

# Klasifikasi Berita Menggunakan Metode Naive Bayes pada Surat Kabar Harian Radar Tarakan

Anny Susilowati, Aida indriani, Muhammad Hafid

**Abstraksi**— Mengarsipkan berita secara konvensional biasanya hanya sekadar disimpan dalam file, Compact Disk (CD) dan atau media internal server perusahaan. Sebagai perusahaan yang terus berkembang dalam menyajikan berita, maka mengarsipkan sekaligus mengklasifikasikan dalam kategori tertentu mutlak diperlukan. Mengingat selain tuntutan, juga keperluan akan berita-berita yang terkategori dan tersistem secara otomatis akan memudahkan dalam pengelolaannya.

Disini penulis menggunakan berita dari data arsip Redaksi Surat Kabar Harian Radar Tarakan tahun 2015 sebanyak 100 data. Seratus data ini dapat digunakan sebagai data training dan data testing. Proses diawali dengan menghitung probabilitas tiap-tiap kategori berita, yaitu politik, ekonomi, hiburan, kriminal, dan olahraga. Setelah itu baik data training maupun data testing melewati tahapan praproses, yaitu *case folding*, *tokenizing*, *stoplist* dan *stemming*. Data testing yang telah melewati praproses kemudian dilakukan perhitungan Naive Bayes Classifier (NBC). Hasilnya masih diuji lagi untuk mengetahui seberapa besar persentase yang dihasilkan.

Penulis telah melakukan pengujian sebanyak 10 kali dengan jumlah data training 75 dan data testing 25. Pada pengujian ini diperoleh tingkat persentase akurasi tertinggi 76% atas semua kategori, baik politik, ekonomi, hiburan, kriminal dan olahraga. Sedangkan rata-rata persentase akurasi sebesar 62%. Penulis juga melakukan sejumlah pengujian lainnya, sehingga diperoleh kesimpulan bahwa berita yang dapat diklasifikasi tergantung pada pola awal dari data berita sampel yang disimpan dalam database. Sehingga semakin banyak data berita maka semakin banyak pula pola data berita yang terbentuk.

**Kata Kunci**— klasifikasi, berita, naive bayes, radar tarakan

## I. PENDAHULUAN

Industri media berkembang dengan pesat bersamaan dengan perkembangan teknologi informasi. Dengan teknologi semakin memudahkan manusia untuk mengetahui sumber-sumber informasi. Tak terkecuali pada industri media cetak berupa surat kabar harian. Informasi yang disajikan setiap hari, tak urung menjadi bahasan menarik untuk diulas, dianalisa dan dibicarakan berbagai kalangan masyarakat.

Surat Kabar Harian Radar Tarakan adalah salah satu media cetak yang jangkauannya meliputi wilayah Kalimantan Utara (Kaltara) yang terdiri dari Kota Tarakan, Kabupaten Bulungan, Kabupaten Nunukan, Kabupaten Malinau dan Kabupaten Tana Tidung. Setiap hari setidaknya memproduksi 100 berita yang terdiri dari berita politik, ekonomi, sosial budaya, hiburan, pembangunan, teknologi dan sebagainya. Pengelolaan yang baik terhadap berita-berita yang dihasilkan Surat Kabar Harian

Radar Tarakan membuat media ini menjadi tujuan penelitian berbagai kalangan.

Menyadari betapa pentingnya nilai sebuah berita, sistem pengelolaan pengarsipan yang baik juga sudah diterapkan. Meskipun masih berbentuk manual. Sebatas menyimpan data fisik koran, menyimpan berita-berita ke dalam CD dan diurutkan berdasarkan tanggal, bulan dan tahun berita dibuat. Persoalan baru muncul manakala jajaran redaksi berniat untuk mengemas berita lebih mendalam atau *news indepth* dan butuh pendalaman materi dari berita-berita yang pernah diterbitkan Surat Kabar Harian Radar Tarakan. Pencarian yang sulit berdasarkan tema ditambah lagi dengan durasi yang lama memunculkan sejumlah keluhan. Hal ini memunculkan ide penulis untuk membuat sebuah aplikasi yang bisa memudahkan pencarian dan pengklasifikasian berita.

## II. TINJAUAN PUSTAKA

Text mining adalah proses penganalisisan text guna menyoroti informasi yang bermanfaat untuk tujuan tertentu. Proses data mining untuk dokumen atau text memerlukan lebih banyak tahapan, mengingat data text memiliki karakteristik yang lebih kompleks daripada data biasa.

Adapun karakteristik dokumen teks cirinya adalah, database teks yang berukuran besar, memiliki dimensi yang tinggi yaitu satu kata merupakan satu dimensi, mengandung kumpulan kata yang saling terkait (frase) dan antara kumpulan kata yang satu dengan yang lain dapat memiliki arti yang berbeda. Ciri lainnya adalah banyak mengandung kata ataupun arti yang bias (ambiguity) [1].

Berdasarkan ketidakteraturan struktur data teks, maka proses teks mining memerlukan beberapa tahap awal yang pada intinya adalah mempersiapkan agar teks dapat diubah menjadi lebih terstruktur. Bentuk perubahan yang dilakukan adalah ke dalam spreadsheet, kolom menunjuk dokumen dan baris menunjuk kata, sedangkan selnya menunjuk frekuensi kata dalam dokumen.

Klasifikasi adalah suatu proses pengkategorian yang dilakukan terhadap sekumpulan dokumen. Klasifikasi sangat penting untuk kemudahan pengguna dalam melakukan pencarian dokumen. Proses klasifikasi diawali dengan membagi koleksi dokumen menjadi dua bagian utama yaitu data latih dan data uji. Dari data latih dengan menggunakan metode tertentu diperoleh model klasifikasi yang nantinya akan digunakan untuk penentuan kelas terhadap data uji [2]. Banyak sekali metode-metode yang dapat digunakan dalam klasifikasi dokumen, antara lain: K-NN, Rocchio, WAK-NN, Naive Bayes Classifier (NBC) dan lain-lain.

Tugas pokok pada tahapan praproses di dalam klasifikasi adalah membangun index dari koleksi dokumen. Pembuatan inverted index harus melibatkan konsep linguistic processing yang bertujuan mengekstrak term term penting dari dokumen yang direpresentasikan sebagai bag-of-words. Ekstraksi term biasanya melibatkan empat operasi utama berikut:

1. Case folding

Case Folding adalah mengubah semua huruf dalam teks dokumen menjadi huruf kecil. Karakter selain huruf dihilangkan serta dianggap delimiter [3].

2. Pemisahan rangkaian term (tokenization).

Tokenizing (tokenisasi) yaitu memisahkan kata per kata pada kalimat menjadi token atau potongan kata tunggal. Tahapan ini juga menghilangkan karakter tertentu seperti tanda baca ~@#%&\* \_=+|\\[]{} <>.!?:/()+;0 dan mengubah semua token ke bentuk huruf kecil (lower case) [4].

3. Penghapusan (stop words)

Stop words atau penghapusan kata dilakukan untuk membuang kata-kata yang tidak mempunyai hubungan dengan dokumen yang terdapat pada database. Contoh stop words adalah, ada, adalah, adanya, adapun, agak, dia, saya, mereka, bagaimana, dan lainnya [5].

4. Stemming

Stemming atau pembuangan awalan atau akhiran dari suatu kata sehingga mendapatkan bentuk kata dasarnya. Sebagai contoh, kata mengecam setelah melalui proses stemming menjadi kata kecam. Demikian pula kata dijadikan akan mengambil bentuk kata dasarnya jadi [6].

Dalam penelitian ini stemming yang digunakan adalah stemming Arifin dan Setiono. Proses dalam stemming Arifin dan Setiono didahului dengan pembacaan tiap kata dari file sampel [7]. Sehingga input dari algoritma ini adalah sebuah kata yang kemudian dilakukan:

- a. Pemeriksaan semua kemungkinan bentuk kata. Setiap kata diasumsikan memiliki 2 awalan atau prefix dan 2 akhiran atau sufiks. Sehingga bentuknya menjadi : prefix 1 + prefix 2 + kata dasar + sufiks 3 + sufiks 2 + sufiks 1. Seandainya kata tersebut tidak memiliki imbuhan sebanyak imbuhan diatas, maka imbuhan yang kosong diberi tanda x untuk prefix dan dan diberi tanda xx untuk sufiks.
- b. Pemotongan dilakukan secara berurutan yaitu AW:AW (Awalan), AK : AK (Akhiran), KD : KD (Kata Dasar).
  - AW I, hasilnya disimpan pada p1 (prefiks 1)
  - AW II, hasilnya disimpan pada p2 (prefiks 2)
    - AK I, hasilnya disimpan pada s1 (sufiks 1)
    - AK II, hasilnya disimpan pada s2 (sufiks 2)
    - AK III, hasilnya disimpan pada s3 (sufiks 3)

Pada setiap tahap pemotongan di atas diikuti dengan pemeriksaan di kamus apakah hasil pemotongan tersebut sudah berada dalam bentuk dasar. Jika pemeriksaan berhasil maka proses dinyatakan selesai dan tidak perlu melanjutkan proses pemotongan imbuhan lainnya. Contoh pemenggalan kata “mempermainkannya”:

- Langkah 1  
Cek apakah kata ada dalam kamus Ya : Sukses. Jika Tidak : Lakukan pemotongan AW I, hasilnya “permainkannya”.
- Langkah 2  
Cek apakah kata ada dalam kamus Ya : Sukses. Jika Tidak : Lakukan pemotongan AW II, hasilnya “mainkannya”.

- Langkah 3  
Cek apakah kata ada dalam kamus Ya : Sukses. Jika Tidak : Lakukan pemotongan AK I, hasilnya “mainkan”.
  - Langkah 4  
Cek apakah kata ada dalam kamus Ya : Sukses. Jika Tidak : Lakukan pemotongan AK II, hasilnya “main”.
  - Langkah 5  
Cek apakah kata ada dalam kamus Ya : Sukses. Jika Tidak : Lakukan pemotongan AK III. Dalam hal ini AK III tidak ada, sehingga hasilnya tetap “main”.
- Jika sampai pada pemotongan AK III belum juga ditemukan di kamus, maka dilakukan proses kombinasi. KD yang dihasilkan dikombinasikan dengan imbuhan-imbuhan dalam 12 konfigurasi sebagai berikut:
- a. KD
  - b. KD + AK III
  - c. KD + AK III + AK II
  - d. KD + AK III + AK II + AK I
  - e. AW I + AW II + KD
  - f. AW I + AW II + KD + AK III
  - g. AW I + AW II + KD + AK III + AK II
  - h. AW I + AW II + KD + AK III + AK II + AK I
  - i. AW II + KD
  - j. AW II + KD + AK III
  - k. AW II + KD + AK III + AK II
  - l. AW II + KD + AK III + AK II + AK I

Sebenarnya kombinasi a, b, c, d, h dan l sudah diperiksa pada tahap sebelumnya, karena kombinasi ini adalah hasil pemotongan bertahap tersebut. Dengan demikian kombinasi yang masih perlu dilakukan tinggal 6. Yakni, pada kombinasi-kombinasi yang belum dilakukan e, f, g, i, j, dan k. Tentunya bila hasil pemeriksaan suatu kombinasi ada maka pemeriksaan pada kombinasi lainnya sudah tidak diperlukan lagi.

Pemeriksaan 12 kombinasi ini diperlukan karena adanya over stemming pada algoritma pemotongan imbuhan. Kelemahan ini berakibat pada pemotongan bagian kata yang sebenarnya adalah milik kata dasar itu sendiri yang kebetulan mirip dengan salah satu jenis imbuhan yang ada.

Stemming juga diartikan proses pencarian akar (root) kata dari tiap kata yaitu dengan mengembalikan suatu kata berimbuhan ke bentuk dasarnya (stem). Untuk pemrosesan pada bahasa Indonesia, proses stemming dilakukan dengan menghilangkan imbuhan yang mengawali dan mengakhiri kata, sehingga diperoleh bentuk dasar dari kata tersebut.

Berita didefinisikan sebagai peristiwa yang dilaporkan. Segala yang didapat di lapangan dan sedang dipersiapkan untuk dilaporkan belum dapat disebut berita. Wartawan yang menonton dan menyaksikan peristiwa belum tentu telah menemukan peristiwa. Wartawan harus bisa menemukan peristiwa setelah memahami proses atau jalan cerita, yaitu

- harus tahu apa (*what*) apa yang terjadi,
  - siapa (*who*) yang terlibat.
  - Bagaimana kejadian itu terjadi (*how*),
  - kapan (*when*) terjadi,
  - dimana (*where*) peristiwa itu terjadi dan
  - mengapa (*why*) sampai terjadi.
- Keenam hal itu merupakan unsur berita [8].

Seperti diungkapkan diatas, suatu peristiwa patut diangkat menjadi sebuah berita jika memang memiliki nilai berita. Unsur nilai dari suatu berita antara lain:

1. Kebermaknaan (*significance*). Kejadian yang dapat memengaruhi kehidupan orang banyak atau pembaca. Contoh, kenaikan BBM, tariff TDL, biaya pulsa telepon.
2. Besaran (*magnitude*). Kejadian menyangkut angka angka yang berarti bagi kehidupan orang banyak. Contoh, kasus korupsi triliunan yang merugikan Negara.
3. Kebaruan (*Timeless*). Suatu kejadian menyangkut peristiwa yang baru terjadi. Contoh, gempa bumi dan tsunami yang melanda Nangroe Aceh Darussalam dan Sumatera Utara.
4. Kedekatan (*Proximity*). Suatu kejadian yang berada di dekat pembaca. Kedekatan itu bias secara geografis atau emosional. Contoh, peristiwa tabrakan mobil, kebakaran atau pembunuhan.
5. Kemasyuran/sisi manusiawi (*prominence/human interest*). Suatu kejadian yang memberi sentuhan rasa kepada pembaca. Mengungkap peristiwa orang terkenal, figure public, atau masyarakat biasa dalam peristiwa luar biasa. Contoh, Desi Ratnasari mengalami kecelakaan.

Suatu berita dikatakan bermutu apabila isi beritanya memuat kaidah-kaidah jurnalistik, antara lain:

1. Faktualitas, berisi aktualitas, kelengkapan, kejelasan dan objektivitas.
2. Kaidah nilai berita, berisi besarnya peristiwa, ketermasaan, kedekatan, konflik, konsekuensi, kemenonjolan, seksualitas dan human interest.
3. Kaidah politik pemberitaan, setiap media berisi kepentingan politik pemberitaan yang berbeda.
4. Mutu pemberitaan dilihat dari isi, sistematika dan bahasa. Nilai isi media berkaitan dengan penerapan nilai berita. Nilai sistematika berhubungan dengan struktur dan teknik penyajian. Nilai bahasa berhubungan dengan kelancaran dan segi kenyamanan bagi pembaca.

Saat ini berita jurnalistik terbagi dalam 3 jenis, yaitu

1. Berita langsung (*straight/hard/spot news*). Berita langsung digunakan untuk menyampaikan kejadian penting kepada pembaca.
2. Berita ringan (*soft news*). Berita ringan tidak mengutamakan unsur penting yang hendak diberitakan, tetapi mengenai sesuatu yang menarik. Misalnya kehidupan artis cantik Prilly Latuconsina.
3. Berita kisah (*feature*). Tulisan tentang kejadian yang bisa menyentuh perasaan atau menambah pengetahuan pembaca lewat penjelasan lengkap dan mendalam.

Adapun ciri-ciri berita jurnalistik adalah, singkat, padat, sederhana, lugas, menarik dan jelas.

Naive Bayes Classifier atau NBC merupakan sebuah metode klasifikasi yang berakar pada teorema bayes. Ciri utama dari NBC adalah asumsi yang sangat kuat atau naif akan independensi dari masing-masing kondisi atau kejadian [9].

Metode Naive Bayes Classifier (NBC) menempuh 2 tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan vocabulary, yaitu kata yang muncul dalam koleksi dokumen sampel yang sedapat

mungkin dapat menjadi representasi dokumen. Selanjutnya adalah penentuan probabilitas prior bagi tiap kategori berdasarkan sampel dokumen. Pada tahap klasifikasi ditentukan nilai kategori dari suatu dokumen berdasarkan term yang muncul dalam dokumen yang diklasifikasi [10].

Lebih jelasnya sebagai berikut, bila diasumsikan koleksi dokumen  $D = \{d_i \mid i=1,2,\dots \mid D\} = \{d_1, d_2, \dots, d_{|D|}\}$  dan koleksi kategori  $V = \{v_j \mid j=1,2,\dots \mid V\} = \{v_1, v_2, \dots, v_{|V|}\}$ . Klasifikasi NBC dilakukan dengan cara mencari probabilitas  $P(V=v_j \mid D=d_i)$ , yaitu probabilitas kategori  $v_j$  jika diketahui dokumen di. Dokumen dipandang sebagai tuple dari kata-kata dalam dokumen, yaitu  $\langle a_1, a_2, \dots, a_n \rangle$ , yang frekuensi kemunculannya diasumsikan sebagai variable random dengan distribusi probabilitas Benouli (McCallum and Nigam, 1998). Selanjutnya klasifikasi dokumen adalah mencari nilai maksimum dari :

$$V_{MAP} = \operatorname{argmax} P(v_j \mid a_1, a_2, \dots, a_n) \quad (2.1)$$

Teorema bayes menyatakan tentang probabilitas bersyarat:

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)} \quad (2.2)$$

Dengan menerapkan teorema bayes persamaan (2.1) dapat ditulis:

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(a_1, a_2, \dots, a_n \mid v_j) P(v_j) \quad (2.3)$$

Karena nilai  $P(a_1, a_2, \dots, a_n)$  untuk semua  $v_j$  besarnya sama maka nilainya dapat diabaikan, sehingga persamaan (2.3) menjadi:

$$V_{MAP} = \operatorname{argmax}_{v_j \in V} P(a_1, a_2, \dots, a_n \mid v_j) P(v_j) \quad (2.4)$$

Dengan mengasumsikan bahwa setiap kata dalam  $\langle a_1, a_2, \dots, a_n \rangle$  adalah independent, maka  $P(a_1, a_2, \dots, a_n \mid v_j)$  dalam persamaan (4) dapat ditulis sebagai :

$$P(a_1, a_2, \dots, a_n \mid v_j) P(v_j) = \prod_i P(a_i \mid v_j) P(v_j) \quad (5)$$

Sehingga persamaan (4) dapat ditulis:

$$V_{MAP} = \operatorname{argmax}_{v_j} P(v_j) \prod_i P(a_i \mid v_j) \quad (6)$$

nilai  $P(v_j)$  ditentukan saat pelatihan, yang nilainya didekati dengan :

$$P(v_j) = \frac{|doc_j|}{n + |vocabulary|} \quad (7)$$

dimana  $|doc_j|$  adalah banyaknya dokumen yang memiliki kategori  $j$  dalam pelatihan, sedangkan  $|vocabulary|$  banyaknya dokumen dalam contoh yang digunakan untuk pelatihan.

Untuk nilai  $P(w_k \mid v_j)$  yaitu probabilitas kata  $w_k$  dalam kategori  $j$  ditentukan dengan :

$$P(w_k \mid v_j) = \frac{n_k + 1}{n + |vocabulary|}$$

dimana  $n_k$  adalah frekuensi munculnya kata  $w_k$  dalam dokumen yang berkategori  $v_j$ , sedangkan nilai  $n$  adalah banyaknya seluruh kata dalam dokumen berkategori  $v_j$  dan  $|vocabulary|$  adalah banyaknya kata dalam contoh pelatihan.

Pada tanggal 1 Agustus 2001 Surat Kabar Harian (SKH) Radar Tarakan di-launching perdana. Mengawali kiprahnya sebagai salah satu koran grup Kaltim Post dan berinduk kepada grup Jawa Pos, redaksi Radar Tarakan menempati ruko di kawasan bisnis Jl Yos Sudarso, Tarakan. Bukan hanya awak redaksi, iklan, pemasaran dan percetakan pun menempati areal yang sama.

Seiring waktu koran yang memiliki tagline Koran Harian terbesar di Kalimantan Utara ini pindah ke gedung baru di kawasan yang digadang-gadang sebagai kawasan perkantoran Jl Mulawarman. Persis di depan pintu gerbang Bandara Juwata Tarakan.

Setiap hari Koran Radar Tarakan memproduksi sekitar 100 berita, yang terdiri dari berita lokal, nasional, regional dan internasional. Berita-berita tersebut mulai dari bidang politik, ekonomi, sosial budaya, pertahanan dan keamanan, olahraga, hiburan dan sebagainya.

Di usia Radar Tarakan yang kini memasuki 16 tahun, Radar Tarakan bukan hanya sekadar koran tetapi telah memiliki beberapa anak perusahaan seperti Kaltara Pos (koran harian kriminal terbesar), Radio RTFM 98.7 MHz, Event Organizer (EO) Radar Tarakan, serta Kaltara Tour and Travel.

### III. ANALISA DAN DESAIN SISTEM

#### A. Penerapan Metode

Pada penelitian ini, untuk proses klasifikasi dokumen menggunakan data berita sebanyak 100 data yang bersumber dari dokumen berita redaksi Surat Kabar Harian Radar Tarakan. 100 data ini dapat menjadi data latih sekaligus data uji.

TABEL I  
TABEL ATRIBUT EVALUASI

---

---

Kode berita : D1
Judul berita : Aktor Asia Penentu Sukses Film Hollywood
Isi berita :
DUA blockbuster Hollywood tayang pada bulan yang sama di tanah air yakni Furious 7 dan Avengers Age Of Ultron. Keduanya sama-sama dinanti dan keduanya pun berhasil membuat bioskop penuh sesak dijejali penonton yang mengantre.
Satu kesamaan yang terlihat dari keduanya tidak lain adalah adanya wajah-wajah Asia yang membintanginya. Seperti yang sudah kita tahu, Fast & Furious 6 di 2013 sudah hadir terlebih dulu dengan Joe Taslim sebagai bintangnya. Mulai dari Sung Kang, Joe Taslim, Tony Ja dan Ali Fazal. Tentunya kehadiran mereka memperkaya tatanan pemain Fast And Furious sekaligus memperbesar pendapatan mereka.
Terbukti penggabungan mereka secara berturut-turut meningkatkan pendapatan Universal studio. Menurut data Box office Mojo sebelumnya pendapatan sekuel Fast And Furious hanya sebesar USD 236 juta. Setelah penggabungan seluruh aktor pentolannya beserta Sung Kang (Korsel) dalam Fast & Furious di 2009, sukses memperoleh pendapatan USD 363 juta dari bioskop seluruh dunia.
Kehebatan wajah Asia kembali dibuktikan dengan kehadiran aktor Indonesia, Joe Taslim dalam Fast And Furious 6. Pendapatan Fast and Furious di 2013 memuai hingga mencapai USD 626 juta. Indonesia menyumbang angka USD 9,1 juta, yang sebelumnya hanya berhasil mengeruk USD 3,1 juta dalam Fast Five. Bertambah tiga kali lipat.
Langkah tersebut kini diikuti Avengers Age Of Ultron. Sutradara Joss Whedon pun menggaet salah satu bintang Asia untuk bermain di dalamnya. Lalu ada juga rumor bahwa salah satu aktor asal Indonesia

---

---

---

---

sudah menyatakan diri bermain dalam film Captain America Civil War yang baru akan tayang 2016 mendatang. (muv/dka/yan)
Kategori : Hiburan

---

---

Praproses data berita dilakukan setelah pengumpulan dokumen data berita. Tahapan-tahapan dalam praproses data meliputi case Praproses data berita dilakukan setelah pengumpulan dokumen folding, tokenizing, stop words dan stemming. Tahapan praproses data berita dapat dijelaskan sebagai berikut:

1. Case Folding adalah mengubah semua huruf dalam teks dokumen menjadi huruf kecil. Karakter selain huruf dihilangkan serta dianggap delimiter. Berikut bentuk dokumen data berita sebelum proses case folding dilakukan:

TABEL II  
TABEL DOKUMEN DATA BERITA SEBELUM CASE FOLDING

---

---

ID Berita: D1
Judul Berita: Donald Trump Investasi di Indonesia
Isi Berita:
JAKARTA - Ketua Umum DPP Partai Persatuan Pembangunan (PPP) hasil Mukhtar Surabaya, M Romahurmuziy mengancam Donald Trump yang menyerukan pencegahan total dan menyeluruh orang-orang muslim memasuki Amerika Serikat (AS) dan pengawasan terhadap masjid-masjid di AS. Menurut Rommy, sapaan Romahurmuziy, pernyataan Trump itu menyedihkan. Pernyataan Trump itu karena sikap apriori terhadap umat Islam. Pernyataan tersebut dapat memancing kemarahan umat Islam di seluruh dunia. Bahkan bisa memantik timbulnya benturan peradaban, sebagaimana ramalan Samuel P Huntington.
Kategori: Politik
ID Berita: D2
Judul Berita:
Isi Berita:
JAKARTA - Para pengusaha kelapa sawit menjerit. Pasalnya, harga crude palm oil (CPO) di pasaran dunia terus mengalami penurunan. Penurunan harga tersebut, selain disebabkan harga komoditas lainnya seperti kedelai dan sun flower yang sedang merangkak naik di negara penghasil, juga dipicu harga minyak dunia yang terus merosot. Kondisi ini juga membuat harga BBM (premium dan solar) di Indonesia juga mengalami penurunan.
Kategori : Ekonomi
ID Berita : D3
Judul Berita : Jadi Aktivist, Hitamkan Kulit
Isi Berita :
CHELSEA Islan akan hadir lagi di awal tahun ini lewat film terbarunya bersama MNC Pictures, Di Balik 98. Di film tersebut, Chelsea berperan sebagai seorang aktivis mahasiswa bernama Diana.
Mengingat pada saat tragedi Mei 98 Chelsea masih sangat belia, aktris yang membintangi film Merry Riana (2014) ini pun harus melakukan observasi demi menjalani karakternya tersebut.
Kategori : Hiburan
ID Berita : D4
Judul Berita : Jual Sabu, Mantan Siswi SMK Dibekuk
Isi Berita :
TARAKAN-Mantan siswi salah satu Sekolah Menengah Kejuruan (SMK) berinisial In (16), dibekuk unit Reskoba Polres Tarakan, Selasa (14/4) dinihari.
In ditangkap di Jl Halmahera, Kelurahan Pamusian, Kecamatan Tarakan Tengah, terkait kepemilikan sabu bersama rekannya berinisial Ay dan Le.
Kapolres Tarakan AKBP Sarif Rahman melalui Perwira Urusan (Paur) Subbag Humas Iptu Hadi Sucipto menjelaskan, awalnya Unit Reskoba mendapat informasi dari masyarakat yang memberitahukan bahwa di kawasan telaga air, tepatnya di rumah kontrakan belakang Rumah Sakit Umum Tarakan seringkali dijadikan tempat transaksi narkoba.
Kategori : Kriminal
ID Berita : D5
Judul Berita : Wali Kota Cup Meriahkan HUT Tarakan
Isi Berita :

---

---

TARAKAN-Kerinduan pecinta bola voli di Tarakan untuk menyaksikan lagi turnamen bola voli bergengsi di Tarakan, bakal terobati di tahun ini. Dalam rangka memeriahkan HUT Tarakan, Dinas Kebudayaan, Pariwisata, Pemuda dan Olahraga (Disbudparpora) akan menggelar open tournament bertajuk Wali Kota Tarakan Cup. Turnamen ini bernilai plus. Sebab, tidak hanya tim-tim di Kaltara saja yang boleh ambil bagian, panitia juga membuka kesempatan bagi tim-tim di luar Kaltara untuk ikut meramaikan persaingan memperebutkan gelar juara, seperti dari Berau dan Tawau (Malaysia).  
 Kategori : Olahraga

TABEL III  
 TABEL DOKUMEN HASIL CASE FOLDING

I	jakarta ketua umum dpp partai persatuan pembangunan (ppp) hasil muktamar surabaya, m romahurmuzyi mengancam donald trump yang menyerukan pencegahan total dan menyeluruh orang-orang muslim memasuki amerika serikat (as) dan pengawasan terhadap masjid-masjid di as menurut rommy, sapaan romahurmuzyi, pernyataan trump itu menyedatkan pernyataan trump itu karena sikap apriori terhadap umat islam pernyataan tersebut dapat memancing kemarahan umat islam di seluruh dunia bahkan bisa memantik timbulnya benturan peradaban, sebagaimana ramalan samuel p huntington
II	jakarta - para pengusaha kelapa sawit menjerit. pasalnya, harga crude palm oil (cpo) di pasaran dunia terus mengalami penurunan penurunan harga tersebut, selain disebabkan harga komoditas lainnya seperti kedelai dan sun flower yang sedang merangkak naik di negara penghasil, juga dipicu harga minyak dunia yang terus merosot kondisi ini juga membuat harga bbm (premium dan solar) di indonesia juga mengalami penurunan.
III	jakarta - para pengusaha kelapa sawit menjerit. pasalnya, harga crude palm oil (cpo) di pasaran dunia terus mengalami penurunan penurunan harga tersebut, selain disebabkan harga komoditas lainnya seperti kedelai dan sun flower yang sedang merangkak naik di negara penghasil, juga dipicu harga minyak dunia yang terus merosot kondisi ini juga membuat harga bbm (premium dan solar) di indonesia juga mengalami penurunan.
IV	jakarta - para pengusaha kelapa sawit menjerit. pasalnya, harga crude palm oil (cpo) di pasaran dunia terus mengalami penurunan penurunan harga tersebut, selain disebabkan harga komoditas lainnya seperti kedelai dan sun flower yang sedang merangkak naik di negara penghasil, juga dipicu harga minyak dunia yang terus merosot kondisi ini juga membuat harga bbm (premium dan solar) di indonesia juga mengalami penurunan.
V	tarakan - kerinduan pecinta bola voli di tarakan untuk menyaksikan lagi turnamen bola voli bergengsi di tarakan, bakal terobati di tahun ini dalam rangka memeriahkan hut tarakan, dinas kebudayaan, pariwisata, pemuda dan olahraga (disbudparpora) akan menggelar open tournament bertajuk wali kota tarakan cup turnamen ini bernilai plus sebab, tidak hanya tim-tim di kaltara saja yang boleh ambil bagian, panitia juga membuka kesempatan bagi tim-tim di luar kaltara untuk ikut meramaikan persaingan memperebutkan gelar juara, seperti dari berau dan tawau (malaysia).

Pada Tabel II sebelum case folding dilakukan bentuk dari data berita belum huruf kecil semua. Pada Tabel III setelah dilakukan tahapan case folding bentuk data berita pun berubah menjadi kalimat-kalimat huruf kecil semua.

Tokenizing (tokenisasi) yaitu memisahkan kata per kata pada kalimat menjadi token atau potongan kata tunggal. Tahapan ini juga menghilangkan karakter tertentu seperti tanda baca ~@#%&\* =+\[\]\<>.!?:-/+(0). Bentuk hasil dari tokenisasi ditunjukkan pada tabel berikut ini.

TABEL IV  
 TABEL DOKUMEN HASIL CASE TOKENIZING

I	jakarta ketua umum dpp partai persatuan pembangunan ppp hasil muktamar surabaya m romahurmuzyi mengancam donald trump yang menyerukan pencegahan total dan menyeluruh orang-orang muslim memasuki amerika serikat as dan pengawasan terhadap masjid
---	---

II	masjid di as menurut rommy sapaan romahurmuzyi pernyataan trump itu menyedatkan pernyataan trump itu karena sikap apriori terhadap umat islam pernyataan tersebut dapat memancing kemarahan umat islam di seluruh dunia bahkan bisa memantik timbulnya benturan peradaban sebagaimana ramalan samuel p huntington jakarta para pengusaha kelapa sawit menjerit. pasalnya harga crude palm oil cpo di pasaran dunia terus mengalami penurunan penurunan harga tersebut selain disebabkan harga komoditas lainnya seperti kedelai dan sun flower yang sedang merangkak naik di negara penghasil juga dipicu harga minyak dunia yang terus merosot kondisi ini juga membuat harga bbm premium dan solar di indonesia juga mengalami penurunan
III	chelsea islan akan hadir lagi di awal tahun ini lewat film terbarunya bersama mnc pictures di balik 98 di film tersebut, chelsea berperan sebagai seorang aktivis mahasiswa bernama diana mengingat pada saat tragedi mei 98 chelsea masih sangat belia aktris yang membintangi film merry riana 2014 ini pun harus melakukan observasi demi menjalani karakternya tersebut
IV	tarakan mantan siswi salah satu sekolah menengah kejuruan smk berinisial in 16 dibekuk unit reskoba polres tarakan, selasa 14 4 dinihari in ditangkap di jl halmahera kelurahan pamusian kecamatan tarakan tengah terkait kepemilikan sabu bersama rekannya berinisial ay dan le kapolres tarakan akbp sarif rahman melalui perwira urusan paur subbag humas iptu hadi suctipto menjelaskan awalnya unit reskoba mendapat informasi dari masyarakat yang memberitahukan bahwa di kawasan telaga air tepatnya di rumah kontrakan belakang rumah sakit umum tarakan seringkali dijadikan tempat transaksi narkoba
V	tarakan - kerinduan pecinta bola voli di tarakan untuk menyaksikan lagi turnamen bola voli bergengsi di tarakan, bakal terobati di tahun ini dalam rangka memeriahkan hut tarakan, dinas kebudayaan, pariwisata, pemuda dan olahraga (disbudparpora) akan menggelar open tournament bertajuk wali kota tarakan cup turnamen ini bernilai plus sebab, tidak hanya tim-tim di kaltara saja yang boleh ambil bagian, panitia juga membuka kesempatan bagi tim-tim di luar kaltara untuk ikut meramaikan persaingan memperebutkan gelar juara, seperti dari berau dan tawau (malaysia).

Pada Tabel III meski semua huruf sudah kecil semua, tetapi masih terdapat tanda baca. Setelah proses tokenisasi hasilnya dapat dilihat pada Tabel IV.

Stop words atau penghapusan kata dilakukan untuk membuang kata-kata yang tidak mempunyai hubungan dengan dokumen yang terdapat pada database. Contoh stop words adalah, ada, adalah, adanya, adapun, agak, dia, saya, mereka, bagaimana, dan lainnya. Bentuk hasil penghapusan stop words ditunjukkan pada table V sebagai berikut:

TABEL V  
 TABEL DOKUMEN HASIL STOP-WORDS

I	jakarta ketua umum dpp partai persatuan pembangunan ppp hasil muktamar surabaya romahurmuzyi mengancam donald trump menyerukan pencegahan total menyeluruh orang orang muslim memasuki amerika serikat pengawasan masjid masjid rommy sapaan romahurmuzyi pernyataan trump menyedatkan pernyataan trump karena sikap apriori terhadap umat islam pernyataan memancing kemarahan umat islam seluruh dunia memantik benturan peradaban ramalan samuel Huntington
II	jakarta pengusaha kelapa sawit menjerit harga crude palm oil cpo pasaran dunia terus mengalami penurunan penurunan harga disebabkan harga komoditas kedelai sun flower merangkak naik negara penghasil dipicu harga minyak dunia terus merosot kondisi membuat harga bbm premium solar indonesia mengalami penurunan
III	chelsea islan hadir awal tahun lewat film terbarunya mnc pictures dibalik 98 film chelsea berperan aktivis mahasiswa bernama diana mengingat tragedi mei 98 chelsea belia aktris membintangi film merry riana 2014 harus melakukan observasi menjalani karakternya
IV	tarakan mantan siswi sekolah menengah kejuruan smk berinisial in 16 dibekuk unit reskoba polres tarakan selasa 14 4 dinihari in ditangkap jl halmahera kelurahan pamusian kecamatan tarakan tengah terkait kepemilikan sabu bersama rekannya berinisial ay le kapolres tarakan akbp sarif rahman perwira urusan paur subbag humas iptu hadi suctipto

V	menjelaskan unit reskoba mendapat informasi masyarakat memberitahukan kawasan telaga air rumah kontrakan belakang rumah sakit umum tarakan seringkali dijadikan tempat transaksi narkoba tarakan kerinduan pecinta bola voli tarakan menyaksikan turnamen bola voli bergengsi tarakan bakal terobati memeriahkan hut tarakan dinas kebudayaan pariwisata pemuda olahraga disbudparpora menggelar open tournament wali kota tarakan cup turnamen ini bernilai plus sebab tidak hanya tim tim kaltara ambil bagian panitia membuka kesempatan tim tim luar kaltara meramaikan persaingan memperebutkan gelar juara berau tawau Malaysia
---	---

Stemming atau pembuangan awalan atau akhiran dari suatu kata sehingga mendapatkan bentuk kata dasarnya. Sebagai contoh, kata mengecam setelah melalui proses stemming menjadi kata kecam. Demikian pula kata dijadikan akan mengambil bentuk kata dasarnya jadi. Bentuk hasil stemming ditunjukkan pada tabel VI

TABLE VI  
 TABEL DOKUMEN HASIL STEMMING

I	jakarta ketua umum dpp partai satu bangun ppp hasil muktamar surabaya romahurmuziyi kecam donald trump seru cegah orang orang muslim masuk amerika serikat awas masjid masjid rommy sapa romahurmuziyi nyata trump sesat nyata trump sikap apriori umat islam nyata pancing marah umat islam dunia pantik bentur adab ramal samuel Huntington.
II	jakarta usaha kelapa sawit jerit harga crude palm oil cpo pasaran dunia alami turun harga harga komoditas kedelai sun flower rangkai naik negara hasil picu harga minyak dunia terus merosot kondisi buat harga bbm premium solar indonesia alami turun.
III	chelsea islan hadir awal tahun lewat film terbaru mnc pictures dibalik 98 film chelsea peran aktris mahasiswi nama diana ingat tragedi mei 98 chelsea belia aktris bintang film merry riana 2014 laku observasi jalani karakter.
IV	tarakan mantan siswi sekolah tengah juru smk inisial in beku unit reskoba polres tarakan selasa dinihari in tangkap jl halmahera lurah pamusian camat tarakan tengah kait milik sabu rekan inisial ay le kapolres tarakan akbp sarif rahman perwira urusan paur subbag humas iptu hadi sucipto jelas unit reskoba informasi masyarakat beritahu kawasan telaga air rumah kontrak belakang rumah sakit umum tarakan jadi tempat transaksi narkoba.
V	tarakan rindu cinta bola voli tarakan saksikan turnamen bola voli gengsi tarakan bakal obat meriah hut tarakan dinas budaya pariwisata pemuda olahraga disbudparpora gelar open tournament wali kota tarakan cup turnamen nilai plus tim tim kaltara ambil bagian panitia buka kesempatan tim tim luar kaltara ramai saing rebut gelar juara berau tawau Malaysia.

Tabel VI menunjukkan bahwa banyak kata-kata mengalami proses stemming. Misalnya, untuk kata *membuat* mengalami pemotongan awalan *me + m* sehingga hanya tersisa kata *buat*. Demikian pula kata *menggelar* mengalami pemotongan awalan *me +ng* sehingga hanya tersisa *gelar*

Setelah melalui praproses data berita, langkah berikutnya adalah pembagian data berita. Seperti dijelaskan pada poin gambar 3.2.1 dokumen data berita setelah melewati praproses data dibagi menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk mendapatkan model klasifikasi dengan menggunakan metode NBC, sedangkan data uji digunakan untuk mendapatkan klasifikasi dokumen dengan menggunakan model klasifikasi yang dihasilkan oleh metode NBC. Pembagian data yang dilakukan untuk data latih dan data uji yaitu 100%. Pada penelitian ini digunakan 100 data dokumen berita, dokumen tersebut digunakan sebagai data latih sekaligus data uji.

Tahapan awal dalam proses NBC yaitu dengan menghitung probabilitas masing-masing kelas terhadap keseluruhan data latih. Pada table sebelumnya diketahui bahwa data latih sebanyak 100 data berita dengan jumlah kelas sebanyak 5

kategori, yakni politik, ekonomi, kriminal, olahraga, hiburan. Berikut adalah perhitungan probabilitas untuk kelas politik.

Probabilitas disimbolkan sebagai p.  
 $p(\text{politik}) = \frac{\text{Jumlah kelas politik}}{\text{Jumlah data latih}}$

$$p(\text{politik}) = \frac{20}{2208} = 0,0091$$

$$p(\text{ekonomi}) = \frac{20}{2728} = 0,0073$$

$$p(\text{hiburan}) = \frac{20}{3090} = 0,0065$$

$$p(\text{kriminal}) = \frac{20}{2277} = 0,0088$$

$$p(\text{olahraga}) = \frac{20}{2709} = 0,0074$$

Dari perhitungan probabilitas untuk kelas politik diperoleh nilai probabilitas yaitu 0,0091, sedangkan kelas ekonomi sebesar 0,0073, kelas kriminal sebesar 0,0088, kelas olahraga sebesar 0,0074 dan kelas hiburan sebesar 0,0065. Selengkapny terdapat dalam table berikut ini:

TABLE VII  
 TABEL PERHITUNGAN PROBABILITAS

Perhitungan Probabilitas	Politik	Ekonomi	Kriminal	Olahraga	Hiburan
	0,0091	0,0073	0,0088	0,0074	0,0065

Langkah berikutnya adalah melakukan pengklasifikasian data uji dengan menggunakan nilai probabilitas dari setiap kelas yang telah diperoleh. Berikut adalah contoh data forum yang akan diklasifikasikan secara otomatis dengan menggunakan NBC. Data berita forum uji ditunjukkan pada table berikut ini.

TABLE VIII  
 TABEL KOLEKSI DATA BERITA FORUM UJI

Ini Dia International Trailer Fantastic Four TIDAK ada yang meragukan kehebatan Twentieth Century Fox, dalam mengolah setiap film-film yang berada di bawah naungannya. Kini, rumah produksi itu kembali dengan sekuel terbaru dengan salah satu superhero terbaik mereka. Film 'Fantastic Four' yang saat ini tengah digarap di bawah asuhan Josh Trank dan Simon Kinberg itu kembali dengan sesuatu yang baru. Ya, trailer versi Jepang dari film ini resmi dirilis. Tentu saja, hal ini membuat para fans semakin gregetan. Kali ini, sisi alter para superhero ini sengaja ditonjolkan sebagai salah satu konflik yang ada di film ini.
---

Pada tabel diatas terdapat 1 data uji yang belum memiliki kelas. Untuk tahapan awal, data uji juga melalui tahapan praproses seperti yang dilalui data latih. Data uji yang telah melewati tahapan praproses ditunjukkan pada tabel dibawah berikut ini.

TABLE IX  
 TABEL CASE FOLDING

TIDAK ada yang meragukan kehebatan Twentieth Century Fox, dalam mengolah setiap film-film yang berada di bawah naungannya. Kini, rumah produksi itu kembali dengan sekuel terbaru dengan salah satu superhero terbaik mereka Film 'Fantastic Four' yang saat ini tengah digarap di bawah asuhan Josh Trank dan Simon Kinberg itu kembali dengan sesuatu yang baru Ya, trailer versi Jepang dari film ini resmi dirilis Tentu saja, hal ini membuat para fans semakin gregetan Kali ini, sisi alter para superhero ini sengaja ditonjolkan sebagai salah satu konflik yang ada di film ini
--

TABLE X  
 TABEL TOKENIZING

tidak ada yang meragukan kehebatan twentieth century fox dalam mengolah setiap film film yang berada di bawah naungannya kini rumah produksi itu kembali dengan sekuel terbaru dengan salah satu superhero terbaik mereka film fantastic four yang saat ini tengah digarap di bawah asuhan josh trunk dan simon kinberg itu kembali dengan sesuatu yang baru ya trailer versi jepang dari film ini resmi dirilis tentu saja hal ini membuat para fans semakin gregetan kali ini sisi alter para superhero ini sengaja ditonjolkan sebagai salah satu konflik yang ada di film ini

TABLE XI  
 TABEL TOKENIZING

tidak ada meragukan kehebatan twentieth century fox mengolah film film dibawah naungannya rumah produksi kembali sekuel terbaru superhero terbaik film fantastic four digarap dibawah asuhan josh trunk simon kinberg sesuatu baru trailer versi jepang film resmi dirilis membuat fans gregetan sisi alter superhero sengaja ditonjolkan konflik film

TABLE XII  
 TABEL STEMMING

ragu hebat twentieth century fox olah film film dibawah naungan rumah produksi kembali sekuel baru superhero baik film fantastic four garap bawah asuhan josh trunk simon kinberg sesuatu baru trailer versi jepang film resmi rilis buat fans gregetan sisi alter superhero sengaja tonjol konflik film

TABLE XIII  
 TABEL PERHITUNGAN NBC

ragu hebat twentieth century fox olah film film dibawah naungan rumah produksi kembali sekuel baru superhero baik film fantastic four garap bawah asuhan josh trunk simon kinberg sesuatu baru trailer versi jepang film resmi rilis buat fans gregetan sisi alter superhero sengaja tonjol konflik film

Tabel diatas merupakan 1 data forum uji yang akan diklasifikasikan secara otomatis dengan menggunakan NBC. Langkah awal yaitu dengan menentukan nilai n, nc, p dan m untuk setiap kelas yang ada dan untuk setiap term yang terdapat pada data uji seperti dijelaskan pada poin sebelumnya.

TABLE XIV  
 TABEL MENENTUKAN NILAI PERSAMAAN I

Dokumen	Term	Jumlah Kata	Kelas
D1	ragu;1, hebat;1, twentieth;1, century;1, fox;1, olah;1, film;5, bawah;2, naungan;1, rumah;1, produksi;1, kembali;1, sekuel;1, baru;1, superhero;1, baik;1, fantastic;1, four;1, garap;1, asuhan;1, josh;1, trunk;1, simon;1, kinberg;1, sesuatu;1, baru;1, trailer;1, versi;1, jepang;1, resmi;1, rilis;1, buat;1, fans;1, gregetan;1, sisi;1, alter;1, superhero;1, sengaja;1, tonjol;1, konflik;1.	45	Hiburan

Perhitungan perbandingan antara term pada data testing dengan setiap kelas yang ada dengan menggunakan persamaan.

$$P(a_i | v_j) = \frac{nc + mp}{n + m}$$

n = jumlah term pada data latih dimana v = v<sub>j</sub>

nc = jumlah term dimana v = v<sub>j</sub> dan a = a<sub>i</sub>

p = probabilitas setiap kelas dalam data latih

m = jumlah term pada data uji.

Sedangkan untuk menentukan klasifikasi pada data uji digunakan persamaan

$$V_{nb} = \text{argmax } v_{jv} P(v_j) \text{ II } P(a_i | v_j)$$

Berikut adalah contoh nilai-nilai untuk kelas hiburan.

Dari nilai-nilai yang diperoleh dan dengan menggunakan persamaan 1 diperoleh perhitungan sebagai berikut:

$$P(\text{hiburan/ragu}) = \frac{1 + (45 \times 0,0065)}{3,090 + 45} = \frac{1,2925}{3,135} = 0,0004$$

$$P(\text{hiburan/hebat}) = 0,0003$$

$$P(\text{hiburan/twentieth}) = 0,0004$$

$$P(\text{hiburan/century}) = 0,0004$$

$$P(\text{hiburan/fox}) = 0,0004$$

$$P(\text{hiburan/olah}) = 0,0004$$

$$P(\text{hiburan/film}) = \frac{5 + (45 \times 0,0065)}{3,090 + 45} = \frac{5,2925}{3,135} = 0,0016$$

$$P(\text{hiburan/bawah}) = \frac{2 + (45 \times 0,0065)}{3,090 + 45} = \frac{2,2925}{3,135} = 0,0007$$

$$P(\text{hiburan/naungan}) = 0,0004$$

$$P(\text{hiburan/rumah}) = 0,0004$$

$$P(\text{hiburan/produksi}) = 0,0004$$

$$P(\text{hiburan/kembali}) = 0,0004$$

$$P(\text{hiburan/sekuel}) = 0,0004$$

$$P(\text{hiburan/baru}) = 0,0004$$

$$P(\text{hiburan/superhero}) = 0,0004$$

$$P(\text{hiburan/baik}) = 0,0004$$

$$P(\text{hiburan/fantastic}) = 0,0004$$

$$P(\text{hiburan/four}) = 0,0004$$

$$P(\text{hiburan/garap}) = 0,0004$$

$$P(\text{hiburan/asuhan}) = 0,0004$$

$$P(\text{hiburan/josh}) = 0,0004$$

$$P(\text{hiburan/trank}) = 0,0004$$

$$P(\text{hiburan/simon}) = 0,0004$$

$$P(\text{hiburan/kinberg}) = 0,0004$$

$$P(\text{hiburan/sesuatu}) = 0,0004$$

$$P(\text{hiburan/baru}) = 0,0004$$

$$P(\text{hiburan/trailer}) = 0,0004$$

$$P(\text{hiburan/versi}) = 0,0004$$

$$P(\text{hiburan/jepang}) = 0,0004$$

$$P(\text{hiburan/resmi}) = 0,0004$$

$$P(\text{hiburan/rilis}) = 0,0004$$

$$P(\text{hiburan/buat}) = 0,0004$$

$$P(\text{hiburan/fans}) = 0,0004$$

$$P(\text{hiburan/gregetan}) = 0,0004$$

$$P(\text{hiburan/sisi}) = 0,0004$$

$$P(\text{hiburan/alter}) = 0,0004$$

$$P(\text{hiburan/superhero}) = 0,0004$$

$$P(\text{hiburan/sengaja}) = 0,0004$$

$$P(\text{hiburan/tonjol}) = 0,0004$$

$$P(\text{hiburan/konflik}) = 0,0004$$

Untuk perhitungan p(ai/vj) yang lainnya dilakukan proses yang sama seperti pada perhitungan p(hiburan/selebritas). Dengan menggunakan persamaan 2, yaitu mencari nilai maksimal dari hasil perkalian nilai probabilitas dan nilai p untuk setiap kelas diperoleh hasil sebagai berikut:

$$V(\text{Politik}) = 5,6732E-185$$

$$V(\text{Ekonomi}) = 5,6732E-185$$

$$V(\text{Kriminal}) = 5,6732E-185$$

$$V(\text{Hiburan}) = \mathbf{5,6326E-155}$$

$$V(\text{Olahraga}) = 5,6732E-185$$

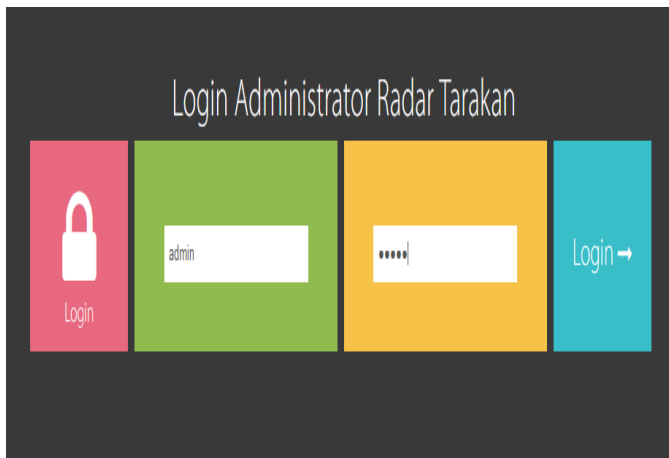
Vnb = argmax ( v(politik) | v(ekonomi) | v(criminal) | v(hiburan) | v(olahraga))  
 Vnb = argmax ( 5,6732E-185 | 5,6732E-185 | 5,6732E-185 | 5,6326E-155 | 5,6732E-185 | 1 )  
 Vnb = 5,6326E-155

TABLE XV  
 TABEL PERSAMAAN I DAN PERSAMAAN II

Term Hiburan	Kelas				
	Hiburan	Politik	Ekonomi	Kriminal	Olahraga
Ragu	0,0004	0,00009	0,00009	0,00009	0,00009
Hebat	0,0004	0,00009	0,00009	0,00009	0,00009
Twentieth	0,0004	0,00009	0,00009	0,00009	0,00009
Century	0,0004	0,00009	0,00009	0,00009	0,00009
Fox	0,0004	0,00009	0,00009	0,00009	0,00009
Olah	0,0004	0,00009	0,00009	0,00009	0,00009
.....	....	....	....	...	..
Gregetan	0,0004	0,00009	0,00009	0,00009	0,00009
Sisi	0,0004	0,00009	0,00009	0,00009	0,00009
Alter	0,0004	0,00009	0,00009	0,00009	0,00009
Superhero	0,0004	0,00009	0,00009	0,00009	0,00009
Sengaja	0,0004	0,00009	0,00009	0,00009	0,00009

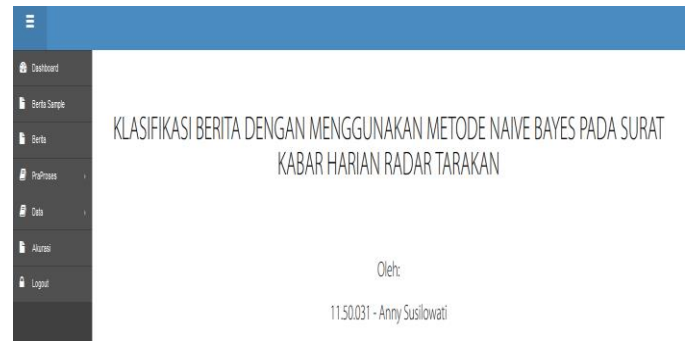
**B. Uji Coba**

Untuk masuk dalam program klasifikasi berita terlebih dahulu masuk ke menu Login Administrator Radar Tarakan. Disini masukkan username dan password, selanjutnya klik tombol login. Jika yang diinput benar, maka akan masuk ke menu admin. Bila salah, maka akan muncul pemberitahuan, "maaf, username dan password anda tidak terdaftar, silah coba kembali!".



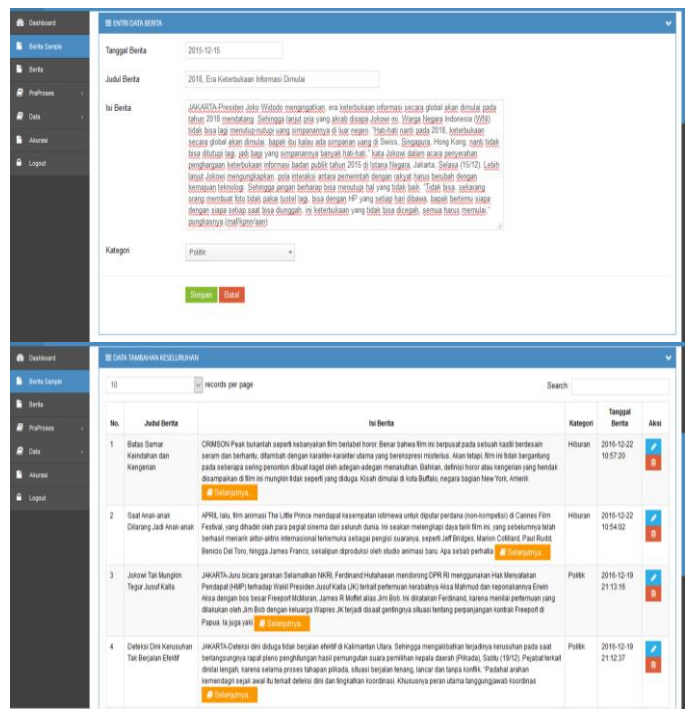
Gambar 3.1 Form Login Admin

Setelah masuk ke menu login dan berhasil, selanjutnya akan tampil menu program. Menu ini terdiri dari berita sample, berita, praproses training dan testing, data training dan testing, akurasi dan log out. Berikut interface dari menu:



Gambar 3.2 Form Menu

Langkah pertama untuk proses klasifikasi adalah dengan memasuki form berita sample. Disini data berita diinput dan ditampilkan di layar. Data ini bisa diedit dan dihilangkan.

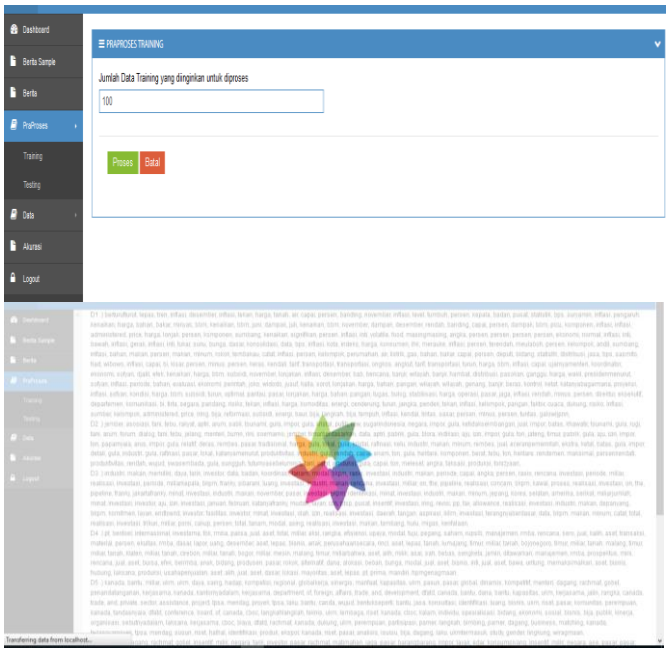


Gambar 3.3 Form Berita

Dalam form ini diinputkan tanggal berita, judul berita, isi berita dan kategorinya. Bila tombol simpan diklik maka pada layar akan muncul kata "Simpan Sukses..!". Selanjutnya bila di-scroll di layar di bawahnya akan muncul berita yang baru saja diinput, termasuk sejumlah berita yang telah diinput sebelumnya. Bila ada berita yang hendak diedit tinggal klik tombol edit dan mulai untuk mengedit judul berita, isi berita atau kategorinya lalu klik kembali tombol simpan dan dilayar tampil kata-kata "Edit Sukses..!". Adapun tombol hapus akan seketika menghilangkan berita yang di-input dari layar. Bila hapus sukses akan ditandai dengan kata-kata "Hapus Sukses..!".

Dalam form praproses training diinputkan berapa banyak data yang dikehendaki untuk kemudian diproses.

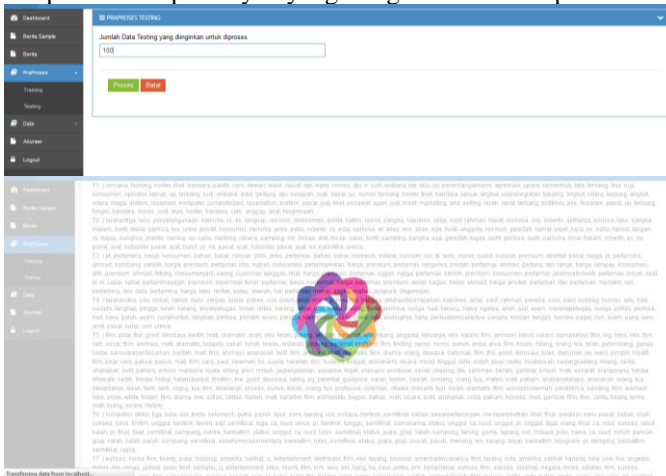




Gambar 3.4 Form Praproses Training

Awalnya data training diambil secara random, kemudian melewati tahapan praproses dimulai dengan case folding, tokenizing, stoplist dan stemming. Agar proses pemilihan kata dasar itu cepat dan akurat, maka stemming yang digunakan adalah stemming arifin. Hasilnya sebuah berita yang terdiri dari kata-kata yang tinggal kata dasarnya saja sebagai hasil praproses tadi.

Demikian pula dengan praproses testing, dimana data diinputkan berupa banyak yang diinginkan untuk diproses.



Gambar 3.5 Form Praproses Testing

Saat proses berlangsung ditandai dengan kipas pelangi yang berputar, tampak menjadi latar belakang adalah praproses testing. Sama halnya dengan praproses training, dalam praproses testing data juga melewati proses case folding, tokenizing, stoplist dan stemming.

No	M	Kolom	Kategori	NBC
1	T1	JAKARTA Rencana Kementerian Perhubungan menaruh kontrol ketat di Bandara Merpati karena dari Dewan Perwakilan Rakyat DPR RI Wakil Ketua Komisi V DPR RI.	Ekonomi	Ekonomi
2	T2	TENGGARONG petinggi partai persidangan narabisa jawa Tj Eci dan Ar yang dituntut Rencan Kemendagri Pula.Kabem rewa jadi narabisa Kaprodes Tarakan NBC	Kemana	Kemana
3	T3	JAKARTA PT Pertamina non kasasiang kasasiang baban basu narabisa BEM jawa perform Baban Basu dengan Reseach Odora Humber RCH di main tel	Ekonomi	Ekonomi
4	T4	TABURAN(Dua parrada beseser Ar dan An dituntut itu baru sepagi beser Patera Tarakan ungu mencan pular ekor an di J.Kasuma Bangsa Jurnal dot.	Kemana	Kemana
5	T5	DALAM film Pinar bertano The Good Doctor baru ter yang menyedikan teraji Dalah kabura adikan kembar dramatis apah Ais saat ditelan pular dot.	Hiburan	Puika
6	T6	TABURAN Komptas Densi Utama Laga Bina iak Tarakan Lenda di kelompak pular telah memasuki fase baru baru dit di lapangan uji imbua pular dot.	Chakra	Chakra
7	T7	SUKSES di Korea Kim Twenty basar dipul di Sinsag narabisa Serikat Baban Cii Entertainment sebagai distributor film besabur telah memis informas dot.	Hiburan	Puika
8	T8	JAKARTAAnggapa Komisi V DPR RI Aisa Bima mengatikan melarisa panemban penganggri Horvate mengaji tradid Agara Lasono Kisa gan pular dot.	Hiburan	Puika
9	T9	SITUS ream film animasi Boro: Harato the Movie telah memis baser trailer pertama di dunia masa Jurnal Laci Ci alu terlatik animo terasau dot.	Hiburan	Puika
10	T10	JAKARTA Para penguaha kelapa saat menang Pasatima harga crude pular di CPO di pasaran dunia terus mengampai panemban Panemban harga terasau dot.	Ekonomi	Puika

Gambar 3.6 Form Perhitungan NBC

#### IV. KESIMPULAN

Berdasarkan pembahasan pada bab-bab sebelumnya dapat disimpulkan bahwa :

1. Data training dalam sebuah klasifikasi sangat penting karena sangat mempengaruhi hasil akurasi.
2. Semakin banyak data berita sample yang di-inputkan, maka saat proses klasifikasi banyak pola yang akan terbentuk dan akan semakin baik pula hasil klasifikasinya
3. Pemilihan fitur penyeleksi data berita yang tepat juga sangat penting, karena akan meningkatkan kualitas klasifikasi.

#### REFERENSI

- [1] Susanto, Budi. *Materi Kuliah Text dan Web Mining*. Teknik Informatika UKDW Yogyakarta.Xu, Wunsch, 2009 “ Pengertian Clustering dalam komputasi “
- [2] Indriani, Aida. *Klasifikasi Data Forum dengan Menggunakan Metode Naïve Bayes Classifier*. STMIK PPKIA Tarakanita Rahmawati. Tarakan. 2015
- [3] Chandra, Denny Nathaniel. Gede Irawan.I Nyoman Sukajaya.*Klasifikasi Berita Lokal Radar Malang dengan Menggunakan Naïve Bayes dengan Fitur N-Gram*. Universitas Pendidikan Ganesha. Jurnal Ilmiah Teknologi dan Informasia Asia (Jitika). Vol 10. No 1, Februari 2016. ISSN 0852-730X
- [4] Indriani, Aida. *Weight Adjusted K-Nearest Neighbour dan Minimum Spanning Tree untuk Information Retrieval System Perpustakaan STMIK PPKIA Tarakanita Rahmawati Tarakan*. Tesis. 2013.
- [5] Indriani, Aida. *Weight Adjusted K-Nearest Neighbour dan Minimum Spanning Tree untuk Information Retrieval System Perpustakaan STMIK PPKIA Tarakanita Rahmawati Tarakan*. Tesis. Tarakan. 2013.
- [6] Indriani, Aida. *Weight Adjusted K-Nearest Neighbour dan Minimum Spanning Tree untuk Information Retrieval System Perpustakaan STMIK PPKIA Tarakanita Rahmawati Tarakan*. Tesis.
- [7] Anggeriani, Rina. *Pemanfaatan Vector Space Model untuk Peringkasan Berita*. Skripsi. STMIK PPKIA Tarakanita Rahmawati Tarakan. Tarakan. 2016
- [8] Setiati, Eni. *Ragam Jurnalistik Baru dalam Pemberitaan. Strategi Wartawan Menghadapi Tugas Jurnalistik*. Yogyakarta. Andi. 2005
- [9] Natalius, Samuel. *Metoda Naïve Bayes Classifier dan Penggunaannya pada Klasifikasi Dokumen*. ITB. Bandung
- [10] Hamzah, Amir. *Klasifikasi Teks dengan Naïve Bayes Classifier (NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis*. Prosiding Seminar Nasional Aplikasi Sains dan Teknologi (SNAST) Periode III, Yogyakarta, 2012. ISSN: 1979-911X.

**Penulis I**, Anny Susilowati, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK PPKIA Tarakanita Rahmawati Tarakan, lulus tahun 2017.

**Penulis II**, Aida Indriani, memperoleh gelar Magister Komputer (M.Kom), Universitas Gajah Mada. Saat ini menjadi Dosen di STMIK PPKIA Tarakanita Rahmawati.

**Penulis III**, Muhammad Hafid., memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK PPKIA Tarakanita Rahmawati Tarakan, lulus tahun 2014. Saat ini menjadi Dosen di STMIK PPKIA Tarakanita Rahmawati.