

Klasifikasi SMS Spam Menggunakan Algoritma Support Vector Machine (SVM)

Green Arther Sandag^{*1)}, Raymond J. Sambur²⁾, Jebriella Bororing³⁾

^{1,2,3} Fakultas Ilmu Komputer, Universitas Klabat
Jl. Arnold Mononutu, Airmadidi, Minahasa Utara
E-mail*: greensandag@unklab.ac.id

Abstrak

Informasi merupakan kumpulan data yang memiliki beragam konten yang dapat bersifat positif dan negatif. Konten-konten yang tersebar dalam informasi tersebut dapat dengan mudah kita peroleh lewat berbagai macam jenis media contohnya SMS. Banyak informasi yang dikirim lewat SMS merupakan spam dimana konten yang dikirim berisi informasi yang tidak relevan dengan pengguna. Dengan fakta tersebut, perlu dibuat suatu system dalam mengklasifikasikan Spam dengan menggunakan text mining menggunakan metode Support Vector Machine (SVM), yang nantinya diharapkan dapat memilah komentar spam dalam SMS agar informasi yang diperoleh lebih akurat dan terpercaya. Data SMS yang didapat dari database Kaggle diolah terlebih dahulu dengan menggunakan teknik tokenizing, normalisasi kata, filtering, dan stemming. Selanjutnya peneliti menggunakan cross validation untuk menguji data training yang nantinya digunakan dalam proses klasifikasi. Algoritma SVM mampu mengklasifikasi spam dalam SMS dengan akurasi sebesar 96.72% dibanding naive bayes.

Kata kunci: Spam, SVM, Naive Bayes, tokenizing, filtering, stemming

1. Pendahuluan

Informasi merupakan data yang terdiri dari berbagai konten sehingga konten yang terdapat dalam informasi dapat memberikan arti bagi pengguna dan bisa saja berisi konten positif dan negatif. Untuk memperoleh sebuah informasi, manusia dapat berinteraksi dengan cara berkomunikasi satu dengan yang lain lewat berbagai media [1]. Teknologi yang semakin canggih memungkinkan manusia untuk berkomunikasi dengan cepat, salah satunya Short Message Service (SMS) dan media sosial. Media sosial adalah aplikasi berbasis internet yang memungkinkan para pengguna untuk membuat, berbagi, atau bertukar informasi, ide, dan gambar/video dalam suatu jaringan virtual [2]. Berbagai media sosial digunakan oleh pengguna untuk saling bertukar informasi seperti Facebook, Twitter, Path dan

Instagram, yang dapat kita lihat dalam bentuk foto, tautan, hingga komentar.

Spam merupakan informasi yang dikirimkan ke beberapa pengguna dimana informasi yang dikirim tidak berhubungan dengan penerima informasi tersebut [3]. Tujuan pembuatan spam sendiri adalah promosi dan iklan yang dapat berupa link, berita, email, tulisan dan bahkan komentar [3]. Media social adalah aplikasi internet yang memungkinkan penggunaannya untuk membuat, berbagi, atau bertukar informasi, ide, dan gambar/video dalam suatu jaringan virtual [4]. Spam pada media sosial merupakan komentar yang berisi konten yang dikirim oleh spammer menggunakan fitur comment, dimana konten yang dikirimkan berisi informasi yang tidak relevan dengan pengguna yang menerima komentar tersebut [5]. Munculnya spam membuat pengguna Instagram merasa terganggu, khususnya para artis yang akunnnya sering dijadikan sasaran oleh para spammer karena memiliki jumlah follower yang banyak. Bahkan beberapa pengguna pernah menjadi korban penipuan yang mengakibatkan terjadinya kerugian [6].

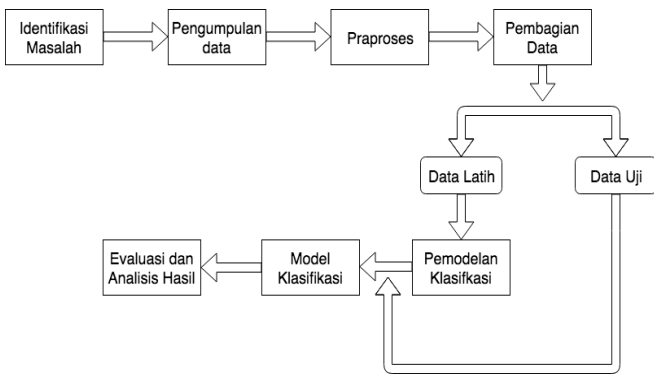
Dari permasalahan berkembangnya kemunculan spam, dibuatlah berbagai penelitian yang berhubungan dengan pengklasifikasian teks untuk mendeteksi apakah suatu data merupakan spam atau bukan spam. Salah satu penelitian yang dibuat yang berhubungan dengan pengklasifikasian spam adalah penelitian dari Yuan Lukito dkk [7] dengan objek penelitian yaitu Instagram, metode yang digunakan dalam penelitian tersebut hanya satu yaitu naive bayes dengan jumlah data yang dikumpulkan berjumlah 17,007 dari 10 akun aktor/artis Indonesia yang memiliki jumlah follower terbanyak. Hasil yang didapat dari penelitian tersebut adalah Accuracy 77.25%, Classification Error 22.75%, Recall 57.21%, Specificity 97.29%, Precision 95.48% dan F-Measure 71.5%. Penelitian yang sejenis dalam mengklasifikasi email spam dengan algoritma SVM digabungkan dengan teknik RBF yang dilakukan Ulama[8] mendapatkan hasil akurasi sebesar 96.6%

dengan error rate 3.4% dibanding dengan KNN sebesar 92.28%.

SVM merupakan metode klasifikasi yang dikenal sangat efisien dalam klasifikasi maupun regresi. Pengklasifikasian metode ini menggunakan teknik hyperplane terbaik yang berguna untuk memisahkan dua buah class pada input space[9-11]. Diharapkan dengan menggunakan metode text mining dalam mengklasifikasi spam, dapat menambah pengetahuan dan pengalaman dalam menganalisa informasi yang penting dan tidak penting [12].

Penelitian ini bertujuan untuk menerapkan algoritma text mining yaitu SVM dalam pendeteksian spam SMS.

2. Pembahasan



Gambar 1. Desain penelitian

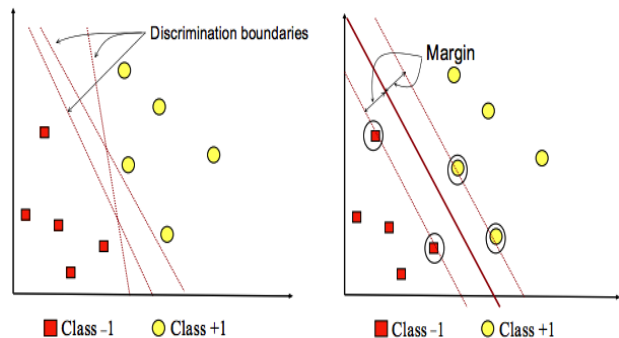
Gambar 1 merupakan tahapan – tahapan yang ada dalam penelitian ini. Tahapan ini meliputi [11]:

1. Identifikasi Masalah
Tahap ini merupakan tahap pertama dalam penelitian ini, dimana peneliti mengidentifikasi masalah – masalah yang muncul dan juga menentukan tujuan dan manfaat dari penelitian yang diteliti.
2. Pengumpulan Data
Pengujian terhadap implementasi metode SVM akan dilakukan terhadap seluruh data uji yang telah diambil dari data SMS spam berbahasa Inggris yang diunduh dari KAGGLE. Data SMS spam yang diunduh tersimpan dalam format csv yang berisikan dua kelas yaitu spam dan ham.
3. Praproses
Data yang telah diberi label kemudian diproses untuk menjadi suatu dataset yang siap digunakan dalam penelitian ini. Data ini diproses menggunakan tools RapidMiner dengan cara:
 - Tokenizing, proses mengubah semua karakter huruf menjadi huruf kecil sehingga proses selanjutnya tidak perlu lagi memperhatikan huruf besar dan huruf kecil. Proses ini juga menghapus karakter – karakter, simbol, URL,

angka yang tidak digunakan dalam pembentukan dataset.

- Normalisasi Kata, proses mengubah kata yang tidak baku menjadi kata baku.
 - Filtering, proses menghapus stopword yang ada di komentar Instagram.
 - Stemming, mengubah kata kedalam bentuk dasar
4. Pembagian Data
Dataset yang telah ada kemudian dibagi menjadi dua data yaitu data training dan data testing dengan rasio 80:20.
 5. Pembuatan Model Klasifikasi
Setelah tahap pembagian data selesai maka data training akan diuji menggunakan teknik cross-validation dengan 10 fold cross-validation. Data ini kemudian digunakan sebagai data eksperimen sebelum diuji menggunakan data testing. Data ini akan diuji menggunakan algoritma SVM. Hasil dari data training akan menjadi acuan bagi peneliti untuk menguji data testing.
 6. Evaluasi dan Analisis
Pengujian dilakukan terhadap data testing menggunakan algoritma SVM. Hasil pengujian diukur menggunakan tiga parameter performance measure (accuracy, precision, recall) yang diukur dengan menggunakan confusion matrix. Hasil dari pengujian ini kemudian dievaluasi dan analisis untuk menjadi output dalam penelitian ini.

Support vector machine (SVM) merupakan metode pembelajaran machine yang menerapkan prinsip structural risk minimization (SRM) untuk menentukan hyperplane terbaik seperti pada Gambar 2 [9].



Gambar 2. Menentukan hyperplae terbaik dari dua class

Gambar 2 memperlihatkan beberapa pattern dari dua class: +1 dan -1. Class -1 digambarkan dengan kotak berwarna merah dan class +1 digambarkan dengan lingkaran berwarna kuning. Teknik SVM berusaha untuk mencari hyperplane yang memisahkan dua class tersebut. Margin merupakan jarak garis dengan pattern terdekat, pattern tersebut adalah support vector, garis merah tersebut adalah hyperplane terbaik yang memisahkan dua buah class. Persamaan untuk mencari hyperplane memenuhi persamaan berikut:

$$f(x) = \sum_{i=1}^{SV} \alpha_i \cdot K(SV_i, x) + b \dots\dots\dots 1$$

Dengan keterangan:

α_i = vektor koefisien untuk *lagrange multiplier*

$K(SV_i, x)$ = fungsi kernel yang digunakan

b = error atau bias

Performance Measure digunakan untuk mengevaluasi akurasi sistem dalam mengklasifikasikan spam pada dataset. Berikut ini merupakan perhitungan performance measure berdasarkan Tabel 1 Confusion Matrix dibawah ini.

Tabel 1. Confusion Matrix

Actual/Prediction	Negatif	Positif
Negatif	True Negatif (TN)	False Negatif (FN)
Positif	False Positif (FP)	True Positif (TP)

Di mana:

True negative = jumlah data negatif yang benar dikategorikan sebagai class negatif

False negatif = jumlah data negatif yang dikategorikan sebagai class positif

False positif = jumlah data positif yang dikategorikan sebagai class negatif

True positif = jumlah data positif yang benar dikategorikan sebagai class positif

Pengukuran hasil dari model yang dibuat diukur menggunakan 3 parameter performance evaluation, yang terdiri dari [13];

- a. Accuracy, persentase dari jumlah dataset yang benar berdasarkan metode yang digunakan.

$$Accuracy = (TN + TP) / (TN + FP + FN + TP) \dots\dots\dots 2$$

- b. Recall, tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

$$Recall = TP / (FN + TP) \dots\dots\dots 3$$

- c. Precision, persentase ketepatan informasi yang diberikan oleh komputer yang sesuai dengan permintaan pengguna.

$$Precision = TP / (FP + TP) \dots\dots\dots 3$$

Cross Validation adalah suatu metode statistik dari mengevaluasi dan membandingkan algoritma pembelajaran dengan cara membagi data menjadi dua segmen: satu digunakan sebagai data pembelajaran suatu model atau data training dan satu untuk validasi model atau disebut data testing.

Proses k-fold cross-validation adalah sebagai berikut; pertama data dibagi sebanyak k yang sudah ditentukan, selanjutnya k diiterasi dan setiap k berbeda digunakan sebagai validasi k yang lain dan hal ini diteruskan sampai k yang terakhir. Didalam data mining dan machine

learning 10-fold cross-validation adalah yang paling umum digunakan [14].

Data yang digunakan dalam penelitian ini merupakan kumpulan data yang dikhususkan untuk penelitian SMS spam. Jumlah data yang dikumpulkan berjumlah 5,574 SMS berbahasa Inggris dan juga telah diberi label *spam* dan *ham* berdasarkan konten dari data tersebut. Tabel 2 merupakan contoh data yang dikumpulkan dari KAGGLE.

Tabel 2. Contoh Data dari KAGGLE

NO	V1	V2
1	Ham	Go untul jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...
2	Ham	Ok lar...joking wif u oni...
3	Spam	Free entry in 2 a wkly comp to win FA Cup finat tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate) T&C's apply 08452810075over18's
4	Ham	U dun say so early hor... U c already then say...
5	Ham	Nah i don't think he goes to usf, he lives around here though
...
5565	Ham	Pity, * was in mood for that. So...any other suggestions?
5566	Ham	Rofl. Its true to its name

Proses pada tahap ini adalah proses pembentukan *dataset* berdasarkan data SMS yang telah dikumpulkan. Proses ini dilakukan dalam empat tahap, *Tokenizing*, *Normalisasi Kata*, *Filtering*, *Stemming* [7].

- a. *Tokenizing*, proses ini mencakup penghapusan kata yang termasuk angka, tanda baca, URL dan mengubah semua huruf menjadi huruf kecil.

Tabel 3. Praproses Tokenizing

Sebelum	Sesudah
Free entry in 2 a wkly comp to win FA Cup finat tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate) T&C's apply	Free entry in a wkly comp to win fa cup finat tkts st may text fa to to receive entry question std txt rate

08452810075over18's	tcs apply overs
---------------------	-----------------

- b. Normalisasi Kata, proses ini merupakan proses perbaikan kata yang dilakukan dengan mengubah kata tidak baku menjadi kata baku.

Tabel 4. Praproses Normalisasi Kata

Sebelum	Sesudah
free entry in a wkly comp to win fa cup finat tkts st may text fa to to receive entry question std txt rate tcs apply overs	free entry in a weekly competition to win fa cup final tickets st may text fa to to receive entry question std text rate tcs apply overs

- c. Filtering, menghapus kata yang masuk dalam daftar stopwords.

Tabel 5. Praproses Filtering

Sebelum	Sesudah
free entry in a weekly competition to win fa cup final tickets st may text fa to to receive entry question std text rate tcs apply overs	free entry weekly competition win fa cup final tickets st may text fa receive entry question std text rate tcs apply overs

- d. Stemming, mengubah kata ke dalam bentuk dasar.

Tabel 6. Praproses Stemming

Sebelum	Sesudah
free entry weekly competition win fa cup final tickets st may text fa receive entry question std text rate tcs apply overs	free entry week compete win fa cup final tickets st may text fa receive entry question std text rate tcs apply overs

Data yang digunakan dalam pengujian *cross-validation* adalah sebanyak 80% dari total dataset. Jumlah total data keseluruhan adalah 5,574, sehingga jumlah data yang digunakan dalam pengujian *cross-validation* adalah sebanyak 4,391 data dengan pembagian yang dilakukan secara acak. Selanjutnya merupakan proses penerapan algoritma SVM untuk menjadi hasil akhir dalam pengujian ini. Tabel 7 menunjukkan hasil pengujian dalam bentuk tabel menunjukkan perbandingan dengan menggunakan algoritma lain. SVM memberikan akurasi terbaik dibanding dengan naive bayes sebesar 95.95%, precision sebesar 100%, dan recall sebesar 70.24%.

Tabel 7. Performance Measure Cross-Validation

Hasil Klasifikasi	Naive Bayes	SVM
True Positive (TP)	526	420
True Negative (TN)	3135	3793
False Positive (FP)	72	0
False Negative (FN)	658	178
Accuracy	83.37%	95.95%
Precision	44.54%	100.00%
Recall	87.96%	70.24%

Hasil Pengujian Independen adalah hasil pengujian yang dilakukan dengan menggunakan data uji yang telah ditentukan berdasarkan pembagian dataset yaitu, 20% dari total data. Jumlah data keseluruhan adalah 5,574, sehingga jumlah data yang digunakan dalam pengujian independen adalah sebanyak 1,097 data dengan pembagian yang dilakukan secara acak. Data kemudian diterapkan algoritma SVM untuk menjadi hasil akhir dalam penelitian ini. Tabel 8 menunjukkan hasil klasifikasi algoritma berdasarkan performance measure yang dijadikan sebagai parameter pengujian menunjukkan hasil perbandingan dengan menggunakan algoritma yang lain. Algoritma SVM memberikan hasil terbaik dibanding naive bayes dengan akurasi sebesar 96.72%.

Tabel 8. Performance Measure Independent

Hasil Klasifikasi	Naive Bayes	SVM
True Positive (TP)	136	114
True Negative (TN)	761	947
False Positive (FP)	13	1
False Negative (FN)	187	35
Accuracy	81.77%	96.72%
Precision	42.11%	99.13%
Recall	91.28%	76.51%

3. Kesimpulan

Penelitian ini berhasil mengimplementasikan algoritma SVM untuk mendeteksi spam pada SMS berbahasa Inggris dengan tingkat akurasi diatas 90%, dan hasil pada presisi dapat dikatakan bagus karena memiliki nilai 99.13% untuk independent dataset. Begitu juga dengan teknik cross validation, SVM memberikan hasil akurasi yang terbaik dengan nilai 95.95% dan precision sebesar 100%. Perbandingan yang dilakukan juga membuktikan

bahwa algoritma SVM lebih baik daripada Naive Bayes untuk mendeteksi SMS *spam*. Saran – saran yang masih perlu dikerjakan dalam penelitian selanjutnya adalah:

1. Melakukan perbandingan dengan menggunakan metode/algoritma yang lain
2. Menggunakan metode pembersihan *dataset* yang lebih baik.

Dapat dikembangkan lebih lanjut menjadi software/plugin yang dapat mendeteksi SMS *spam* secara otomatis

Daftar Pustaka

- [1]. M. B. Romney & P. J. Steinbart, *Accounting Information Systems*, 13th penyunt., Pearson, 2015.
- [2]. J. M. Rewah, “Pengaruh Promosi Berbasis Teknologi Informasi Terhadap Pengusaha Muda Universitas Klabat,” *COGITO*, vol. 4, no. 1, pp. 24-36, 2018.
- [3]. M. Webster, “Definition of SPAM,” 2016. [Online]. Available: <https://www.merriam-webster.com/dictionary/spam..> [Diakses 04 October 2017].
- [4]. H. Marti, “What Is Text Mining?,” SIMS, UC Berkeley, 2003.
- [5]. M. McCord & M. Chuah, “Spam Detection on Twitter Using Traditional Classifier,” *CSE*, 2011.
- [6]. S. Geerthik, “Survey on Internet Spam: Classification and Analysis,” *International Journal Computer Technology & Application*, vol. IV, no. 3, pp. 384-391, 2013.
- [7]. L. Yuan & C. A. Rahmat, “Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes,” *Ultimatics UMN*, vol. IX, no. 1, p. 58, 2017.
- [8]. S. N. D. Pratiwi & B. S. S. Ulama, “Klasifikasi Email Spam dengan Menggunakan Support Vector Machine dan k-Nearest Neighbor,” *Journal Sains dan Seni ITS*, Vol 5. No. 2. 2016
- [9]. K. Tsuda, “Overview of Support Vector Machine”, *Journal of IEICE*, Vol.83, No.6, 2000, pp.460-466
- [10]. D. E. Sondakh, “COMPARATIVE STUDY OF CLASSIFICATION,” *COGITO*, vol. I, no. 1, pp. 13-22, 2015.
- [11]. J. Han & M. Kamber, *Data Mining: Concepts and Techniques*, 2nd penyunt., San Fransisco, United States of America: Morgan Kaufmann Publishers, 2006.
- [12]. Z. Khan & U. Qamar, “Text Mining Approach to Detect Spam in Emails,” dalam *The International Conference on Innovations in Intelligent Systems and Computing Technologies*, Philippines, 2016.
- [13]. Z. Khan & U. Qamar, “Text Mining Approach to Detect Spam in Emails,” dalam *The International Conference on Innovations in Intelligent Systems and Computing Technologies*, Manila, 2016.
- [14]. R. Rafaelzadeh, L. Tang & H. Liu, “Cross-Validation,” dalam *Encyclopedia of Database System*, New York, Springer, 2009, pp. 532-538.