IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI KOMENTAR SPAM PADA INSTAGRAM

Uro Abdurohim[1], Dedy Apriyadi[2], Arya Devi Listiani[3]

[1,2,3]STMIK BANDUNG
Sekolah Tinggi Manajemen dan Informatika Bandung
JL. Cikutra No.113, Bandung 40124, INDONESIA
Contact address:
[1]uro.abdulrohim@gmail.com

ABSTRAK

Instagram adalah sebuah aplikasi berbagi foto dan video yang memungkinkan pengguna mengambil foto, mengambil video, menerapkan *filter digital* dan membagikannya ke berbagai layanan jejaring sosial. Para tokoh masyarakat, terutama aktor, aktris, *influencer* dan selebritas instagram menggunakan Intagram sebagai salah satu *platform* yang digunakan dalam mempromosikan berbagai kegiatan atau karya mereka kepada para penggemar atau pengikutnya. Tetapi, banyak sekali komentar spam yang ada pada postingan tersebut baik berupa komentar baik ataupun buruk. Penelitian ini memperoleh data komentar dari Instagram untuk dijadikan dataset yaitu dari *public figure*, aktris dan aktor yang memiliki lebih dari 5 juta pengikut. Dengan menggunakan *preprocessing* (*cleansing data, tokenization, casefolding* dan *stopword removal)* dan TF-IDF untuk perbobotan, pemilihan kernel yang ingin dipakai dan dilakukan pelatihan model SVM pada proses klasifikasi tersebut. Sehingga diperlukan sebuah sistem yang dapat mengklasifikasikan bahwa komentar tersebut spam atau tidak. Metode yang digunakan adalah *Support Vector Machine (SVM). Support Vector Machine* adalah suatu metode *machine learning* yang memiliki prinsip kerja *Structural Risk Minimization* bertujuan mencari *hyperplane* atau pemisah antar dua buah kelas pada suatu *input space.* Dari hasil pengujian diperoleh bahwa performansi algoritma yang diusulkan menghasilkan tingkat akurasi sebesar 96.82%, presisi 94%, recall 92%, dan F1-Score 93%.

Kata Kunci : klasifikasi spam, algoritma support vector machine, data mining, machine learning, instagram

ABSTRACT

Instagram is a photo and video sharing application that allows users to take photos, take videos, apply digital filters, and share them on various social networking services. Public figures, especially actors, actresses, influencers and Instagram celebrities, use Instagram as a platform to promote their various activities or works to their fans or followers. However, there are a lot of spam comments on these posts, both good and bad. This research obtained comment data from Instagram to be used as a dataset, namely from public figures, actresses and actors who have more than 5 million followers. By using preprocessing (data cleansing, tokenization, case folding and stopword removal) and TF-IDF for weighting, selecting the kernel you want to use and training the SVM model in the classification process. We need a system that can classify whether comment are spam or not spam. The method used is Support Vector Machine is a machine learning method that has the working principle of structural risk minimization aimed at finding hyperplanes or separators between two classes in an input space. From the test results it was found that the performance of the proposed an accuracy level of 96.82%, precision 94%, recall 92%, and f1-score 93%.

Keywords: spam classification, support vector machine algorithm, data mining, machine learning, instagram

1. Pendahuluan

Di era teknologi, internet membuat segalanya jadi mudah, seperti halnya media sosial banyak digunakan oleh masyarakat sebagai media komunikasi secara daring dan sebagai media hiburan. Media sosial yang sering digunakan adalah salah satunya Instagram. Instagram adalah *platform* media sosial yang digunakan sebagai tempat bersosialisasi secara *online* dengan membagikan foto, membuat video atau reels. Aplikasi Instagram banyak digunakan oleh *public figure*, artis, influencer dan kalangan masyarakat.

Instagram tidak hanya digunakan sebagai sarana bersosialisasi *online*, tetapi juga sebagai sarana hiburan, pemasaran bisnis atau produk, dan media kreatif. Untuk mendapatkan informasi yang lebih detail dari sebuah postingan atau konten dari pengguna atau ingin memberikan umpan balik terkait postingan tersebut bisa dilakukan dengan mengomentari postingan yang diunggah pada kolom komentar. Tetapi pada kolom komentar tersebut terdapat beberapa komentar yang mengandung spam, sehingga informasi yang didapatkan bisa tidak relevan. Untuk mengatasi hal tersebut dibutuhkan suatu sistem yang dapat mengklasifikasikan komentar ke dalam dua kelas spam dan non spam.

Berikut ringkasan penelitian terkait algoritma yang digunakan dengan nilai akurasi yang didapatkan berbeda – beda. Syam, dkk. [1] melakukan Klasifikasi Komentar Spam pada Instagram menggunakan metode *Support Vector Machine.* Hasil penelitian yang didapat dengan kernel linier memberikan ketetapan nilai akurasi 97,33%. Agus Setiyono dan Hilman F. Pardede [2] telah melakukan Klasifikasi SMS Spam menggunakan *Support Vector Machine* dengan hasil yang didapatkan nilai akurasi 98,33%.

Pada penelitian ini menggunakan algoritma *Support Vector Machine* untuk mengidentifikasi komentar spam dan tidak spam. Penelitian ini dilakukan dengan mengambil komentar dari Instagram untuk nantinya dijadikan sebagai dataset.

1.1. Overview Data Mining dan Machine Learning

Data Mining merupakan suatu proses penggalian informasi dan pola yang bermanfaat dari sebuah data yang sangat besar. Data mining mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data

Dapat disimpulkan dari kedua pendapat tersebut, *Data mining* adalah suatu proses untuk mendapatkan informasi dengan melakukan pengolahan data yang memiliki ukuran besar dan perlu dilakukan ekstraksi data untuk menjadi sebuah informasi baru

Machine Learning merupakan bidang studi yang befokus pada analisis dan desain algoritma yang memungkinkan komputer memecahkan masalah dengan cara belajar seperti manusia. Sebagai contoh komputer dapat mendeteksi email yang berisi spam dan non spam melalui algoritma pengenalan tulisan tangan tanpa menggunakan pemrograman tradisional. Algoritma yang sama ketika diberikan data pelatihan yang berbeda menghasilkan logika klasifikasi yang berbeda.

Machine Learning dapat dikelompokkan menjadi tiga, yaitu [3]:

1) *Supervised Learning* adalah pembelajaran teraraj atau terawasi, dimana proses pembelajaran, komputer dan mesin akan mempelajari *data training* yang berisi label. Adapun contoh – contoh algoritma yang termasuk ke dalam *supervised learning* adalah *Linear Regression, Logistic Regression, Linear Discriminant Analysis, k-Nearest Neighbors, Support Vector Machines, Random Forests,* dan *Neural networks.*
2) *Unsupervised Learning* adalah proses pembelajaran tanpa petunjuk. Algoritma dalam komputer yang akan menemukan pola-pola di dalam data. *Unsupervised learning* terjadi ketika memiliki jumlah data *input* (x) dan tanpa variabel *output* yang berhubungan. *Unsupervised Learning* ini dibagi menjadi dua kelas asosiasi dan *clustering*.
3) *Reinforcement Learning* berbeda dengan *supervised* dan *unsupervised learning*. Komputer akan diminta untuk memecahkan suatu tugas dengan berinteraksi dengan suatu lingkungan. Mesin akan mempelajari cara mengambil keputusan yang spesifik berdasarkan lingkungan yang dinamis. Salah satu contoh implementasi metode ini adalah

dengan menggunakan *reward and punishment*.

1.2. Penelitian Terkait

Penelitian terkait dengan riset ini adalah sebagai berikut:

Penelitian perbandingan kinerja algoritma *Naïve Bayes* dan *C.45* dalam klasifikasi spam email oleh Sulaeman, dkk. pada tahun 2022 [4]. Penelitian ini menggunakan *dataset spam* dan *not spam* berasal dari https://www.kaggle.com/ dengan menggunakan algoritma *Naïve Bayes* dan *C.45*, dataset menggunakan aplikasi excel dan tool RapidMiner dan *Cross Validation* untuk model uji tiap algoritma. Kemudian pada penerapan model algoritma *Naïve Bayes* menjelaskan bahwa tingkat hasil akurasi yang didapatkan dari algoritma *Naïve Bayes* dengan nilai sebesar 96.79%.

Pada penelitian Klasifikasi SMS Spam dengan Komparasi Metode SVM dan *Naïve Bayes* pada tahun 2023 oleh [5]. Perbandingan hasil dari presisi, *recall*, dan *f1-score* dengan pemrograman bahasa python, hasil penelitian yang didapat nilai akurasi tertinggi diperoleh pada Algoritma *Naïve Bayes* dengan nilai akurasi sebesar 94% dan Algoritma *Support Vector Machine* 93%.

Penelitian pada tahun 2021 telah mengimplementasikan algoritma *Multinomial Naïve Bayes* untuk mengklasifikasikan SMS spam berbahasa Indonesia menggunakan dataset dari Kaggle [6]. Setelah melakukan *preprocessing data*, termasuk *case folding*, *stemming*, *tokenizing*, dan penghapusan *stopwords*, model *Multinomial Naïve Bayes* dilatih dengan partisi data 75:25 dan parameter-parameter tertentu. Hasil evaluasi menunjukkan bahwa model yang diperoleh memiliki nilai *precision*, *recall*, *F1-score*, dan akurasi rata-rata sebesar 0.93, 0.92, 0.93, dan 0.94.

Pada penelitian Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan *Support Vector Machine* pada tahun 2021 oleh [7]. Untuk Pengambilan dataset dari Twitter dilakukan dengan menggunakan *library* yang mampu untuk melakukan *crawling data*. Untuk praproses menggunakan tahapan *case folding*, *tokenizing, filtering,* dan *stemming*. Penelitian ini dilakukan pengujian dengan membandingkan tiga buah kernel yang umum dipakai, dan hasilnya menunjukkan bahwa sistem dapat melakukan klasifikasi ujaran kebencian untuk membantu menganalisa perilaku masyarakat di dunia maya dengan penggunaan kernel RBF yang menghasilkan nilai akurasi paling tinggi 93% dengan komposisi data latih sebanyak 1000 data dengan perbandingan 7:3 data latih dan data uji.

Pada penelitian Perbandingan Klasifikasi SMS Berbasis *Support Vector Machine, Naïve Bayes Classifier*, *Random Forest* dan *Bagging Classifier* Pada tahun 2021 oleh [8]. Hasil penelitian yang didapatkan *Performance Score SVM* dengan nilai 86.5%, *Naive Bayes* 96.4%, *Random Forest* 96.8 %, dan *Bagging Classifier* 97.4%. Sehingga *Bagging Classifier* memiliki nilai tertinggi dan dapat dipergunakan sebagai sarana untuk memfiltrasi SMS yang masuk ke dalam *inbox* pengguna, dan memberikan hasil akurat untuk menyaring SMS yang masuk.

2. Metodologi

Metodologi yang digunakan pada penelitian ini adalah:

1) Metode Pengumpulan Data

Pada tahap pengumpulan data, dilakukan pengambilan data pelatihan dari Instagram dengan menggunakan Selenium. Data akan di ambil dari 10 akun dengan pengikut minimal 5 Juta, semakin banyak pengikut maka akan semakin banyak komentar spam-nya. Apabila data sudah didapatkan, maka akan dilakukan *preprocessing* data dan pelabelan data secara manual oleh ahli bahasa Indonesia dan kemudian melakukan implementasi dengan algoritma *Support Vector Machine* untuk klasifikasi teks.

2) Metode Analisis

Pada tahap ini melibatkan langkah – langkah atau tahapan yang dilakukan untuk menerapkan SVM dalam mengklasifikasikan komentar spam. Akan menguraikan bagaimana dalam mengumpulkan data, memilih parameter SVM dan mengidentifikasi langkah – langkah penting dalam proses pelatihan dan pengujian model.

3) Metode Perancangan

Pada tahap ini membantu mengatur langkah – langkah dan strategi yang diperlukan dalam membangun dan mengimplementasikan algoritma SVM secara efektif. Meliputi pemilihan fitur, pra-pemprosesan data, pengaturan parameter

SVM, pengujian model, dan evaluasi kinerja.

4) Implementasi

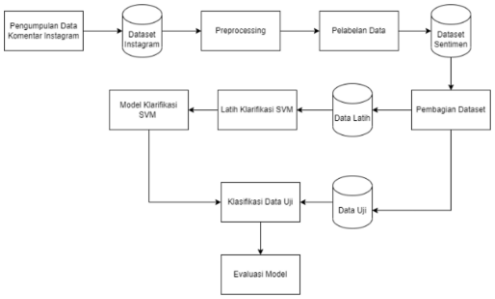
Pada tahap ini akan dilakukan pembuatan kode program sampai proses klasifikasi sentimen dokumen.

5) Pengujian

Pada tahap ini dilakukan perhitungan akurasi dari implementasi yang telah dilakukan. Pengujian yang dilakukan adalah perbandingan akurasi klasifikasi yang dihasilkan SVM.

3. Pembahasan

Algoritma *Support Vector Machine* (SVM) merupakan metode pembelajaran mesin yang berlandaskan prinsip *Structural Risk Minimization*. Tujuan utama SVM adalah menemukan *hyperplane* optimal yang memisahkan data dari dua kelas berbeda dalam ruang input. SVM sangat efektif dalam masalah klasifikasi biner, di mana data pelatihan terdiri dari contoh-contoh positif dan negatif. Untuk mengatasi permasalahan data yang *non-linear separable*, SVM memperkenalkan konsep fungsi kernel yang memetakan data ke ruang fitur berdimensi tinggi. Dalam penelitian ini, kami akan membahas langkah-langkah implementasi SVM untuk mengklasifikasi komentar pada platform Instagram.



Gambar 1. Arsitektur Sistem

3.1. Preprocessing

Data preprocessing merupakan tahap dalam melakukan data mining, sebelum menuju tahap pemprosesan. Dalam proses ini data mentah akan diubah dalam bentuk yang akan lebih mudah dipahami. Pada proses *preprocessing* memiliki beberapa sub-proses yaitu *case folding, cleansing data, normalization, tokenization, stopwords removal*, dan *stemming.*

1) *Case Folding*

Pada tahap ini seluruh teks diubah menjadi *lower case* (huruf kecil). Dilakukan untuk menghindari masalah sensitivitas huruf besar atau kecil dalam analisis teks.

Sebelum	Sesudah
Semoga allah kasih jodoh yg terbaik buat kalian	semoga allah kasih jodoh yg terbaik buat kalian

2) *Data Cleansing*

Dalam tahap ini teks yang diperoleh akan dibersihkan dengan menghapus seluruh karakter huruf, link url, username, tanda baca.

Sebelum	Sesudah
semoga allah kasih jodoh yg terbaik buat kalian	semoga allah kasih jodoh yg terbaik buat kalian

3) *Normalization*

Tahapan ini akan dilakukan pengubahan kata pada kata yang tidak baku sesuai dengan kata yang ada di kamus.

Sebelum	Sesudah
semoga allah kasih jodoh yg terbaik buat kalian	semoga allah kasih jodoh yang terbaik buat kalian

4) *Tokenization*

Pada tahap ini dilakukan pemecahan kalimat menjadi kata.

Sebelum	Sesudah
semoga allah kasih jodoh yang terbaik buat kalian	['semoga', 'allah', 'kasih', 'jodoh', 'yang', 'terbaik', 'buat', 'kalian']

5) *Stopwords Removal*

Merupakan tahapan filterisasi, untuk menghapus kata-kata yang tidak memiliki arti, sehingga tidak mempengaruhi proses pembobotan.

Sebelum	Kata dihapus	Sesudah
['semoga', 'allah', 'kasih', 'jodoh', 'yang', 'terbaik', 'buat', 'kalian']	['yang', 'buat', 'kalian']	['semoga', 'allah', 'kasih', 'jodoh', 'terbaik']

6) *Stemming*

Pada tahap ini adalah mengubah data menjadi *root* atau kata dasar.

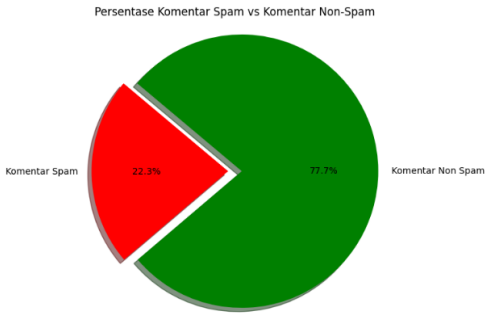
Sebelum	Sesudah
['semoga', 'allah', 'kasih', 'jodoh', 'terbaik']	['moga', 'allah', 'kasih', 'jodoh', 'baik']

3.2. Pembentukan Dataset

Membentuk dataset untuk pelatihan dan pengujian model *Support Vector Machine*. Ini melibatkan pemberian label pada setiap komentar yang menunjukkan apakah komentar tersebut spam atau non-spam. Pada penelitian ini dihasilkan dataset dengan jumlah 2800 data komentar yang ditunjukkan pada Tabel 1.

Tabel 1. Jumlah Dataset

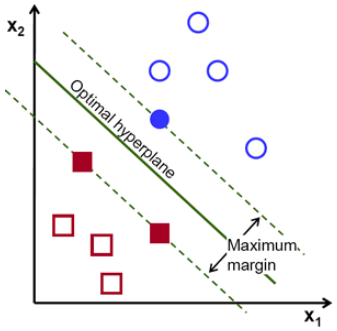
	Label / Kategori	**Jumlah**
1	Komentar Spam	625
2	Komentar Non spam	2175
Total Komentar		**2800**



Gambar 2. Persentase Dataset

3.3. Algoritma yang diusulkan

Algoritma ini menerapkan *Support Vector Machine*, sebuah teknik yang mampu menangani masalah klasifikasi baik *linier* maupun *non-linier*. Untuk masalah *non-linier*, SVM memanfaatkan konsep *kernel* untuk memetakan data ke *hyperplace*. Tujuannya adalah menemukan *hyperplane* optimal yang memisahkan data dari dua kelas dengan margin maksimal. Margin ini adalah jarak terdekat antara *hyperplane* dan data dari masing-masing kelas. Data yang terletak pada margin inilah yang disebut sebagai *Support Vector*.



Gambar 3. Pemisah *Hyperplane* (Sumber: medium.com)

Asumsikan kedua kelas -1 dan +1 dapat terpisah secara sempurna oleh *hyperplane* berdimensi d yang didefinisikan pada persamaan (1).

$$\vec{w} \cdot x + b = 0 \quad \textbf{(1)}$$

Pola $\vec{xi}$ termasuk kelas 1 dapat dirumuskan sebagai pola yang memenuhi pertidaksamaan (2).

$$\vec{w} \cdot \vec{xi} + b \leq -1 \quad \textbf{(2)}$$

Pola $\vec{xi}$ termasuk kelas 0 akan dirumuskan sebagai pola yang memenuhi pertidaksamaan (3).

$$\vec{w} \cdot \vec{xi} + b \geq +1 \quad \textbf{(3)}$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya persamaan (4).

$$1 \, \|\vec{w}\| \quad \textbf{(4)}$$

Berdasarkan persamaan tersebut, metode yang digunakan adalah *hard margin classification*. Pendekatan ini memiliki keterbatasan karena mencari *hyperplane* yang memisahkan data dengan margin maksimal secara sempurna. Hal ini menyebabkan model menjadi kaku dan rentan terhadap *noise* dalam data. Selain itu, *hard margin* tidak efektif untuk data yang *non-linear separable*.

Untuk mengatasi permasalahan tersebut, *Support Vector Machine* memanfaatkan fungsi kernel. Fungsi kernel berperan dalam memetakan data ke ruang fitur

berdimensi tinggi sehingga data *non-linear* dapat dipisahkan secara *linear*. Beberapa jenis fungsi kernel yang umum digunakan adalah *kernel polynomial*, *radial basis function* (RBF), dan *sigmoid*. Persamaan umum fungsi kernel dituliskan pada persamaan 5, 6, 7 dan 8.

Linear:

$$K(x_i, x_d) = X_i^T X_j \quad \textbf{(5)}$$

Polynomial:

$$K(x_i, x_d) = (X_i^T X_j + 1)d, \boldsymbol{\gamma} > \mathbf{0} \quad \textbf{(6)}$$

Radial Basis Function (RBF):

$$K(x_i, x_d) = exp\ exp\ (-\boldsymbol{\gamma}\ \|\boldsymbol{Xi} - \boldsymbol{Xj}\|\mathbf{2})\ , \boldsymbol{\gamma} > \mathbf{0} \quad \textbf{(7)}$$

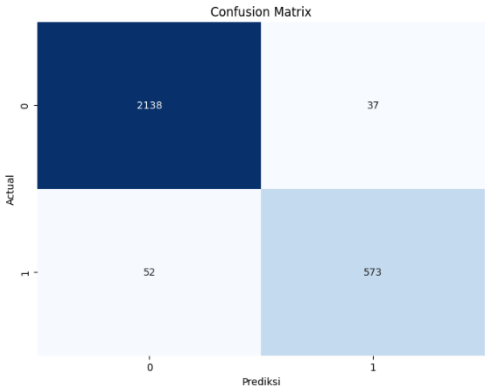
Sigmoid:

$$K(x\_i\ \ [\![, x]\!]\ \_d\ ) = \ tan\ \gamma\ Xi\ T\ Xj\ + r \quad \textbf{(8)}$$

Pada sistem klasifikasi ini akan diterapkan kernel *Radial Basis Function* (RBF) dikarenakan kernel fungsi basis radial biasanya digunakan dalam klasifikasi SVM, dapat memetakan ruang dalam dimensi tak terbatas. Pada penelitian ini juga akan menggunakan perpustakaan *scikit-learn* dalam melakukan pelatihan menggunakan *support vector machine*.

3.4. Pengujian

Model yang didapatkan dari tahap pelatihan algoritma SVM, akan dilakukan diuji menggunakan data *testing.* Pengujian dilakukan untuk mengetahui performa model. Pada penelitian ini dilakukan percobaan menggunakan metode kernel *Radial Basis Function* (RBF). Beberapa metrik evaluasi umum yang digunakan dalam evaluasi model klasifikasi meliputi: Akurasi *(Accuracy)*, Presisi *(Precision)*, *Recall (Sensitivitas atau True Positive Rate)*, *F1-Score* dan *Confusion Matrix.* Hasil yang didapatkan pada evaluasi model dengan menggunakan data pengujian di dapat nilai akurasi 96,82%, presisi 94%, recall 92% f1-score 93%.



Gambar 4. Hasil evaluasi *confusion matrix*

4. Penutup

4.1. Kesimpulan

Penelitian ini melakukan pengujian model menggunakan *Support Vector Machine (SVM)* dengan menggunakan kernel RBF *(Radial Basis Function)* dan parameter Gamma untuk melakukan klasifikasi komentar spam pada Instagram. Dari model svm yang sudah dilatih sebelumnya dan dilakukan prediksi, sistem dapat melakukan prediksi terkait komentar spam dan non spam.

Adapun proses evaluasi dan pengujian yang dilakukan dalam penelitian ini didapatkan dari melakukan evaluasi sistem menggunakan *confusion matrix* dengan nilai akurasi 96.82% yang didapatkan, demikian disimpulkan bahwa algoritma *Support Vector Machine* cocok digunakan dalam hal pengklasifikasikan komentar spam pada Instagram.

4.2. Saran

Pada penelitian ini, terdapat keterbatasan dan kekurangan. Kekurangan dan keterbatasan ini bisa dijadikan acuan dan pertimbangan untuk penelitian selanjutnya. Adapun saran yang dihasilkan dalam penelitian ini yaitu sebagai berikut:

1) Untuk *data training* dan *data testing* bisa menggunakan dataset yang berbeda.
2) Diharapkan menggunakan algoritma lain supaya hasil akan dapat dibandingkan untuk mendapatkan hasil yang terbaik.
3) Sistem dapat dikembangkan dengan pengembangan antarmuka untuk dapat digunakan oleh pengguna lain secara mudah.

Daftar Pustaka

[1] A. T. Syam *dkk.*, "Klasifikasi Komentar Spam Pada Instagram Menggunakan Metode Support Vector Machine," *Jurnal Buffer Informatika*, vol. 6, no. 2, hlm. 1–5, 2020, [Daring]. Tersedia pada: https://journal.uniku.ac.id/index.php/buffer

[2] A. Setiyono dan H. F. Pardede, "KLASIFIKASI SMS SPAM MENGGUNAKAN SUPPORT VECTOR MACHINE," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, hlm. 275–280, Sep 2019, doi: 10.33480/pilar.v15i2.693.

[3] I. D. Id, *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*, 1st ed. Gobah Pekanbaru: UR PRESS, 2021.

[4] Sulaeman, N. Suarna, A. Ajiz, A. Bahtiar, dan Fathurrohman, "Perbandingan Kinerja Algoritma Naïve Bayes Dan C.45 Dalam Klasifikasi Spam Email," *KOPERTIP : Jurnal Ilmiah Manajemen Informatika dan Komputer*, vol. 6, no. 1, hlm. 8–14, Jun 2022, doi: 10.32485/kopertip.v6i1.130.

[5] M. H. S. Ajat, "KLASIFIKASI SMS SPAM DENGAN KOMPARASI METODE SVM DAN NAÏVE BAYES," *METHODIKA: Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 1, hlm. 31–34, Mar 2023, doi: 10.46880/mtk.v9i1.1694.

[6] H. Herwanto, N. L. Chusna, dan M. S. Arif, "Klasifikasi SMS Spam Berbahasa Indonesia Menggunakan Algoritma Multinomial Naïve Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 4, hlm. 1316–1325, Okt 2021, doi: 10.30865/mib.v5i4.3119.

[7] O. H. Rahman, G. Abdillah, dan A. Komarudin, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, hlm. 17–23, Feb 2021, doi: 10.29207/resti.v5i1.2700.

[8] D. Irawan, E. B. Perkasa, Y. Yurindra, D. Wahyuningsih, dan E. Helmud, "Perbandingan Klassifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 10, no. 3, hlm. 432–437, Des 2021, doi: 10.32736/sisfokom.v10i3.1302.