

ALGORITMA NAIVE BAIYES CLASSIFIER PADA APLIKASI DATA MINING BERBASIS WEB

Akbar Octaviari Murdiansyah¹⁾, Ir.Siswanto,MM²⁾

¹⁾Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur

²⁾Jl. Raya Ciledug, Petukangan Utara, Kebayoran Lama, Jakarta Selatan 12260

E-mail : ariakbar1995@gmail.com¹⁾, siswanto@budiluhur.ac.id²⁾

Abstrak

Perusahaan pada PT. Telkom Indonesia lebih tepatnya (Indihome) adalah salah satu produk layanan group berupa memberi paket atau data layanan yang terpadu dengan berbagai macam layanan. Dalam hal ini dengan salah satu penyedia layanan big data dan salah satu situs layanan mikro blog yang seharusnya oleh masing-masing pemiliknya untuk membuat suatu notes yaitu bermaksud untuk mencari atau mendekripsikan tentang berbagai macam topik dan membahas sebuah komentar-komentar dengan berbagai macam keluhan yang terjadi pada saat ini. Maka pengguna inisiatif melakukan postingan pendapat mereka terhadap kinerja dari perusahaan indihome itu sendiri. Hal ini tentunya sangat positif untuk divisi humas dan manager consumer Indihome. Untuk mengambil suatu keputusan atau strategi sosialisasi yang lebih tepat dan menjadi bahan evaluasi eksternal ataupun internal. Maka perusahaan indihome tersebut dapat mengolah data perusahaan secara baik. Salah satu cara untuk mengolah suatu data tersebut adalah dengan cara melakukan opinion mining, yaitu yaitu bermaksud dengan penggalian opini atau pendataan secara otomatis. Pada proses ini banyak sekali keluhan dan tanggapan negative dari client sangatlah pesat dan membuat data menjadi besar atau banyaknya data, karena salah satu masalah tersebut dibiarkan akan sangat disayangkan jika data tersebut tidak digunakan dengan baik akan berdampak besar data. Seiring perkembangannya teknologi yang semakin pesat dapat banyaknya permintaan client untuk mengubah data tersebut menjadi data informasi yang berguna adalah hal yang sangat dibutuhkan saat ini. Hal ini adanya penelitian bahwa penulis dapat menggunakan data yang sekiranya begitu tinggi 1350 sampai 1450 data tweet sentimen, sehingga data tersebut diklasifikasi secara manual dalam masing-masing dan diolah kedalam beberapa data lagi dengan mulai dari 450 ataupun sampai 550 data untuk sentimen positif atau negatif dan untuk netralisir. Hal ini dapat diproses testing suatu data kepada penulis menyediakan testing menggunakan Twitter search API maupun datatest yang sebelumnya sudah pernah disimpan. Klasifikasi yang dilakukan menggunakan algoritma Naive Bayes Classifier secara otomatis. Dalam hal ini dapat menghasilkan pengujian 100 sampai dengan 200 tweet rata-rata akurasi mencapai 83.00% hingga sampai lebih beberapa persen.

Kata kunci: Opinion Mining, Teks Mining, Twitter, Naive Baiyes Classifier, Data Mining

1. PENDAHULUAN

Kemajuan teknologi yang sangat cepat mendorong setiap perusahaan PT. Telkom Indonesia untuk tetap mengikuti perkembangan teknologi dan terus meningkatkan kemampuannya dalam mengelola data-data dan informasi yang lebih aman, akurat, dan efisien yang dibutuhkan suatu instansi. yaitu data yang masih tidak terstruktur dan masih tidak optimal, yaitu data yang tidak asing lagi seperti Twitter sebuah informasi saat ini juga tidak terlepas dari kecanggihan sebuah teknologi. Selama ini masih banyak perusahaan yang belum menggunakan secara maksimal teknologi yang ada, dimana sebenarnya teknologi ini sangat berguna bagi perusahaan untuk menunjang kinerja operasional baik perusahaan maupun pegawainya.

Sehingga perkembangan teknologi informasi berpengaruh besar pada banyaknya perubahan aktifitas manusia. Selama ini jumlah tweet pada twitter mencapai 500 juta tweets perhari, tweet juga menyediakan sebuah aplikasi program interface (API) yang memungkinkan perhari mendapatkan data mereka. Hal ini data yang didapatkan dan juga dimanfaatkan dari lewat sebuah salah satu media social dengan sebuah aplikasi twitter itu sendiri yaitu

dengan berupa status-status atau yang lebih dikenal dengan kata tweets.

Dalam hal ini beberapa tahun terakhir perkembangan terus meningkat dari berbagai aplikasi web adalah salah satu kunci dari media sosial. Data tersebut tentunya mempunyai manfaat masing-masing, manfaat yang sangat penting jika digunakan dengan cara lebih baik. Dalam pengetahuan inidari suatu data mining sangat memperhatikan dengan mendessaknya suatu kebutuhan client untuk mengubah data diman data sebut data informasi untuk pengetahuan yang berguna. Dengan banyak sekali aplikasi-aplikasi yang sangat diuntungkan jika data tersebut dapat diubah menjadi suatu informasi yang berguna, seperti aplikasi analisa untuk mendekripsikan sebuah pasar dan manajemen bisnis atau dengan aplikasi lain dengan cara membuat aplikasi penilaian akan sangat berarti bagi instansi.

Ada bebarap aplikasi dengan bernam database berfungsi untuk penyimpanan sebuah informasi dalam berbentuk teks. Aplikasi ini sangat sudah tidak asing lagi dan sangat familiar. Aplikasi ini juga dapat diolah secara tidak rumit. Dengan hal ini bermaksud penyimpanan dalam bentuk teks dan teks mining adalah salah satu area untuk riset yang

menanganidata tersebut. Text mining adalah menjadi salah satu factor yang sangat penting untuk melakukan proses menentukan ataupun menemukan pengetahuan bisa disebut dalam Bahasa luar (*Knowledge discovery*). *Text mining* tersebut untuk mengambil suatu sampel untuk informasi dari beberapa sumber yaitu sumber dengan tertulis atau juga bisa dari informasi yang sebelumnya belum diketahui dari kumpulan data yang tidak terstruktur.

Karena aplikasi peninjauan untuk kinerja perusahaan IndiHome ini dianggap sangat penting oleh penulis sebab sangat berguna untuk melihat atau repond penggunanya twitter se-indonesia terhadap kinerja perusahaan tersebut. Aplikasi tersebut kedepannya dapat memudahkan bagian sales indihome dan utnuk mengambil suatu strategi atau sosialisasi terhadap produknya. Dalam hal ini dapat disimpulkan penulis dapat menggunakan atau mengimplementasikan dengan algoritma *Navie Bayes Classifier* karena pada penelitian ini terlebih dahulu menggunakan algoritma ini karena mempunyai nilai akurasi yang sangat tinggi..

Oleh sebab itu masalah keamanan suatu data merupakan aspek penting dari suatu sistem informasi. Teknologi data mining membantu untuk mengimpelmentasikan informasi yang berguna dari berbagai *database*.

2. METODE PENELITIAN

Dapat disimpulkan dan bahwa penelitian ini ada beberapa metode dalam pengembangannya yang digunakan oleh penulis untuk menyusun suatu penelitian yang sedang dilakukan meliputi:

2.1 Metode Pustaka

Mencari dan memahami serta mempelajari dalam skema dari berbagai jurnal atau buku dan macam artikel berkaitan dengan data mining, serta algoritma *Naïve BayesClassifier* serta semua yang berkaitan dengan topic ini. Serta mencari sumber-sumber pustaka lainnya yang juga mendukung seperti jurnal , forum-forum diskusi, pendapat ahli daniiii sebagainya baik media cetak elektronik.

2.2 Motode Prototyping

a. Menentukan Kebutuhan

dengan kebutuhan yang akan kami buatkan yaitu dengan pegguan yang akan menggunakan aplikasi mengimplementasikan dan juga mengidentifikasi kebutuhan secara bersama-sama menentukan secara gratis besar konsep system yang akan dikembangkan.

b. Membuat Prototype

dari segi besar dalam konsep sudah disepakati, setelah itu membuat suatu *prototype* dengan membuat suatu rancangan aplikasi sementara yang difokuskan pada garis besar konsep yang sudah disepakati.

c. Evaluasi Prototype

Pada proses ini *user* melakukan evaluasi dari *prototype* aplikasi yang sudah sesuai dengan keinginan dengan yang dibutuhkan oleh pengguna. Jika *prototype* sudah selesai maka langkah keempat akan diambil, namun jika tidak maka *prototype* akan diperbaiki dengan mengulai langkah pertama dan selanjutnya.

d. Mengkodeka Aplikasi

Dalam hal ini, *prototyping* yang sudah disepakati oleh pengguna dalam pemrograman yan sesuai.

e. Evaluasi Aplikasi

Pengguna melakukan evaluasi terhadap aplkasi yang dikembangkan, apakah aplikasi sudah sesuai apa yang diharapkan atau mungkin ridak diharapkan. Jika sudah maka langkah-langkah yangakan dilakukan, namu jika belum aplikasi akan dibuat kembali dengan mengulang langkah keempat dan selanjutnya.

3. ANALISA MASALAH DAN PERANCANGAN APLIKASI

3.1 Permasalahan

Perkembangan teknologi sangat ini dengan adanya suatu data berupa data gital yang sangat pesat dapat membuat data perusahaan pada PT. Telkom Indonesia(IndioHome) tersebut terjadi menumpuk dan tidak tertata rapih data tersebut, hal ini dapa membuat timbulnya data dapat tidak teratur dan tidak sesuai yang diinginkan pada saat mengubah data tersebut menjadi data informasi pengetahuan yang bermanfaat untuk instnsni. Banyaknya sekali keuntungan-keuntungan yang bisa dikatagorikan jika data tersebut sangat besar diolah menjadi sangat berarti.Hal ini seperti aplikasi yang dapat mengembangkan dan juga mengevaluasi dari suatu pekerjaan dengan menganalisis sentimen.Salah satu aplikasi tersebut adalah aplikasi media sosial yaitu *twitter*.

Dengan adanya permasalahan tersebut maka dibutuhkan sebuah wadah penyimpanan data tersebut yang bisa digunakan oleh pegawai PT Telkom Indonesia (IndiHome) secara *Web*.yangnantinya saat data yang dibutuhkan tersimpan dalam satu tempat dan juga dapat diakses oleh *manager* atau perusahaan yang dikategorikan juga sebagai pemilik aplikasi tersebut. Dengan demikian tersedia sistem yang bisa mengevaluasi kinerja instansi secara internal, akan tetapi belum ada sistem yang bekerja untuk mengevaluasi kinerja aplikasi secara eksternal, hal ini sangat dibutuhkan karena perusahaan tersebut yang dinilai lansung oleh masyarakat sekitarnya agar dapat meminimalisir penilaian dari masyarakat sangat paling berarti kedepannya untuk perusahaan IndiHome sendiri.

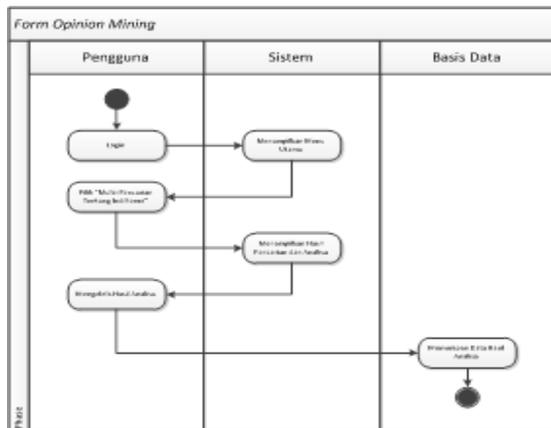
3.2 Solusi Penyelesaian Masalah

Dari permasalahan diatas, maka dibuatlah sebuah aplikasi yang dapat menguasai kinerja dari perusahaan secara eksternal melalui media sosial yaitu berbagai macam twitter secara flexsibel. Dengan cara menggunakan teknik yang disediakan oleh *user* untuk membuat aplikasi tersebut dengan *text mining*, hal ini hampr sama dari bagian data mining. Pada proses ini berbagai macam data training masih dalam pembuatan manual oleh *user* dan masih tersimpan didalam sistem *database* tersebut yang nantinya dapat digunakan sebagai model dalam proses klasifikasi *text* secara otomatis.

Dengan hal ini penulis dapat memberikan suatu ide untuk membuat aplikasi dengan menggunakan algoritma *klasifikasi naive bayesclassifier* dengan menggunakan algoritma ini dapat berguna untuk perusahaan tersebut, dan juga aplkasi tersebut mempunyai kelebihan dengan pemrosesan teks bisa seleksi untuk fitur-fiturnya, dan juga stopword, tokenizing yaitu dnegan teknik pengambilan morfologi kata-kata yang berupa teks dan juga steming menggunakan algoritma nazief dan sebagainya.

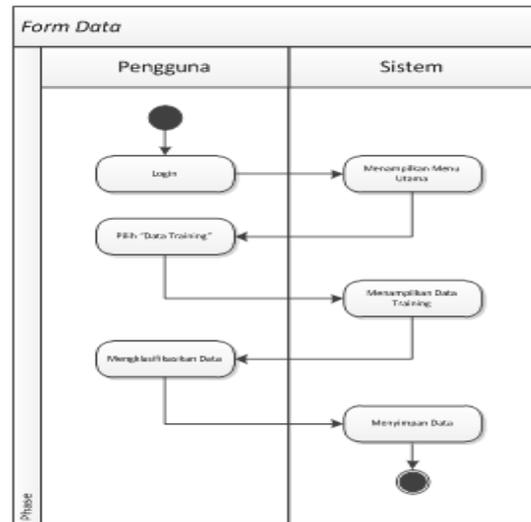
3.3 Pencarian Proses Opinion Mining

Pada proses ini menjelaskan tentang alur proses bisnis usulan, yaitu dnegan menggambarkan skema aplikasi berguna untuk menjalankan alur aplikasi tersebut dan dapat bisa dipahami. Hal ini dibutlah suatu *activity diagram* proses bisnis usulan pada sistem tersebut. Seperti gambar 1.



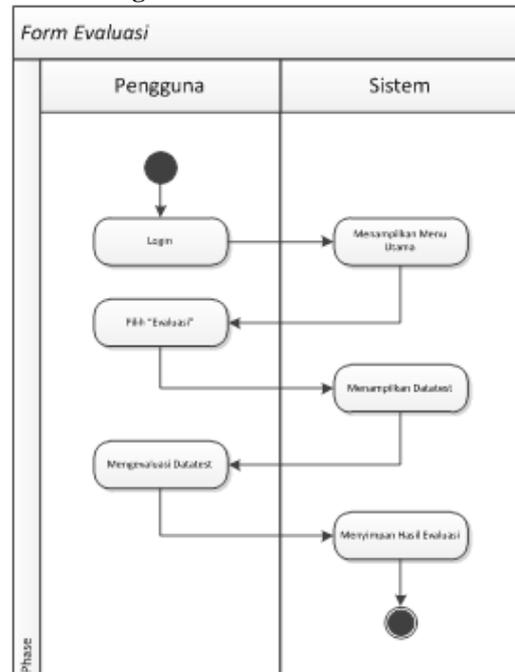
Gambar 1. Activity Diagram Proses Opinion Mining.

3.4 Proses Pengklasifikasi Datasets



Gambar 2. Activity Diagram Klasifikasi Datasets.

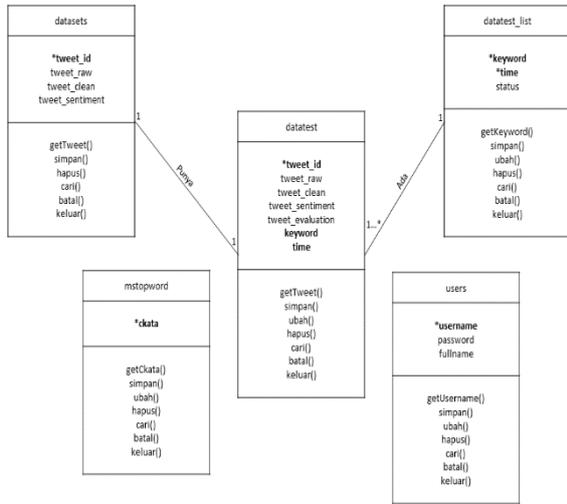
3.5 Proses Pengevaluasian Datasets



Gambar 3. Activity Diagram evaluasi Datasets.

3.6 Rancangan Class Diagram

Pada proses ini menjelaskan tentang alur *class diagram*, dengan adanya tabel ini dapat disimpulkan bahwa sebagian tabel-tabel dapat teratur karena masing-masing tabel mempunyai *primary key* yang bermasuk dan bertujuan untuk mengatur sebuah data agar data atau tabel tersebut dapat terelasi dengan benar. Seperti gambar 4.

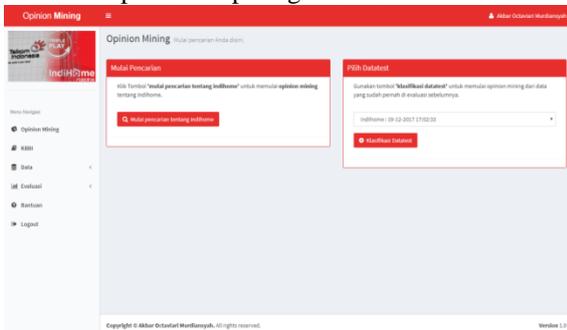


Gambar 4. Proses Class Diagram.

4. HASIL DAN PEMBAHASAN

4.1 Tampilan Layar Menu Login

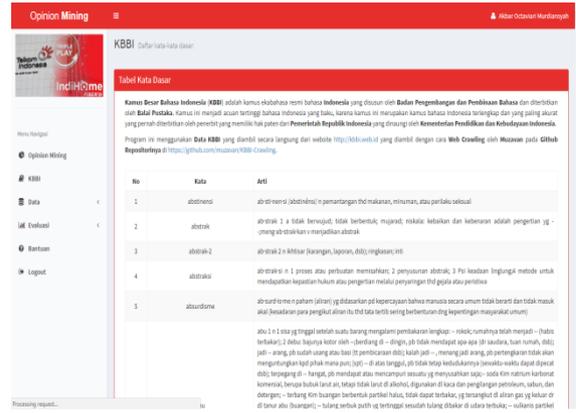
Pada tampilan layar menu *login* adalah halaman yang akan muncul ketika *user* membuka suatu aplikasi tersebut. Pada halaman ini terdapat menu bar *user* dapat memilih salah satu menu tersebut. Dengan memilih salah satu data *training* atau juga bisa memilih data testing (*opinion mining*). Dan masih banyak menu yang seperti *KBBI*, *stopword*, *datasets* dan juga menu tambahan yaitu menu bantuan dan evaluasi aplikasi. Seperti gambar 5.



Gambar 5. Tampilan Layar Menu Login.

4.2 Tampilan Layar Form Data KBBI

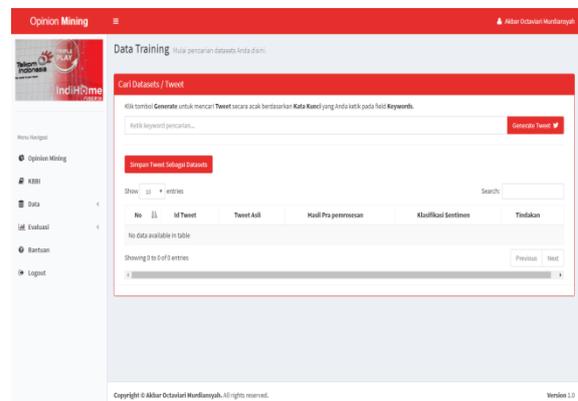
Pada tampilan layar *form KBBI* adalah terdapat penjelasan KBBI, yaitu berguna pada saat *user* melakukan proses daftar atau melihat *view* dari kata dasar pada KBBI yang telah disediakan agar mempermudah *user* untuk melakukan proses pendataan pada *form* tersebut. Seperti gambar 6.



Gambar 6. Tampilan Layar Form Data KBBI.

4.3 Tampilan Form Data Training

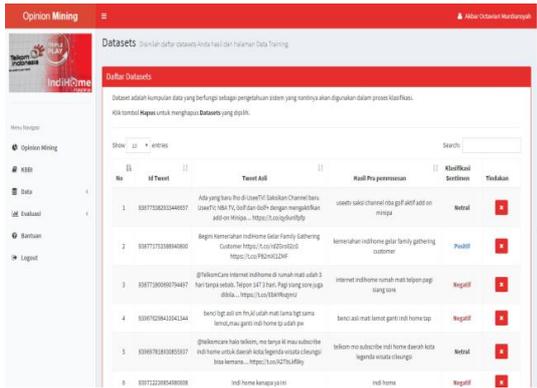
Pada tampilan layar *form* ini adalah menggambarkan halaman yang diproses oleh *user* untuk mengubah atau meng-*copy* data tweeter yaitu nantinya akan diproses dan *disave* dan mengubahnya menjadi data penting dengan bermaksud untuk data training setelah ditandai dengan manual oleh *user*. Hal ini pada saat memasuki aplikasi tersebut pada halaman ini muncul pop up yang dimana data tersebut adalah data training. Seperti gambar 7.



Gambar 7. Tampilan Layar Form Data Training.

4.4 Tampilan Layar Form Datasets

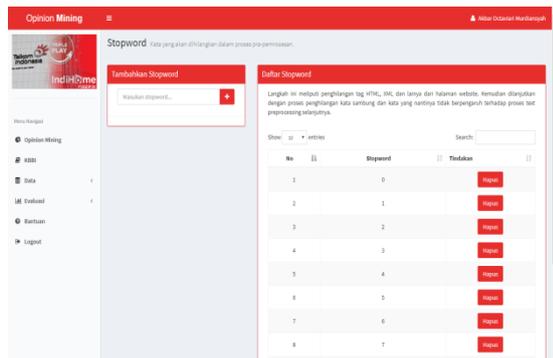
Pada tampilan layar *datasets* adalah pada saat *user* ingin melihat atau proses menghapus data training yang sebelumnya sudah pernah melakukan *input* dan tersimpan di dalam *database* yang tersambung oleh halaman training. Seperti gambar 8.



Gambar 8. Tampilan Layar Form Datasets

4.5 Tampilan Layar Form Data Stopwords

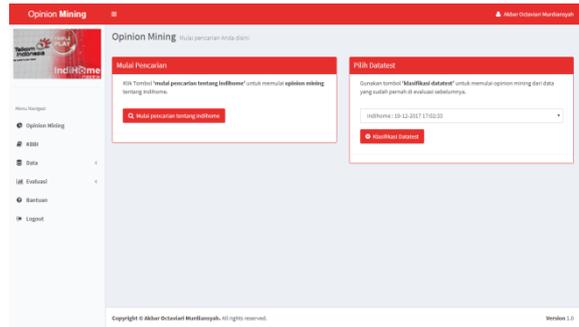
Pada tampilan layar *stopwords* adalah pada saat *user* ingin melakukan proses menghapus data atau menambahkan *datastopword* yang nantinya pada saat memilih dalam tahapan data-data untuk pra pemrosesan teks-teks tersebut. Seperti gambar berikut ini:



Gambar 9. Tampilan Layar Form Data Stopwords.

4.6 Tampilan Layar Form Data Opinion Mining

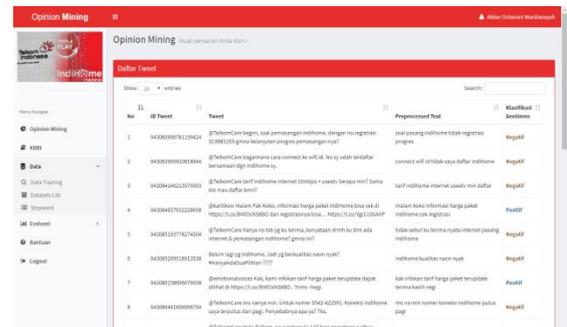
Pada tampilan layar data opinion mining adalah ketika *user* ingin melakukan proses klasifikasi dengan pengguna. Selain itu *user* dapat *request* klasifikasi tersebut melalui via media sosial yaitu menggunakan *twittersearch* API yang sudah pernah dibuta lalu tersimpan di dalam sistem *database* tersebut. Seperti gambar 10.



Gambar 10. Tampilan Layar Form Data Opinion Mining.

4.7 Tampilan Layar Form Data Klasifikasi

Pada tampilan layar data klasifikasi menjelaskan mengenai proses penyimpanan ke dalam *database*, jika pada saat melalui proses step pertama yang bermaksud step meng-*input* data-data awal, maka hal ini akan tersambung di *form* ini. Pasa proses ini *user* dapat memilih dan menyimpan dengan ketentuan tertentu yang nantinya akan hasil dari klasifikasi lalu proses analisis data dan proses *save* data tersebut. Halaman ini adalah yang menggunakan *twitter search* API. Seperti gambar 11.



Gambar 11. Tampilan Layar Form Data Klasifikasi.

4.8 Tabel Hasil Pengujian Data

Pada tabel pengujian ini adalah hasil dilakukan pada saat mengimplementasikan aplikasi opinion mining ini, dapat dilakukan dengan menguji data testing pada beberapa kali dengan jumlah data training yang berbeda. Seperti pada tabel 7.

Tabel 1. Hasil Uji Data (opinion mining).

Positif	Data Training		Data Testing	Akurasi	Sumber
	Negatif	Netral			
200	200	200	100	30.85%	Twitter Search API
250	250	250	100	78.00%	Twitter Search API
350	350	350	100	86.41%	Twitter Search API
400	400	400	100	71.76%	Twitter Search API
450	450	450	100	83.00%	Twitter Search API
450	450	450	100	83.00%	Twitter Search API

5. EVALUASI PROGRAM

Evaluasi adalah salah satu hal yang dilakukan dalam setiap pengembangan aplikasi untuk mengetahui kelebihan dan kekurangannya. Dalam

evaluasi ini ditemukan beberapa kelebihan dan kekurangan antara lain:

a. Kelebihan Program

- 1) Aplikasi ini mampu mengklasifikasi *datatest* secara otomatis dari twitter.
- 2) Aplikasi ini diimplementasikan dengan algoritma *Naïve Bayes Classifier* dengan tambahan media *platform website* sehingga aplikasi tersebut dapat terinstal atau *support* sistem mana saja.
- 3) Aplikasi ini mampu menjumlahkan teks-teks dengan banyak sehingga dapat meminimalisir evaluasi sentiment *client* dihome.
- 4) Terdapat fitur-fitur yang dimana fitur tersebut hanya bisa dibuka oleh *user* seperti *login*.
- 5) Dengan fitur yang elegant juga flexible dan mampu menghasilkan banyaknya evaluasi hasil dari *opinion mining* yang lebih akurat.

b. Kekurangan Program

- 1) Aplikasi ini mampu mengambil data-data *tweet* 100 dan dimaksimalkan 100 kali dalam jangka waktu yang ditentukan oleh *twitter*.
- 2) Karena menggunakan *server* online maka dapat menggunakan atau terkoneksi internet agar aplikasi tersebut berjalan dengan cepat juga untuk proses klasifikasinya menggunakan *twitter search API* dapat otomatis mengambil data sebanyak 100 *tweet*.
- 3) Membutuhkan banyaknya data sehingga untuk morfologi kata-kata karena banyaknya pengguna bahasa tersebut tidak benar.
- 4) Evaluasi *datatest* masih dilakukan secara manual, karena kurangnya data-data tersebut.
- 5) Diperlukan dengan adanya *datasets* yang nantinya akan banyak mendapatkan hasil *opinion mining* yang lebih akurat dan *flexible*.

6. KESIMPULAN

Berdasarkan hasil analisa yang telah kami lakukan terhadap permasalahan dan aplikasi yang dikembangkan, maka dapat ditarik suatu kesimpulan sebagai berikut:

- a. Dengan rata-rata akurasi tertinggi didapat oleh hasil klasifikasinya adalah 86% dari tabel klasifikasinya basis data.
- b. Rata-rata akurasi yang dicapai adalah 80,43% dari penjumlahan tabel.
- c. Masih ada beberapa klasifikasinya atau beberapa kata yang tidak bisa di *stemming* karena memungkinkan pengguna menggunakan bahasa daerah ataupun bahasa luar.
- d. Proses dari data-data disebabkan dari negasi kata membantu untuk menentukan arti sebenarnya dari kata atau teks.

- e. Proses ini untuk morfologi kata-kata membantu dalam proses *stemming* dengan algoritma Nazien & Adriani.

7. SARAN

Selain menarik beberapa kesimpulan, dapat pula diajukan saran-saran yang mungkin bisa dijadikan pertimbangan dalam pengembangan sistem, antara lain:

- a. Aplikasi tersebut diharapkan dapat mengklasifikasi data-data *training* secara tersendiri.
- b. Aplikasi tersebut diharapkan dapat mengevaluasi tersendiri data-data *training* secara flexibel.
- c. Algoritma yang telah dibuat sebaiknya selalu ditingkatkan, karena dengan semakin berkembang ilmu pengetahuan dan perubahan bahasa maka tidak dapat dipastikan program ini akan selalu akurat.
- d. Aplikasi ini diharapkan untuk dapat ditambah fitur POS (*part of speech*) tag untuk.

8. DAFTAR PUSTAKA

- [1] Agusta, L. (2009) 'Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief & Adriani untuk Stemming Dokumen Teks Bahasa Indonesia', in *Konferensi Nasional Sistem dan Informatika*. Bai, pp. 196–201.
- [2] Falahah and Nur, D. D. A. (2015) Pengembangan Aplikasi Sentiment Analysis Menggunakan Metode Naïve Bayes', *Seminar Nasional Sistem Informasi Indonesia*, (November), pp. 335–340.
- [3] Gorunescu, F. (2011) *Data Mining: Concepts, models and techniques*. Springer Science & Business Media.
- [4] ndrayuni, E. (2016) 'Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization', *Jurnal Evolusi*, 4.
- [5] Krikorian, R. (2013) *New Tweets per second record, and how!* Available at: <https://blog.twitter.com/2013/new-tweets-per-second-record-and-how> (Accessed: 18 April 2016).
- [6] Ling, J., Kencana, I. putu E. N. and Oka, T. B. (2014) Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square', *E-Jurnal Matematika*, 3(3), pp. 92–99.

- [7] Nurzahputra, A. and Muslim, M. A. (2016) Analisis Sentimen pada Opini Mahasiswa Menggunakan Natural Language Processing', in *Seminar Nasional Ilmu Komputer*. Semarang, pp. 114–118.