

KLASIFIKASI SMS SPAM DENGAN METODE NAIVE BAYES CLASSIFIER UNTUK MENYARING PESAN MELALUI SELULAR

Budi Indiartho

Teknik Informatika, Fakultas Teknologi Informasi, IKPIA Perbanas
Jl. Perbanas, Karet Kuningan, Setiabudi, Jakarta Selatan, 12940
budi.indiartho@perbanas.id

ABSTRAK

SMS atau Short Message Service (Layanan Pesan Singkat) merupakan sarana komunikasi yang masih dipergunakan oleh banyak orang yang menggunakan telepon genggam. SMS adalah sebuah layanan pesan singkat yang berisi 160 karakter per pesan yang dikirimkan antar pengguna telepon genggam. Orang dapat dengan mudah mengirimkan dan menerima SMS tanpa ada batasan. SMS tidak hanya digunakan untuk mengirimkan pesan antar pengguna yang sudah mengenal satu sama lain, tetapi juga digunakan oleh mereka yang tidak dikenal untuk berbagai kepentingan, seperti menawarkan berbagai macam produk dan jasa bahkan SMS juga digunakan untuk tindakan yang tidak sesuai dengan norma hukum yang mengarah pada tindak pidana. SMS yang tidak diinginkan oleh penerima dikategorikan sebagai spam atau junk SMS atau SMS sampah. Untuk menyaring SMS yang masuk ke inbox SMS, filter spam bisa dibuat. Salah satu algoritma yang dapat dipergunakan untuk menyaring SMS adalah algoritma Naive Bayes Classifier. Dengan menggunakan algoritma penyaring SMS spam diharapkan dapat digunakan untuk menyaring SMS yang masuk ke inbox pengguna, sehingga pengguna dapat memilih SMS yang bermanfaat saja yang akan dibaca dan/atau ditindaklanjuti.

Kata kunci: SMS, spam, penyaringan, naive bayes classifier

I. PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi telekomunikasi sudah sedemikian pesat sehingga perangkat selular yang tadinya berfungsi sebagai alat untuk berkomunikasi suara sudah bertambah menjadi komunikasi tulisan dan terakhir kali sudah berubah menjadi perangkat pintar atau diistilahkan dengan *smartphone* karena perangkat genggam sudah bisa berfungsi layaknya sebagai sebuah komputer. Namun dibalik kecanggihan perangkat genggam yang ada sekarang masih ada layanan yang bertahan sampai sekarang berupa layanan pesan singkat atau lebih dikenal dengan istilah SMS atau *short message service*.

Berdasarkan data dari ITU [1], pada tahun 2013 jumlah nomor selular yang dipakai di Indonesia sudah melebihi jumlah penduduk Indonesia atau sekitar 313 juta. Perangkat selular yang dipergunakan saat ini semuanya memiliki fasilitas untuk mengirimkan pesan singkat atau SMS. Sekalipun sudah ada sarana lain yang dapat dipergunakan untuk berkirim pesan selain menggunakan sms seperti *email*, *instant messaging*, dan sebagainya tetapi dibutuhkan perangkat genggam khusus untuk memanfaatkannya, sehingga SMS masih tetap bertahan sampai dengan sekarang.

Dengan banyaknya jumlah pengguna perangkat genggam yang memiliki fasilitas SMS tersebut maka tidak mengherankan jika SMS masih menjadi salah satu media yang dipergunakan untuk berkomunikasi antar pengguna. Namun dalam perkembangannya SMS tidak saja dipergunakan untuk mengirimkan pesan kepada pengguna yang sudah saling kenal tetapi dipergunakan juga untuk mengirimkan pesan kepada orang yang belum dikenal dengan tujuan untuk menawarkan

produk maupun jasa. Bahkan sudah mulai banyak dipergunakan oleh orang tidak bertanggungjawab untuk melakukan kejahatan. Hal tersebut berarti juga bahwa pengguna sering menerima SMS yang tidak dikehendaki atau diistilahkan sebagai spam yang masuk ke dalam inbox penggunanya.

Naive Bayes merupakan suatu metode *machine learning* yang menggunakan perhitungan probabilitas. Konsep dasar yang dipergunakan dalam Naive Bayes adalah Teorema Bayes (atau aturan Bayes) dengan asumsi independensi (ketidakbergantungan) yang kuat (naif). Teorema tersebut dipergunakan dalam statistika untuk menghitung suatu peluang dari satu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal.

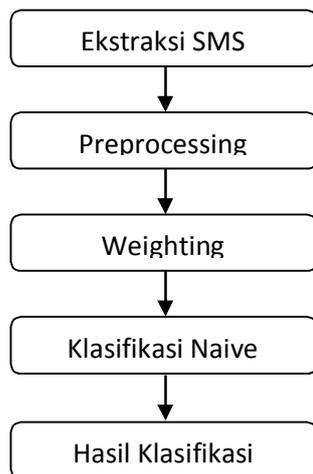
1.2. Identifikasi Masalah

Berdasarkan uraian latar belakang masalah tersebut di atas, maka permasalahan yang akan diteliti dapat diidentifikasi sebagai berikut :

Bagaimana membangun klasifikasi SMS spam dengan metode *Naive Bayes Classifier* untuk menyaring pesan melalui selular ?

1.3. Kerangka Pemikiran

Kerangka berpikir penelitian ini dapat digambarkan sebagai berikut :



Gambar 1. Kerangka Pemikiran

Proses pertama adalah proses ekstraksi SMS dari inbox pengguna, kemudian dilanjutkan dengan pra proses terhadap text yang diperoleh yang bertujuan memperoleh kata-kata yang diidentifikasi, selanjutnya adalah pembobotan yang berupa fekwensi untuk mencari peluang Bayes. Pada tahap selanjutnya dari hasil tahap sebelumnya dipergunakan untuk melakukan klasifikasi yang akan menentukan apakah SMS berupa spam atau bukan.

II. LANDASAN TEORI

2.1 Short Message Service (SMS) dan Spam

Short Message Service atau disingkat dengan SMS merupakan komponen layanan pesan tekstual sistem komunikasi telepon, web atau telepon bergerak. Komunikasi menggunakan protokol komunikasi standar yang memungkinkan perangkat telepon tetap dan telepon bergerak untuk bertukar pesan singkat. Pesan singkat yang dikirimkan maksimal 160 karakter per pengiriman [2].

Mobile phone spam SMS spam adalah SMS yang masuk ke dalam inbox pengguna seluler tetapi SMS tersebut merupakan SMS yang tidak dikehendaki oleh penggunanya. Istilah lainnya adalah mobile spamming, SMS spam, text spam, m-spam atau mspam [3].

2.2 Data Mining dan Text Mining

Data mining sebagai proses untuk mendapatkan informasi yang berguna dari gudang data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga sebagai knowledge discovery [4].

Text Mining didefinisikan sebagai menambang data yang berbentuk teks dimana sumber datanya didapat dari suatu dokumen atau tulisan dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen[5].

2.3 Algoritma Naive Bayes Classifier

Algoritma Naive Bayes Classifier merupakan metode terbaru yang digunakan untuk mengklasifikasikan sekumpulan dokumen. Algoritma ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [6].

Dasar dari teorema naive Bayes yang digunakan dalam pembahasan ini adalah sebagai berikut :

$$P(a|b) = (P(b|a) * P(a)) / P(b)$$

Peluang kejadian a sebagai b ditentukan dari peluang b saat a, peluang a, dan peluang b. Dalam pengaplikasiannya rumus tersebut akan berubah menjadi :

$$P(c_i|d) = (P(d|c_i)*P(c_i)) / P(d)$$

Naive Bayes Classifier atau dapat juga disebut dengan multinomial naive bayes merupakan model penyederhanaan dari algoritma bayes yang sesuai untuk pengklasifikasian text. Persamaannya adalah sebagai berikut :

$$V_{map} = \arg \max P(v_j | a_1, a_2, \dots, a_n)$$

Berdasarkan persamaan tersebut di atas, maka persamaan pertama dapat dituliskan dengan :

$$V_{map} = \arg \max_{V_j \in V} \frac{P(a_1, a_2, \dots, a_n | v_j) P(v_j)}{P(a_1, a_2, \dots, a_n)}$$

Dalam hal P(a₁, a₂, ... a_n) konstan dapat dihilangkan, maka persamaan menjadi :

$$V_{map} = \arg \max_{V_j \in V} \prod_i P(a_i | v_j)$$

P(a₁, a₂, ... a_n | v_j) sulit dihitung, dalam konteks tersebut maka dianggap setiap kata pada SMS tidak memiliki saling keterkaitan (naive).

$$V_{map} = \arg \max_{V_j \in V} P(V_j) \prod_i P(a_i | v_j)$$

Keterangan :

$$P(V_j) = \frac{|dok_j|}{|total|}$$

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

Dimana :

P(v_j) adalah probabilitas setiap dokumen terhadap sekumpulan dokumen

P(w_k | v_j) adalah probabilitas kemunculan kata w_k pada suatu dokumen dengan kategori kelas v_j

|dok| adalah frekwensi dokumen pada setiap kategori

|total| adalah jumlah dokumen yang ada

n_k adalah frekwensi kata ke – k pada setiap kategori

|kosakata| : jumlah kata pada dokumen test

2.4 Penelitian Terdahulu

- 1) Rujukan penelitian pertama adalah penelitian Prasetyo Anugroho, Idris Winarno, Nur Rosyid dengan judul Klasifikasi Email Spam dengan Metode Naive Bayes Classifier Menggunakan Java Programming. Dalam penelitian tersebut peneliti menggunakan naive bayes classifier untuk meyeleksi email spam [5].

- 2) Rujukan penelitian selanjutnya adalah penelitian Amir Hamzah dengan judul Klasifikasi Teks dengan Naive Bayes Classifier (NBC) untuk Pengelompokan teks berita dan abstrak akademis. Dalam penelitian tersebut naive bayes classifier dipergunakan untuk mengklasifikasi dokumen berita maupun dokumen akademis [7].
- 3) Rujukan penelitian berikutnya adalah penelitian Selvia Lorena Br Ginting dan Reggy Pasya Trinanda yang berjudul Penggunaan Metode Naive Bayes Classifier pada Aplikasi Perpustakaan. Dalam penelitian ini naive bayes classifier dipergunakan untuk mengklasifikasikan beberapa judul dan kategori yang terdapat pada database perpustakaan [8].

III. METODE PENELITIAN

3.1 Metode Penelitian

Penelitian yang dilakukan adalah *action research* yang merupakan bentuk penelitian terapan (*applied research*) yang bertujuan untuk mencari cara yang efektif dalam menyaring informasi yang masuk dalam komunikasi seluler melalui layanan pesan singkat atau SMS sehingga dapat memisahkan informasi yang dikehendaki dengan informasi yang tidak dikehendaki.

3.2 Obyek Penelitian

Obyek dari penelitian ini adalah SMS yang masuk dalam inbox pengguna selular operator telkomsel dengan nomor 0812 830XXXX dalam kurun waktu tahun 2010 - 2014.

3.3 Metode Pengumpulan Data

Dalam penelitian ini data dikumpulkan dari SMS yang masuk ke dalam inbox pengguna dalam kurun waktu tahun 2010 sampai dengan tahun 2014 yang dibackup secara berkala setiap tahun. Data tersebut dibackup dalam bentuk file dan disimpan dalam komputer.

Data SMS selama kurun waktu tersebut kemudian dibuka dan dijadikan satu kesatuan dalam dua buah tabel yang berisi data yang terkait dengan nomor seluler dan data yang terkait dengan isi SMS.

3.4 Teknik Analisa Data

Analisis dilakukan dengan menggunakan metode data mining, khususnya adalah text mining dengan menggunakan algoritma *naive bayes classifier*. Dasar teorema naive Bayes yang dipergunakan adalah sebagai berikut :

1) Learning (Pembelajaran)

Naive Bayes merupakan algoritma yang termasuk ke dalam kelompok *supervised learning*, dimana dibutuhkan pengetahuan awal untuk dapat mengambil keputusan. Langkah-langkah yang diperlukan adalah sebagai berikut :

- a) Langkah 1 : bentuk vocabulary pada setiap dokumen *training*
- b) Langkah 2 : hitung probabilitas pada setiap kategori $P(v_j)$
- c) Langkah 3 : tentukan frekwensi setiap kata w_k yang ada pada setiap kategori $P(w_k | v_j)$

2) Classify (Pengklasifikasian)

Naive Bayes merupakan algoritma yang termasuk ke dalam kelompok *supervised learning*, dimana dibutuhkan pengetahuan awal untuk dapat mengambil keputusan. Langkah-langkah yang diperlukan adalah sebagai berikut :

- a) Langkah 1 : Hitung $P(v_j) \prod P(a_i | v_j)$ untuk setiap kategori
- b) Langkah 2 : Tentukan kategori dengan nilai $P(v_j) \prod P(a_i | v_j)$ maksimal

IV. HASIL DAN ANALISIS

4.1 Ekstraksi SMS

Sesuai dengan tahapan penelitian, maka langkah pertama adalah mempersiapkan dokumen atau file yang akan dijadikan sebagai sumber data untuk melakukan penelitian. Dalam penelitian penulis mengambil data SMS yang berasal dari inbox HP Nokia E71 dengan menggunakan perangkat lunak Nokia Suite yang mengubah file SMS menjadi file dengan format CSV. File SMS hasil ekstraksi tersebut dapat dilihat dalam tabel berikut ini :

Tabel 1. Contoh Data SMS hasil ekstraksi

| | | | | |
|-----|---------------|----------|---------------------|---|
| sms | READ,RECEIVED | CITIBANK | 2011.09.07 19:05 | Payment for your CITIBANK Card VISA 4513 amount Rp 3.065.225 received 07-09-2011 19:04. For info type balance(space) last 6 digit card number sent to 69999 |
| sms | READ,RECEIVED | 6.28E+11 | 2011.09.07 18:58 | ok insya allah jumat jam 10 sy ke id sirti |
| sms | SENT,READ | 6.28E+11 | 2011.09.07 18:56 | Wass, minal aidin wal faidzin, jum'at dari pagi sampai siang Insyah'Alloh bisa |
| sms | READ,RECEIVED | 6.28E+11 | 2011.09.07 17:41 | ass p budi..minal aidin wal faizin..mohon maaf lahir batin btw..apa jumat sy bisa ketemu? jam berapa? tks |
| sms | READ,RECEIVED | 6.28E+12 | 2011.09.07 14:46 | Cicilan 3.166.667/Bln Anda Dpt KTA 100jt,DASYAT!!FROVISI 0% 5jt s/d 200jt,heboh HADIAHnya TV LCD 24 Inc Di setujui >100jt,Syrt |

| | | | | |
|-----|---------------|----------|---------------------|--|
| | | | | KTP & CC,Hub Ari 087877556087 |
| sms | READ,RECEIVED | 6.29E+12 | 2011.09.07 14:43 | Selamat siang., Kredit Tnp Agunan dg Bunga 1.5%-1.79%, FREE Provisi dari 20-150jt,Proses 3-5hr, Syarat: copy KTP+Kartu Krdt, Usia Max.63th &Tenor 1-5th.(ArrA) |
| sms | READ,RECEIVED | BCA | 2011.09.07 11:58 | PAULANER BRAUHAUS Special DISCOUNT 20% on WEEKDAY LUNCH (MON-FRI) with BCA Gold,Platinum & World CreditCards until October 31, 2011. T&C apply.RSVP 021-23583871 |
| sms | SENT,READ | 6.28E+12 | 2011.09.07 09:58 | Pesawat Emirates - EK 357, depart : jakarta 25 sept, 17.55 arrive : dubai : 22.45, pulang : dubai 29 sept, 04.15, jakarta : 15.40 |

File dalam format CSV tersebut akan diekstraksi lagi untuk diambil isinya yang dipilih adalah SMS dikirimkan oleh pihak lain. Tanggal dan nomor HP yang ada dalam file tersebut bukan merupakan bagian dari penelitian ini, sehingga datanya diabaikan. Berikut adalah contoh hasil ekstraksi file CSV yang diubah dalam bentuk file TXT.

Tabel 2. Contoh Teks SMS hasil ekstraksi

| |
|--|
| <p>Selamat!!! simcard anda mendapatkan Hadiah Cek Tunai Rp.15jt dari kukubima Ener-G! U/info Hub: PT.SIDOMUNCUL 021-32793481 Penting/Rahasia, Cek Hadiah Anda PIN - Bx777AA INFO Lengkap: www.TELKOMSEL777.tk Selamat!!Anda men-dpt Hadiah 1unit Mobil AVANZA dari TELKOMSELpoin 2013 Pin Anda.xy1i94b U/Info Buka Website www.telkomselpoint-2013.webs.com Pengirim: 777 INFO RESMIPT.INDOFOOD Anda pmenang POP MIE 2013 M-ndpt hadiah PIN; pmenang Anda; 25E57 U/INFO; HUB 082223475555 www.kejutanberhadiah-indofood.blogspot.c0m</p> |
|--|

| |
|--|
| <p>Selamat! No.Hp Anda terpilih 081283xxxx dapat Hadiah THR dr TELKOMSEL No.PIN Anda b89c7h9 Info klik,di www.undian-semarak-telkomselpoin.blogspot.com PESAN RESMI dr PT.TELKOMSEL Anda mendapat HADIAH 1 unit MOBIL dari TELKOMSELpoin NO.Pin.b89c7h9 Info;klik www.telkomselpoin222.hol.es SELAMAT!! Anda Mendapt'HADIAH Cek.Rp.75.jt dr PT.TELKOMSEL PIN,277fg49 Info.hb: 085377177138 Ir.Serianto Klik: www.telkomselpoin18th.blogspot.com km dr BANK BRI sdh tlpn Anda tp tdk trsbng jd km SMS No.PIN.Pemenang Anda kg65h8k9 M-dpt hadiah dr untung bliung britAma info klik: www.Gebyar-BRI.blogspot.com Berdasarkan hasil seleksi PIN BlackBerry 2013 periode 5, no RKP ID anda (8496127CV) Resmi mendapat Hadiah Rp 57jt U/info Klik; www.RIM-BlackBerryPROMO.webs.com Nasabah MANDIRI Yth anda m-dptkan Cek Tunai Rp.30jt dr Bank MANDIRI fiesta ID:658QAWK KLIK: www.pestahadiah.webs.com info lengkap Hub: 021-3460-7596</p> |
|--|

4.2 Preprocessing SMS

Dari sejumlah teks SMS yang sudah diperoleh akan diseleksi beberapa teks yang akan dijadikan sebagai dokumen latih dalam tahapan awal *text mining*. Tahapan yang selanjutnya dilakukan adalah proses tokenisasi, dimana teks akan dipecah menjadi penggalan kata-kata sebagaimana terlihat dalam contoh tabel berikut ini :

Tabel 3. Contoh Tokenisasi teks SMS

| |
|--|
| <p>Selamat!!! simcard anda mendapatkan Hadiah Cek Tunai Rp.15jt dari kukubima Ener-G! U/info Hub: PT.SIDOMUNCUL 021-32793481 Penting/Rahasia, Cek Hadiah Anda PIN - Bx777AA INFO Lengkap: www.TELKOMSEL777.tk Selamat!!Anda men-dpt Hadiah 1unit</p> |
|--|

Mobil
 AVANZA
 dari
 TELKOMSELpoin
 2013
 Pin
 Anda.xy1i94b
 U/Info
 Buka
 Website
 www.telkomselpoint-2013.webs.com
 Pengirim:
 777

Pada tahapan berikutnya yang dilakukan adalah dengan menghapus *stopword*. Proses ini dilakukan untuk menghilangkan kata-kata yang kurang memiliki arti. Contoh *stopwords* adalah kata depan seperti : di, ke, pada, dari, untuk dan sebagainya. Dalam penelitian ini, *stopword* yang akan dijadikan sebagai bagian kata yang dibuang adalah : di, ke, dari, pada, kepada, untuk.

Pada tahapan berikutnya adalah dilakukan proses *stemming*, dimana untuk menghasilkan kata dasar yang ada dalam bahasa Indonesia proses ini diperlukan perubahan kata-kata yang diperoleh menjadi kata-kata dasarnya. Dalam hal ini ada dokumen yang dibutuhkan, yaitu kata dasar yang berasal dari Kamus Besar Bahasa Indonesia, yang memiliki kurang lebih 28.526 kata dasar.

4.3 Weighting

Pada tahapan berikutnya adalah dilakukan proses *stemming*, dimana untuk menghasilkan kata dasar yang ada dalam bahasa Indonesia proses ini diperlukan perubahan kata-kata yang diperoleh menjadi kata-kata dasarnya. Dalam hal ini ada dokumen yang dibutuhkan, yaitu kata dasar yang berasal dari Kamus Besar Bahasa Indonesia, yang memiliki kurang lebih 28.526 kata dasar.

Setelah tahapan tersebut selesai, maka dapat dilakukan proses selanjutnya yang berupa pembobotan. Pembobotan yang dipergunakan adalah *term frequency* atau *tf*. Pembobotan tersebut dipergunakan karena pada klasifikasi *Naive Bayes* diperlukan frekuensi dari kata-kata untuk mencari peluang *Bayes*-nya.

Dari hasil analisis terhadap data SMS digunakan kurang lebih 120 SMS sebagai data training, maka diperoleh informasi sebagai berikut :

Tabel 4. Data Training dan tingkat error

| Jml Data | Jml Spam | Jml Bukan Spam | Jumlah keyword pada spam | Jumlah keyword pada bukan spam | Nilai Error |
|----------|----------|----------------|--------------------------|--------------------------------|-------------|
| 20 | 10 | 10 | 124 | 104 | 2 |
| 40 | 20 | 20 | 345 | 267 | 2 |
| 60 | 30 | 30 | 367 | 304 | 4 |
| 80 | 40 | 40 | 456 | 412 | 4 |
| 100 | 50 | 50 | 524 | 536 | 4 |
| 120 | 60 | 60 | 611 | 657 | 4 |

4.4 Klasifikasi dengan Naive Bayes

Dari tabel tersebut di atas dapat dilihat hasil klasifikasi SMS untuk 120 data latih yang telah diklasifikasi. Terdapat 4 kesalahan yang diperoleh, yaitu 2 data yang merupakan spam ternyata bukan dan 2 data bukan spam ternyata spam. Sehingga prosentase kesalahan adalah = $4/120 * 100\% = 3,33\%$.

Tabel 5. Hasil Klasifikasi Data Training

| No | SMS yang masuk | Hasil klasifikasi SMS | Nilai Kebenaran |
|----|--------------------|-----------------------|-----------------|
| 1 | Spam 1 | SPAM | Benar |
| 2 | Spam 2 | SPAM | Benar |
| 3 | Spam 3 | SPAM | Benar |
| 4 | Spam 4 | BUKAN SPAM | Salah |
| 5 | Spam 5 - 55 | SPAM | Benar |
| 6 | Spam 56 | BUKAN SPAM | Salah |
| 7 | Spam 57 - 60 | SPAM | Benar |
| 8 | Bukan Spam 1 - 11 | BUKAN SPAM | Benar |
| 9 | Bukan Spam 12 | BUKAN SPAM | Salah |
| 10 | Bukan Spam 13 - 47 | BUKAN SPAM | Benar |
| 11 | Bukan Spam 48 | BUKAN SPAM | Salah |
| 12 | Bukan Spam 49 - 60 | BUKAN SPAM | Benar |

V. PENUTUP

5.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa :

- a. Naive bayes classifier algorithm dapat dipergunakan sebagai sarana untuk memfiltrasi SMS yang masuk ke dalam inbox pengguna
- b. Naive bayes classifier algorithm dapat memberikan hasil filtrasi yang cukup akurat untuk menyaring SMS yang masuk

5.2. Saran

Saran yang dapat diberikan untuk pengembangan penelitian ini adalah sebagai berikut :

- a. Mengingat SMS masih merupakan sarana yang dapat dipergunakan oleh masyarakat untuk mengirimkan pesan dengan mudah, dan masih banyaknya kemungkinan SMS spam yang akan muncul di kemudian hari, maka diperlukan lagi data uji baru sehingga dapat dipergunakan untuk meningkatkan kualitas filtrasi SMS yang masuk ke dalam inbox pengguna.
- b. Penelitian dapat dilanjutkan dengan implementasi ke dalam salah satu perangkat telepon genggam yang berbasis android, iOS maupun yang lain.

DAFTAR PUSTAKA

- [1] ITU, "Mobile-cellular subscriptions 2000-2013", [online], <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>, diakses 20 Pebruari 2014.
- [2] http://en.wikipedia.org/wiki/Short_Message_Service, Short message service, akses 2 Pebruari 2015
- [3] http://en.wikipedia.org/wiki/Mobile_phone_spam, Mobile phone spam, akses 2 Pebruari 2015.
- [4] Tan, P. et al. (2006). Introduction to Data Mining. Boston : Pearson Education
- [5] Prasetyo Anugroho, Idris Winarno, Nur Rosyid. (2010). Klasifikasi Spam dengan Metode Naive Bayes Classifier menggunakan Java Programming. PENS Surabaya
- [6] Dr. Ghayda A. Al-Talib1, Hind S. Hassan, A . (2013). Study on Analysis of SMS Classification Using TF-IDF Weighting
- [7] Amir Hamzah. (2012). Klasifikasi Teks Dengan Naive Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita dan Abstrak Akademis. SNAST Periode III
- [8] Selvia Lorena Br Ginting, Reggy Pasya Trinanda. (2013). Penggunaan Metode Naive Bayes Classifier Pada Aplikasi Perpustakaan. Unikom Bandung