**Infotek : Jurnal Informatika dan Teknologi**
Vol. 4 No. 2, Juli 2021
Hal. 249-258

**e-ISSN 2614-8773**

**DOI : 10.29408/jit.v4i2.3665** **Link : https://dx.doi.org/10.29408/jit.v4i2.3665**

# Komparasi Algoritma Naïve Bayes dan Support Vectors Machine pada Analisis Sentimen SMS HAM dan SPAM

**Lila Dini Utami[1], Lestari Yusuf[2], Dini Nurlaela[3]**
[1]Program Studi Informasi Akuntansi Kampus Kota Bogor, Universitas Bina Sarana Infromatika
[2]Program Studi Sistem Informasi, Universitas Nusa Mandiri
[3]Program Studi Sistem Informasi, Universitas Bina Sarana Informatika
lila.ldu@bsi.ac.id

## Abstrak

SMS merupakan bentuk komunikasi berupa SMS yang dikirimkan menggunakan handphone antar nomor yang di tuju. SMS saat ini sudah jarang digunakan karena fungsinya banyak berubah digantikan oleh aplikasi chat. Tetapi ditur SMS tidak dihilangkan karena satu dah lain hal, SMS resmi dari berbagai aplikasi untuk melakukan verifikasi ataupun info-info resmi lainnya masih menggunakan SMS sebagai tanda nomer telepon yang digunakan itu ada. Tetapi sejak 2011 banyak sekali penyelahgunaan fungsi tersebut sehingga disinyalir banyak penipuan yang menggunakan SMS sebagai alat mempengaruhi korban. Kategori penyalahgunaan sms ini masuk kepada SMS spam. Maka dari itu SMS perlu diklasifikasikan agar pengguna dapat mengetahui SMS tersebut termasuk kedalam kategori Spam atau ham (kebalikan dari spam). Dengan menggunakan 400 dataset yang diambil dari UCI repository yang dibagi kedalam dua class yaitu spam dan ham kami membandingkan dua metode klasifikasi yaitu Naive Bayes dan Support vector Machine agar dapat mendapatkan filtering sms dengan benar. Dan setelah dilakukan perhitungan didapatkan accurasy yang akurat pada naive Bayes Yaitu sebesar 90.00% sedangkan Support Vector Machine 81.00%..

**Kata Kunci :** Klasifikasi, Naive bayes, SMS, Support Vector Machine

## Abstract

*SMS is a form of communication in the form of messages sent using mobile phones between the designated numbers. SMS is now rarely used because many of the features that have changed are used by chat applications. However, the SMS feature was not removed for one thing, official messages from various applications for leveraging or other official information still use SMS as a sign that the phone number used is there. However, since 2011 there have been so many misuses of this function, so it is suspected that many frauds use SMS as a tool to influence victims. This sms category goes to SMS spam. Therefore, SMS needs to be classified so that users can find out that the SMS is included in the category of Spam or ham (the opposite of spam). Using 400 datasets taken from the UCI repository which is divided into two classes, namely spam and ham, we compare two classification methods, namely Naive Bayes and Support vector Machine in order to get SMS filtering correctly. And after the calculations are done, the accuracy is obtained in Naive Bayes, which is 90.00% Support Vector Machine 81.00%.*

***Keyword**: Classification, Naive bayes, SMS, Support Vector Machine*

## 1. Pendahuluan

Sebelum banyak sekali aplikasi chat, bentuk komunikasi pada handphone sebelumnya menggunakan *Short Message Servise*(SMS). Sebuah komunikasi menggunakan bentuk alphanumeric yang digunakan sebagai pengirim

Infotek : Jurnal Informatika dan Teknologi
Vol. 4 No. 2, Juli 2021
Hal. 249-258

e-ISSN 2614-8773

DOI : 10.29408/jit.v4i2.3665 Link : https://dx.doi.org/10.29408/jit.v4i2.3665

SMS antar terminal tanpa menggunakan kabel [1]. Peran SMS saat ini tidak bisa diabaikan, dikarenakan banyak organisasi ataupun bagian yang menggunakan SMS sebagai sarana penting untuk menympaikan SMS resmi sebuah perusahaan atau bahkah penggunaan untuk pemberitahuan lanjutan sebuah apliaksi pendeteksi gerak [2]. Begitu juga dalam dunia hukum SMS digunakan sebagai bukti tindak pidana [3].

Tetapi SMS pun menjadi suatu sarana yang rentan akan penyalahgunaan [4]. Dilangsir dalam[5] bahwa jumlah trafik didunia, SMS lajunya meningkat sampai 6.9 Triliun semenjak 2010 dan menyebabkan spammers untuk mengirimkan SMS spam (yang tidak diminta). Spam diartikan sebagai pemanfaatan SMS dengan tujuan penyalahgunaan oleh orang yang hendak melakukan kejahatan berkedok penawaran produk maupun jasa [6]. Bahkan pada taun 2011 di Indonesia tercatat terjadi kenaikan tindak kejahatan sebanyak 60 kasus yang mengakitbatkan penyitaan barang bukti handphone yang didalamnya terdapat SMS spam sebagai barang bukti [7].

Maka sebagai jalan keluar proses pengklasifikasian SMS spam yang dijadikan sebagai proses utama secara signifikan dapat membantu pengguna untuk membedakan SMS antara ham dan spam[8].

Beberapa peneliti menyarakan agar dilakukannya filtering menggunakan dua atau lebih teknik yang berbeda agar meningkatkan akurasi filtering SMS spam[9]. Karena pada penelitian sebelumnya beberapa peneliti pun mengatasi masalah ini dengan berbagai cara seperti penelitian untuk pengklasifikasian SMS spam [10] juga dibuatnya berbagai macam algoritma untuk klasifikasi dan pengelompokkan SMS spam menggunakan perbandingan Rapidminer dan Weka [8].

Dengan menggunakan data public yang diambil dari UCI repository mengenai SMS spam penelitian ini akan melakukan komparasi metode Naive Bayes dan Support Vector Mechine. Dengan melakukan perbandingan tersebut akan diketahui bahwa SVM unggul akan tingkat accuracy sebesar 90,50% dibandingkan dengan naive bayes sebesar 59,98% [11]. Perbandingan ini pun pernah dibandingkan dalam menganalisa sentimen perusahaan listrik negara cabang ambon dengan akurasi untuk naive bayes sebesar 62,4% dan svm sebesar 76,42%[12].

Maka pada penelitian ini akan dibuat komparasi pengklasisfikasian data SMS Spam antara metode Support Vector Mechine dan Naive Bayes yang ditujukan agar menghasilkan accuracy algorithma terbaik untuk pengklasifikasian SMS Spam.

Infotek : Jurnal Informatika dan Teknologi
Vol. 4 No. 2, Juli 2021
Hal. 249-258

e-ISSN 2614-8773

DOI : 10.29408/jit.v4i2.3665 Link : https://dx.doi.org/10.29408/jit.v4i2.3665

## 2. Tinjauan Pustaka

### 2.1 Penelitian Terkait

Pada penelitian sebelumnya pengklasifikasian SMS spam di hitung menggunakan metode Support vector mechine untuk melihat tingkat akurasi algoritmanya dan mendapatkan ilai sebesar 98,9% [9]. Dan juga [13] menyelesaikan permasalahan spam yang terlallu mengganggu ini menggunakan teknik filtering deep learningn yang menghasilkan accurasi 99.44%. dan dengan masalah yang sama peneliti sebelumnya menggunakan bi-level untuk mengidentifikasi sms spam dan menghasilkan metode ini dapat digunakan untuk dibuat keberbagai aplikasi [5]. Pembuatan apliaksi android untuk identifikasi sms spampun pernah dibuat oleh [7] dan penyaringan sms spam pun pernah di analisa dan dilakukan klasifikasi menggunakan naive bayes yang menghasilkan metode naive bayes dapat digunakan sebagai filtrasi sms spam yang cukup akurat [6]

### 2.2 Landasan Teori

1. *Text Mining* atau *Natural Language Processing (NLP)*

*Natural Language Processing* (NLP) dan *Text Mining* (TM) merupakan bentuk data mining yang mengacu pada algoritma oromatis otomatis yang digerakkan mesin untuk pemetaan semantik, penggalian informasi, dan information pemahaman tentang bahasa manusia (alami). Ini melibatkan ekstraksi informasi penting dari sejumlah besar teks tidak terstruktur. Untuk melakukannya, diperlu membangun algoritma pemetaan semantik dan sintaksis untuk pemrosesan berat yang efektif dri sebuah teks. Tekait proses ini pengklasifikasian text yang kuat ditunjukkan saat menggunakan algoritma naive Bayes [14]

2. *Naive Bayes*

Dikarenakan dalam text mining didalamnya memiliki hal yang tidak mudah seperti mengekstraksi bahasa manusia menjadi sebuah informasi penting dengan menggunakan metode naive bayes ini menjadikan yang mendasari secara inheren dalam algoritma pembelajaran adalah bahwa atribut item data independen satu sama lain. Karena hubungan semantik yang kuat antara kata-kata yang dipilih sebagai fitur yang ada, asumsi justru bertentangan dengan kenyataan, terutama dalam kategorisasi text [15]

$$P(a_i = x_i|C_k) = P(x_i|C_k) = \frac{N_{ki}}{N_k}$$

Gambar 1. Contoh perhitungan kemungkinan nilai individu 2 atribute

3. *Support Vector Machine*

Tidak seperti pendekatan pembelajaran mesin generatif, yang memerlukan perhitungan distribusi probabilitas bersyarat, fungsi klasifikasi diskriminan mengambil titik data x dan menetapkannya ke salah satu kelas berbeda

Infotek : Jurnal Informatika dan Teknologi
Vol. 4 No. 2, Juli 2021
Hal. 249-258

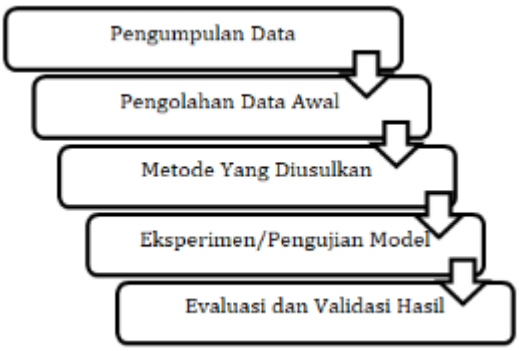
e-ISSN 2614-8773

DOI : 10.29408/jit.v4i2.3665 Link : https://dx.doi.org/10.29408/jit.v4i2.3665

yang merupakan bagian dari tugas klasifikasi. Kurang kuat daripada pendekatan generatif, yang sebagian besar digunakan ketika prediksi melibatkan deteksi outlier, pendekatan diskriminan membutuhkan lebih sedikit sumber daya komputasi dan lebih sedikit data pelatihan, terutama untuk ruang fitur multidimensi dan ketika hanya probabilitas posterior yang diperlukan [16].

## 3. Metode Penelitian

Proses penelitian yang terdiri dari langkah-langkah penelitian berupa pengolahan data berpedoman pada desin penelitian. Berikut langkah-langkah pengolahan data di jelaskan pada gambar 2.




Gambar 2. Tahapan Penelitian

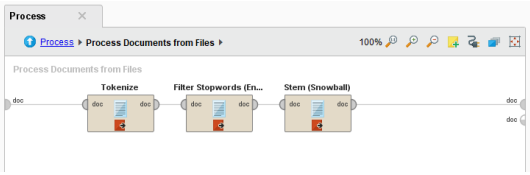
Sumber: Septiani dalam [17]

1. Pengumpulan Data

Dalam penelitian ini, dilakukan tahap awal penelitian, yakni pengumpulan data. Adapun cara pengumpulan data tersebut adalah mengambil contoh isi sebuah SMS HAM dan SPAM dari UCI Repository (https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection) sebanyak 400 SMS yang terdiri dari 200 SMS HAM dan 200 SMS SPAM.

2. Pengolahan Data Awal

Dalam tahap pengolahan data, terdapat 3 tahapan *pre processing*, yakni:




Gambar 3. Tahapan *Pre-Processing* Pada Rapid Miner 7.2

a. *Tokenization*

Merupakan sebuah langkah dimana kata-kata dikumpulkan lalu tanda baca maupun simbol dihilangkan [18]

Tabel 1. Hasil *Review Tokenization*

| *Text* | *Tokenization* |
|---|---|
| No. I meant the calculation is the same. That <#> units at <#> . This school is really expensive. Have you started practicing your accent. Because its important. And have you decided if you are doing 4years of dental school or if you'll just do the nmde exam. | No I meant the calculation is the same That lt gt units at lt gt This school is really expensive Have you started practicing your accent Because its important And have you decided if you are doing years of dental school or if you ll just do the nmde exam |

Infotek : Jurnal Informatika dan Teknologi
Vol. 4 No. 2, Juli 2021
Hal. 249-258

e-ISSN 2614-8773

DOI : 10.29408/jit.v4i2.3665 Link : https://dx.doi.org/10.29408/jit.v4i2.3665

b. *Stopword*

Langkah berikutnya adalah stopword, stopword merupakan langkah pembuangan imbuhan-imbuhan yang ada didalam kata-kata atau biasa disebut stoplist, karena kata-kata imbuhan tersebut tidak bisa diklasifikasikan kedalam sebuah class [19]. Langkah ini dijelaskan pada table 2

Tabel 2. Hasil *Review Stopwords*

| *Text* | *Stopwords* |
|---|---|
| No. I meant the calculation is the same. That <#> units at <#> . This school is really expensive. Have you started practicing your accent. Because its important. And have you decided if you are doing 4years of dental school or if you'll just do the nmde exam. | I meant calculation lt gt units lt gt school expensive started practicing accent decided years dental school ll nmde exam |

c. *Stemming*

Dalam langkah stemming kata-kataa mulai di kembalikan ke kata-kata awal menjadi bentuk asal yang di sesuaikan kepada aturan suatu bahasa [20]

Tabel 3. Hasil *Review Stemming*

| *Text* | *Stemming* |
|---|---|
| No. I meant the calculation is the same. That <#> units at <#> . This school is really expensive. Have you started practicing your accent. Because its important. And have you decided if you are doing 4years of dental school or if you'll just do the nmde exam. | i meant calcul lt gt unit lt gt school expens start practic accent decid year dental school ll nmde exam |

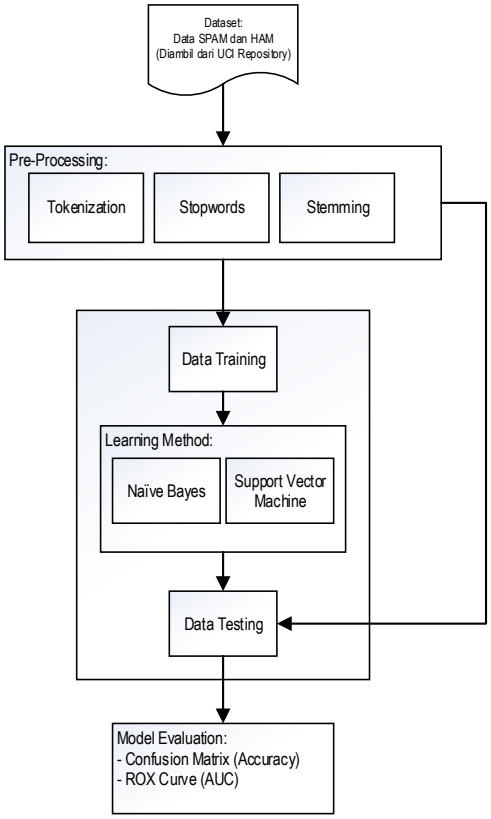
3.1 Model Penelitian

Dalam penelitian ini menggunakan 200 data SMS SPAM dan 200 data SMS HAM. Data tersebut dibuat secara terpisah dalam masing-masing file berbentuk .txt. Data SMS SPAM dijadikan satu dalam folder yang diberi nama SPAM, sementara data SMS HAM dijadikan satu dalam folder yang diberi nama HAM. Penelitian ini juga menggunakan tidak Pre-Processing, yakni: Tokenization, Stopwords, dan Stemming. Kemudian diolah menggunakan Rapid Miner Versi 7.2 dengan menggunakan perbandingan algoritma Naïve Bayes dengan Support Vector Machine agar dapat diketahuin akurasi dan AUC (Lihat Gambar 2).

Infotek : Jurnal Informatika dan Teknologi
Vol. 4 No. 2, Juli 2021
Hal. 249-258

e-ISSN 2614-8773

DOI : 10.29408/jit.v4i2.3665 Link : https://dx.doi.org/10.29408/jit.v4i2.3665



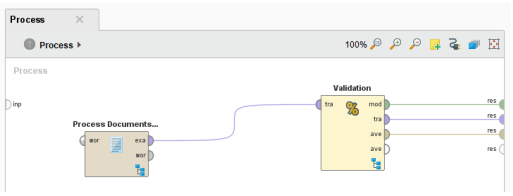

Gambar 4. Model Penelitian Yang Diusulkan

## 4. Hasil dan Pembahasan

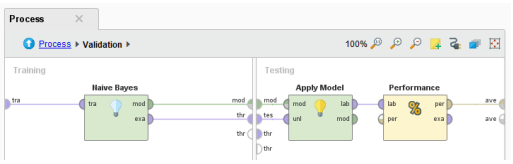
Pengujian dilakukan menggunakan Rapid Miner 7.2 dan dataset berupa 200 data SMS HAM dan 200 data SMS SPAM, yang dimana data tersebut dipisahkan dalam sebuah folder dan masing-masing diberi nama SPAM dan HAM.

1. Naïve Bayes

   Menurut Alfa Saleh dalam [21], Naive bayes merupakan probabilistik klasifikasi sederhana yang perhitungan probabilitasnya merupakan penjumlahan frekuensi dan kombinasi dari nilai sebuah dataset. Proses pengolahan data awal di jelaskan pada gambar 5.




Gambar 5. Proses Pengolahan Awal Pada Rapid Miner 7.2




Gambar 6. Proses Validasi Algoritma *Naïve Bayes* Pada Rapid Miner 7.2

Dari sebanyak 200 data SMS HAM dan 200 data SMS SPAM, sebanyak 170 data SMS yang diprediksi sesuai yakni data SMS HAM, dan sebanyak 30 data SMS yang diprediksi sebagai SMS HAM, tetapi ternyata adalah SMS SPAM. Sebaliknya, sebanyak 190 data SMS SPAM yang diprediksi sesuai yakni data SMS SPAM, dan sebanyak 10 data SMS yang diprediksi sebagai SMS SPAM, tetapi ternyata adalah SMS HAM. Hasil akurasi yang diperoleh menggunakan algoritma Naïve Bayes adalah 90.00% (Lihat Tabel. 4) dan AUC: 0.833 (Lihat Gambar. 6)

Tabel 4. *Confusion Matrix* Algoritma Naïve Bayes

| ***Accuracy:* 90.00% +/- 5.36% (mikro: 90.00%)** | | | |
|---|---|---|---|
| | ***True HAM*** | ***True SPAM*** | ***Class Precision*** |
| **Pred. HAM** | 170 | 10 | 94.44% |

Infotek : Jurnal Informatika dan Teknologi
Vol. 4 No. 2, Juli 2021
Hal. 249-258

e-ISSN 2614-8773

DOI : 10.29408/jit.v4i2.3665 Link : https://dx.doi.org/10.29408/jit.v4i2.3665

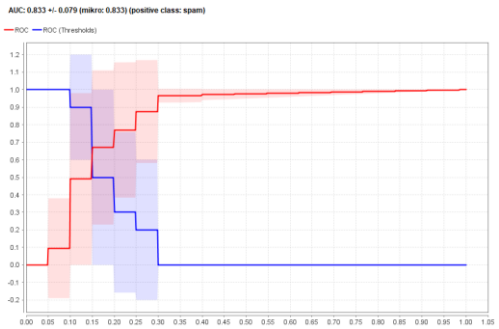
| | | | |
|---|---|---|---|
| **Pred. SPAM** | 30 | 190 | 86.36% |
| **Class Recall** | 85.00% | 95.00% | |

Nilai akurasi dari *Confusion Matrix* diatas adalah sebagai berikut:

$$Accuracy = \frac{(TN + TP)}{(TN + FN + TP + FP)}$$

$$Accuracy = \frac{(170 + 190)}{(170 + 30 + 190 + 10)}$$

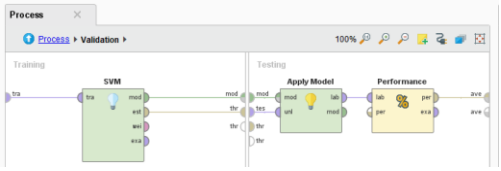
$$Accuracy = \frac{360}{400} = 0{,}9 = 90\%$$



Gambar 7. Grafik *Area Under Curve* (AUC) Algoritma Naïve Bayes

2. *Support Vector Machine*

Menurut Prabowo dalam [22], berbeda dengan *Naive Bayes, pengklasifikasian pada metode ini* merupakan jenis terpadu (*supervised*) meskipun begitu dalam penanganan model nonlinear termasuk kedalm metode yang sangat akurat. Validasi SVM pada rapid miner digambarkan pada gambar 8.



Gambar 8. Proses Validasi Algoritma *Support Vector Machine* Pada Rapid Miner 7.2

Dari sebanyak 200 data SMS HAM dan 200 data SMS SPAM, sebanyak 200 data SMS yang diprediksi sesuai yakni data SMS HAM, dan tidak ada data SMS yang diprediksi sebagai data SMS SPAM. Sebaliknya, sebanyak 124 data SMS SPAM yang diprediksi sesuai yakni data SMS SPAM, dan sebanyak 76 data SMS yang diprediksi sebagai SMS SPAM, tetapi ternyata adalah SMS HAM. Hasil akurasi yang diperoleh menggunakan algoritma *Support Vector Machine* adalah 81.00% (Lihat Tabel. 5) dan AUC: 0.9.87 (Lihat Gambar. 8)

Tabel 5. *Confusion Matrix* Algoritma *Support Vector Machine*

| ***Accuracy:* 81.00% +/- 3.91% (mikro: 81.00%)** | | | |
|---|---|---|---|
| | ***True HAM*** | ***True SPAM*** | ***Class Precision*** |
| **Pred. HAM** | 200 | 76 | 72.46% |
| **Pred. SPAM** | 0 | 124 | 100.00% |
| **Class Recall** | 100.00% | 62.00% | |

Nilai akurasi dari *Confusion Matrix* diatas adalah sebagai berikut:

$$Accuracy = \frac{(TN + TP)}{(TN + FN + TP + FP)}$$

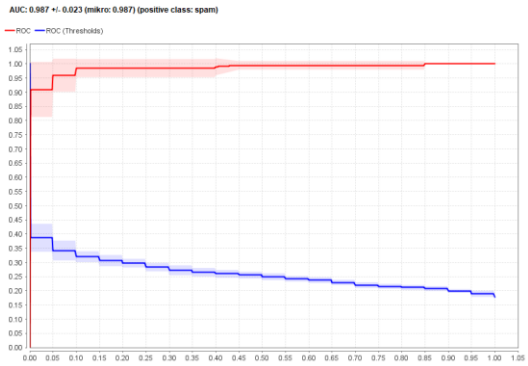
$$Accuracy = \frac{(200 + 124)}{(200 + 0 + 124 + 76)}$$

Infotek : Jurnal Informatika dan Teknologi
Vol. 4 No. 2, Juli 2021
Hal. 249-258

e-ISSN 2614-8773

DOI : 10.29408/jit.v4i2.3665 Link : https://dx.doi.org/10.29408/jit.v4i2.3665

$$Accuracy = \frac{324}{400} = 0{,}81 = 81\%$$



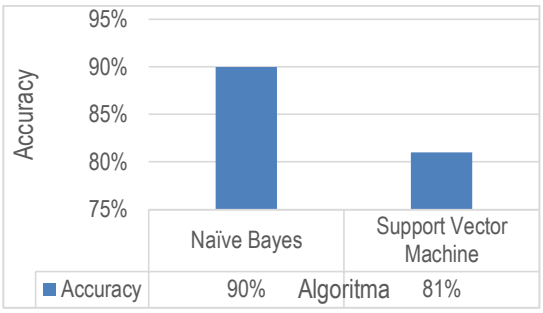

Gambar 9. Grafik *Area Under Curve* (AUC) Algoritma *Support Vector Machine*

## 5. Kesimpulan

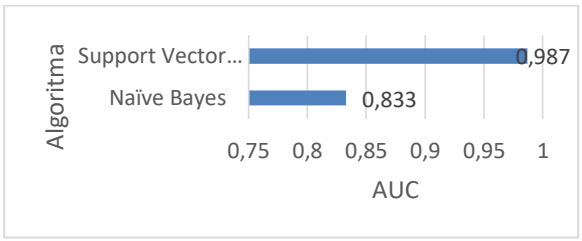
Komparasi antara algoritma *Naïve Bayes* dan *Support Vector Machine* dengan menggunakan Rapid Miner 7.2 dan dataset berupa data SMS HAM dan SPAM menghasilkan sebuah akurasi yang berbeda. Pengolahan menggunakan algoritma *Naive Bayes* menghasilkan akurasi sebesar 90% dan nilai AUC sebesar 0.833, sedangkan pengolahan menggunakan algoritma *Support Vector Machine* menghasilkan akurasi sebesar 81% dan nilai AUC sebesar 0.987 (Lihat Tabel 6)

Tabel 6. Perbandingan Hasil Pengolahan *Naïve Bayes* dan *Support Vector Machine*

| Algoritma | *Accuracy* | AUC |
|---|---|---|
| ***Naïve Bayes*** | 90.00% | 0.833 |
| ***Support Vector Machine*** | 81.00% | 0.987 |




Gambar 10. Grafik Perbandingan Akurasi Algoritma *Naïve Bayes* dan *Support Vector Machine*




Gambar 11. Grafik Perbandingan AUC Algoritma *Naïve Bayes* dan *Support Vector Machine*

## 6. Daftar Pustaka

[1] I. MUTIA, "Perancangan Sistem Informasi Akademik Dengan Teknologi Short Message Service (Sms) Pada Xyz," *Tek. Inform.*, vol. 7, no. 3, p. 13, 2014.

[2] R. Toyib, I. Bustami, D. Abdullah, and O. Onsardi, "Penggunaan Sensor Passive Infrared Receiver (PIR) Untuk Mendeteksi Gerak Berbasis Short Message Service Gateway," *Pseudocode*, vol. 6, no. 2, pp. 114–124, 2019, doi: 10.33369/pseudocode.6.2.114-124.

[3] R. R. dkk Lumbanraja, "PERTANGGUNGJAWABAN PIDANA PELAKU PENGHINAAN MELALUI LAYANAN SMS SINGKAT ATAU SMS

Infotek : Jurnal Informatika dan Teknologi
Vol. 4 No. 2, Juli 2021
Hal. 249-258

e-ISSN 2614-8773

DOI : 10.29408/jit.v4i2.3665 Link : https://dx.doi.org/10.29408/jit.v4i2.3665

(SHORT MESSAGE SERVICE)," *USU Law J.*, vol. 5, no. 1, pp. 107–118, 2017.

[4] J. Miranda, "SHORT MESSAGE SERVICE (SMS) DITINJAU MENURUT PASAL 184 KITAB UNDANG- UNDANG HUKUM PIDANA UU NO. 8 TAHUN 1981," *Lex Soc.*, vol. 5, no. 9, pp. 14–21, 2017.

[5] N. K. Nagwani and A. Sharaff, "SMS spam filtering and thread identification using bi-level text classification and clustering techniques," *J. Inf. Sci.*, vol. 43, no. 1, pp. 75–87, 2017, doi: 10.1177/0165551515616310.

[6] B. Indiarto, "Klasifikasi Sms Spam Dengan Metode Naive Bayes Classifier Untuk Menyaring SMS Melalui Selular," *J. Telemat. MKOM*, vol. 8, no. 2, pp. 167–172, 2016.

[7] A. Mair, Zaid Romegar; Ashari, "Aplikasi untuk Identifikasi Short Message Service (SMS) Spam Berbasis Android," *Bimipa*, vol. 24, no. 3, pp. 257–262, 2017.

[8] K. Zainal, N. F. Sulaiman, and M. Z. Jali, "An Analysis of Various Algorithms For Text Spam Classification and Clustering Using RapidMiner and Weka," *Int. J. Comput. Sci. Inf. Secur.*, vol. 13, no. 3, pp. 66–74, 2015.

[9] N. N. A. Sjarif, Y. Yahya, S. Chuprat, and N. H. F. M. Azmi, "Support vector machine algorithm for SMS spam classification in the telecommunication industry," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 10, no. 2, pp. 635–639, 2020, doi: 10.18517/ijaseit.10.2.10175.

[10] K. S. Kawade, D. R., & Oza, "SMS Spam Classification using WEKA," *Int. J. Electron. Commun. Comput. Technol.*, vol. 5, no. April, pp. 43–47, 2015.

[11] andi nurul Hidayat, "Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine Dan Naive Bayes," *J. Elektron. Sistim Inf. Dan Komput.*, vol. 1, no. 1, pp. 1–7, 2015.

[12] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *J. Inform. J. Pengemb. IT*, vol. 3, no. 3, pp. 394–401, 2018, doi: 10.30591/jpit.v3i3.977.

[13] P. K. Roy, J. P. Singh, and S. Banerjee, "Deep learning to filter SMS Spam," *Futur. Gener. Comput. Syst.*, vol. 102, pp. 524–533, 2020, doi: 10.1016/j.future.2019.09.001.

[14] I. D. Dinov, *Data science and predictive analytics: Biomedical and health applications using R*. 2018.

[15] T. Jo, *Text Mining Concept, Implementation and Big Dta Challenge*, vol. 36, no. 2. 2019.

[16] G. H. Lewes, "Support Vector Machines for Classification," pp. 39–66.

[17] L. D. Utami, "Inti nusa mandiri," *Inti Nusa Mandiri*, vol. 14, no. 2, pp. 133–138, 2020.

[18] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel," *Khatulistiwa Inform.*, vol. VI, no. 1, 2018.

[19] Emrald and K. M. Lhaksmana, "Klasifikasi Kategori Hadits Menggunakan Naive Bayes Classifier," *e-Proceeding Eng.*, vol. 6, no. 2, pp. 9848–9858, 2019.

[20] Jayanta, H. Mahfud, and T. Pramiyati, "Analisis Pengukuran Self Plagiarism Menggunakan Algoritma Rabin-Karp dan

e-ISSN 2614-8773
DOI : 10.29408/jit.v4i2.3665 Link : https://dx.doi.org/10.29408/jit.v4i2.3665

Jaro-Winkler Dictance dengan Stemming Tala," *Semin. Nas. Teknol. Inf. dan Multimed.*, vol. 5, no. 1, pp. 1–6, 2017.

[21] E. Manalu, F. A. Sianturi, and M. R. Manalu, "Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan Pada CV. Papadan Mama Pastries," *J. Mantik Penusa*, vol. 1, no. 2, pp. 16–21, 2017.

[22] R. I. Nurachim, "Pemilihan Model Prediksi Indeks Harga Saham Yang Dikembangkan Berdasarkan Algoritma Support Vector Machine (SVM) Atau Multilayer Perceptron (MLP) Studi Kasus : Saham PT Telekomunikasi Indonesia Tbk," *J. Teknol. Inform. Komput.*, vol. 5, no. 1, pp. 29–35, 2019.