KLASIFIKASI SPAM PADA EMAIL MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN DETEKSI ANOMALY

OLEH :

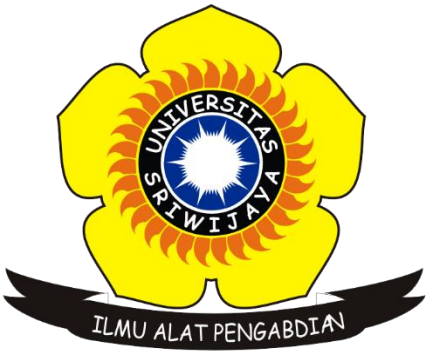
SITI AISYAH
09011181621024

JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2020

KLASIFIKASI SPAM PADA EMAIL MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN DETEKSI ANOMALY

SKRIPSI

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer



OLEH :

SITI AISYAH
09011181621024

JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2020

LEMBAR PENGESAHAN

KLASIFIKASI SPAM PADA EMAIL MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN DETEKSI ANOMALY

SKRIPSI

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh :

SITI AISYAH
09011181621024

Inderalaya, Juli 2020

Mengetahui

Pembimbing I/Tugas Akhir	**Pembimbing II Tugas Akhir**
Deris Stiawan, M.T., Ph.D.	**Huda Ubaya, S.T., M.T.**
NIP. 197806172006041002	**NIP. 198106162012121003**

Ketua Jurusan Sistem Komputer

Dr. Ir. Sukemi. M.T.
NIP. 196612032006041001

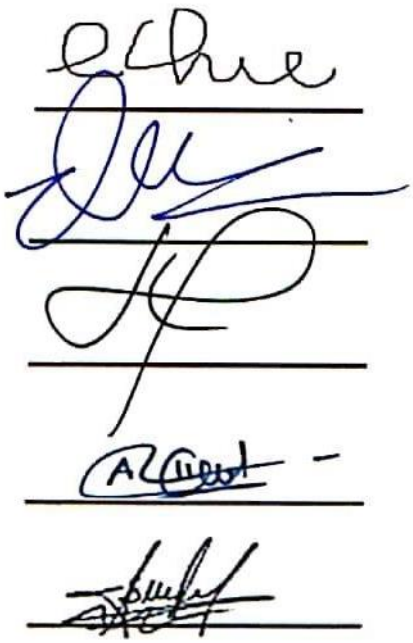
HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Jum'at
Tanggal : 19 Juni 2020

Tim Penguji :

1. **Ketua** : **Sri Desy Siswanti, S.T., M.T.**
2. **Sekretaris I** : **Deris Stiawan, M.T., Ph.D.**
3. **Sekretaris II** : **Huda Ubaya, S.T., M.T.**
4. **Anggota I** : **Ahmad Heryanto, S.Kom., M.T.**
5. **Anggota II** : **Sarmayanta Sembiring, S.SI., M.T.**



Mengetahui,
Ketua Jurusan Sistem Komputer

Dr. Ir. Sukemi. M.T.
NIP. 196612032006041001

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Siti Aisyah

NIM : 09011181621024

Judul : Klasifikasi Spam pada Email Menggunakan Metode *Support Vector Machine* dan Deteksi *Anomaly*

Hasil Pengecekan *Software iThenticate/Turnitin* : 6%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan / plagiat dari penelitian orang lain. Apabila ditemukan unsur penjiplakan / plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan.

Inderalaya, Juli 2020

Yang menyatakan,

Siti Aisyah

HALAMAN PERSEMBAHAN

" Sesungguhnya bersama kesulitan pasti ada kemudahan. Maka apabila engkau telah selesai (dari suatu urusan), tetaplah bekerja keras (untuk urusan yang lain)." (Q.S Al-Insyirah : 6-7)

"Dia mengajarkan manusia apa yang tidak diketahuinya." (Q.S Al-Alaq : 5)

"Niscaya Allah akan mengangkat (derajat) orang-orang yang beriman diantaramu dan orang-orang yang diberi ilmu beberapa derajat." (Q.S Al-Mujadilah : 11)

"Sesungguhnya Allah tidak akan mengubah keadaan suatu kaum sebelum mereka mengubah keadaan diri mereka sendiri." (Q.S Ar-Rad : 11)

Tugas Akhir ini kupersembahkan kepada Ayah dan Ibunda tercinta yang setiap saat selalu mendoakan, memberi semangat serta dukungan. Semoga Allah mengampuni semua dosanya dan memberikan kesahatan serta selalu dalam lindungan-Nya. Adik-adik tersayang yang selalu memberi motivasi. Keluarga besar dan saudara-saudara yang selalu memberikan doa, semangat maupun dukungan. Sahabat seperjuangan dalam susah maupun senang. Serta Keluarga besar Sistem Komputer dan Civitas Akademika Universitas Sriwijaya

KATA PENGANTAR

Puji syukur atas kehadirat Allah SWT, atas segala karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan penulisan Proposal Tugas Akhir ini dengan judul **"Klasifikasi Spam pada *Email* Menggunakan Metode *Support Vector Machine* dan Deteksi *Anomaly*"**.

Penulisan Proposal Tugas Akhir ini dilakukan untuk melengkapi salah satu syarat memperoleh gelar Sarjana Komputer di Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya. Adapun sebagai bahan penulisan, penulis mengambil berdasarkan hasil penelitian, observasi dan beberapa sumber literatur yang mendukung dalam penulisan proposal ini. Pada kesempatan ini juga, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu baik dari segi moril ataupun materil serta memberikan kemudahan, dorongan, saran dan kritik selama dalam proses penulisan Proposal Tugas Akhir ini.

Oleh karena itu, pada kesempatan ini penulis mengucapkan rasa syukur kepada Allah SWT. dan mengucapkan terima kasih kepada yang terhormat :

1. Orang Tua serta keluarga penulis tercinta, yang telah memberikan doa dan restu serta dukungan yang sangat besar selama mengikuti dan melaksanakan perkuliahan di Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya hingga dapat menyelesaikan Proposal Tugas Akhir ini.
2. Bapak Jaidan Jauhari, S.Pd., M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Dr. Ir. H. Sukemi, M.T. selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Ahmad Zarkasi, S.T., M.T., selaku Dosen Pembimbing Akademik di Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Bapak *Deris* Stiawan, M.T., Ph.D., selaku Dosen Pembimbing Satu Tugas Akhir di Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

6. Bapak Huda Ubaya, S.T., M.T., selaku Dosen Pembimbing Dua Tugas Akhir di Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Mba Winda selaku admin Jurusan Sistem Komputer yang telah membantu mengurus seluruh berkas.
8. Seluruh dosen, staff, serta karyawan Fakultas Ilmu Komputer Universitas Sriwijaya.
9. Seluruh teman-teman seperjuangan angkatan 2016 Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
10. Almamater.

Penulis menyadari bahwa masih banyak kekurangan dalam penulisan Proposal Tugas Akhir ini. Karena sesungguhnya tak ada yang sempurna didunia ini. Untuk itu, segala saran dan kritik sangatlah penting bagi penulis. Akhir kata, semoga Tugas Akhir ini dapat bermanfaat dan berguna bagi khalayak.

Palembang, Juli 2020

Penulis

Spam Classification in Email Using the Support Vector Machine Method and Anomaly Detection

Siti Aisyah (09011181621024)
Department of Computer Systems, Faculty of Computer Science,
Sriwijaya University
Email: aaisyahichaa@gmail.com

Abstract

Email is a written communication tool commonly used in everyday life. The problem with e mail is spam. This study includes a machine learning approach, Support Vector Machine, which is used for spam classification on e-mail. Using two datasets, data that has not been vectorized and data that has been vectorized. For data that has not been vectorized, the first step taken is processing the text so that the data becomes numeric. After the data has been vectorized, the next step for these two data is to detect anomalies using Isolation Forest for removal of outliers in the data. The next step will be the data resampling using SMOTE so that the data becomes balanced. Then the last step is classification using the Support Vector Machine method by sharing data using K-Fold Cross Validation and normalizing using Min Max Scaler. In the research the best validation value for the Emails dataset obtained an average accuracy value of 96.80%, Recall 98.70%, Precision 95.12%, F1 Score 96.88%, FPR 5.11%, AUC 96.79%, Error 3.19%. The best validation values for the Spambase dataset obtained an average accuracy value of 94.08%, Recall 92.55%, Precision 95.31%, F1 Score 93.91%, FPR 4.42%, AUC 94.06%, Error 5 91%. Based on the results, it means that the method used in spam classification on e-mail is the right method.

Keywords: Email Spam, Support Vector Machine, Isolation Forest, SMOTE, Classification.

Klasifikasi Spam pada Email Menggunakan Metode Support Vector Machine dan Deteksi Anomaly

Siti Aisyah (09011181621024)
Jurusan Sistem Komputer, Fakultas Ilmu Komputer,
Universitas Sriwijaya
Email: aaisyahichaa@gmail.com

Abstrak

Email merupakan sebuah alat komunikasi tertulis yang biasa digunakan dalam kehidupan sehari-hari. Permasalahan yang ada pada email adalah spam. Penelitian ini memuat metode pendekatan machine learning yaitu Support Vector Machine yang digunakan untuk klasifikasi spam pada email. Menggunakan dua dataset yaitu data yang belum di vektorisasi dan data yang telah di vektorisasi. Untuk data yang belum di vektorisasi, langkah pertama yang dilakukan adalah pengolahan teks agar data menjadi angka. Setelah data selesai di vektorisasi langkah selanjutnya untuk kedua data ini adalah deteksi anomali menggunakan Isolation Forest untuk penghapusan outlier yang ada pada data. Tahap selanjutnya data akan di resampling menggunakan SMOTE agar data menjadi seimbang. Kemudian tahap terakhir adalah klasifikasi menggunakan metode Support Vector Machine dengan pembagian data menggunakan K-Fold Cross Validation dan normalisasi menggunakan Min Max Scaler. Pada penelitian nilai validasi terbaik untuk dataset Emails didapat rata-rata nilai Akurasi 96,80%, Recall 98,70%, Presisi 95,12%, F1 Score 96,88%, FPR 5,11%, AUC 96,79%, Error 3,19% . Nilai validasi terbaik untuk dataset Spambase didapat rata-rata nilai Akurasi 94,08%, Recall 92,55%, Presisi 95,31%, F1 Score 93,91%, FPR 4,42%, AUC 94,06%, Error 5,91%. Berdasarkan hasil yang ada, berarti metode yang digunakan dalam klasifikasi spam pada email merupakan metode yang tepat.

Kata Kunci: Spam Email, Support Vector Machine, Isolation Forest, SMOTE, Klasifikasi.

DAFTAR ISI

Halaman

BAB II TINJAUAN PUSTAKA

BAB III METODOLOGI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1. Latar Belakang

Spam merupakan aktivitas kejahatan yang cukup mudah seperti terlihat pada definisinya yang merupakan sebuah tindakan yang dilakukan berulang-ulang. Dengan artian, pengirim informasi yang dikatakan melakukan *spam* (*spammer*) bisa terjadi secara semgaja dengan mengirimkan *spam* untuk berbuat kejahatan atau pengirim *spam* yang tidak disengaja sehingga tidak mengetahui bahwa diri nya telah melakukan *spam* [1]. Seiring dengan pertumbuhan internet dan *Email*, telah terjadi pertumbuhan *spam* dalam beberapa tahun terakhir. *Spam* dapat berasal dari setiap lokasi di seluruh dunia dimana akses internet tersedia. Jumlah pesan *spam* terus meningkat pesat [2]. Pada saat ini, lebih dari 85% dari total *Email* yang masuk adalah *spam* [3].

Dalam rangka untuk melawan masalah yang berkembang, penentuan teknik terbaik untuk melawan *spam* dengan berbagai alat yang tersedia harus dianalisa oleh organisasi [4]. Dalam *dataset* biasanya terdapat sebuah data yang mempunyai karakteristik berbeda dengan kebanyakan data dan merupakan suatu kejadian yang jarang muncul pada suatu kasus pada sebuah sumber data [5]. Item yang tidak terduga atau tidak normal pada *dataset* ini bisa menurunkan hasil performansi dalam suatu penelitian, oleh sebab itu deteksi anomali adalah proses yang bisa dilakukan untuk mengidentifikasi item tak terduga atau yang berbeda dari normal pada *dataset*.

Deteksi anomali merupakan sebuah proses untuk menemukan pola dalam sebuah *dataset* yang perilakunya tidak normal seperti yang diharapkan. Perilaku tak terduga juga disebut sebagai anomali atau *outlier*. Anomali tidak selalu bisa dikategorikan sebagai serangan tetapi bisa menjadi perilaku mengejutkan yang sebelumnya tidak diketahui. Deteksi anomali memberikan informasi yang sangat signifikan dan penting dalam berbagai aplikasi [6]. Terdapat beberapa algoritma deteksi anomali tanpa pengawasan yang terdiri dari empat kelompok utama

diantaranya *Nearest-neighbor based, Clustering based, Statistical Subspace based* dan *Classifier based/Other*.

Metode SMOTE (*Synthetic Minority Over-sampling Technique*) merupakan metode yang populer untuk menangani data yang tidak seimbang yang merupakan masalah inti dalam deteksi anomali. Teknik ini mensintesis sampel baru dari kelas minoritas untuk menyeimbangkan *dataset* dengan cara sampling ulang sampel kelas minoritas. SMOTE berhasil meningkatkan performa klasifikasi dengan hasil terbaik didapatkan dengan metode Naive Bayes sebesar 90,7% [7].

KNN dengan kombinasi teknik partisi data K-*Fold Cross Validation* dalam mengklasifikasikan *spam* pada *Email* menghasilkan akurasi sebesar 92,28%, metode naive bayes menghasilkan akurasi sebesar 84,30%. Banyak metode yang dilakukan dalam mengklasifikasikan *spam* pada *Email* tetapi yang menggunakan kombinasi teknik partisi data K-*Fold Cross Validation* sedikit sehingga akurasi yang didapatkan tidak terlalu tinggi. SVM akan memberikan hasil ketepatan klasifikasi yang lebih baik jika dikombinasikan dengan teknik partisi data K-*Fold Cross Validation* [8].

SVM adalah model klasifikasi yang berfungsi untuk mesin inferensi dari sistem pakar konstruksi sebuah *hyperplane* keputusan untuk memisahkan dua kelas *spam* dan sah [9]. pengklasifikasi SVM adalah *classifier* menonjol untuk penyaringan *spam* atau klasifikasi [10]. SVM sudah dikenal sebagai algoritma pembelajaran terbaik untuk klasifikasi pada data biner [11]. Keuntungan dari SVM adalah bahwa akurasinya tidak menurunkan bahkan ketika banyak fitur yang hadir. Oleh karena itu, pendekatan tersebut telah diadopsi untuk penyaringan *email spam* [3].

Vanitha, Devaraj, dan Venkatesulu (2015) telah melakukan penelitian klasifikasi data gen *microarray* dengan memperoleh tingkat akurasi tertinggi yaitu 97,77%. Pada penelitian tersebut metode yang digunakan adalah *Support Vector Machine* (SVM) [12]. Itulah yang melatar belakangi penulis mengambil judul "Klasifikasi *Spam* pada *Email* Menggunakan Metode *Support Vector Machine* dan Deteksi *Anomaly*", dikarenakan metode SVM dianggap memperoleh tingkat akurasi yang baik dalam mengklasifikasikan sebuah data.

1.2. Rumusan Masalah

Dataset yang digunakan berbatas pada *spam*. Data pertama menggunakan *dataset* Spambase. Dalam *dataset* Spambase, data sudah berupa angka dan siap diolah. Pada data ini total kasus *Email* adalah 4601. 1813 dari contoh *Email* ini ditandai sebagai *spam* (39,4%) dan sisanya adalah non-*spam*. *Dataset* Spambase terdiri dari 57 fitur dan 1 atribut klasifikasi, yang merupakan label kelas yang menunjukkan status setiap contoh *Email* apakah itu adalah *spam* (1) atau non-*spam* (0). Sebagian besar fitur (1-54) menunjukkan karakter tertentu atau kata-kata yang berulang kali terjadi di *Email* atau tidak. Fitur 55-57 menyajikan pengukuran untuk panjang huruf kapital berturut-turut[13].

Dataset kedua yang digunakan adalah *dataset Emails*. Pada data ini total kasus *Email* adalah 5728. 1368 dari contoh *Email* ini ditandai sebagai *spam* dan 4360 adalah non-*spam*. Label kelas *spam* ditandai dengan angka (1) dan non-*spam* ditandai dengan angka (0). Data *Emails* ini berbentuk teks sehingga apabila dilakukan klasifikasi harus melewati tahap *text mining*.

Adapun ruang lingkup dan perumusan masalah dalam penulisan Proposal Tugas Akhir ini adalah sebagai berikut:

1.2.1. Perumusan Masalah

Rumusan masalah dalam penulisan Proposal Tugas Akhir ini sebagai berikut:

1. Bagaimana cara mengubah data teks menjadi angka menggunakan *count vecorizer*?
2. Bagaimana cara deteksi anomali pada sebuah *dataset* yang digunakan untuk mengklasifikasi *spam* pada *Email* menggunakan metode *Isolation Forest*?
3. Bagaimana cara mengklasifikasikan *spam* pada *Email* menggunakan metode *Support Vector Machine*?
4. Apa saja *output* yang dihasilkan dari klasifikasi *spam* pada *Email* menggunakan metode *Support Vector Machine*?
5. Apa *software* atau *tools* yang digunakan untuk mengklasifikasikan *spam* pada *Email* menggunakan metode *Support Vector Machine*?

1.2.2. Batasan Masalah

Batasan masalah dalam penulisan Proposal Tugas Akhir ini adalah sebagai berikut:

1. Metode klasifikasi yang digunakan adalah *Support Vector Machine* dari toolbox python.
2. Klasifikasi *spam* yang diteliti hanya pada *Email.*
3. Data yang digunakan adalah *dataset* Spambase dan *dataset Emails*.

1.3. Tujuan dan Manfaat

Adapun tujuan dan manfaat dari penulisan Proposal Tugas Akhir ini adalah sebagai berikut :

1.3.1. Tujuan

1. Mempelajari konsep *text mining* yang digunakan untuk mengubah data *text* menjadi angka.
2. Mempelajari konsep algoritma *Isolation Forest* yang digunakan untuk mendeteksi anomali.
3. Mengimplementasikan metode *Support Vector Machine* untuk klasifikasi *spam* dan non-*spam*.
4. Melakukan pegujian dan validasi pada model yang dipilih.
5. Menganalisa *performance* dari metode yang digunakan dalam mendeteksi dan klasifikasi *spam* menggunakan metode *Isolation Forest* dan *Support Vector Machine*.

1.3.2. Manfaat

a. Dapat menerapkan algoritma *text ming* untuk mengolah data *text* menjadi angka.
b. Dapat menerapkan algoritma *Isolation Forest* untuk deteksi anomali.
c. Dapat menerapkan metode *Support Vector Machine* dalam mengklasifikasikan *spam* pada *Email.*
d. Membantu mempermudah penelitian mengenai *spam* pada *Email.*

1.4. Metodologi Penelitian

Dalam penulisan tugas akhir ini Metodologi yang digunakan akan melewati beberapa tahapan sebagai berikut :

1.4.1. Metode Literatur

Pada literatur, penulisan dilakukan dengan cara mencari sumber informasi yang dibutuhkan sebagai media pembelajaran yang diantaranya adalah berasal dari internet seperti artikel yang terkait, jurnal ilmiah, buku yang merupakan peran penting dalam penulisan skripsi ini.

1.4.2. Metode Konsultasi

Pada tahap metode konsultasi, dilakukan tanya jawab kepada orang-orang baik secara online maupun tatap muka. Yang dianggap memiliki pengetahuan dan wawasan yang cukup terhadap permasalahan yang akan dibahas dalam pembuatan Proposal Tugas Akhir.

1.4.3. Metode Pengumpulan Data

Dalam metode ini, pengumpulan data dilakukan dengan mencari sebuah *dataset* yang terdapat di *website* seperti UCI, Kaggle dan *website dataset* lainnya sehingga *dataset* yang dipakai dalam penelitian ini adalah Spambase *dataset* dan *Emails* dataset.

1.4.4. Metode Observasi

Pada metode ini, observasi yang dilakukan adalah dengan mengamati, mencatat, dan menganalisa terhadap data yang diperoleh.

1.4.5. Metode Perancangan *Software*

Pada tahap ini perancangan serta pembuatan sistem (*software*) akan dilakukan agar bisa melakukan penelitian untuk klasifikasi *spam* pada *Email* dengan bahasa pemrograman Python di Jupyter dan di spyder.

1.4.6. Metode Analisa dan Kesimpulan

Hasil dari penelitian yang telah di uji akan dilakukan dianalisa yang bertujuan agar dapat mengetahui letak kekurangan pada perancangan dan apa faktor yang menyebkannya sehingga bisa digunakan untuk mengembangkan penelitian selanjutnya dan dibuat kesimpulan dari hasil penelitian.

1.5. Sistematika Penulisan

Sistematika penulisan pada Proposal Tugas Akhir ini adalah sebagai berikut:

BAB I PENDAHULUAN

Pada bab I akan berisikan latar belakang masalah, tujuan dan manfaat serta metodologi penelitian dan sistemaika penulisan.

BAB II TINJAUAN PUSTAKA

Pada Bab II akan berisi dasar teori *Machine Learning*, *Spam*, *Email*, *Text Mining*, Anomali, *Isolation Forest*, Smote, *Support Vector Machine* dan K-*Fold Cross Validation.*

BAB III METODOLOGI

Pada Bab III berisi uraian mengenai kerangka kerja, perancangan sistem dan penjelasan *flowchart* pada tiap tahap yang dilakukan dipenelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Pada Bab IV membahas proses membahas mengenai proses, hasil dan analisa dari penerapan metode *Isolation Forest* dan *Support Vector Machine* pada data *spam Email* dengan melalui tahap pelatihan dan pengujian.

BAB V KESIMPULAN DAN SARAN

Pada bab V berisi kesimpulan dari bab-bab yang sudah dicantumkan mengenai hasil dari cara mengubah data *text* menjadi angka, pengimplementasian metode *Support Vector Machine* dalam

pengklasifikasian *spam* pada *Email* dan pendeteksian anomali menggunakan algoritma *Isolation Forest*. Pada bab ini juga akan berisi saran yang diharapkan dapat digunakan untuk penelitian selanjutnya.

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR PUSTAKA

[1] E. N. Putra, "Pengiriman E-mail Spam sebagai Kejahatan Cyber di Indonesia," vol. 7, no. 2, pp. 169–182, 2016.

[2] V. Christina, S. Karpagavalli, and G. Suganya, "Email Spam Filtering using Supervised Machine Learning Techniques," vol. 02, no. 09, pp. 3126–3129, 2010.

[3] I. Santos, C. Laorden, X. Ugarte-pedrero, and B. Sanz, "Anomaly Based Spam Filtering," no. january, pp. 124–133, 2011.

[4] S. K. Tuteja and N. Bogiri, "Email Spam Filtering using BPNN Classification Algorithm," pp. 915–919, 2016.

[5] N. Cholis, "Pencarian Anomali Berdasarkan Kerapatan Data Menggunakan Algoritma Shared Nearest Neighbor Density Anomaly Detection Based on Data Density Using Shared Nearest Neighbors Density Algorithm," vol. 1, 2006.

[6] S. Agrawal and J. Agrawal, "Survey on Anomaly Detection using Data Mining Techniques," *Procedia - Procedia Comput. Sci.*, vol. 60, pp. 708–713, 2015.

[7] R. Siringoringo, "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma Smote dan K-Nearest Neighbor," vol. 3, no. 1, pp. 44–49, 2018.

[8] S. Novelia, D. Pratiwi, B. Sutijo, and S. Ulama, "Klasifikasi Email Spam dengan Menggunakan Metode Support Vector Machine dan k- Nearest Neighbor," vol. 5, no. 2, pp. 344–349, 2016.

[9] G. Sanghani and K. Kotecha, "Incremental personalized E-mail spam filter using novel TFDCR feature selection with dynamic feature update," *Expert Syst. Appl.*, vol. 115, pp. 287–299, 2019.

[10] J. Ara, "A Survey of Existing E-Mail Spam Filtering Methods Considering Machine Learning Techniques," vol. 1, no. 2, 2018.

[11] T. Shon, "A hybrid machine learning approach to network anomaly detection," vol. 177, pp. 3799–3821, 2007.

[12] W. Agustina, M. T. Furqon, and B. Rahayudi, "Implementasi Metode

Support Vector Machine (SVM) Untuk Klasifikasi Rumah Layak Huni (Studi Kasus : Desa Kidal Kecamatan Tumpang Kabupaten Malang),” vol. 2, no. 10, pp. 3366–3372, 2018.

[13] L. M. El Bakrawy, “Hybrid Particle Swarm Optimization and Pegasos Algorithm for Spam Email Detection,” pp. 11–22, 2019.

[14] R. Saravanan and P. Sujatha, “A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification,” *2018 Second Int. Conf. Intell. Comput. Control Syst.*, no. Iciccs, pp. 945–949, 2018.

[15] J. R. Méndez, T. R. Cotos-yañez, and D. Ruano-ordás, “A new semantic-based feature selection method for spam filtering,” *Appl. Soft Comput. J.*, vol. 76, pp. 89–104, 2019.

[16] C. Varol and H. M. T. Abdulhadi, “Comparision of String Matching Algorithms on Spam Email Detection,” *2018 Int. Congr. Big Data, Deep Learn. Fight. Cyber Terror.*, pp. 6–11, 2018.

[17] B. Sarrafzadeh, A. H. Awadallah, C. H. Lin, C. Lee, M. Shokouhi, and S. T. Dumais, “Characterizing and Predicting Email Deferral Behavior,” pp. 627–635, 2019.

[18] A. Kajian, K. Puspiptek, K. P. Serpong, and K. Kunci, “Media Komunikasi Efektif pada Layanan Jasa Informasi : Studi Kasus di Kawasan Pusat Penelitian Ilmu Pengetahuan dan Teknologi (Puspiptek),” vol. 26, no. 2, pp. 109–117, 2018.

[19] N. F. Shah and P. Kumar, “A Comparative Analysis of Various Spam Classification,” pp. 265–271, 2018.

[20] P. Soepomo, “Penerepan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes,” *Jural Sarj. Tek. Inform.*, vol. 2, pp. 73–83, 2014.

[21] M. Luo, K. Wang, Z. Cai, A. Liu, and Y. Li, “Using Imbalanced Triangle Synthetic Data for Machine Learning Anomaly Detection,” vol. 58, no. 1, pp. 15–26, 2019.

[22] D. Xu, Y. Wang, Y. Meng, and Z. Zhang, “An improved data anomaly detection method based on isolation forest,” *Proc. - 2017 10th Int. Symp.*

Comput. Intell. Des. Isc. 2017, vol. 2, pp. 287–291, 2018.

[23] G. A. Sandag, R. J. Sambur, and J. Bororing, "Klasifikasi SMS Spam Menggunakan Algoritma Support Vector Machine (SVM)," pp. 291–295, 2018.

[24] S. O. Olatunji, "Improved email spam detection model based on support vector machines," *Neural Comput. Appl.*, vol. 31, pp. 691–699, 2017.

[25] J. M. Sma, "Analisa Perbandingan Tingkat Performansi Metode Support Vector Machine dan Naive Bayes Classifier untuk Klasifikasi Jalur Minat SMA," no. March, 2018.

[26] S. Nurhayati, E. T. Luthfi, and U. Y. Papua, "Prediksi Mahasiswa Drop Out Menggunakan Metode Support Vector Machine," vol. x, no. x, pp. 82–93, 1978.

[27] M. Rangga, A. Nasution, and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," vol. 6, no. 2, pp. 226–235, 2019.

[28] K. Kotipalli, N. Carolina, and S. Suthaharan, "Modeling of Class Imbalance using an Empirical Approach with Spambase Dataset and Random Forest Classification," pp. 75–80, 2014.

[29] H. Adil, A. Algafore, and S. H. Hashem, "Spam Filtering based on Naïve Bayesian with Information Gain and Ant Colony System," vol. 57, no. 1, pp. 719–727, 2016.