

Analisa Pola Penyalahgunaan *Facebook* Sebagai Alat Kejahatan *Trafficking* Menggunakan *Data Mining*

Linda Jayanti, Steven Ray Sentinuwo, Oktavian A. Lantang, Agustinus Jacobus
Teknik Informatika Universitas Sam Ratulangi Manado, Indonesia.

110216014@student.unsrat.ac.id, steven@unsrat.ac.id, Oktavianlantang@gmail.com, a.jacobus@unsrat.ac.id

Abstrak - *Facebook* memungkinkan penggunaanya berinteraksi dengan orang yang kita kenal maupun orang yang tidak kita kenal, dimana hal tersebut dapat membuka peluang bagi kejahatan dunia maya seperti, penculikan, perdagangan manusia (*trafficking*), hingga pembunuhan. IOM mencatat bahwa korban perdagangan orang atau *trafficking* di Indonesia mencapai 74.616 hingga 1 juta per tahun, dimana tindak kejahatan tersebut banyak dilakukan melalui *facebook* sebagai medianya. Data teks (*status*) yang berada di halaman *facebook* sangat besar. Dengan menggunakan Teknik pengolahan data dari ilmu *Data Mining*, terutama di bidang *text mining*, penulis memanfaatkannya untuk mengidentifikasi data teks (*status facebook*) yang terindikasi sebagai proses kejahatan *trafficking* dengan memakai salah satu teknik klasifikasi dengan teorema *naïve bayes classifier (NBC)*.

Kata kunci : *facebook, trafficking, data mining, text mining, klasifikasi, naïve bayes classifier.*

I. PENDAHULUAN

Facebook merupakan situs media sosial yang dimana penggunaanya bisa berinteraksi, berbagi data dan informasi, serta menjalin relasi. Saat ini banyak sekali bentuk penyalahgunaan *facebook* di kalangan masyarakat, seperti kasus penculikan, kasus penipuan, kasus pembunuhan, hingga kasus perdagangan manusia (*human trafficking*). Dari beberapa bentuk penyalahgunaan *facebook* diatas, kasus *trafficking* adalah kasus yang paling banyak terjadi di Indonesia, dengan jumlah kasus 74 ribu hingga 1 juta jiwa pertahunnya, Indonesia merupakan salah satu Negara terbesar sumber tempat transit dan penerimaan *trafficking* yang menggunakan media sosial terutama *facebook* sebagai alat menjalani aksi kejahatan tersebut.

Status *facebook* merupakan salah satu alat bukti yang sah di dalam pemeriksaan perkara pidana, khususnya dalam pemeriksaan perkara tindak pidana *cyber crime* menurut UU No 11 Tahun 2008 Pasal 5 Ayat 1 [6]. Selain itu, status *facebook* juga merupakan data terbanyak yang di *posting* pengguna di halaman *facebook*. Peneliti menggunakan alasan tersebut untuk menganalisa pola status yang terindikasi merupakan proses dari kejahatan *trafficking* di media sosial *facebook*, dengan memanfaatkan pendekatan *data mining*, terutama bidang pengolahan teks (*text mining*).

II. LANDASAN TEORI

A. *Facebook*

Facebook merupakan sebuah situs jejaring sosial yang diluncurkan pada bulan Februari 2004, dimiliki dan dioperasikan oleh *Facebook, Inc.* Pada September 2012, *Facebook* memiliki lebih dari satu miliar pengguna aktif, lebih dari separuhnya menggunakan handphone. *Facebook* memiliki beberapa fitur, yaitu : *wall, status, chat, friend request, inbox, notification, search bar, games.*

B. *Trafficking*

Pengertian *trafficking* dalam Undang-undang Pemberantasan Tindak Pidana Perdagangan Orang (UU No 21 Tahun 2007) lebih luas dibandingkan dengan KUHP [4], dengan memasukkan proses dan definisi korban maupun pelaku. Pasal 1 angka (1) UU No. 21 Tahun 2007 mendefinisikan *trafficking* sebagai berikut “*Perdagangan Orang (trafficking) adalah tindakan perekrutan, pengangkutan, penampungan, pengiriman, pemindahan, atau penerimaan seseorang dengan ancaman kekerasan, penggunaan kekerasan, penculikan, penyekapan, pemalsuan, penipuan, penyalahgunaan kekuasaan atau posisi rentan, penjeratan utang atau memberi bayaran atau manfaat, sehingga memperoleh persetujuan dari banyak orang yang memegang kendali atas orang lain tersebut, baik yang dilakukan di dalam Negara maupun antar Negara, untuk tujuan eksploitasi atau mengakibatkan orang tereksplorasi*” Dari definisi tersebut, terdapat tiga unsur utama *trafficking*, yaitu [2] :

1. Memindahkan orang, baik di dalam maupun di luar batas Negara (termasuk perekrutan, pengangkutan, penampungan, pengiriman, pemindahan atau penerimaan).
2. Cara-caranya melawan hukum (termasuk ancaman, penggunaan kekuasaan, penculikan, penyekapan, pemalsuan, penipuan, penyalahgunaan kekuasaan atau posisi rentan, penjeratan utang atau memberi bayaran atau manfaat sehingga memperoleh persetujuan dari orang yang memegang kendali atas orang lain tersebut).
3. Tujuannya eksploitasi atau menyebabkan orang tereksplorasi.

C. *Text Mining*

Text mining merupakan variasi dari *data mining* yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar [1]. Langkah yang dilakukan dalam *text mining* adalah sebagai berikut [3]:

1. *Text Preprocessing*

Text preprocessing adalah tahapan awal dari *text mining* yang bertujuan untuk mempersiapkan teks menjadi data

yang akan mengalami pengolahan pada tahapan berikutnya. Tindakan yang dilakukan pada tahap ini adalah :

- *Case folding*, yaitu proses untuk mengubah huruf didalam dokumen menjadi huruf kecil semua
- *Cleansing*, yaitu proses untuk menghapus hastag (#) atau *link* yang terdapat dalam status.
- *Tokenisasi*, yaitu proses memisahkan setiap kata penyusun data status..

2. Text Transformation

Pada tahap ini hasil yang diperoleh dari tahap *text preprocessing* akan melalui proses transformasi ini dilakukan dengan mengurangi jumlah kata-kata yang ada dengan penghilangan *stopword*.

Stopword adalah kata yang bukan merupakan ciri (kata unik) dari suatu dokumen seperti kata sambung, kata kepunyaan. Tindakan yang dilakukan pada tahap ini adalah :

- *Stopword removal*, yaitu proses untuk menghapus kata yang termasuk *stopword* yang terdapat dalam data status.
- *Convert emoticon*, yaitu proses untuk mengubah *emoticon* yang terdapat dalam data status menjadi sebuah string.
- *Convert negation*, yaitu proses untuk mengubah singkatan kata negasi yang terdapat dalam data status menjadi kata negasi yang utuh.

3. Pattern Discovery

Tahap penemuan pola atau *pattern discovery* adalah tahap terpenting dari seluruh proses *text mining*. Tahap ini berusaha menemukan pola atau pengetahuan dari keseluruhan teks. Dalam *text mining* terdapat dua teknik pembelajaran pada tahap *pattern discovery*, yaitu *unsupervised* dan *supervised learning*. Adapun perbedaan antara keduanya adalah pada *supervised learning* terdapat label atau nama kelas pada data latih atau data *training* dan data baru diklasifikasikan berdasarkan data latih. Sedangkan pada *unsupervised learning* tidak terdapat label atau nama kelas pada data latih, data latih dikelompokkan berdasarkan ukuran kemiripan pada suatu kelas.

D. Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek [8]. Dalam model klasifikasi, peneliti menggunakan algoritma *naïve bayes classifier* (NBC).

1. Algoritma Naïve Bayes Classifier (NBC)

Algoritma *naïve bayes classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat [6]. Dalam penelitian ini yang menjadi data uji adalah dokumen teks digital dari social media. Ada dua tahap pada klasifikasi data teks. Tahap pertama adalah pelatihan (training) terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah pengujian (testing), yaitu 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 data teks. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diuji (V_{MAP}) [8], dimana persamaannya adalah sebagai berikut :

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \frac{P(V_j)P(x_1, x_2, x_3, \dots, x_n | V_j)}{P(x_1, x_2, x_3, \dots, x_n)} \quad (1)$$

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

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} P(V_j)P(x_1, x_2, x_3, \dots, x_n | V_j) \quad (2)$$

Persamaan diatas dapat disederhanakan menjadi sebagai berikut :

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \prod_{i=1}^n P(x_i | V_j)P(V_j) \quad (3)$$

Dimana,

V_j = Kategori status facebook $j = 1, 2, 3, \dots, n$. Dimana dalam penelitian ini

j_1 = kategori status positif, j_2 = kategori status negative.

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

$P(V_j)$ = Probabilitas dari V_j

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

$$P(V_j) = \frac{|\text{docs } j|}{|\text{contoh}|} \quad (4)$$

$$P(x_i | V_j) = \frac{n_k + 1}{n + |\text{kosakata}|} \quad (5)$$

Dimana,

$|\text{docs } j|$ = jumlah dokumen setiap kategori j

$|\text{contoh}|$ = jumlah dokumen dari semua kategori

n_k = jumlah frekuensi kemunculan tiap kata

n = jumlah frekuensi kemunculan kata dari setiap kategori

$|\text{kosakata}|$ = jumlah semua kata dari semua kategori

III. METODOLOGI PENELITIAN

A. Data Penelitian

Dalam sub bab data penelitian ini akan dibahas, mengenai kategori data yang akan dipakai dalam proses analisa serta teknik sampling dalam Pengumpulan data tersebut.

1. Kategori Data

Dalam penelitian ini, peneliti membagi kategori (*class*) status ke dalam 2 kategori, yaitu kategori positif dan kategori negatif. Kategori positif, yaitu status yang dinyatakan berindikasi merupakan proses kejahatan *trafficking* yang mengandung beberapa komponen

trafficking yang terkandung dalam definisi *trafficking* menurut UU TPPO pasal 1 ayat 1, dimana komponen tersebut yaitu : perekrutan, pengangkutan, penampungan, pengiriman, penerimaan, eksploitasi, dan eksploitasi seks. Sedangkan, kategori negative adalah status yang tidak berindikasi sebagai proses kejahatan *trafficking*, tidak mengandung komponen *trafficking* diatas.

2. Teknik Sampling

Teknik sampling yang dipakai disini adalah teknik sampling nonprobabilitas, yaitu *purposive sampling*, yaitu teknik sampling dimana pengambilan sample secara sengaja sesuai dengan kriteria data sampel yang dibutuhkan. Jumlah data sampel (status) tergantung dari kebutuhan pengguna [5]. Dalam penelitian ini data status *facebook* dibagi kedalam 2 bagian, yaitu data *training* dan data *testing*.

TABEL 1. TABEL DATA TRAINING

id	Status	Kategori
1	buat teman cewek, kalau bisa balas inbox, karena kita butuh 150 orang untuk tempat baruku, nanti kalau OK bisa ketemu langsung dengan saya.	Positif
2	untuk cewek manado yang mau promosi bisa inbox saya, harga diatas 15 juta per bulan	Positif
3	Hindari sex bebas yang tidak aman jika tidak ingin terjangkit HIV	Negatif
4	Jangan mengeluh atas masalah dalam hidupmu. Tuhan memberikan cobaan yang berat padamu, karena Dia tahu kamu mampu mengatasinya	Negatif

Dalam penelitian ini, peneliti menggunakan jumlah data sebanyak 320 data status, yang dibagi menjadi 50:50 untuk data *training* dan data *testing*.

Tabel 1 merupakan data *training* berisi kumpulan status yang akan dijadikan sebagai data *training* atau data latihan untuk sistem, dalam data *training*, data status sudah memiliki kategori positif atau negatif. Dalam penelitian ini, peneliti Menggunakan data *training* dengan jumlah 160 data status yang disimpan di *database* dengan nama tabel *set_training*. Tabel *set_training* terdiri dari *field* id, status, dan kategori.

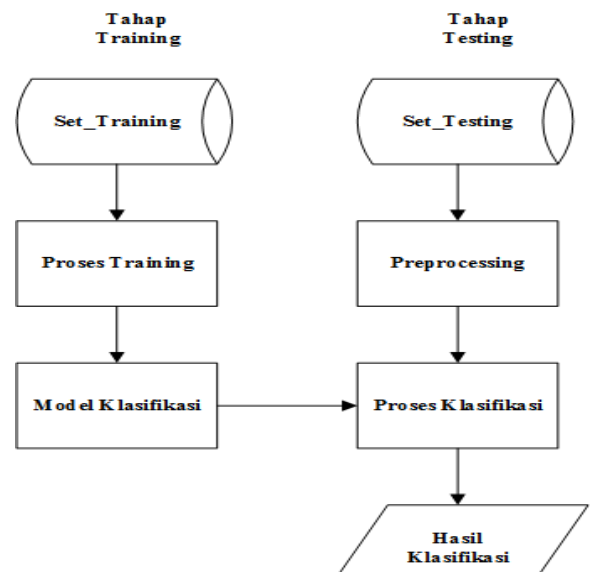
Tabel 2. Tabel Data Testing

id	Status
1	untuk cewek manado yang mau promosi bisa inbox saya, harga diatas 15 juta per bulan.
2	Dibutuhkan pegawai serabutan. Wanita single, mau belajar, ada mess, tanpa surat lamaran. Langsung sms ke 082232366105
3	OPEN BO 700 permalam DP 300 PIN 25A9E916
4	Setiap orang menginginkan kebahagiaan. Tidak ada yang menginginkan kepedihan atau kesedihan.

Tabel 2 merupakan data *testing* berisi kumpulan status baru yang belum memiliki kategori, yang nantinya akan dianalisa oleh sistem apakah status tersebut berkategori positif atau negatif berdasarkan data *training*. Dalam penelitian ini, peneliti Menggunakan data *testing* dengan jumlah 160 data status yang disimpan dalam *database* dengan nama tabel *set_testing*.

B. Flowchart sistem klasifikasi

Flowchart sistem didefinisikan sebagai penguraian dari sistem utama ke dalam sub-sub sistem dengan tujuan untuk mengidentifikasi permasalahan-permasalahan yang ada dan kebutuhan-kebutuhan yang diperlukan agar dapat diusulkan dan diciptakan sistem baru yang lebih baik. Dalam mengimplementasikan metode *Naive Bayes Classifier* akan dibangun sebuah aplikasi dengan gambaran sistem sesuai pada gambar 1. Sistem akan dibangun memiliki 3 tahap yaitu praproses, training data, testing data



Gambar 1. Flowchart Sistem Klasifikasi.

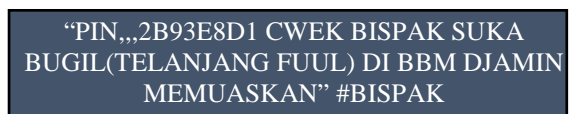
IV. HASIL DAN PEMBAHASAN

A. Analisa Pola Status

Dalam analisa pola status, ada beberapa tahap yang harus dilakukan untuk menentukan hasil klasifikasi pola dari data status yang ada. Ada beberapa tahap dalam analisa pola status :

1. Text Preprocessing

Tahap text preprocessing adalah tahap awal dalam model text mining, dimana data status yang berupa data teks yang tidak terstruktur akan diubah menjadi data teks yang terstruktur atau semi terstruktur. Misalnya terdapat data input sebagai berikut :



Gambar 2. Inputan Status

Dalam tahap *text preprocessing* terdapat beberapa proses, yakni :

- Pada proses *case folding*, data inputan yang ada pada gambar 2 akan berubah menjadi seperti pada gambar 3.

pin cwek bispak suka bugil (telanjang fuul)
di bbm dijamin memuaskan
bkn tipu #bispak☺

Gambar 3. Output Proses Case Folding

- Pada proses *cleansing*, data status yang ada pada gambar 3 akan berubah menjadi seperti pada gambar 4.

pin cwek bispak suka bugil (telanjang fuul)
di bbm dijamin memuaskan bkn tipu ☺

Gambar 4. Output Proses Cleansing

- Pada proses *stopword removal*, data status pada gambar 4 akan berubah menjadi seperti pada gambar 5.

pin cwek bispak suka bugil (telanjang fuul) bbm
jamin memuaskan bkn tipu ☺

Gambar 5. Output Proses Stopword Removal

- Pada proses *convert emoticon*, data status hasil dari proses *stopword removal* pada gambar 5 akan berubah menjadi seperti pada gambar 6.

Pin cwek bispak suka bugil (telanjang fuul)
bbm jamin memuaskan bkn tipu emotsenang

Gambar 6. Output Proses Convert Emoticon

- Pada proses *convert negation*, data status hasil proses *convert emoticon* pada gambar 6 akan berubah menjadi seperti pada gambar 7.

pin cwek bispak suka bugil (telanjang fuul) bbm
jamin memuaskan bukan tipu emotsenang

Gambar 7. Output Proses Convert Negation

- Pada proses *tokenisasi*, data status hasil proses *convert negation* yang ada pada gambar 7 berubah menjadi seperti pada gambar 8.

2. Tahap *training*

Tahap *training* adalah tahap dimana peneliti akan *train* data status yang ada, yang nantinya akan dijadikan sebagai data acuan atau data latih untuk menentukan kategori dalam proses klasifikasi terhadap data baru yang

belum berkategori. Misalnya terdapat data status *training* sebagai berikut pada tabel 3.

pin
cwek
bispak
suka
bugil
telanjang
fuul
bbm
jamin
memuaskan
bukan
tipu
emotsenang

Gambar 8. Output Proses Tokenisasi

Tabel 3. Status *training*

Status Facebook	Kategori	Fitur
Status1	Positif	sherry imuet mencari dikamar sendiri menemani
Status2	Positif	pin cwek bispak bugil telanjang full bbm djamin memuaskan
Status3	Positif	pilihan harus hadapi hidup pahit
Status4	Negatif	senyum tawamu kebahagiaan terbesar hidupku
Status5	Negatif	positif hidup panjang jalani ikhlas masa depan indah
Status6	Negatif	kini mengulangi sudah

Dari kumpulan data status *training* yang berjumlah 160 data status, akan di cari nilai probabilitas positif dan negatifnya dengan Menggunakan persamaan 5, penjabarannya seperti dibawah ini :

$$\begin{aligned}
 P(W_{bugil} | c_{positif}) &= \frac{\#words(w_{bugil}, c_{positif}) + 1}{\#words(c_{positif}) + |V|} \\
 &= \frac{1 + 1}{20 + 37} = \frac{2}{57} = 0,0350877193
 \end{aligned}$$

$$\begin{aligned}
 &P(W_{bugil} | c_{negatif}) \\
 &= \frac{\#words(w_{bugil}, c_{negatif}) + 1}{\#words(c_{negatif}) + |V|} \\
 &= \frac{0 + 1}{17 + 37} = \frac{1}{54} = 0,0185185185
 \end{aligned}$$

Jika dibentuk sebuah tabel, maka hasil dari perhitungan probabilistik setiap kata pada data training terdapat pada Tabel 4.

Tabel 4. Detail data *training*

Kategori	Positif	Negatif
P(c)	1/2	1/2
P(W _{ki} c)		
sherry	0,0350877193	0,0185185185
imuet	0,0350877193	0,0185185185
mencari	0,0350877193	0,0185185185
dikamar	0,0350877193	0,0185185185
sendiri	0,0350877193	0,0185185185
menemani	0,0350877193	0,0185185185
pin	0,0350877193	0,0185185185
cwek	0,0350877193	0,0185185185
bispak	0,0350877193	0,0185185185
panjang	0,0175438596	0,037037037
jalani	0,0175438596	0,037037037
ikhlas	0,0175438596	0,037037037
masa	0,0175438596	0,037037037
depan	0,0175438596	0,037037037
indah	0,0175438596	0,037037037
kini	0,0175438596	0,037037037
mengulangi	0,0175438596	0,037037037
sudah	0,0175438596	0,037037037

3. Tahap Klasifikasi

Dalam tahap klasifikasi, terdapat data status baru yang belum memiliki kategori yang nantinya akan ditentukan kategori dari masing-masing status, apakah kategori positif atau kategori negatif berdasarkan pada data *training* atau data latih yang telah ada. Misalnya terdapat sebuah status sebagai berikut :

Tabel 5. Status baru (belum berkategori)

Status facebook	Kategori	Fitur
Status1	?	canda tawa hapuskan canda yg gagal
Status2	?	berkunjung dikamar
Status3	?	hidup mencari kerja susah

Untuk menentukan kategori dari ketiga status baru yang belum berkategori tersebut, digunakan persamaan (2) dan (3), dengan mencari nilai probabilitas positif dan negatif untuk tiap status kemudian dari nilai probabilitas positif dan negatif ditentukan nilai probabilitas tertinggi untuk kategori status tersebut. Contoh penjabarannya seperti dibawah ini :

$$P(\text{Status1} | C_{positif})$$

$$\begin{aligned}
 &= P(W_{canda} | C_{positif}) \times P(W_{tawa} | C_{positif}) \times P(W_{hapuskan} | C_{positif}) \\
 &\quad \times P(W_{yg} | C_{positif}) \times P(W_{gagal} | C_{positif}) \times P(C_{positif}) \\
 &= 1 \times 1 \times 1 \times 1 \times 1 \times \frac{1}{57} \times \frac{1}{2} = 0,0087719298245614
 \end{aligned}$$

$$P(\text{Status1} | C_{negatif})$$

$$\begin{aligned}
 &= P(W_{canda} | C_{negatif}) \times P(W_{tawa} | C_{negatif}) \times P(W_{hapuskan} | C_{negatif}) \\
 &\quad \times P(W_{yg} | C_{negatif}) \times P(W_{gagal} | C_{negatif}) \times P(C_{negatif}) \\
 &= 1 \times 1 \times 1 \times 1 \times 1 \times \frac{2}{54} \times \frac{1}{2} = 0,0185185185185185
 \end{aligned}$$

Dari kedua hasil perhitungan probabilitas daridata status1 tersebut, ditentukan nilai probabilitas tertinggi. Antara probabilitas positif 0,0087719298245614 dan probabilitas negatif 0,0185185185185185, maka kategori data status1 adalah negatif. Begitupun seterusnya untuk menentukan kategori untuk data status2 dan data status3, berikut peneliti menampilkan hasil perhitungan probabilitas untuk ketiga data status yang akan di uji.

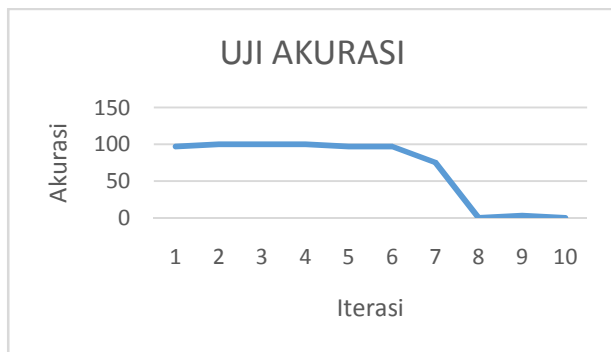
Tabel 6. Hasil pengklasifikasian status

Status	Positif	Negatif
Status1	0,008771929824561	0,018518518518518
Status2	0,017543859649122	0,009259259259259
Status3	0,000615574022776	0,000342935528120

Untuk dokumen status1 sebagai data baru pertama dapat disimpulkan bahwa termasuk kedalam kategori negatif, dilihat dari nilai yang terbesar yaitu terdapat pada kategori negatif. Sedangkan untuk status 2 termasuk kedalam kategori positif karena dilihat dari hasil yang didapatkan, nilai dengan kategoripositif lebih besar dibandingkan dengan kategori

negatif dan status 3 merupakan kategori positif karena nilai yang terbesar yaitu pada perhitungan dengan kategori positif.

4. Tahap Pengujian



Gambar 9. Hasil pengujian akurasi

Gambar 9 menggambarkan tentang hasil dari proses pengujian, yang ditampilkan dalam bentuk grafik. Dalam teknik pengujian akurasi dengan *10-fold cross validation* ini dataset dibagi menjadi sejumlah 10 bagian partisi secara acak. Kemudian dilakukan sejumlah 10 kali (10 iterasi) pengukuran akurasi, dimana masing iterasi menggunakan data partisi sebagai data *testing* dan memanfaatkan sisanya sebagai data *training*. Untuk mendapatkan nilai akurasi dari keseluruhan iterasi, dapat diambil nilai rata dari seluruh iterasi yang ada.

Dari hasil pengujian akurasi untuk data dengan jumlah 320 menghasilkan akurasi rata-rata 57.1875 %. Dimana untuk setiap iterasi dalam 10 iterasi menggunakan data test sebanyak 32 dan data *training* 288.

V. PENUTUP

A. Kesimpulan

Berdasarkan dari pembahasan-pembahasan pada bab-bab sebelumnya, maka dapat ditarik kesimpulan dari Skripsi “Analisa Pola Penyalahgunaan Facebook Sebagai Alat Kejahatan Trafficking menggunakan Data Mining”, sebagai berikut :

1. Hasil akurasi penggunaan metode *naïve bayes classifier* pada sistem klasifikasi dengan menggunakan dataset sejumlah 320, menghasilkan akurasi 57.1875%
2. Hasil akurasi terhadap penggunaan metode *naïve bayes classifier* (NBC) dipengaruhi oleh beberapa hal, yaitu :
 - Komposisi jumlah data *training* dan *testing*
 - Jumlah dataset yang digunakan

B. Saran

Saran yang ingin disampaikan untuk pengembangan lebih lanjut penelitian ini adalah sebagai berikut :

1. Pengembangan sistem online yang dapat mengambil dataset dan menghasilkan analisis klasifikasi secara realtime.
2. Perlu dibuat kamus khusus untuk penanganan kasus yang menggunakan data Facebook berbahasa Indonesia, yang cenderung menggunakan bahasa yang tidak baku, agar performansi menjadi lebih baik.

DAFTAR PUSTAKA

- [1] Fieldman, Ronen.2011. *The Text Mining Handbook*. Cambdrige University
- [2] Kusumawardhani. *Human Trafficking : Pola Pencegahan dan Penanggulangan Terpadu Terhadap Perdagangan Perempuan*. Lembaga Ilmu Pengetahuan Indonesia. 2010.
- [3] Manalu, Boy Utomo. 2014. Analisis Sentimen Pada Twitter Menggunakan *Text Mining*. Program Studi Teknologi Informasi. Fakultas Ilmu Komputer Teknologi Informasi. Universitas Sumatera Utara. Medan.
- [4] Muflichah, Hj. Siti, dkk. Fakultas Hukum Universitas Jendral Soedirman Purwokerto. Trafficking : suatu studi tentang perdagangan perempuan dari aspek social, budaya dan ekonomi dikabupaten banyumas.
- [5] Mulyatiningsih, Endang. Metode Penentuan Sampel.
- [6] Ramadhan, R Anna. 2013. Keabsahan Status I Facebook Sebagai Alat Bukti Dalam Perkara Tindak Pidana. Fakultas Hukum. Universitas Mataram.
- [7] Saraswati, Ni Wayan Sumartini. 2011. *Text Mining Dengan Metode Naïve Bayes Classifier dan Support Vector Machines Untuk Sentiment Analysis*. Program Pascasarjana. Universitas Udayana Denpasar.
- [8] Shadiq, Ammar.2012.*Keoptimalan Naïve Bayes Dalam Klasifikasi*. Program Ilmu Komputer. Universitas Pendidikan Indonesia.
- [9] Subiyakto, A'ang. Penggunaan Algoritma Klasifikasi Dalam *Data Mining*. Program Studi Sistem Informasi Fakultas Sains dan Teknologi UIN Jakarta.



Sekilas dari penulis dengan nama lengkap Linda Jayanti, lahir di kota Berau, Provinsi Kalimantan Timur. Anak ke-1 dari 2 bersaudara. Dengan pendidikan Sekolah Dasar Negeri 001 Batu Putih. Kemudian melanjutkan ke Sekolah Menengah Pertama Negeri SMPN 019 Berau. Kemudian melanjutkan ke Sekolah Menengah Atas SMAN 1 Berau. Setelah lulus tahun 2011 melanjutkan ke Perguruan Tinggi di Universitas Sam Ratulangi Manado dengan mengambil Jurusan Teknik Informatika. Pada tahun 2015 bulan april, penulis membuat Skripsi demi memenuhi syarat Sarjana (S1) dengan penelitian berjudul Analisa Pola Penyalahgunaan Facebook Sebagai Alat Kejahatan Trafficking Menggunakan Data Mining yang dibimbing oleh dua dosen pembimbing yaitu Dr. Eng Steven R. Sentinuwo, ST., MTI dan Oktavian Abraham Lantang, ST., MTI sehingga pada tanggal 6 April 2016 penulis resmi lulus di Teknik Informatika Universitas Sam Ratulangi Manado dan menyandang gelar Sarjana Komputer dengan predikat Cum Laude.