

KLASIFIKASI SMS SPAM MENGGUNAKAN *SUPPORT VECTOR MACHINE*

Agus Setiyono¹, Hilman F. Pardede^{1,2}

¹Magister Ilmu Komputer

^{1,2}Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri

www.nusamandiri.ac.id

¹ agasasutadewa@gmail.com

²Pusat Penelitian Informatika LIPI

Lembaga Ilmu Pengetahuan Indonesia (LIPI)

<http://informatika.lipi.go.id/>

² hilman@nusamandiri.ac.id

Abstract—It is now common for a cellphone to receive spam messages. Great number of received messages making it difficult for human to classify those messages to Spam or no Spam. One way to overcome this problem is to use Data Mining for automatic classifications. In this paper, we investigate various data mining techniques, named Support Vector Machine, Multinomial Naïve Bayes and Decision Tree for automatic spam detection. Our experimental results show that Support Vector Machine algorithm is the best algorithm over three evaluated algorithms. Support Vector Machine achieves 98.33 accuracy.

Intisari—Saat ini, setiap telepon seluler sering memperoleh pesan spam. Dengan banyaknya pesan yang diterima setiap telepon seluler membuat sulit melakukan klasifikasi spam dan bukan spam secara manual dengan tenaga manusia. Oleh karena itu diperlukan alat identifikasi spam otomatis menggunakan data mining. Pada penelitian ini, kami melakukan investigasi terhadap beberapa teknik data mining yaitu *support vector machine*, *multinomial naive bayes* dan *decision tree*. Dari hasil pengujian yang dilakukan dapat diketahui bahwa *Support Vector Machine* menghasilkan akurasi tertinggi dari ketiga metode yang diuji. Akurasi dari *Support Vector Machine* sebesar 98.33%.

Kata Kunci : *decision tree*, *multinomial naive bayes*, *sms spam detection*, *support vector machine*.

PENDAHULUAN

Saat teknologi komunikasi dan informasi semakin berkembang dan begitu pula kemampuan seseorang dalam berkomunikasi dan mengolah informasi tersebut sehingga informasi yang terkumpul merupakan aset yang dapat

dimanfaatkan untuk dianalisis yang hasilnya berupa pengetahuan atau informasi berharga untuk masa mendatang. Terkadang informasi tersebut bisa berupa pesan yang bermanfaat dan juga pesan yang tidak bermanfaat. Karena semakin maraknya pesan di handphone banyak mengandung *spam*, sehingga perlu cara untuk menanganinya.

Di era digital yang berkembang begitu cepat dan semakin canggihnya perangkat telepon seluler semakin mempermudah komunikasi manusia. Telepon seluler digunakan untuk komunikasi telepon dan komunikasi pesan. Pemanfaatan pesan SMS digunakan untuk mengirim pesan ataupun menerima pesan. Penerimaan pesan SMS sekarang ini semakin tidak terkendali, terutama pesan SMS dari nomor yang tidak dikenal. Karena semakin banyaknya pesan SMS yang tidak berguna sehingga sangat mengganggu oleh pengguna perangkat telepon seluler tersebut. Maka dari itu digunakanlah suatu teknik *data mining* yang berfungsi untuk mengklasifikasikan pesan SMS termasuk *spam* atau *ham*.

Berikut ringkasan penelitian terkait tentang algoritma yang digunakan, memiliki model pencarian dengan nilai akurasi yang hampir sama dengan hasil yang berbeda. Hasil akurasi di dapat dari penggunaan algoritma *Support Vector Machine*, *Multinomial Naïve Bayes*, dan *Decision Tree* pada penelitian tentang klasifikasi SMS *spam*. *Spam Message Classification Based on the Naïve Bayes Classification Algorithm* pada *IAENG International Journal of Computer Science* dengan hasil akurasi *Naïve Bayes* 91.39% dan *Support Vector Machine* 91.02% (Ning, Junwei, & Feng, 2019) *SMS Spam Filtering Using Machine Learning Techniques: A Survey* pada *Machine Learning Research* dengan hasil akurasi *Naïve Bayes* 98.12% dan *Support Vector Machine* 97.64% (Sajedi, Zarghami Parast, & Akbari, 2016). *Text Normalization and Semantic Indexing to Enhance Instant Messaging and SMS Spam Filtering* pada *Knowledge-Based Systems* dengan hasil akurasi

Naïve Bayes 86.4%, *Support Vector Machine* 92.9% dan *Decision Tree* 80.2% (Almeida, Silva, Santos, & Gómez Hidalgo, 2016). *An Analysis of Various Algorithms For Text Spam Classification and Clustering Using RapidMiner and Weka* pada *International Journal of Computer Science and Information Security* dengan hasil akurasi *Naïve Bayes* 84.79%, *Support Vector Machine* 96.64% dan *Decision Tree* 94.74% (Zainal, Sulaiman, & Jali, 2015). *SMS Spam Filterinig Using Keyword Frequency Ratio* pada *International Journal of Security and Its Applications* dengan hasil akurasi *Naïve Bayes* 96.92% dan *Decision Tree* 96.33% (Kim, Jo, & Choi, 2015).

Pada penelitian ini menggunakan metode klasifikasi untuk menganalisis SMS spam atau ham yaitu *Support Vector Machine*, *Multinomial Naïve Bayes* dan *Decision Tree*. Dipilihnya tiga metode ini untuk mengetahui perbandingan dari masing-masing algoritma, yang mana dari ketiga algoritma tersebut mana yang terbaik. *Support Vector Machine* lebih menekankan pada korelasi hubungan kata, *Multinomial Naïve Bayes* tidak saling ketergantungan kata satu dengan yang lain sedangkan *Decision Tree* lebih kedalam pembagian kedalam kelas-kelas.

BAHAN DAN METODE

Data mining adalah suatu proses menemukan hubungan yang berarti dan kecenderungan dengan memeriksa sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Utama, Sihwi, & Doewes, 2014).

Support Vector Machine adalah metode yang menganalisis data dan mengenali pola yang digunakan untuk klasifikasi data. Sistem kerja *Support Vector Machine* dengan cara memasukan data tertentu kedalam sebuah sistem kemudian memprediksikannya. Dari data tersebut akan di kelompokkan kedalam dua kelas yang berbeda. Data baru tersebut akan diklasifikasikan kedalam kelas yang satu atau kelas yang lain. Dari hal tersebut maka akan jelas bahwa data tersebut akan dikelompokkan ke dalam kelas yang sesuai (Lidya, Sitompul, & Efendi, 2015).

Support Vector Machine adalah metode terbaik yang bisa dipakai dalam permasalahan klasifikasi. Konsep *Support Vector Machine* bermula dari masalah klasifikasi dua kelas sehingga membutuhkan *training set* positif dan negatif. *Support Vector Machine* berusaha menemukan pemisah terbaik untuk memisahkan ke dalam dua kelas dan memaksimalkan margin antara dua kelas. Pada banyak kasus, data tidak bisa diklasifikasi menggunakan metode *linier*

Support Vector Machine, sehingga dikembangkanlah fungsi *kernel* untuk mengklasifikasikan data dalam bentuk *non-linier* (Pratama, Wihandika, & Ratnawati, 2018).

Fitur ruangnya sangat besar dan tak terhingga sangat sulit untuk ditransformasikan sehingga berakibat fitur jadi tidak dikenali. Untuk mengatasi masalah tersebut, pada *Support Vector Machine* menggunakan *kernel trick*. Fungsi yang dihasilkan dari percobaan tersebut adalah sebagai berikut:

$$f(x_d) = \sum_{i=1}^{n_s} a_i y_i K(x_i, x_d) + b(x_i sv) \dots\dots\dots (1)$$

Sebuah fungsi bisa menjadi fungsi kernel jika memenuhi Teorema Mercer, yang mana menyatakan bahwa matriks kernel yang dihasilkan harus bersifat semi *positive* semi *definite*. Berikut adalah beberapa fungsi kernel yang umum digunakan yaitu :

a. Kernel linier

$$K(x_i, x) = x_i^T x \dots\dots\dots (2)$$

b. Polynomial

$$K(x_i, x) = (\gamma \cdot x_i^T x + r)^p, \gamma > 0 \dots\dots\dots (3)$$

c. Radial basis function

$$K(x_i, x) = \exp(-\gamma |x_i - x|^2), \gamma > 0 \dots\dots\dots (4)$$

d. Sigmoid kernel

$$K(x_i, x) = \tanh(\gamma \cdot x_i^T x + r) \dots\dots\dots (5)$$

Pada penelitian ini akan diterapkan kernel *Radial Basis Function* dengan parameter C dan Gamma.

Pada prinsipnya *Support Vector Machine* bekerja secara *linear*, dan dikembangkan untuk dapat diterapkan pada masalah *non-linear*. Dengan menggunakan metode *kernel trick* yang mencari pemisah dengan cara mentransformasi *dataset* ke ruang vektor yang berdimensi lebih tinggi, kemudian proses klasifikasi dilakukan pada *feature space* tersebut. Penentuan fungsi kernel yang digunakan akan sangat berpengaruh pada hasil prediksi. Misalkan $\{x_1, \dots, x_n\}$ 1 adalah *dataset* dan $y_i \in \{+1, -1\}$ adalah label kelas dari data x_i . Bidang pemisah yang terbaik tidak hanya dapat memisahkan data tetapi juga memiliki *margin* yang paling besar. Data yang berada tepat pada bidang pemisah disebut sebagai *support vector*.

Algoritma *Naïve Bayes* adalah salah satu metode klasifikasi terbaik dari penerapan pembelajaran Bayes. Kemampuannya bisa dibandingkan dengan algoritma *decision tree* dan *neural network* dalam beberapa aplikasi lain, tetapi kompleksitas komputasinya jauh lebih sedikit dari algoritma yang lain (Ning et al., 2019).

Multinomial Naïve Bayes digunakan untuk mengasumsikan independensi kemunculan kata dalam dokumen. Metode *Multinomial Naïve Bayes* tidak memperhitungkan urutan kata dan *information-context* dalam dokumen, namun

memperhitungkan jumlah kata dalam dokumen. *Multinomial Naïve Bayes* mengambil jumlah kata yang muncul pada dokumen, didalam *Multinomial Naïve Bayes*, sebuah dokumen terdiri dari beberapa kejadian kata dan di asumsikan panjang dokumen tidak bergantung pada kelasnya. Dengan menggunakan asumsi Bayes yang sama bahwa kemungkinan tiap kejadian kata dalam sebuah dokumen adalah bebas tidak terpengaruh dengan konteks kata dan posisi kata dalam dokumen tersebut (Marta, Yuridis, Butar, & Fauzi, 2019).

Bayes mudah dihitung untuk fitur bertipe kategori *Multinomial Naïve Bayes* yang merupakan model pengembangan dari algoritma *bayes* yang sangat cocok dalam pengklasifikasian dokumen. Pada formula *Multinomial Naïve Bayes*, kelas dokumen tidak hanya ditentukan dengan kata yang muncul tetapi juga jumlah kemunculannya dalam dokumen tersebut (Marta et al., 2019).

Decision tree adalah diagram alir yang serupa dengan struktur pohon, di mana setiap cabang menotasikan atribut yang diujikan, setiap cabangnya dapat menampilkan hasil dari atribut tes, dan daun menampilkan kelas-kelas tertentu atau distribusi dari kelas-kelas tertentu (Andri, Yesi Novaria Kunang, 2013).

Decision tree sangat didasarkan pada pendekatan *divide-and-conquer* untuk mengklasifikasikan suatu masalah. *Decision tree* bekerja dari atas ke bawah, mencari pada setiap tahap atribut untuk membaginya ke dalam bagian terbaik *class* tersebut, dan memproses secara rekursif submasalah yang dihasilkan dari pembagian tersebut. Strategi ini dapat menghasilkan sebuah pohon keputusan yang dapat diubah menjadi satu set *classification rules* (Andriani, 2013).

Metode *TF-IDF* (*Term Frequency Inverse Document Frequency*) merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen (Amrizal, 2018). Metode *TF-IDF* menggabungkan dua konsep untuk perhitungan bobot yaitu, frekuensi kemunculan sebuah kata didalam sebuah dokumen tertentu dan inverse frekuensi sebuah dokumen yang mengandung kata tersebut.

Inverse Document Frequency (idf) dihitung dengan menggunakan formula sebagai berikut (Amrizal, 2018) :

$$IDF = \text{Log}\left(\frac{D}{DF}\right) \dots\dots\dots (6)$$

Sedangkan formula TF-IDF :

$$W = TF \times IDF \dots\dots\dots (7)$$

Dimana :

IDF = *inverse document frequency*

D = jumlah dokumen
 DF = frekuensi dokumen
 W = TF-IDF
 TF = frekuensi kata

Pada Penelitian *Spam Message Classification Based on the Naïve Bayes Classification Algorithm* (Ning et al., 2019) terdapat kekurangan dalam mengklasifikasikan SMS *spam*, karena *Naïve Bayes* tidak cocok untuk mengklasifikasikan *text*. Metode *Naïve Bayes* tidak memperhitungkan urutan kata dan *information-context* dalam dokumen, namun memperhitungkan jumlah kata dalam dokumen. Dengan menggunakan asumsi *Naïve Bayes* yang sama bahwa kemungkinan tiap kejadian kata dalam sebuah dokumen adalah bebas tidak terpengaruh dengan konteks kata dan posisi kata dalam dokumen.

Maka dari itu dipilihlah metode klasifikasi *Support Vector Machine* karena bisa mengkorelasikan antar kata satu dengan yang lain, yang mana *Support Vector Machine* bisa membedakan kata sehingga bisa dikelompokkan pada *decision boulder* kanan ataupun kiri. *Support Vector Machine* berasal dari masalah kalsifikasi dua kelas sehingga memerlukan data *training set* positif dan data *training set* negative. *Support Vector Machine* berusaha untuk menemukan pemisah terbaik, untuk dipisahkan kedalam dua kelas tersebut. Sehingga bisa digunakan untuk klasifikasi SMS *spam* atau *ham*.

Metode penelitian yang digunakan untuk menganalisis SMS *spam* atau *ham* adalah *Support Vector Machine*, *Multinomial Naïve Bayes* dan *Decision Tree*. Metode *Support Vector Machine* ini merupakan salah satu metode yang ada pada teknik klasifikasi dalam *data mining*. Pada penelitian ini penerapan data mining menggunakan metode *Support Vector Machine* untuk mengetahui klasifikasi SMS *spam* atau *ham* karena *Support Vector Machine* mempunyai bidang pemisah yang terbaik dalam memisahkan data dan juga bisa mengkorelasikan kata satudengan yang lain. Proses analisis *data mining* dalam penelitian ini menggunakan perangkat lunak aplikasi *datamining* Python.

HASIL DAN PEMBAHASAN

Penelitian ini dilakukan pada bulan April 2019, data dalam penelitian ini dengan mengambil data dari UCI Repository berupa data set SMS *spam*. Data berjumlah 4813 sms *spam*. Parameter yang digunakan yaitu *spam* atau *ham* di lihat berdasarkan isi SMS. Data set sampel SMS *spam* dapat dilihat pada tabel 1. berikut :

Tabel 1. Data set sampelSMSspam

category	message
ham	Go until jurong point
ham	Ok lar... Joking wifu oni...
spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's
ham	U dun say so early hor... U c already then say...
ham	Nah I don't think he goes to usf
spam	FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std chgs to send
ham	Even my brother is not like to speak with me. They treat me like aids patent.
ham	As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Callers. Press *9 to copy your friends Callertune
spam	WINNER!! As a valued network customer you have been selected to receivea ??900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
spam	Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030

Sumber: (Setiyono & Pardede, 2019)

Hasil rata-rataakurasi TF-IDF dengancross validation= 10dapat dilihat pada tabel 2. dibawah ini:

Tabel 2. Hasil akurasi TF-IDF dengancv= 10

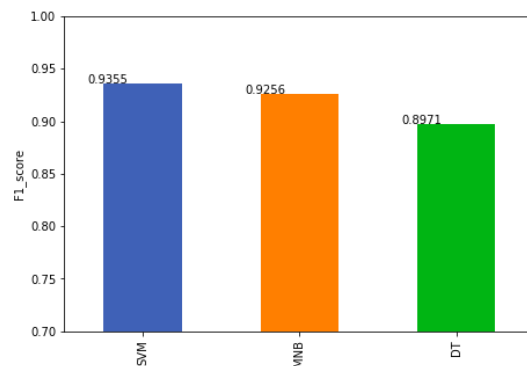
TF-IDF cv = 10	SVM		MNB	DT
	rbf	sigmoid		
Feature 1600	0.9833	0.9813	0.9813	0.9710
Feature 1500	0.9833	0.9813	0.9813	0.9710
Feature 1000	0.9793	0.9771	0.9772	0.9647
Feature 800	0.9793	0.9792	0.9689	0.9647
Feature 400	0.9709	0.9688	0.9689	0.9606

Sumber :HasilPenelitian (2019)

Tahap evaluasi digunakan *10-fold cross validation* untuk dapat menghasilkan nilai akurasi dan nilai *Area Under Curve* sebagai bahan perbandingan dalam menentukan algoritma terbaik yang digunakan. *10-fold cross validation* bekerja dengan membagi *dataset* masukan menjadi 10 bagian yang sama rata. 9 bagian kemudian di-training sedangkan yang 1 bagian lainnya digunakan untuk testing. Proses ini diulang sebanyak 10 kali untuk setiap bagian sehingga setiap bagian dari kesepuluh bagian pernah menjadi data untuk testing. *Operator Cross Validation* melakukan proses *10-fold cross validation* ini untuk ketiga algoritma yang digunakan. Untuk setiap percobaan akan dihitung akurasinya. Akurasi akhir adalah nilai rata-rata dari akurasi sepuluh percobaan tersebut. Hasilnya dapat disajikan dalam bentuk *confusion matrix*.

Pada pengujian ini, data digunakan adalah data bersih yang telah melalui *preprocessing*. Data tersebut diambil dari operator Read Excel, hal ini dilakukan karena *dataset* disimpan dalam bentuk Excel. *Process documents from files* untuk mengkonversi *files* menjadi dokumen. *Process validasi* terdiri dari *data training* dan *data testing*. Pada tahap ini juga menggunakan *Set Role* yang berfungsi untuk menentukan field pada kelas kemudian menggunakan *TF-IDF* agar akurasi yang dihasilkan lebih tinggi. Dari hasil pemodelan yang telah dilakukan sebelumnya. Berikut ini akan dijelaskan *F1 Score* dan *Confusion Matrix* dari algoritma *Support Vector Machine* dengan jumlah *feature* yang digunakan 1600 *feature*.

Berikut ini akan dijelaskan *F1 Score* dari algoritma *Support Vector Machine*, *Multinomial Naïve Bayes* dan *Decision Tree* :



Sumber: (Setiyono & Pardede, 2019)

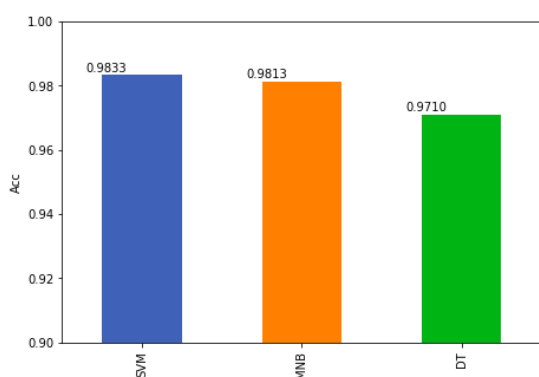
Gambar 1. F1 Score SVM, MNB, DT

KategoriKlasifikasi :

1. 0.90 - 1.00 = *excellent classification*
2. 0.80 - 0.90 = *good classification*
3. 0.70 - 0.80 = *fair classification*
4. 0.60 - 0.70 = *poor classification*
5. 0.50 - 0.60 = *failure*

Hasil *F1 Score Support Vector Machine* dengan nilai yang dihasilkan dari gambar 1 diatas sebesar 0.9355 dimana diagnose hasilnya adalah *Excellent Classification*. Hasil *F1 Score Multinomial Naïve Bayes* dengan nilai yang dihasilkan dari gambar 1 diatas sebesar 0.9256 dimana diagnosah asilny adalah *Excellent Classification*. Hasil *F1 Score Decision Tree* dengan nilai yang dihasilkan dari gambar 1 diatas sebesar 0.8971 dimana diagnosa hasilnya adalah *Good Classification*.

Berikut ini akan dijelaskan *Accuracy dan Confusion Matrix* dari algoritma *Support Vector Machine, Multinomial Naïve Bayes* dan *Decision Tree* :



Sumber: (Setiyono & Pardede, 2019)
Gambar 2. Accuracy SVM, MNB, DT

Confusion Matrix memberikan penilaian *performance* klasifikasi berdasarkan objek benar dan salah. *Confusion Matrix* berisi informasi actual dan prediksi pada sistem klasifikasi.

Tabel 3. *Confusion Matrix SVM*

Accuracy : 98.33%	class precision		
pred. Positif (ham)	4133	46	98.90%
pred. Negatif (spam)	34	600	
class recall	98.90%		

Sumber: (Setiyono & Pardede, 2019)

$$Recall = \frac{TP}{TP + FP} = \frac{4133}{4133 + 46} = \frac{4133}{4179} = 0.9890$$

$$Precision = \frac{TP}{TP + FN} = \frac{4133}{4133 + 34} = \frac{4133}{4167} = 0.9918$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{4133 + 600}{4133 + 600 + 46 + 34} = \frac{4733}{4813} = 0.9833$$

Akurasi yang diperoleh dari algoritma *Support Vector Machine* yaitu 98.33% dari 4813 SMS *spam* dan *ham* yang di ambil dari UCI Repository. Data SMS *ham* yang sesuai prediksi yaitu 4133 data. Data SMS *spam* yang termasuk ke dalam prediksi *ham* yaitu 46 data. Data SMS *ham* yang termasuk kedalam prediksi *spam* yaitu 34 data dan data SMS *spam* yang sesuai prediksi yaitu 600 data.

KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dijelaskan maka dapat disimpulkan bahwa nilaia kursi tertinggi padaa lgoritma *Support Vector Machine* dengan akurasi 98.33%. Hal ini menunjukkan bahwa algoritma *Support Vector Machinememiliki* performa yang handal dalam melakukan klasifikasi SMS *spam* di banding metode *Multinomial Naïve Bayes* dan *Decision Tree*.

REFERENSI

Almeida, T. A., Silva, T. P., Santos, I., & Gómez Hidalgo, J. M. (2016). Text normalization and semantic indexing to enhance Instant Messaging and SMS spam filtering. *Knowledge-Based Systems, 108*(November 2017), 25–32. <https://doi.org/10.1016/j.knosys.2016.05.001>

Amrizal, V. (2018). Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim). *Jurnal Teknik Informatika, 11*(2), 149–164. <https://doi.org/10.15408/jti.v11i2.8623>

Andri, Yesi Novaria Kunang, S. M. (2013). Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan, *2013*(June 2016), 56–63. <https://doi.org/10.13140/RG.2.1.4212.1845>

Andriani, A. (2013). Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus : Amik “ Bsi Yogyakarta .” *Seminar Nasional Teknologi Informasi Dan Komunikasi 2013 (SENTIKA 2013)*,

- 2013(Sentika), 163–168.
- Kim, S. E., Jo, J. T., & Choi, S. H. (2015). SMS Spam filtering using keyword frequency ratio. *International Journal of Security and Its Applications*, 9(1), 329–336. <https://doi.org/10.14257/ijisia.2015.9.1.31>
- Lidya, S. K., Sitompul, O. S., & Efendi, S. (2015). Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (Svm). *Seminar Nasional Teknologi Dan Komunikasi 2015*, 2015(Sentika), 1–8. <https://doi.org/10.1016/j.eswa.2013.08.047>
- Marta, T., Yuridis, E., Butar, B., & Fauzi, M. A. (2019). Penentuan Rating Review Film Menggunakan Metode Multinomial Naïve Bayes Classifier dengan Feature Selection berbasis Chi-Square dan Galavotti-Sebastiani-Simi Coefficient. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTI IK) Universitas Brawijaya*, 3(1), 447–453.
- Ning, B., Junwei, W., & Feng, H. (2019). Spam message classification based on the naïve Bayes classification algorithm. *IAENG International Journal of Computer Science*, 46(1).
- Pratama, A., Wihandika, R. C., & Ratnawati, D. E. (2018). Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer E-ISSN, 2548, 964X*, 2(4), 1704–1708.
- Sajedi, H., Zarghami Parast, G., & Akbari, F. (2016). To cite this article: Hedieh Sajedi, Golazin Zarghami Parast, Fatemeh Akbari. SMS Spam Filtering Using Machine Learning Techniques: A Survey. *Machine Learning Research*, 1(1), 1–14. <https://doi.org/10.11648/j.mlr.20160101.1>
- Setiyono, A., & Pardede, H. F. (2019). *Laporan Akhir Penelitian: Klasifikasi Sms Spam Menggunakan Support Vector Machine*. Jakarta.
- Utama, T. D., Sihwi, S. W., & Doewes, A. (2014). Implementasi Algoritma Iterative Dichotomiser 3 Pada Penyeleksian Program Mahasiswa Wirausaha Uns. *Jurnal Itsmart*, 3(2), 74–82.
- Zainal, K., Sulaiman, N. F., & Jali, M. Z. (2015). An Analysis of Various Algorithms For Text Spam Classification and Clustering Using RapidMiner and Weka. *International Journal of Computer Science and Information Security (IJCSIS)*, 13(3), 66–74.