge Analysis*) adalah sebuah paket *tools machine learning* praktis, yang dibuat di Universitas Waikato, New Zealand untuk penelitian, pendidikan dan berbagai aplikasi. Weka mampu menyelesaikan masalah-masalah *data mining* di dunia nyata, khususnya klasifikasi yang mendasari pendekatan-pendekatan *machine learning*. Perangkat lunak ini ditulis dalam hirarki *class Java* dengan metode berorientasi obyek dan dapat berjalan hampir di semua *platform*.

Weka mudah digunakan dan diterapkan pada beberapa tingkatan yang berbeda. Tersedia pilihan algoritma pembelajaran terbaru yang dapat kita terapkan pada dataset yang kita miliki. Weka memiliki perangkat untuk melakukan pra-proses data, klasifikasi, regresi, klustering, aturan asosiasi, dan visualisasi. Pengguna dapat melakukan tahap pra proses pada data, memasukkannya dalam sebuah skema pembelajaran, untuk selanjutnya menganalisa klasifikasi dan performansi yang dihasilkan tanpa melakukan pemrograman.



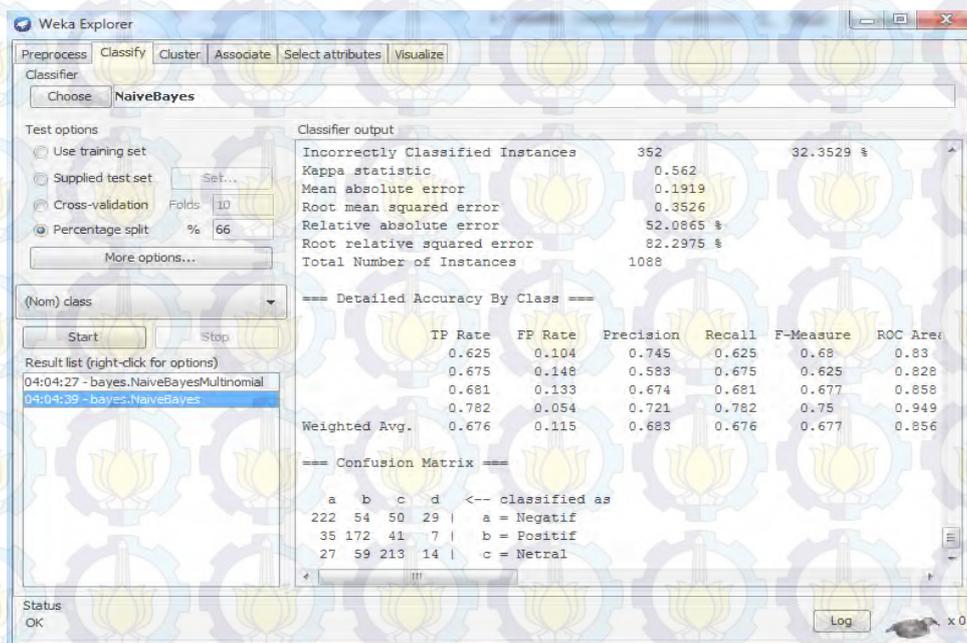
Gambar 2.4 Weka GUI Chooser

Tools yang dapat digunakan untuk pre-processing dataset membuat user dapat berfokus pada algoritma yang digunakan tanpa terlalu memperhatikan detail seperti pembacaan data dari file-file, implementasi algoritma filtering, dan penyediaan kode untuk evaluasi hasil.



Gambar 2.5 Weka Explorer (Preprocess)

Contoh penggunaan WEKA adalah dengan menerapkan sebuah metode pembelajaran ke dataset dan menganalisa hasilnya untuk memperoleh informasi tentang data, atau menerapkan beberapa metode dan membandingkan performansinya untuk dipilih.



Gambar 2.6 Weka Explorer (Classify)

2.15 Penelitian Sebelumnya

Penelitian mengenai klasifikasi sentimen telah dilakukan oleh Bo Pang (2002). Pada papernya, Bo Pang melakukan klasifikasi sentimen terhadap review film dengan menggunakan berbagai teknik pembelajaran mesin. Teknik pembelajaran mesin yang digunakan yaitu Naïve Bayes, Maximum Entropy, dan Support Vector Machines (SVM).

Suhaad Prasad (2011) mencoba untuk menggunakan Naïve Bayes dengan berbagai macam pendekatan yakni, Bernoulli, Bernoulli Chi Square, Multinomial Unigram, Linear Bigram, Back off Bigram, Empirical Bigram, dan Weighted- Normalized Complement Naïve Bayes (WCNB).

Penelitian Analisis Sentimen pada Opini Terhadap Tokoh Publik dilakukan oleh Ismail Sunni dan Dwi Hendratmo Widyantoro (2012). Mereka menggunakan F3 (F3 is Factor Finder) yang memiliki beberapa metode praproses yang diperkirakan mampu menangani permasalahan model bahasa yang ditemukan. F3 menggunakan Naïve Bayes untuk melakukan analisis sentimen karena telah teruji di berbagai penelitian. Sedangkan untuk mengetahui perubahan sentimen, F3 akan menampilkan perubahan sentimen dalam bentuk kurva menggunakan

metode Tf-Idf dengan discounted-cumulative untuk menangani karakter topik yang muncul di Twitter yang berkelanjutan. Hasil analisis dan pengujian menunjukkan tahapan praproses tidak memiliki pengaruh yang signifikan terhadap akurasi (69.4%-72.8%) klasifikasi sentimen. Sedangkan untuk pengestrakan topik menunjukkan bahwa penggunaan Tf-Idf dengan discounted cumulative mampu meningkatkan jumlah topik terekstrak yang sesuai.

Penelitian yang serupa juga dilakukan oleh Paulina Aliandu (2013). Penelitian ini melakukan eksperimen untuk melakukan klasifikasi sentimen terhadap data yang diperoleh dari Twitter dengan mengambil Tweet akun Presiden RI @SBYudhoyono baik sentimen positif, negatif ataupun netral. Aliandu menerapkan Naive Bayes Method untuk klasifikasi sentimen tersebut dan dapat mengklasifikasi dengan baik dengan akurasi 79,42% (Aliandu, 2013).

Tabel 2.8 Penelitian Sebelumnya

| No | Peneliti/Tahun | Judul |
|----|----------------------------|---|
| 1. | Pang, 2002 | Sentiment Classification Using Machine Learning Techniques |
| 2. | Prasad, 2011 | Microblogging Sentiment Analysis Using Bayesian Classification Methods |
| 3. | Sunni & Widyantoro (2012). | Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik |
| 4. | Aliandu, 2013 | Twitter Used by Indonesian President: An Sentiment Analysis of Timeline |

BAB III METODOLOGI

3.1 Diagram Alir Metodologi Penelitian

Berikut ini adalah diagram alir yang akan digunakan untuk menyelesaikan atau mencapai tujuan penelitian:



Gambar 3.1 Metodologi penelitian

3.2 Penentuan Kebijakan Publik yang Akan Dianalisa

Kebijakan publik mempunyai ranah yang sangat luas, untuk itu diperlukan suatu contoh kebijakan publik yang dapat digunakan untuk melakukan simulasi penentuan prioritas kebijakan publik berdasarkan opini masyarakat yang berasal dari Twitter. Sebagai contoh, kebijakan publik yang akan dibuat prioritasnya dalam penelitian ini adalah MDGs (*Millenium Development Goals*) yang mempunyai delapan tujuan. Di bawah ini adalah delapan kebijakan publik yang ada dalam MDGs:

1. Memberantas Kemiskinan dan Kelaparan Ekstrem
2. Mewujudkan Pendidikan Dasar untuk Semua
3. Mendorong Kesetaraan Gender dan Pemberdayaan Perempuan
4. Menurunkan Angka Kematian Anak
5. Meningkatkan Kesehatan Ibu
6. Memerangi HIV dan AIDS Malaria Serta Penyakit Lainnya
7. Memastikan Kelestarian Lingkungan
8. Mengembangkan Kemitraan Global untuk Pembangunan

3.3 Penentuan Kata Kunci

Setelah menentukan kebijakan publik yang akan dibuat prioritasnya, langkah selanjutnya adalah menentukan kata kunci (*key word*) yang dapat mewakili masing-masing kebijakan dalam MDGs. Kata kunci berfungsi untuk melakukan pencarian opini masyarakat melalui *Twitter* (*tweet harvesting*) yang diharapkan hasil pencariannya bersesuaian dengan kebijakan publik yang telah ditetapkan.

Untuk mencegah sedikitnya jumlah *tweet* yang akan didapatkan dalam proses *tweet harvesting* karena kesalahan dalam penentuan kata kunci, terlebih dahulu dilakukan pencarian awal *tweet* dengan menggunakan kata kunci yang telah ditetapkan melalui situs-situs pencarian *tweet* yang ada di internet. Apabila kata kunci tersebut dapat mewakili poin-poin kebijakan dalam MDGs dan cukup populer di *twitter* maka dapat ditetapkan sebagai kata kunci untuk proses selanjutnya. Setiap kebijakan dalam MDGs, diwakili oleh dua kata kunci. Daftar kata kunci dan kebijakan publik yang diwakilinya dapat dilihat pada tabel 3.1.

Tabel 3.1 Kata kunci

| No | Kebijakan Publik (MDGs) | Kata Kunci |
|----|--|---|
| 1. | Memberantas Kemiskinan dan Kelaparan Ekstrem | Kemiskinan Kelaparan |
| 2. | Mewujudkan Pendidikan Dasar untuk Semua | Pendidikan Buta huruf |
| 3. | Mendorong Kesetaraan Gender dan Pemberdayaan Perempuan | Kesetaraan gender Pemberdayaan perempuan |
| 4. | Menurunkan Angka Kematian Anak | Kematian bayi Imunisasi |
| 5. | Meningkatkan Kesehatan Ibu | Kesehatan ibu Kesehatan reproduksi |
| 6. | Memerangi HIV dan AIDS Malaria Serta Penyakit Lainnya | Cegah hiv Cegah penyakit |
| 7. | Memastikan Kelestarian Lingkungan | Keanekaragaman hayati Kelestarian lingkungan |
| 8. | Mengembangkan Kemitraan Global untuk Pembangunan | Akses internet Perdagangan bebas |

3.4 Penentuan Kata Negatif dan Positif

Penentuan kata positif dan negatif maupun emoticon positif dan negatif digunakan untuk membantu menentukan apakah suatu tweet nantinya memiliki kecenderungan sentimen positif atau sentimen negatif pada saat proses pelabelan data training secara manual. Berikut ini adalah daftar kata positif dan negatif maupun emoticon positif dan negatif:

Tabel 3.2 Kata positif, negatif dan emoticon

| Kata Positif | Kata Negatif |
|--|--|
| baik, bagus, berhasil, positif, layak, nyaman, efektif, efisien, maju, cepat, sukses, bisa, oke, lancar, mudah, mampu, dapat | Buruk, jelek, payah, gagal, negatif, masalah, kurang, susah, mundur, sulit, lambat, bohong, ruwet, parah, kacau, bodoh, tolol, gamang, galau, sampah |
| Emoticon Positif | Emoticon Negatif |
| :), :), =), :D | :(, :(, =(, ;(|

3.5 Pengumpulan Data

Proses pengumpulan data (*tweet harvesting*) dilakukan dengan memanfaatkan *Twitter Streaming APIs*. Pencarian dan pengumpulan opini masyarakat di *Twitter* didasarkan pada kata kunci yang telah ditetapkan sebelumnya dalam kurun waktu dua bulan. Data yang didapat dari hasil *harvesting* disimpan ke dalam database.

3.6 Pra Proses

Sebelum dilakukan proses seleksi fitur terhadap tweet yang telah didapatkan dan untuk mendapatkan hasil yang lebih akurat untuk analisa sentimen tweet, dilakukan pra proses (*pre processing*) terhadap data tweet yang ada yang meliputi:

1. *Cleansing*

Pada proses *cleansing* dilakukan penghapusan URL, @mention, #hashtag dan delimiter (karakter angka & simbol)

2. *Case Folding*

Pada tahap ini dilakukan perubahan semua karakter huruf menjadi huruf kecil.

3. *Parsing*

Pemisahan sebuah tweet menjadi kata atau penguraian kalimat menjadi per kata dilakukan pada tahap ini

3.7 Seleksi Fitur

Seleksi fitur dilakukan sebelum proses pembelajaran dan klasifikasi. Pada tahap ini ada dua proses yang dilakukan yaitu:

1. *Stop Word Removal*

Penghapusan kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen (mis: “di”, “oleh”, “pada”, “sebuah”, “karena”)

2. *Stemming*

Proses pemetaan dan penguraian berbagai bentuk (variants) dari suatu kata menjadi bentuk kata dasarnya (stem), dengan cara menghilangkan menghilangkan imbuhan-imbuhan baik itu berupa prefiks, sufiks, maupun konfiks yang ada pada setiap kata

3.8 Pembelajaran dan Klasifikasi

Dari hasil seleksi fitur yang telah dilakukan, selanjutnya dilakukan proses pembelajaran dan klasifikasi dengan menggunakan algoritma naïve bayes yang dibagi menjadi dua tahap yaitu:

1. *Tahap pertama*

Pelatihan terhadap dokumen tweet yang sudah diketahui kategorinya (sentimen positif, negatif, netral dan tanya).

2. *Tahap kedua*

Proses klasifikasi dokumen yang belum diketahui kategorinya (sentimen positif, negatif, netral dan tanya).

3.9 Validasi dan Evaluasi

Tahap ini diperlukan untuk memvalidasi dan melakukan evaluasi sejauh mana tingkat keakuratan (akurasi) proses pembelajaran dan klasifikasi dengan menggunakan algoritma naïve bayes yang telah dilakukan. Untuk memperoleh

hasil yang maksimal dan akurasi yang terbaik, dilakukan berbagai variasi parameter pra-proses dan klasifikasi.

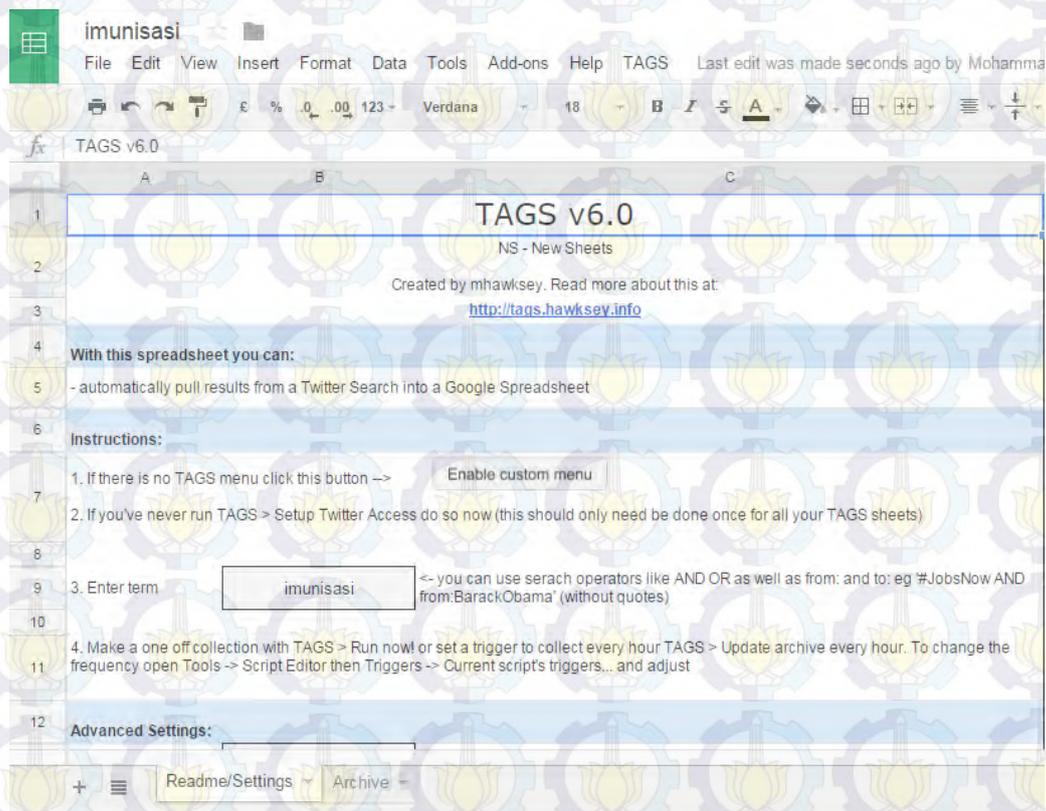
3.10 Penentuan Prioritas Kebijakan

Dari hasil analisa sentimen *tweet* dengan menggunakan algoritma naïve bayes yang telah didapatkan, proses selanjutnya adalah menentukan prioritas kebijakan publik dengan menggunakan algoritma AHP (*Analytical Hierarchy Process*) dengan pembentuk hirarki berupa jumlah tweet sentimen positif, jumlah tweet sentiment negatif, jumlah tweet netral, jumlah tweet tanya, jumlah direct tweet dan jumlah retweet untuk masing-masing kebijakan publik.

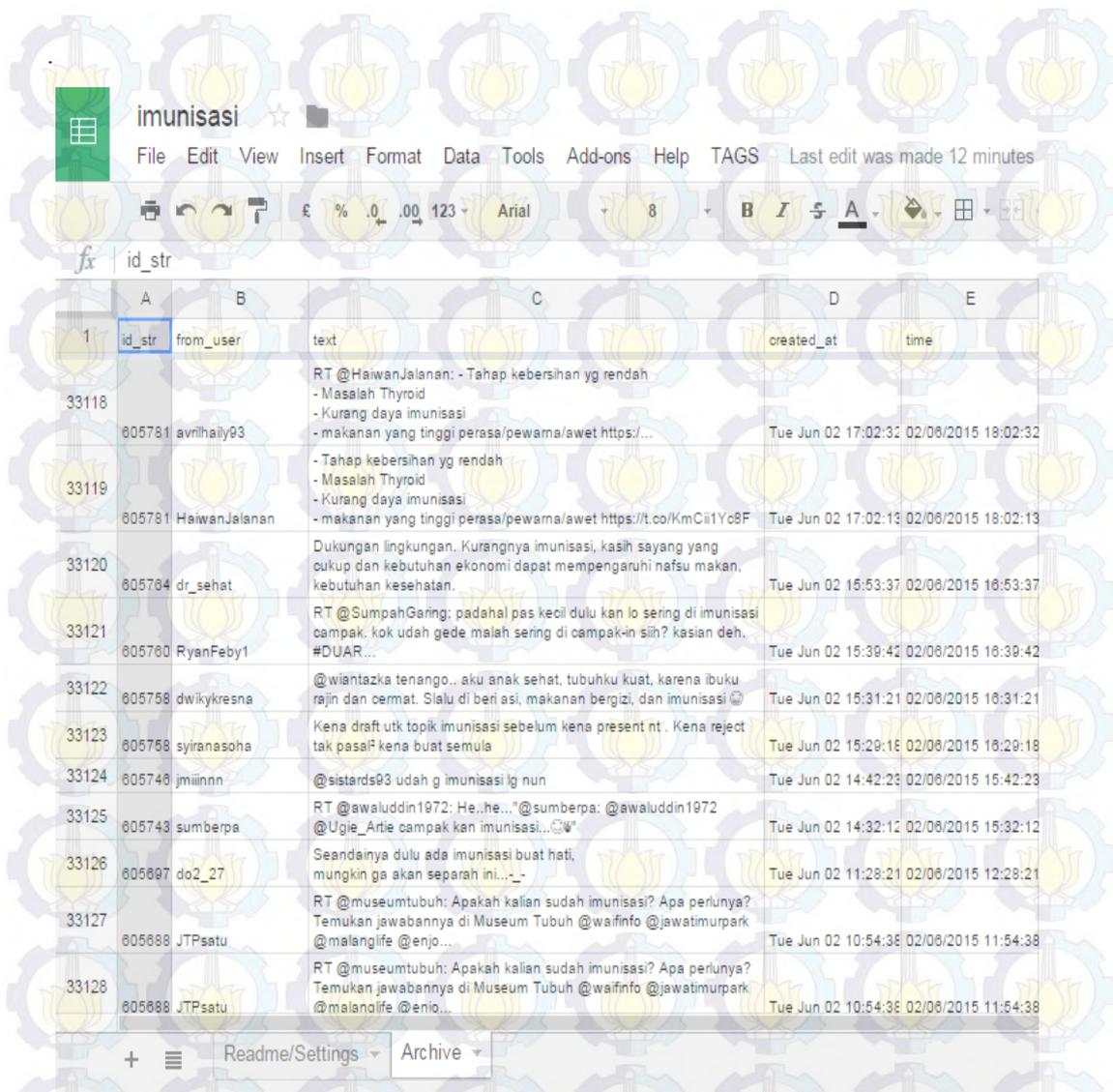
BAB IV HASIL DAN PEMBAHASAN

4.1 Pengumpulan data

Proses pengumpulan data tweet (*tweet harvesting*) berdasarkan kata kunci yang telah ditentukan sebelumnya, dilakukan antara bulan Juni dan Juli tahun 2015. Tweet yang diambil adalah semua tweet berbahasa Indonesia yang berkaitan dengan kata kunci yang telah ditetapkan. Untuk melakukan proses *tweet harvesting* digunakan perangkat lunak Tags V 6.0 yang langsung terhubung dengan *Twitter Streaming APIs*, sehingga data *tweet* dapat diperoleh secara *realtime*. Untuk mencegah *tweet boot* yang akan mengurangi bobot data, tweet yang diambil adalah tweet yang dikirimkan atau dikirim ulang (*retweet*) oleh akun yang minimal memiliki lima orang pengikut (*follower*)



Gambar 4.1 Perangkat lunak Tags V.6.0



| | A | B | C | D | E |
|-------|--------|--------------|---|---------------------|---------------------|
| 1 | id_str | from_user | text | created_at | time |
| 33118 | 605781 | avrihally93 | RT @HawanJalanan: - Tahap kebersihan yg rendah - Masalah Thyroid - Kurang daya imunisasi - makanan yang tinggi perasa/pewarna/awet https://t.co/KmCii1Yc8F | Tue Jun 02 17:02:32 | 02/06/2015 18:02:32 |
| 33119 | 605781 | HawanJalanan | - Tahap kebersihan yg rendah - Masalah Thyroid - Kurang daya imunisasi - makanan yang tinggi perasa/pewarna/awet https://t.co/KmCii1Yc8F | Tue Jun 02 17:02:13 | 02/06/2015 18:02:13 |
| 33120 | 605784 | dr_sehat | Dukungan lingkungan. Kurangnya imunisasi, kasih sayang yang cukup dan kebutuhan ekonomi dapat mempengaruhi nafsu makan, kebutuhan kesehatan. | Tue Jun 02 15:53:37 | 02/06/2015 16:53:37 |
| 33121 | 605760 | RyanFeby1 | RT @SumpahGaring: padahal pas kecil dulu kan lo sering di imunisasi campak, kok udah gede malah sering di campakin sih? kasian deh. #DUAR... | Tue Jun 02 15:39:42 | 02/06/2015 16:39:42 |
| 33122 | 605758 | dwikykesna | @wiantazka tenang.. aku anak sehat, tubukku kuat, karena ibuku rajin dan cermat. Slalu di beri asi, makanan bergizi, dan imunisasi 🍀 | Tue Jun 02 15:31:21 | 02/06/2015 16:31:21 |
| 33123 | 605758 | syiranasoha | Kena draft utk topik imunisasi sebelum kena present nt . Kena reject tak pasal? kena buat semula | Tue Jun 02 15:29:18 | 02/06/2015 16:29:18 |
| 33124 | 605748 | jmiinnn | @sistards93 udah g imunisasi lg nun | Tue Jun 02 14:42:23 | 02/06/2015 15:42:23 |
| 33125 | 605743 | sumberpa | RT @awaluddin1972: He..he.. " @sumberpa: @awaluddin1972 @Ugie_Artie campak kan imunisasi...🍀" | Tue Jun 02 14:32:12 | 02/06/2015 15:32:12 |
| 33126 | 605697 | do2_27 | Seandainya dulu ada imunisasi buat hati, mungkin ga akan separah ini..._- | Tue Jun 02 11:28:21 | 02/06/2015 12:28:21 |
| 33127 | 605688 | JTPsatu | RT @museumtubuh: Apakah kalian sudah imunisasi? Apa perlunya? Temukan jawabannya di Museum Tubuh @waifinfo @jawatimurpark @malanglife @enjo... | Tue Jun 02 10:54:38 | 02/06/2015 11:54:38 |
| 33128 | 605688 | JTPsatu | RT @museumtubuh: Apakah kalian sudah imunisasi? Apa perlunya? Temukan jawabannya di Museum Tubuh @waifinfo @jawatimurpark @malanglife @enio... | Tue Jun 02 10:54:38 | 02/06/2015 11:54:38 |

Gambar 4.2 Contoh hasil *Tweet Harvesting* dengan Tags V.6.0

Adapun hasil dari proses pengumpulan data *tweet* yang telah dilakukan ditunjukkan oleh tabel 4.1:

Tabel 4.1 Perolehan hasil *tweet harvesting*

| No | Kebijakan Publik | Kata Kunci | Jumlah Tweet |
|--------------------|--|------------------------|---------------|
| 1. | Memberantas Kemiskinan dan Kelaparan Ekstrem | Kemiskinan | 50226 |
| | | Kelaparan | 72404 |
| 2. | Mewujudkan Pendidikan Dasar untuk Semua | Pendidikan | 96060 |
| | | Buta huruf | 10350 |
| 3. | Mendorong Kesetaraan Gender dan Pemberdayaan Perempuan | Kesetaraan gender | 1636 |
| | | Pemberdayaan Perempuan | 7444 |
| 4. | Menurunkan Angka Kematian Anak | Kematian bayi | 1766 |
| | | Imunisasi | 11620 |
| 5. | Meningkatkan Kesehatan Ibu | Kesehatan ibu | 3609 |
| | | Kesehatan reproduksi | 2712 |
| 6. | Memerangi HIV dan AIDS Malaria Serta Penyakit Lainnya | Cegah HIV | 345 |
| | | Cegah penyakit | 4082 |
| 7. | Memastikan Kelestarian Lingkungan | Keanekaragaman hayati | 2509 |
| | | Kelestarian lingkungan | 1703 |
| 8. | Mengembangkan Kemitraan Global untuk Pembangunan | Akses internet | 12804 |
| | | Perdagangan bebas | 2854 |
| Total Tweet | | | 282124 |

4.2 Data Training

Dari hasil *tweet harvesting* yang telah dilakukan pada proses pengumpulan data, diambil 3200 tweet yang akan digunakan sebagai data

training dan testing. Pengambilan data training dengan mempertimbangkan prosentase perolehan masing-masing kata kunci agar terdapat unsur keterwakilan.

Selanjutnya data training tersebut dilabeli secara manual apakah termasuk dalam tweet yang memiliki sentimen negatif, sentimen positif, netral atau berupa pertanyaan. Berikutnya data yang telah dilabeli secara manual diimport ke dalam database MySQL agar dapat dilakukan tahap praproses.

Format data training seperti terlihat pada tabel 4.2.

Tabel 4.2 Contoh pelabelan *tweet* secara manual

| No | Tweet | Klasifikasi |
|----|--|-------------|
| 1. | Ujian Baru GunungMeletus & Kemarau. Itu blm sbrp, klu sdh kelaparan Kematian pemerintah msh woles??? :) | Negatif |
| 2. | Sukses Perangi Kelaparan, Indonesia Raih Penghargaan di Italia: INDONESIA dinilai berhasil men... http://t.co/0uyUG58dhk | Positif |
| 3. | Kelaparan menyebabkan turunnya tingkat serotonin dalam tubuh yang mengakibatkan timbul rasa gelisah, stres, marah, dan sedih. | Netral |
| 4. | Apa resep sukses pendidikan Vietnam? http://t.co/BYaFx14JuJ http://t.co/B4Qda19kl1 | Tanya |
| 5. | Pemerintah: tenang kita akan kuat semada pangan tahun 2018 rakyat : kalau begitu kami akan kelaparan duluan. | Negatif |
| 6. | "Sukses Perangi Kelaparan, Indonesia Raih Penghargaan di Italia: INDONESIA dinilai berhasil men... http://t.co/0uyUG58dhk #presidenjokowi" | Positif |

4.3 Pra Proses

Sebelum melakukan seleksi fitur terhadap 3200 tweet data training, terlebih dahulu dilakukan tahap pra proses yang meliputi:

1. Case Folding

Pada tahap ini dilakukan perubahan semua karakter huruf menjadi huruf kecil pada tweet data training.

2. Cleansing

Pada proses cleansing dilakukan penghapusan URL, @mention, #hashtag, RT @ dan delimiter (karakter angka & simbol).

3. Parsing

Pemisahan sebuah tweet menjadi kata atau penguraian kalimat menjadi per kata dilakukan pada tahap ini.

Tabel 4.3 Hasil tahap pra proses data training

| Tweet | Case Folding | Cleansing | Parsing |
|----------------|---------------------|------------------|------------------------|
| Semoga | semoga | semoga | word [1] = semoga |
| pemerintah | pemerintah tak | pemerintah | word [2] = pemerintah |
| tak | membiarkan | tak | word [3] = tak |
| membiarkan | rakyatny | membiarkan | word [4] = membiarkan |
| rakyatny | kelaparan.kasi | rakyatny | word [5] = rakyatny |
| kelaparan.kas | han petani | kelaparan | word [6] = kelaparan |
| ihan petani | butuh uluran | kasihan | word [7] = kasihan |
| butuh uluran | tangan | petani butuh | word [8] = petani |
| tangan | pemerintah | uluran | word [9] = butuh |
| pemerintah | @pak_jk | tangan | word [10] = uluran |
| @Pak_JK | @jokowi | pemerintah | word [11] = tangan |
| @jokowi | https://t.co/ok | | word [12] = pemerintah |
| https://t.co/O | md8y8fhg | | |
| kmD8y8fHG | | | |

4.4 Seleksi Fitur

Tahap selanjutnya adalah melakukan seleksi fitur pada data training yang telah melalui tahap pra proses. Adapun proses yang dijalankan adalah:

1. Stop Word Removal

Pada proses ini daftar kata yang ada pada sebuah dokumen tweet data training dihilangkan bagian yang tidak mempunyai makna. Daftar kata yang tidak mempunyai makna diambil dari hasil penelitian tala (Tala, F. Z. (2003))

2. Stemming

Dokumen *tweet* data training yang telah dipilih melalui proses *stop word removal* diproses dengan menggunakan PHP library Sastrawi (Andylibrian, 2005) yang didasarkan pada algoritma stemming Nazief dan Andriani.

Tabel 4.4 Hasil stemming dengan menggunakan PHP Library Sastrawi

| Hasil Parsing Tweet | Stemming |
|-----------------------|---------------------|
| word [1] = soal | word [1] = soal |
| word [2] = kelaparan | word [2] = lapar |
| word [3] = warga | word [3] = warga |
| word [4] = disana | word [4] = sana |
| word [5] = sudah | word [5] = sudah |
| word [6] = biasa | word [6] = biasa |
| word [7] = pemerintah | word [7] = perintah |
| word [8] = saja | word [8] = saja |
| word [9] = yg | word [9] = yg |
| word [10] = nggak | word [10] = nggak |
| word [11] = tanggap | word [11] = tanggap |

Untuk melakukan tahap praproses dan seleksi fitur dibuat script PHP seperti tampak pada gambar 4.3 dengan hasil akhir seperti tampak pada gambar 4.4.

```

24
25 if($hasil){
26     while($baris=mysql_fetch_array($hasil)){
27         //Tweet
28         $tweet=$baris['tweet'];
29
30         //proses mengubah huruf menjadi huruf kecil semua
31         $lower=strtolower($tweet);
32
33         //menghilangkan hastag
34         $removehashtags = preg_replace('#([\#-]+)/\w+', '', $lower);
35
36         //menghilangkan mention
37         $removemention = preg_replace('/@([\#-]+)/\w+', '', $removehashtags);
38
39         //menghilangkan url
40         $removeurl=preg_replace('/(http|https|ftp|ftps)\:\/\/[a-zA-Z0-9\-\.\.]+\.[a-zA-Z]{2,3}\/\S*?\/', '', $removemention);
41
42         //proses menghilangkan tanda baca
43         $removepunctuation = trim(str_replace("","",preg_replace('/[^\w\@-]+/',' ', $removeurl)));
44
45         //proses stemming
46         $output = $stemmer->stem($removepunctuation);
47
48         //echo $output . "\n";
49         // ekonomi indonesia sedang dalam tumbuh yang bangga
50
51         echo "<td>".$tweet."</td>";
52         echo "<td>".$lower."</td>";
53

```

Gambar 4.3 Script PHP untuk melakukan tahap praproses dan seleksi fitur

| Tweet | Case Folding | Cleansing | Parsing | Stemming |
|---|---|--|---|---|
| 'Jokowi Ingin Perdagangan Bebas RI-Turki Segera Mulai http://t.co/Hip154Dc1 #CieboGroup https://t.co/UMRYSEGTV http://t.co/Ingr3uo6ePO' | 'jokowi ingin perdagangan bebas ri-turki segera mulai http://t.co/Hip154Dc1 #CieboGroup https://t.co/UMRYSEGTV http://t.co/Ingr3uo6ePO' | jokowi ingin perdagangan bebas ri turki segera mulai | word [1] = jokowi word [2] = ingin word [3] = perdagangan word [4] = bebas word [5] = ri word [6] = turki word [7] = segera word [8] = mulai | jokowi ingin dagang bebas ri turki segera mulai |
| 'Jokowi Ingin Perdagangan Bebas RI-Turki Segera Mulai http://t.co/IdnWYERtb #CieboGroup https://t.co/UMRY9Wiw http://t.co/qnnyOFlsX' | 'jokowi ingin perdagangan bebas ri-turki segera mulai http://t.co/IdnWYERtb #CieboGroup https://t.co/UMRY9Wiw http://t.co/qnnyOFlsX' | jokowi ingin perdagangan bebas ri turki segera mulai | word [1] = jokowi word [2] = ingin word [3] = perdagangan word [4] = bebas word [5] = ri word [6] = turki word [7] = segera word [8] = mulai | jokowi ingin dagang bebas ri turki segera mulai |
| 'Jokowi Ingin Perdagangan Bebas RI-Turki Segera Mulai http://t.co/CRy4QmJNl #CieboBiz https://t.co/Lst07LZF http://t.co/0nNH4UM' | 'jokowi ingin perdagangan bebas ri-turki segera mulai http://t.co/CRy4QmJNl #CieboBiz https://t.co/Lst07LZF http://t.co/0nNH4UM' | jokowi ingin perdagangan bebas ri turki segera mulai | word [1] = jokowi word [2] = ingin word [3] = perdagangan word [4] = bebas word [5] = ri word [6] = turki word [7] = segera word [8] = mulai | jokowi ingin dagang bebas ri turki segera mulai |
| 'Incar Perdagangan Bebas Kedua Negara http://t.co/sAW5Nlnmf http://t.co/Hp2lef6ts' | 'incar perdagangan bebas kedua negara http://t.co/sAW5Nlnmf http://t.co/Hp2lef6ts' | incar perdagangan bebas kedua negara | word [1] = incar word [2] = perdagangan word [3] = bebas word [4] = kedua word [5] = negara | incar dagang bebas dua negara |
| 'Incar Perdagangan Bebas Kedua Negara http://t.co/1R00QbG0XV' | 'incar perdagangan bebas kedua negara http://t.co/1r00qgvv' | incar perdagangan bebas kedua negara | word [1] = incar word [2] = perdagangan word [3] = bebas word [4] = kedua word [5] = negara | incar dagang bebas dua negara |
| 'Incar Perdagangan Bebas Kedua Negara http://t.co/aLWZ5F17 Riau Pos' | 'incar perdagangan bebas kedua negara http://t.co/aLwv1zsf7? riau pos' | incar perdagangan bebas kedua negara riau pos | word [1] = incar word [2] = perdagangan word [3] = bebas word [4] = kedua word [5] = negara word [6] = riau word [7] = pos | incar dagang bebas dua negara riau pos |

Gambar 4.4 Hasil praproses dan seleksi fitur

4.5 Pembelajaran dan Klasifikasi

Dari hasil seleksi fitur yang telah dilakukan, selanjutnya dilakukan proses pembelajaran dan klasifikasi dengan menggunakan algoritma naïve bayes yang dibagi menjadi dua tahap yaitu:

1. Tahap pertama

Pada tahap ini dilakukan proses pembelajaran pada data training yang telah mengalami seleksi fitur dan telah diketahui klasifikasinya (negatif, positif, netral, tanya). Perangkat lunak yang digunakan untuk melakukan proses pembelajaran adalah Weka. Sebelum dapat menggunakan perangkat lunak Weka, data training harus diubah terlebih dahulu ke dalam format CSV atau ARFF seperti tampak pada gambar 4.5.

Langkah selanjutnya adalah melakukan konversi data *string* ke dalam bentuk *word vector* seperti tampak pada gambar 4.6. Proses ini harus dilakukan terlebih dahulu agar data training dapat dianalisa dengan menggunakan algoritma Naïve bayes atau algoritma lainnya.

```

@relation Twitter
@attribute body string
@attribute class {Negatif,Positif,Netral,Tanya}
@data
'Ujian Baru GunungMeletus Gamp Kemarau. Itu blm sbkr. klu sdh kelaparan.Kematian. pemerintah msh wales??? :)',Negatif
'Semoga pemerintah tak membiarkan rakyatny kelaparan.kasihlah petani butuh uluran tangan pemerintah @Pak_UK @jokowi https://t.co/OkmD9y6fHG',Ne
'4. Pemerintah harus bisa mewujudkan zero hunger. Jangan ada lagi yang kelaparan di negeri yang berkelimpahan ini Ketum Bpk @Hary_Tanoe',Negat
'Etampodotco Soal kelaparan warga disana sudah biasa. Pemerintah saja yg nggak tanggap.',Negatif
'Tolak HAMAS. Langkah Pemerintah Dipuji PDIP http://t.co/ND0rDqVLoK Tiga kampung di Papua terancam kelaparan http://t.co/XlffVrjX3t',Negatif
'Dia menduga pemerintah daerah setempat menutupi warganya yang mengalami kelaparan. Dprd said.Citra lebih penting nyampe tingkat daerah',
'NTT kelaparan sejak awal Januari 2015. pemerintah apatis. kemana aja????? Hei presiden @jokowi',Negatif
'kondisi perekonomian Indonesia di bawah pemerintahan Presiden Joko Widodo memburuk.banyak kelaparan lapangan kerja modal tidak ada semuanya',N
'Masyarakat Timor Tengah Selatan.NTT kelaparan un pemerintah?@jokowi @setkabgoind',Negatif
'Makan? Itu barang mewah!! Kelaparan dimana2. Kaum pejuang semena2. pemerintahan tidak ada gunanya.',Negatif
'Resaful kabinet ga perlu.pesiden nya yg hrs di resafful...kl pemerintah ini ampe 5 thn bisa mati kelaparan kt semua',Negatif
'Pemerintah daerah Kab/Kota d NTT memang kurang gizi makanya rakyatnya gzi buruk dan kelaparan',Negatif
'RT @nayasaa: Cth pemerintahan ga ada keriaan. raja restoran yg buka selama puasa. Tpi ga pernah ngokok warganya kelaparan. Eshol.',Negatif
'Pemerintah: tenang kita akan kuat sambada pangan tahun 2018.. rakyat : kalau begitu kami akan kelaparan duluan.',Negatif
'RT @IndoPolitiki: busung lapar dan orang mati kelaparan karena tidak bisa beli makanan akan meningkat. apabila pemerintah tidak melakukan t',N
'WAHLI PEMERINTAH DAN PEJABAT JANGAN MENGURUSI DIRI ANDA.TAPI LUPA KEPADA RAKYAT YG KELAPARAN',Negatif
'@kompascom pemerintah sekarang memang gak ada solusi.dr kelaparan ntt.sinabung.pssi dan rupiah loyo. semuanya tida ada solusi',Negatif
'@AnakBaranusu @NdoroTweet Sy cc kan tweet ndoro ke anda. Itu knn berhari2 anda teriak2 ttg kelaparan di NTT dn pemerintah tdk bereaksi..?',Nega
'Begini Perjuangan Warga Kelaparan Bertahan Hidup: Mereka menagih janji pemerintah soal penyediaan sumur b... #MSF',Nega
'Rakyat yg kelaparan terombang ambing dalam peperangan korporasi dgn pemerintah',Negatif
'11. D benak pemerintah. kalau subsidi dicabut. akan ada uang buat pembangunan. Bener sah. tp rakyat mati kelaparan duluan. #SudahlahJokowi',Ne
'Separah itukah jokowek. menghubungkan kasus kelaparan NTT yang menewaskan 11 anak ke DPR dan bukannya pemerintah?',Negatif

```

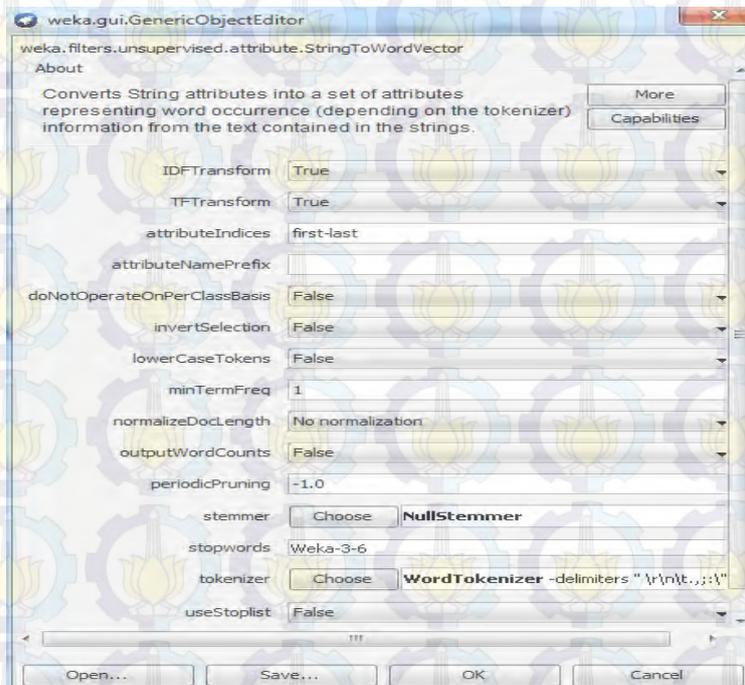
Gambar 4.5 Format data ARFF

Relation: Twitter-weka.filters.unsupervised.attribute.StringToWordVector-R1-W1000-prune-rate-1.0-T1-N0-stemmerweka.core.stemmers.NullStemmer-M1-tokenizerweka.core.tokenizers.WordTok...

| No. | class | abal | aceh | ad | ada | adalah | adli | adli | agadugume | agama | agar | ah | air | aja | ajar | akal | akan | akar | akhi |
|-----|---------|------|------|-----|----------|--------|------|------|-----------|-------|------|-----|----------|-----|------|------|----------|------|------|
| 1 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | Negatif | 0.0 | 0.0 | 0.0 | 1.696... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 4 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 5 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 6 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 7 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.882... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 8 | Negatif | 0.0 | 0.0 | 0.0 | 1.696... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 9 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 10 | Negatif | 0.0 | 0.0 | 0.0 | 1.696... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 11 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 12 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 13 | Negatif | 0.0 | 0.0 | 0.0 | 1.696... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 14 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.055... | 0.0 | 0.0 |
| 15 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.055... | 0.0 | 0.0 |
| 16 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 17 | Negatif | 0.0 | 0.0 | 0.0 | 1.696... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 18 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 19 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 20 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 21 | Negatif | 0.0 | 0.0 | 0.0 | 1.696... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.055... | 0.0 | 0.0 |
| 22 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 23 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 24 | Negatif | 0.0 | 0.0 | 0.0 | 1.696... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 25 | Negatif | 0.0 | 0.0 | 0.0 | 1.696... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 26 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 27 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 28 | Negatif | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

Gambar 4.6 Konversi data *string* ke *word vector*

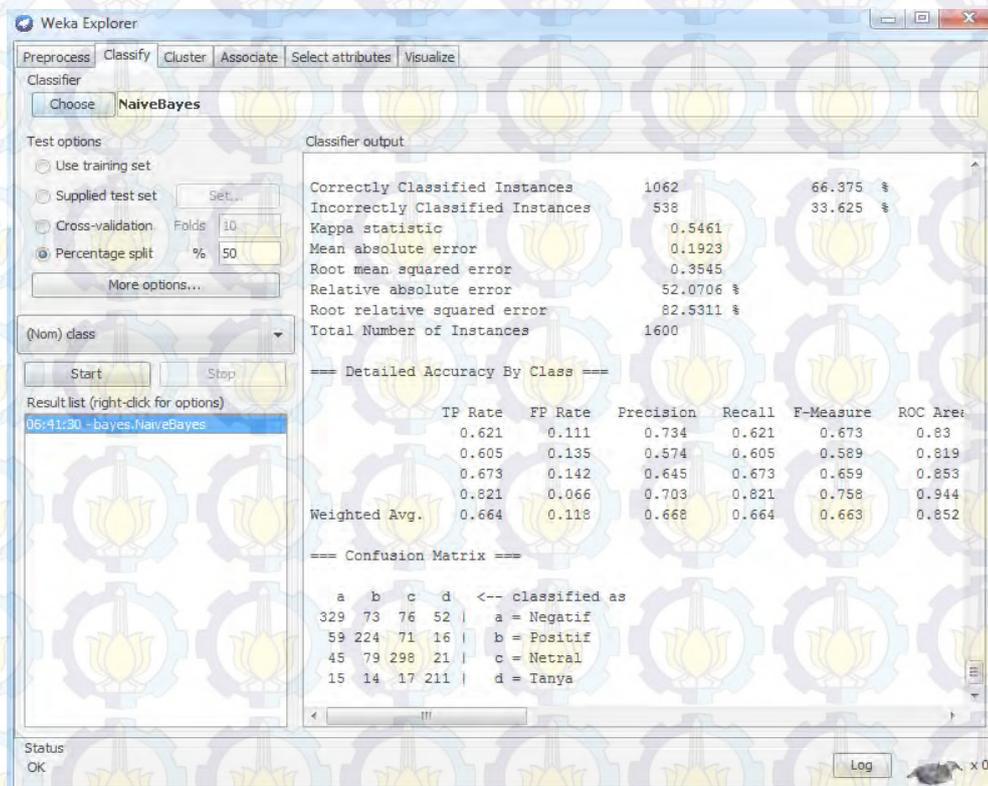
Untuk memperoleh hasil yang optimum dilakukan beberapa kombinasi filtering yang terdiri dari *term weighting*, *min term freq*, *stop word* dan *tokenizer* seperti terlihat pada gambar 4.7.



Gambar 4.7 Beberapa kombinasi pada proses *filtering*

Setelah proses filtering dan konversi *string to word vector* dilakukan, tahap berikutnya adalah melakukan pelatihan data training dengan menggunakan algoritma naïve bayes pada perangkat lunak Weka. Untuk menentukan berapa jumlah data training dan berapa jumlah data testing dapat digunakan menu *percentage split*. Dari hasil klasifikasi dapat diperoleh akurasi dan confusion matrix, seperti tampak pada gambar 4.8.

Gambar 4.8 Contoh hasil klasifikasi pada Weka



Akurasi hasil pembelajaran dengan menggunakan pembobotan TF, IDF dan TF-IDF tampak pada tabel 4.5, 4.6 dan 4.7.

Tabel 4.5 Hasil pembelajaran dengan pembobotan TF

| term weighting | min term freq | stop word | tokenizer | percentage split | accuracy | |
|----------------|---------------|-----------|-----------|------------------|----------|--------|
| TF | 1 | Tala | Word | 50% | 69,89% | |
| | | | | 60% | 69,78% | |
| | | | | 66% | 70,54% | |
| | | Ngram | 50% | 69,14% | | |
| | | | 60% | 69,47% | | |
| | | | 66% | 70,33% | | |
| | | - | - | Word | 50% | 70,14% |
| | | | | | 60% | 71,11% |
| | | | | | 66% | 72,1% |
| | Ngram | - | - | 50% | 69,2% | |
| | | | | 60% | 69,7% | |
| | | | | 70% | 71,06% | |
| | 3 | Tala | Word | 50% | 69,89% | |
| | | | | 60% | 69,84% | |
| | | | | 66% | 70,43% | |
| | | Ngram | Tala | 50% | 69,08% | |
| | | | | 60% | 69,23% | |
| | | | | 66% | 70,22% | |
| - | | - | Word | 50% | 70,14% | |
| | | | | 60% | 71,11% | |
| | | | | 66% | 72,1% | |
| Ngram | - | - | 50% | 69,2% | | |
| | | | 60% | 69,7% | | |
| | | | 66% | 71,06% | | |

Tabel 4.6 Hasil pembelajaran dengan pembobotan IDF

| term weighting | min term freq | stop word | tokenizer | percentage split | accuracy |
|----------------|---------------|-----------|-----------|------------------|----------|
| IDF | 1 | Tala | Word | 50% | 69,89% |
| | | | | 60% | 69,78% |
| | | | | 66% | 70,54% |
| | | Tala | Ngram | 50% | 69,14% |
| | | | | 60% | 69,47% |
| | | | | 66% | 70,33% |
| | | - | Word | 50% | 70,14% |
| | | | | 60% | 71,11% |
| | | | | 66% | 71,99% |
| | - | Ngram | 50% | 69,2% | |
| | | | 60% | 69,7% | |
| | | | 66% | 71,06% | |
| | 3 | Tala | Word | 50% | 69,89% |
| | | | | 60% | 69,86% |
| | | | | 66% | 70,43% |
| | | Tala | Ngram | 50% | 69,08% |
| | | | | 60% | 69,23% |
| | | | | 66% | 70,22% |
| - | | Word | 50% | 70,14% | |
| | | | 60% | 71,11% | |
| | | | 66% | 72,1% | |
| - | Ngram | 50% | 69,2% | | |
| | | 60% | 69,7% | | |
| | | 66% | 71,06% | | |

Tabel 4.7 Hasil pembelajaran dengan pembobotan TF-IDF

| term weighting | min term freq | stop word | tokenizer | percentage split | accuracy |
|----------------|---------------|-----------|-----------|------------------|----------|
| TF-IDF | 1 | Tala | Word | 50% | 69,89% |
| | | | | 60% | 69,78% |
| | | | | 66% | 70,54% |
| | | | Ngram | 50% | 69,14% |
| | | | | 60% | 69,47% |
| | | | | 66% | 70,33% |
| | | - | Word | 50% | 70,14% |
| | | | | 60% | 71,11% |
| | | | | 70% | 71,99% |
| | | | Ngram | 50% | 69,2% |
| | | | | 60% | 69,7% |
| | | | | 66% | 71,06% |
| | 3 | Tala | Word | 50% | 69,89% |
| | | | | 60% | 69,86% |
| | | | | 66% | 70,43% |
| | | | Ngram | 50% | 69,08% |
| | | | | 60% | 69,23% |
| | | | | 66% | 70,22% |
| | | - | Word | 50% | 70,14% |
| | | | | 60% | 71,11% |
| | | | | 66% | 72,1% |
| | | | Ngram | 50% | 69,2% |
| | | | | 60% | 69,7% |
| | | | | 66% | 71,06% |

Dari tabel 4.5, 4.6 dan 4.7 tampak bahwa hasil akurasi optimum pembelajaran dengan menggunakan algoritma Naïve Bayes yang bisa dicapai adalah sebesar 72,1%. Kondisi tersebut tercapai apabila:

- a. Pembobotan yang digunakan adalah TF (*term frequency*) dengan *minimal term frequency*=1, tanpa melakukan proses *stopword*, *tokenizer* yang dipakai adalah= *word tokenizer* dengan *procentage split* sebesar 66%.
- b. Pembobotan yang digunakan adalah TF (*term frequency*) dengan *minimal term frequency*=3, tanpa melakukan proses *stopword*, *tokenizer* yang dipakai adalah= *word tokenizer* dengan *procentage split* sebesar 66%.
- c. Pembobotan yang digunakan adalah IDF (*inverse document frequency*) dengan *minimal term frequency*=3, tanpa melakukan proses *stopword*, *tokenizer* yang dipakai adalah= *word tokenizer* dengan *procentage split* sebesar 66%.
- d. Pembobotan yang digunakan adalah TF-IDF dengan *minimal term frequency*=3, tanpa melakukan proses *stopword*, *tokenizer* yang dipakai adalah= *word tokenizer* dengan *procentage split* sebesar 66%.

2. Tahap kedua

Setelah ditemukan akurasi yang paling optimum dari hasil pembelajaran data training dengan menggunakan algoritma Naïve Bayes, tahap selanjutnya adalah melakukan prediksi terhadap dokumen tweet di luar data training yang belum diketahui klasifikasinya (sentimen negatif, positif, netral, tanya). Proses prediksi ini dilakukan menggunakan model pembelajaran yang akurasinya paling optimum dengan menggunakan perangkat lunak Weka.

Langkah yang harus dilakukan terlebih dahulu adalah melakukan tahap praproses dan stemming pada semua data diluar data training hasil *tweet harvesting* dengan tahapan yang sama dengan data training. Data tersebut dibuat dengan format ARFF atau CSV dengan klasifikasi berupa tanda tanya seperti tampak pada gambar 4.9. Dengan menggunakan menu

supplied test set seperti tampak pada gambar 4.10, kita dapat mengambil file ARFF yang telah kita siapkan sebelumnya untuk melakukan prediksi terhadap data di luar data training. Agar hasil prediksi bisa dilihat harus disi

```

1 @relation Twitter
2
3 @attribute body string
4 @attribute class {Negatif,Positif,Netral,Tanya}
5
6 @data
7
8 'RT @infobencana: #Koruptor bebas bisa menampati gedung2 mewah. rakyat sengsara kelaparan. tak ada jaminan kesejaht
9 '#Koruptor bebas bisa menampati gedung2 mewah. rakyat sengsara kelaparan. tak ada jaminan kesejahteraan sbgmn term
10 'Rakyat kelaparan. bapak bapak malah jalan2 terus https://t.co/zquUSfeDKo',?
11 'Nih hasil kerja menakjubkan 19.4 juta rakyat kelaparan http://t.co/HZM7Bo0Yyp @sijun_zulfiadi @Hary_Tanoe',?
12 'Jgn berteriak Infrastruktur. tapi Rakyat Kelaparan krn tidak bisa beli beras http://t.co/WgUSjEUIGN',?
13 'Melihat harga2 anak dan rakyat menjerit kelaparan..ada baiknya kita mulai saling mengenal tetangga dan sodara..',?
14 'Tiap hari dia plesiran . rakyat meringis kelaparan . uang rakyat dipakai foya - foya si JO-KO-WI http://t.co/IMJH0xf
15 'Hingga Julii5. Tercatat 11.350 Pekerja Kena Pecat http://t.co/vrIudtDWIj http://t.co/HxMPCcwt9 rakyat kelaparan Huru
16 'RT @IDR Kepo: seperti rakyat di Ethiopia sana yang sedang kelaparan .aku jg sama kaya mereka cuma bedanya aku kelapa
17 'RT @princes_lisda: @farhatabbaslaw bantu rakyat miskin sj mas. mah byk rakyat miskin yh kelaparan dmn2.ting pekhata
18 '@mohmahfudmd sungguh yg kami butuh adalah pekerjaan pak .kami kelaparan dijalanan diusir ditendang sm aparat yg mema
19 '@republikonline: Malukah Kita Mengimpor Beras? (2-Habis) http://t.co/pYaBaWlJ38 dan membiarkan rakyat kelaparan?',?
20 'Banyak yg telah dibodohi oleh media.. padahal banyak anak yg kelaparan.. rakyat menderita karna dolar [pic] - https:
21 'SAAT MASJID DIBAKAR &mp RAKYAT KELAPARAN. @jokowi MALAH CENGENGESAN NONTON FILM. PEMIMPIN MACAM APA INI!@ganggy_umba
22 'Papa konflik agama. rakyat NTT kelaparan. sebagian daerah jawa timur mengalami gagal panen.',?
23 '@guazalit Bagaimana bisa dikatakan berkah jika rakyatnya menyandang pakaian kelaparan &mp ketakutan?',Tanya
24 'rakyat kecil miskin diseluruh daerah manapun kuasunya solo raya meradang kelaparan tanpa dana anggaran sosial dan ta
25 '@saatiady mau niku Kusno dengan proyek Mercusuar walo rakyat kelaparan?',?

```

Gambar 4.9 Format data ARFF untuk proses prediksi

The screenshot displays the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' section shows 'Supplied test set' is chosen. The 'Classifier output' section displays the following performance metrics:

| Metric | Value | Percentage |
|----------------------------------|-------|------------|
| Correctly Classified Instances | 1062 | 66.375 % |
| Incorrectly Classified Instances | 538 | 33.625 % |

A 'Test Instances' dialog box is open, showing 'Relation: None', 'Instances: None', and 'Attributes: None'. Below the dialog, a table shows 'Detailed Accuracy by Class':

| | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area |
|---------------|---------|---------|-----------|--------|-----------|----------|
| 0.621 | 0.111 | 0.734 | 0.621 | 0.673 | 0.83 | |
| 0.605 | 0.135 | 0.574 | 0.605 | 0.589 | 0.819 | |
| 0.673 | 0.142 | 0.645 | 0.673 | 0.659 | 0.853 | |
| 0.821 | 0.066 | 0.703 | 0.821 | 0.758 | 0.944 | |
| Weighted Avg. | 0.664 | 0.118 | 0.668 | 0.664 | 0.663 | 0.852 |

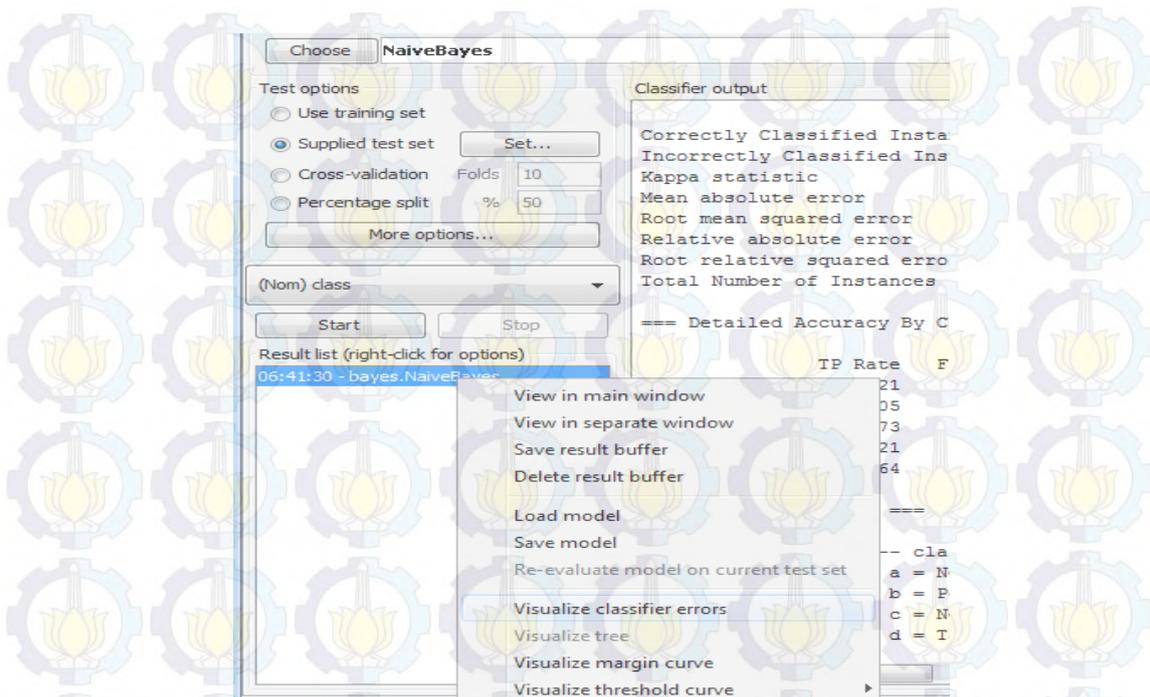
Below the table, a 'Confusion Matrix' is shown:

```

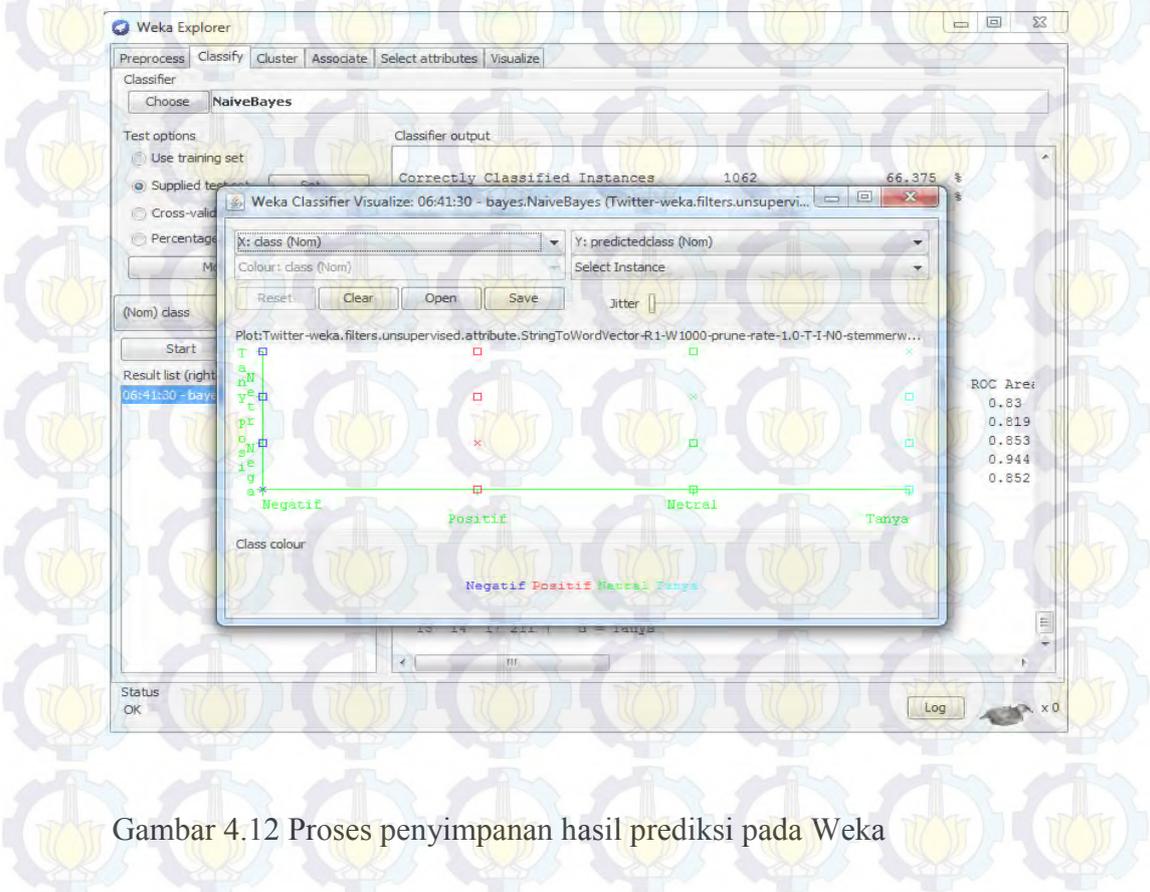
=== Confusion Matrix ===
 a  b  c  d  <-- classified as
329 73 76 52 | a = Negatif
 59 224 71 16 | b = Positif
 45 79 298 21 | c = Netral
 15 14 17 211 | d = Tanya

```

Gambar 4.10 Proses pengambilan data set pada Weka



Gambar 4.11 Proses visualisasi kesalahan klasifikasi pada Weka



Gambar 4.12 Proses penyimpanan hasil prediksi pada Weka

Classifier: NaiveBayes

Test options:

- Use training set
- Supplied test set (Set...)
- Cross-validation (Folds: 10)
- Percentage split (%: 66)

More options... (Nom) class: [dropdown]

Start Stop

Result list (right-click for options):

- 09:46:20 - bayes.NaiveBayes

Classifier output:

| | | | | | | | |
|-----|-----------|-----------|---|--------|--------|--------|--------|
| 965 | 3:Netral | 3:Netral | | 0.043 | 0.462 | *0.48 | 0.015 |
| 966 | 3:Netral | 3:Netral | | 0 | 0.004 | *0.996 | 0 |
| 967 | 4:Tanya | 4:Tanya | | 0 | 0 | 0 | *1 |
| 968 | 1:Negatif | 3:Netral | + | 0.042 | 0.29 | *0.661 | 0.007 |
| 969 | 4:Tanya | 1:Negatif | + | *0.729 | 0.15 | 0.105 | 0.016 |
| 970 | 3:Netral | 1:Negatif | + | *0.567 | 0.074 | 0.294 | 0.064 |
| 971 | 4:Tanya | 4:Tanya | | 0 | 0 | 0 | *1 |
| 972 | 1:Negatif | 1:Negatif | | *0.986 | 0.01 | 0.003 | 0.001 |
| 973 | 2:Positif | 2:Positif | | 0.006 | *0.959 | 0.035 | 0 |
| 974 | 4:Tanya | 4:Tanya | | 0.001 | 0 | 0.004 | *0.995 |
| 975 | 1:Negatif | 2:Positif | + | 0.244 | *0.734 | 0.022 | 0 |
| 976 | 3:Netral | 2:Positif | + | 0.216 | *0.401 | 0.363 | 0.02 |
| 977 | 1:Negatif | 3:Netral | + | 0.164 | 0.312 | *0.472 | 0.052 |
| 978 | 1:Negatif | 1:Negatif | | *0.858 | 0.024 | 0.05 | 0.068 |
| 979 | 1:Negatif | 1:Negatif | | *1 | 0 | 0 | 0 |
| 980 | 3:Netral | 3:Netral | | 0.12 | 0.169 | *0.625 | 0.086 |
| 981 | 2:Positif | 2:Positif | | 0.043 | *0.543 | 0.403 | 0.01 |
| 982 | 4:Tanya | 2:Positif | + | 0.012 | *0.487 | 0.465 | 0.036 |
| 983 | 1:Negatif | 4:Tanya | + | 0.365 | 0.001 | 0.001 | *0.633 |
| 984 | 3:Netral | 3:Netral | | 0.003 | 0.041 | *0.956 | 0.001 |
| 985 | 1:Negatif | 2:Positif | + | 0.052 | *0.633 | 0.3 | 0.015 |
| 986 | 3:Netral | 2:Positif | + | 0.008 | *0.631 | 0.347 | 0.015 |
| 987 | 1:Negatif | 2:Positif | + | 0.115 | *0.533 | 0.351 | 0.001 |
| 988 | 2:Positif | 2:Positif | | 0.277 | *0.52 | 0.19 | 0.012 |
| 989 | 4:Tanya | 4:Tanya | | 0 | 0.001 | 0.002 | *0.998 |
| 990 | 1:Negatif | 3:Netral | + | 0.149 | 0.12 | *0.375 | 0.356 |
| 991 | 3:Netral | 1:Negatif | + | *0.993 | 0.001 | 0.004 | 0.002 |
| 992 | 3:Netral | 2:Positif | + | 0.013 | *0.701 | 0.282 | 0.004 |
| 993 | 3:Netral | 3:Netral | | 0.013 | 0.173 | *0.813 | 0.001 |
| 994 | 2:Positif | 1:Negatif | + | *0.941 | 0.023 | 0.012 | 0.024 |
| 995 | 4:Tanya | 4:Tanya | | 0.003 | 0 | 0 | *0.997 |
| 996 | 1:Negatif | 1:Negatif | | *0.396 | 0.199 | 0.164 | 0.241 |

Gambar 4.13 Contoh hasil prediksi pada Weka

Sedangkan untuk mencari data tweet yang merupakan retweet dilakukan dengan melakukan filtering berdasarkan karakter 'RT @'. Hasil dari klasifikasi dan filterisasi dari keseluruhan tweet tampak pada tabel 4.8:

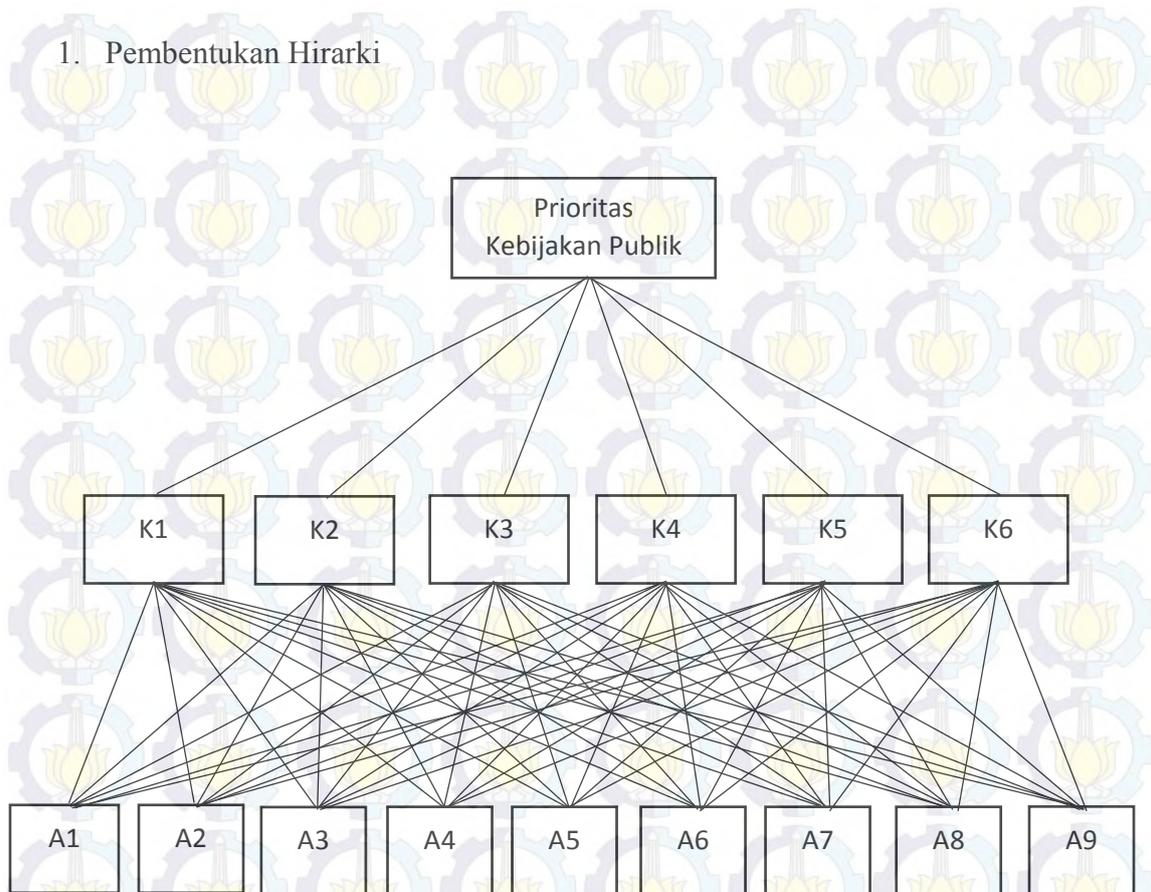
Tabel 4.8 Hasil klasifikasi dan filterisasi

| Kriteria/ Alternatif | Jumlah tweet negatif | Jumlah tweet positif | Jumlah tweet netral | Jumlah tweet tanya | Jumlah retweet | Jumlah direct tweet |
|--|----------------------------|----------------------------|---------------------------|--------------------------|-------------------|---------------------------|
| Memberantas Kemiskinan dan Kelaparan Ekstrem | 11950 | 22297 | 88383 | 8833 | 32048 | 84796 |
| Mewujudkan Pendidikan Dasar untuk Semua | 7198 | 15052 | 84160 | 5990 | 29444 | 73059 |
| Mendorong Kesetaraan Gender dan Pemberdayaan Perempuan | 84 | 908 | 8088 | 145 | 740 | 8225 |
| Menurunkan Angka Kematian Anak | 481 | 1304 | 11601 | 4659 | 3355 | 6890 |
| Meningkatkan Kesehatan Ibu | 154 | 1390 | 4777 | 368 | 1115 | 4864 |
| Memerangi HIV dan AIDS Malaria Serta Penyakit Lainnya | 76 | 1397 | 2954 | 908 | 612 | 2926 |
| Memastikan Kelestarian Lingkungan | 57 | 196 | 3959 | 107 | 1952 | 2218 |
| Mengembangkan Kemitraan Global untuk Pembangunan | 1637 | 5126 | 8895 | 1259 | 2685 | 11805 |

4.6 Penentuan Prioritas Kebijakan Publik

Dari data yang telah diperoleh pada tahap sebelumnya, penentuan prioritas kebijakan publik dengan menggunakan algoritma Analytical Hierarchy Process (AHP) dapat dilakukan dengan dengan langkah-langkah seperti di bawah ini:

1. Pembentukan Hirarki



Gambar 4.1 Pembentukan hirarki

Keterangan:

K1= Jumlah tweet negatif

K2= Jumlah tweet positif

K3= Jumlah tweet netral

K4= Jumlah tweet tanya

K5= Jumlah direct tweet

K6= Jumlah retweet

A1= Memberantas Kemiskinan dan Kelaparan Ekstrem

A2= Mewujudkan Pendidikan Dasar untuk Semua

A3= Mendorong Kesetaraan Gender dan Pemberdayaan Perempuan

A4= Menurunkan Angka Kematian Anak

A5= Meningkatkan Kesehatan Ibu

A6= Memerangi HIV dan AIDS Malaria Serta Penyakit Lainnya

A7= Memastikan Kelestarian Lingkungan

A8= Mengembangkan Kemitraan Global untuk Pembangunan

2. Perbandingan Berpasangan

Dasar utama dari penelitian ini adalah, membuat sebuah pemeringkatan kebijakan publik berdasarkan opini masyarakat terhadap sebuah kebijakan publik yang paling negatif. Selanjutnya dibuat matriks perbandingan berpasangan dengan kriteria di bawah ini:

- a) Jumlah tweet negatif sedikit lebih penting dari jumlah tweet positif.
- b) Jumlah tweet negatif sedikit lebih penting dari jumlah tweet netral.
- c) Jumlah tweet negatif sedikit lebih penting dari jumlah direct tweet.
- d) Jumlah tweet negatif sedikit lebih penting dari jumlah retweet.
- e) Jumlah tweet negatif sedikit lebih penting dari jumlah tweet tanya.
- f) Jumlah tweet positif sedikit lebih penting dari jumlah tweet netral
- g) Jumlah direct tweet sedikit lebih penting dari jumlah tweet tanya

Dengan berpedoman pada skala 1-9 Saaty, L Thomas maka dapat dibuat matriks perbandingan berpasangan seperti tampak pada tabel 4.9:

Tabel 4.9 Matrik perbandingan berpasangan

| Kriteria | Jumlah Tweet Negatif | Jumlah Tweet Positif | Jumlah Tweet Netral | Jumlah Direct Tweet | Jumlah Retweet | Jumlah Tweet Tanya |
|----------------------------|----------------------------|----------------------------|---------------------------|---------------------------|-------------------|--------------------------|
| Jumlah Tweet Negatif | 1,0000 | 3,0000 | 3,0000 | 3,0000 | 3,0000 | 3,0000 |
| Jumlah Tweet Positif | 0,3333 | 1,0000 | 3,0000 | 1,0000 | 1,0000 | 1,0000 |
| Jumlah Tweet Netral | 0,3333 | 0,3333 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| Jumlah Direct Tweet | 0,3333 | 1,0000 | 1,0000 | 1,0000 | 3,0000 | 1,0000 |
| Jumlah Retweet | 0,3333 | 1,0000 | 1,0000 | 0,3333 | 1,0000 | 1,0000 |
| Jumlah Tweet Tanya | 0,3333 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| Jumlah | 2,6667 | 7,3333 | 10,0000 | 7,3333 | 10,0000 | 8,0000 |

3. Normalisasi Matriks Perbandingan Berpasangan

Tahap berikutnya adalah melakukan normalisasi pada matriks perbandingan berpasangan dengan cara membagi setiap nilai pada kolom matrik dengan hasil penjumlahan kolom yang bersesuaian, seperti tampak pada tabel 4.10.

Tabel 4.10 Normalisasi matrik perbandingan berpasangan

| Kriteria | Jumlah Tweet Negatif | Jumlah Tweet Positif | Jumlah Tweet Netral | Jumlah Direct Tweet | Jumlah Retweet | Jumlah Tweet Tanya | Jumlah | Bobot |
|----------------------|----------------------|----------------------|---------------------|---------------------|----------------|--------------------|--------|--------|
| Jumlah Tweet Negatif | 0,3750 | 0,4091 | 0,3000 | 0,4091 | 0,3000 | 0,3750 | 2,1682 | 0,3614 |
| Jumlah Tweet Positif | 0,1250 | 0,1364 | 0,3000 | 0,1364 | 0,1000 | 0,1250 | 0,9227 | 0,1538 |
| Jumlah Tweet Netral | 0,1250 | 0,0455 | 0,1000 | 0,1364 | 0,1000 | 0,1250 | 0,6318 | 0,1053 |
| Jumlah Direct Tweet | 0,1250 | 0,1364 | 0,1000 | 0,1364 | 0,3000 | 0,1250 | 0,9227 | 0,1538 |
| Jumlah Retweet | 0,1250 | 0,1364 | 0,1000 | 0,0455 | 0,1000 | 0,1250 | 0,6318 | 0,1053 |
| Jumlah Tweet Tanya | 0,1250 | 0,1364 | 0,1000 | 0,1364 | 0,1000 | 0,1250 | 0,7227 | 0,1205 |

4. Pemeriksaan Konsistensi Ratio (CR)

Pemeriksaan konsistensi dilakukan untuk melihat apakah matriks berpasangan yang sudah kita buat nilainya konsisten. Hal ini terpenuhi jika nilai $CR \leq 0.1$

Nilai Eigen Maksimum

$$\lambda_{maks} = 6,2889$$

Nilai Konsistensi Index (CI)

$$CI = (\lambda_{maks} - n) / (n - 1)$$

$$CI = 0,057777778$$

Nilai Consistency Ratio (CR)

Nilai RI diambil dari tabel Random Index. Untuk Matriks yang mempunyai ordo 6 nilainya adalah=1,24

$$CR=CI/RI$$

CR= 0,046594982 (Nilai CR \leq 0,1 sehingga konsisten)

5. Evaluasi Bobot Keseluruhan

Tabel 4.11 Tabel evaluasi bobot keseluruhan

| Kriteria/ Alternatif | Jumlah Tweet Negatif | Jumlah Tweet Positif | Jumlah Tweet Netral | Jumlah Direct Tweet | Jumlah Re tweet | Jumlah Tweet Tanya | Nilai akhir | Peringkat/ Prioritas |
|-------------------------|----------------------------|----------------------------|---------------------------|---------------------------|-----------------------|--------------------------|----------------|-------------------------|
| | 0,3614 | 0,1538 | 0,1053 | 0,1538 | 0,1053 | 0,1205 | | |
| A1 | 11950 | 22297 | 88383 | 84796 | 32048 | 8833 | 34533,63 | 1 |
| A2 | 7198 | 15052 | 84160 | 73059 | 29444 | 5990 | 28835,87 | 2 |
| A3 | 84 | 908 | 8088 | 8225 | 740 | 145 | 2381,98 | 5 |
| A4 | 481 | 1304 | 11601 | 6890 | 3355 | 4659 | 3570,06 | 4 |
| A5 | 154 | 1390 | 4777 | 4864 | 1115 | 368 | 1682,21 | 6 |
| A6 | 76 | 1397 | 2954 | 2926 | 612 | 908 | 1177,17 | 7 |
| A7 | 57 | 196 | 3959 | 2218 | 1952 | 107 | 1027,18 | 8 |
| A8 | 1637 | 5126 | 8895 | 11805 | 2685 | 1259 | 4566,40 | 3 |

Dari evaluasi bobot keseluruhan terlihat bahwa urutan atau prioritas kebijakan publik yang dapat diambil berdasarkan algoritma Analytical Hierarchy Process (AHP) adalah sebagai berikut:

1. Memberantas Kemiskinan dan Kelaparan Ekstrem
2. Mewujudkan Pendidikan Dasar untuk Semua
3. Mengembangkan Kemitraan Global untuk Pembangunan
4. Menurunkan Angka Kematian Anak
5. Mendorong Kesetaraan Gender dan Pemberdayaan Perempuan
6. Meningkatkan Kesehatan Ibu
7. Memerangi HIV dan AIDS Malaria Serta Penyakit Lainnya
8. Memastikan Kelestarian Lingkungan

BAB V KESIMPULAN

Dari hasil penelitian yang telah dilakukan dalam tesis ini, dapat diambil kesimpulan sebagai berikut:

1. Penelitian yang telah dilakukan berhasil membuktikan bahwa analisa *microblogging* dapat menjadi bahan pertimbangan dan kajian untuk menentukan prioritas sebuah kebijakan publik yang mendekati aspirasi dan keinginan masyarakat.
2. Dapat dilihat juga bahwa masyarakat sangat mudah beropini terhadap hal-hal yang bersentuhan langsung dan mempengaruhi kehidupan mereka sehari-hari, dibuktikan dengan masalah kemiskinan dan pendidikan menduduki peringkat satu dan dua dalam perolehan tweet yang berkorelasi dengan pemeringkatan prioritas kebijakan publik.
3. Hasil dari penelitian ini menempatkan memberantas kemiskinan dan kelaparan ekstrem menjadi prioritas kebijakan publik yang utama.
4. Akurasi tertinggi klasifikasi tweet ke dalam empat kelas (negatif, positif, netral dan tanya) yang bisa dicapai dalam penelitian ini adalah sebesar 72,1 % dengan menggunakan algoritma Naïve Bayes, pembobotan yang digunakan adalah TF-IDF dengan minimal term frequency=3, tanpa melakukan proses stopwords, tokenizer yang dipakai adalah= word tokenizer dengan percentage split sebesar 66%.
5. Penggunaan *stopword removal* untuk klasifikasi tweet (negatif, positif, netral dan tanya) pada penelitian ini cenderung menurunkan nilai akurasi (1,67%).

DAFTAR PUSTAKA

Aliandu, P. 2013. Twitter Used by Indonesian President: An Sentiment Analysis of Timeline. Dalam Information Systems International Conference (ISICO), 2 – 4 December 2013.al. 713-717. Bali: Indonesia.

Berry, M.W. & Kogan, J. 2010. Text Mining Application and theory. WILEY: United Kingdom.

Feldman, R & Sanger, J. 2007. The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press: New York.

Han, J & Kamber, M.2006 Data Mining: Concepts and Techniques Second Edition. Morgan Kaufmann publisher: San Francisco.

Kusrini. 2007. Konsep dan Aplikasi Sistem Pendukung Keputusan. Yogyakarta: Andi.

Mandala, Rila dan Setiawan, Hendra. 2002. Peningkatan Performansi Sistem Temu-Kembali Informasi dengan Perluasan Query Secara Otomatis, Bandung: Departemen Teknik Informatika Institut Teknologi Bandung.

Mustopadidjaja. 2002. Manajemen Proses Kebijakan Publik. Jakarta: Lembaga Administrasi Negara.

Nazief dan Adriani. 1996. Confix Stripping: Approach to Stemming Algorithm for Bahasa Indonesia. Technical report, Faculty of Computer Science, University of Indonesia, Depok, 1996

Pang, B., Lee, L., & Vithyanathan, S. (2002). SentimentClassification Using Machine Learning Techniques. Dalam Proceedings of The ACL-02 conference on Empirical methods in natural language processing, pp. 79-86. Stroudsburg: Association for computationalLinguistic.

Prasad, S. 2011. Micro-blogging Sentiment Analysis Using Bayesian Classification Methods.

Saraswati, Ni Wayan Sumartini. 2011, “Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis”, Tesis Program Pascasarjana Universitas Udayana, Denpasar

Sunni, I. & Widiantoro, D. H. 2012. Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik

Tala, F. Z. (2003). A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. M.S. thesis. M.Sc. Thesis. Master of Logic Project. Institute for Logic, Language and Computation. Universiteti van Amsterdam the Netherlands

Wang, A. H. 2010. Don't Follow Me: Twitter Spam Detection. Proceedings of 5th International Conference on Security and Cryptography (SECRYPT) Athens 2010: pp. 1-10. California: IEEE.

BIODATA PENULIS



Penulis tesis ini bernama Mohammad Khoiron, anak ketiga dari tujuh bersaudara, lahir 25 Februari 1976 di kota Blitar, kota yang berada 146 km sebelah selatan Surabaya, propinsi Jawa Timur Indonesia. Penulis menghabiskan masa kecilnya dengan sekolah di kota Blitar. Penulis dinyatakan lulus dari SMA Negeri 1 Blitar pada tahun 1994 dan melanjutkan studi ke S1 Teknik Sistem Perkapalan Institut Teknologi Sepuluh Nopember Surabaya hingga lulus pada tahun 2004. Selanjutnya penulis mengambil studi lagi di Institut Sains dan Teknologi Palapa Malang Jurusan Teknik Informatika hingga lulus pada tahun 2007. Saat ini penulis tinggal di kota Mojokerto dan berprofesi sebagai guru di UPT SMK Negeri 1 Mojokerto. Selain mengajar, penulis juga aktif di bidang teknologi informasi maupun komunikasi di kota Mojokerto. Keinginan yang kuat untuk menimba ilmu dan profesi yang penulis jalani, mengantarkan penulis untuk kembali ke bangku kuliah dengan mengambil studi S2 Bidang Keahlian Telematika-CIO, Jurusan Teknik Elektro, Institut Teknologi Sepuluh Nopember di tahun 2014. Penulis dapat dihubungi melalui email di m.khoiron.w@gmail.com.