

**KLASIFIKASI SPAM EMAIL MENGGUNAKAN
ALGORITMA *K-NEAREST NEIGHBOR* DAN *CHI SQUARE***



OLEH

**ULVIYANA
09011281520090**

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2020**

**KLASIFIKASI SPAM EMAIL MENGGUNAKAN
ALGORITMA *PRINCIPAL COMPONENT ANALYSIS*
(PCA) DAN *DECISION TREE***

TUGAS AKHIR

**Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer**



OLEH

SITI PEBSYA ROISATUN SHOLIHAH

09011281520102

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA**

2020

HALAMAN PENGESAHAN

KLASIFIKASI SPAM EMAIL MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR* DAN *CHI SQUARE*

PROPOSAL TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

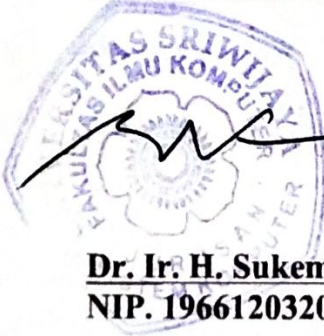
Oleh :

ULVIYANA
09011281520090

Mengetahui,
Ketua Jurusan Sistem Komputer

Indralaya, Agustus 2020

Pembimbing



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

A handwritten signature in black ink, which appears to be 'Deris Stiawan', written over a horizontal line.

Deris Stiawan, M.T., PH.D.
NIP. 197806172006041002

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Senin

Tanggal : 10 Agustus 2020


Tim Penguji :

1. Ketua : Rahmat Fadli Isnanto, M.Sc

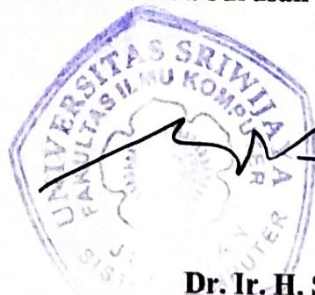
2. Sekretaris : Deris Stiawan, M.T., Ph.D.

3. Anggota I : Dr. Reza Firsandaya Malik, M.T.

4. Anggota II : Aditya Putra Perdana P, M.T.



Mengetahui,
Ketua Jurusan Sistem Komputer



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

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Ulviyana

NIM : 09011281520090

Judul : KLASIFIKASI SPAM EMAIL MENGGUNAKAN ALGORITMA
K-NEAREST NEIGHBOR DAN CHI SQUARE

Hasil Pengecekan *Software iThenticate/Turnitin* : 19%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya saya sendiri dan bukan hasil penjiplakan / plagiat. Apabila ditemukan unsur penjiplakan / plagiat dalam laporan ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

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



Palembang, Agustus 2020



Ulviyana
NIM. 09011281520090

HALAMAN PERSEMBAHAN

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya.”

(Q.S. Al-Baqarah: 286)

“Maka sesungguhnya bersama kesulitan itu ada kemudahan”

(Q.S. Al-Insyirah: 5)

Tugas akhir ini saya persembahkan khusus untuk kedua orang tua ku tercinta. Bunda Ayah, terima kasih banyak karena tak henti-hentinya mendoakan yang terbaik untukku. Terima kasih selalu memberikanku semangat dan nasehat. Teruntuk Abang, panutanku, terima kasih banyak telah menjadi contoh yang baik dan selalu ada untuk mendukungku. Untuk Adik-adikku tersayang, terimakasih telah memberikanku banyak kebahagiaan. Untuk seseorang yang selalu berada disisiku, terima kasih banyak telah menemani, menyemangati, dan membantuku setiap waktu. Dan terakhir terima kasih banyak untuk semua teman-temanku yang telah banyak memberikan bantuan selama perkuliahan ini.

KATA PENGANTAR



Puji dan syukur penulis panjatkan kehadirat Allah SWT, karena berkat karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan proposal tugas akhir ini dengan judul “**Klasifikasi Spam Email Menggunakan Algoritma K-Nearest Neighbor dan Chi Square**”.

Pada penyusunan proposal tugas akhir ini, tidak terlepas dari bantuan, bimbingan serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan rasa syukur dan terimakasih kepada yang terhormat :

1. Allah SWT yang selalu melindungi, membimbing dan memberi kemudahan dalam pelaksanaan pembuatan Tugas Akhir ini.
2. Bunda, Ayah, Abang dan Adik-adik ku tercinta beserta keluarga yang selalu mendoakan, memberi nasehat, serta memberi dukungan.
3. Bapak Jaidan Jauhari, S.Pd., M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
4. Bapak Dr. Ir. H. Sukemi, M.T. selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya
5. Bapak Deris Stiawan, M.T., PH.D. selaku Pembimbing Tugas Akhir Penulis.
6. Bapak Sutarno, S.T., M.T. selaku Pembimbing Akademik Jurusan Sistem Komputer
7. Sahabat-sahabatku Siti Pebsya Roisatun Sholihah, Arfattustary Noorfizir, Dyah Citra Soraya dan Ria Siti Juairiah .
8. Andry Leo Pratama terima kasih telah membantu serta memberikan motivasi serta semangat.
9. Teman-teman seperjuangan angkatan 2015 dan kelas SK15C khususnya yang selalu kebersamai selama perkuliahan dan teman-teman yang telah membantu saya dalam menyelesaikan tugas akhir ini.

10. Almameter Universitas Sriwijaya yang telah memberi kesempatan dan fasilitas selama saya menempuh pendidikan Strata 1 disini.

Penulis menyadari bahwa Laporan ini masih jauh dari kesempurnaan, oleh karena itu penulis sangat mengharapkan kritik dan saran yang bersifat membangun agar lebih baik lagi dikemudian hari.

Akhir kata dengan segala keterbatasan, penulis berharap semoga laporan ini menghasilkan sesuatu yang bermanfaat bagi kita semua khususnya bagi mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya secara langsung ataupun tidak langsung sebagai sumbangan pikiran dalam peningkatan mutu pembelajaran.

Palembang, Agustus 2020

Penulis

EMAIL SPAM CLASSIFICATION USING K-NEAREST NEIGHBOR AND CHI SQUARE ALGORITHM

Ulviyana (09011281520090)

*Departementt of Computer Engineering, Faculty of Computer Science,
Sriwijaya University
Email: ulviyana96@gmail.com*

ABSTRACT

Email spam classification on the Spambase dataset obtained from the UCI Machine Learning Repository using the K-Nearest Neighbor (KNN) algorithm during the classification process of spam and non spam data. This method classifies objects based on their nearest neighbor distance (k). Where the value of k can be determined using distance measurement or the euclidean function. In this study, pre-processing was carried out by dividing the data into training and testing data, normalizing the data, and selecting features by applying the Chi Square algorithm. The results of the accuracy of the classification process using the K-Nearest Neighbor algorithm before applying the Chi Square feature selection when the $k = 1$ value is 89.58%. While the accuracy value when classification uses the K-Nearest Neighbor algorithm after applying the Chi Square algorithm as a feature selection is equal to 90.45%.

Keywords : *Spam Classification, K-Nearest Neighbor, Feature Selection, Chi Square*

KLASIFIKASI SPAM EMAIL MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR* DAN *CHI SQUARE*

Ulviyana (09011281520090)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email: ulviyana96@gmail.com

ABSTRAK

Klasifikasi spam email pada dataset *Spambase* yang diperoleh dari *UCI Machine Learning Repository* dengan menggunakan algoritma *K-Nearest Neighbor* (KNN) saat proses pengklasifikasian data spam dan non spam. Metode ini mengklasifikasikan objek berdasarkan jarak tetangga terdekat (k). Dimana nilai k dapat ditentukan menggunakan pengukuran jarak atau fungsi *euclidean*. Pada penelitian ini dilakukan *pre-processing* dengan melakukan pembagian data menjadi data latih dan data uji, melakukan normalisasi data, dan melakukan seleksi fitur dengan menerapkan algoritma *Chi Square*. Hasil akurasi dari proses klasifikasi menggunakan Algoritma *K-Nearest Neighbor* sebelum diterapkan seleksi fitur *Chi Square* saat nilai $k=1$ yaitu sebesar 89,58%. Sedangkan nilai akurasi saat proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* setelah diterapkan algoritma *Chi Square* sebagai seleksi fitur yaitu sebesar 90,45%.

Kata Kunci : Klasifikasi Spam, *K-Nearest Neighbor*, Seleksi Fitur, *Chi Square*.

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PERNYATAAN.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR	vi
<i>ABSTRACT</i>	vii
ABSTRAK	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xvi
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan dan Batasan Masalah	3
1.3. Tujuan dan Manfaat	3
1.4. Metodologi Penelitian.....	4
1.5. Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Penelitian Terdahulu	7
2.2 <i>Machine Learning</i>	9
2.3 Spam	10
2.4 Spam Email	11
2.5 <i>Feature Selection</i>	12
2.6 <i>Chi Square</i>	13
2.7 <i>K-Nearest Neighbor</i>	13
2.8 <i>Dataset</i>	15
2.10 Performa <i>K-Nearest Neighbor</i>	21

BAB III METODOLOGI	23
3.1 Pendahuluan	23
3.2 Kerangka Kerja	23
3.3 Perancangan Sistem.....	25
3.4 <i>Dataset</i>	26
3.5 <i>Pre-processing</i>	28
3.5.1 <i>Split Data</i>	28
3.5.2 Normalisasi.....	29
3.5.3 <i>Chi Square Feature Selection</i>	30
3.6 <i>Processing</i>	32
3.6.1 Klasifikasi.....	32
BAB IV HASIL DAN ANALISA	35
4.1 Pendahuluan	35
4.2 <i>Pre-Processing</i>	35
4.2.1 <i>Dataset</i>	35
4.2.2 <i>Split Data</i>	36
4.2.3 Normalisasi.....	38
4.2.4 <i>Chi Square Feature Selection</i>	41
4.2.4.1 Seleksi Fitur Persentase 25%	42
4.2.4.1 Seleksi Fitur Persentase 50%	43
4.2.4.1 Seleksi Fitur Persentase 75%	43
4.2.4.1 Seleksi Fitur Persentase 100%	44
4.2.5 Analisis Pengujian Akurasi	44
4.3 <i>Processing</i>	45
4.4.1 Klasifikasi.....	45
4.4 Performansi dan Analisa	47
4.4.1 Analisa Perhitungan <i>Confusion Matrix</i>	47
4.4.2 Analisa <i>K-Nearest Neighbor Behavior</i>	49
4.4.3 Analisa <i>K-Nearest Neighbor with Chi Square Behavior</i>	52
4.5 <i>Detection Result</i>	54
BAB V KESIMPULAN DAN SARAN.....	56
5.1 Kesimpulan	56

5.2 Saran	57
DAFTAR PUSTAKA	58

DAFTAR GAMBAR

Gambar 3.1 Kerangka Kerja Penelitian	24
Gambar 3.2 Perancangan Sistem	26
Gambar 3.3 Algoritma Persiapan <i>Dataset</i>	27
Gambar 3.4 Algoritma <i>Split Data</i>	29
Gambar 3.5 Algoritma Normalisasi.....	30
<i>Gambar 3.6 Algoritma Chi Square Feature Selection</i>	32
Gambar 4.1 Bentuk Dataset	35
Gambar 4.2 <i>Data Training</i>	37
Gambar 4.3 <i>Data Testing</i>	38
Gambar 4.4 Data Sebelum Normalisasi	39
Gambar 4.5 Data Setelah Normalisasi	39
Gambar 4.6 Grafik Nilai Akurasi	44
Gambar 4.7 Grafik Performansi <i>K-Nearest Neighbor</i>	51
Gambar 4.8 Grafik Performansi <i>K-Nearest Neighbor with Chi Square</i>	54
Gambar 4.9 Grafik Perbandingan Performansi <i>KNN</i> dan <i>KNN- Chi Square</i>	55

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	7
Tabel 2.2 Fitur pada Dataset	15
Tabel 2.3 Statistik Atribut	18
Tabel 2.4 <i>Confusion Matrix</i>	21
Tabel 4.1 Performansi Nilai Rata-Rata Sebelum dan Sesudah Normalisasi Data	40
Tabel 4.2 Fitur Data Latih 25%	42
Tabel 4.3 Fitur Data Latih 50%	43
Tabel 4.4 Fitur Data Latih 75%	43
Tabel 4.5 Fitur Data Latih 100%	44
Tabel 4.6 Nilai <i>Confusion Matrix</i>	47
Tabel 4.7 Perhitungan <i>Confusion Matrix</i> pada saat nilai $k=1$	48
Tabel 4.8 Performansi <i>K-Nearest Neighbor</i>	50
Tabel 4.9 Performansi <i>K-Nearest Neighbor with Chi Square</i>	52
Tabel 4.10 Perbandingan Performansi <i>KNN</i> dan <i>KNN-Chi Square</i>	54

DAFTAR LAMPIRAN

LAMPIRAN 1. <i>Code</i> untuk Klasifikasi Tanpa <i>Chi Square</i> pada Dataset <i>Spambase</i>	61
LAMPIRAN 2. <i>Code</i> untuk Klasifikasi dengan <i>Chi Square</i> pada Dataset <i>Spambase</i>	67

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Email adalah sarana alat pengiriman informasi yang digunakan diseluruh bagian dunia, yang dapat mengirim informasi dengan sangat cepat dan hemat biaya serta dapat digunakan melalui PC, *smartphone*, dan gadget elektronik terbaru lainnya. Email digunakan oleh hampir semua manusia, diperkirakan pada akhir 2016, akan ada lebih dari 2,6 miliar pemilik akun email diseluruh dunia dan diperkirakan bahwa hampir setengah dari populasi dunia akan menggunakan email pada akhir 2020. Dengan meingkatnya penggunaan email sebagai alat komunikasi ini telah menyebabkan peningkatan jumlah email spam secara global [1]. Spam merupakan suatu informasi yang disampaikan atau didistribusikan kepada sejumlah besar penerima tanpa memberi tahu penerima tersebut. Adapun spam dapat dikategorikan ke dalam beberapa kelas seperti *WebSpam*, *Voice Over Internet Protocol (VOIP) Spam*, *SMS Spam*, *Social Media Spam*, *Instant Messaging (IM) Spam*, dan *Usenet Newsgroup Spam* akan tetapi *Email Spam* adalah bentuk spam yang paling dikenal dan juga yang menjadi fokus pada penelitian ini [2].

Spam email merupakan suatu masalah yang terus dihadapi pengguna email yang mana hal tersebut dapat mengganggu dan menghabiskan waktu para pengguna email. Pengklasifikasian spam dengan berbagai metode masih banyak dilakukan. Sejauh ini beberapa metode klasifikasi telah diterapkan seperti *K-Nearest Neighbor (KNN)*. KNN adalah sebuah algoritma data pembelajaran yang sangat sederhana.

Untuk tugas pengklasifikasian metode KNN ini dapat mengklasifikasikan objek berdasarkan jarak tetangga terdekat (*nearest neighbor*). Dimana nilai k terdekat dapat ditentukan menggunakan beberapa pengukuran jarak atau fungsi seperti *euclidean*, *minkowsky* atau *mini-max*. Algoritma KNN ini memiliki konsistensi yang tinggi [3].

Pada penelitian ini penulis menggunakan *Spambase dataset* yang diperoleh dari *UCI Machine Learning Repository*. Dalam *dataset* tersebut terdapat 4601 jumlah data, dimana jumlah data spam sebanyak 1813 dan data non-spam sebanyak 2788. Pada penelitian sebelumnya [4], telah dilakukan penelitian menggunakan *dataset* tersebut menggunakan metode *Decision Tree* dan *Logistic Regresion*. Dari penelitian tersebut menghasilkan nilai akurasi sebesar 91,67%. Selain menggunakan metode *Decision Tree*, metode klasifikasi lainnya yang dapat diterapkan adalah metode *K-Nearest Neighbor* (KNN). Sebelumnya [3], Metode KNN telah diterapkan pada penelitian mengenai klasifikasi spam email. Pada penelitian tersebut KNN menghasilkan nilai akurasi yang menjanjikan yaitu sebesar 92,28%.

Jika pada penelitian sebelumnya menggunakan metode *Decision Tree* dengan menerapkan algoritma *Logistic Regresion* yang dapat meningkatkan kinerja dari *Decision Tree* dengan mengurangi *noisy data*, maka pada penelitian ini penulis menggunakan metode KNN dengan menerapkan algoritma *Chi Square*, yang mana pada penelitian sebelumnya [5], *Chi Square* telah berhasil diterapkan pada metode *K-Nearest Neighbor*. Hasil dari penelitian ini menyatakan bahwa metode KNN dengan seleksi fitur *Chi Square* menghasilkan nilai yang lebih efisien dalam proses perhitungan dibandingkan metode KNN yang tanpa menerapkan

seleksi fitur. Algoritma *Chi Square* merupakan salah satu *feature selection*. Seleksi fitur atau *feature selection* adalah salah satu proses yang dapat menghilangkan beberapa fitur yang kurang relevan untuk penentuan topik suatu dokumen. *Chi Square* merupakan *supervised* seleksi fitur yang dapat menyeleksi jumlah fitur tanpa mengurangi nilai akurasi [6].

1.2. Rumusan dan Batasan Masalah

Dari pemaparan latar belakang diatas, maka rumusan masalah yang akan diambil adalah bagaimana cara mengklasifikasikan data spam dari *Spambase dataset* dengan menerapkan algoritma *K-Nearest Neighbor* serta cara untuk menyeleksi fitur pada data spam tersebut dengan menggunakan *Chi Square feature selection*. Adapun batasan masalah pada tugas akhir ini yaitu nilai yang diukur antara lain *Accuracy, Precision, Recall, F-Measure dan Error*. Untuk *dataset* yang digunakan hanya terbatas pada data spam *email* pada *Spambase dataset* yang diambil dari *UCI Machine Learning Repository*.

1.3. Tujuan dan Manfaat

Adapun tujuan dan manfaat dari penelitian ini adalah:

1. Tujuan
 - a. Menerapkan algoritma *Chi Square* sebagai seleksi fitur untuk mengurangi jumlah fitur pada dataset yang digunakan.
 - b. Melakukan klasifikasi data spam dan non-spam dengan menerapkan algoritma *K-Nearest Neighbor*.

- c. Menganalisa serta membandingkan performansi hasil dari proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* tanpa menerapkan *Chi Square* dan algoritma *K-Nearest Neighbor* setelah menerapkan *Chi Square* sebagai seleksi fitur.

2. Manfaat

- a. Dapat mempelajari proses dalam klasifikasi spam email
- b. Dapat memperoleh tingkat akurasi dari klasifikasi spam email menggunakan algoritma *K-Nearest Neighbor* dan *Chi Square*
- c. Dapat menjadi referensi untuk penelitian selanjutnya.

1.4. Metodologi Penelitian

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

1. Studi Pustaka

Pada tahap ini ialah tahap mencari referensi atau *literature* pada *keyword* yang di angkat dari judul yang bertujuan untuk menunjang penelitian yang akan dilakukan.

2. Konsultasi

Pada tahap ini peneliti melakukan konsultasi kepada orang-orang yang dianggap memiliki pengetahuan dan wawasan terhadap permasalahan yang ditemui saat pembuatan tugas akhir.

3. Pengumpulan Data

Pada tahap ini data yang diperoleh adalah *Spambase Dataset* dari UCI *Machine Learning Repository*.

4. Pengolahan Data

Pada tahap ini dilakukan pengolahan data dengan menerapkan algoritma *K-Nearest Neighbor* untuk tahapan klasifikasi data dan *Chi Square* sebagai *feature selection*.

5. Analisis

Pada tahap ini dilakukan pengambilan data dan menganalisa data yang telah dilakukan pengolahan.

6. Kesimpulan dan Saran

Pada tahap ini dilakukan dengan menarik kesimpulan dari analisa serta memberikan saran untuk menjadi bahan referensi bagi penelitian selanjutnya.

1.5. Sistematika Penulisan

Adapun sistematis penulisan laporan sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini dikemukakan latar belakang, tujuan penelitian, batasan masalah, metodologi penelitian dan sistematika penulisan yang akan digunakan di laporan ini.

BAB II TINJAUAN PUSTAKA

Bab ini berisikan uraian singkat tentang teori-teori yang berkaitan dengan penelitian ini, meliputi pembelajaran mesin, spam, *email*, *feature selection*, *Chi Square* dan *K-Nearest Neighbor*.

BAB III METODOLOGI

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

BAB IV ANALISA DAN PEMBAHASAN

Bab ini membahas mengenai proses, hasil dan analisa dari penerapan metode *K-Nearest Neighbor* pada *Spambase dataset* dengan melalui tahap pelatihan dan pengujian.

BAB V KESIMPULAN DAN TINDAK LANJUT

Pada bab ini dikemukakan kesimpulan dari uraian pada bab-bab sebelumnya dan memberikan tindak lanjut untuk penelitian selanjutnya.

DAFTAR PUSTAKA

- [1] M. Abdulhamid, M. Shuaib, and O. Osho, “Comparative Analysis of Classification Algorithms for Email Spam Detection,” no. January, 2018.
- [2] W. Z. Khan, M. K. Khan, F. Bin Muhaya, and M. Y. Aalsalem, “A Comprehensive Study of Email Spam Botnet Detection,” no. c, 2015.
- [3] S. Novelia, D. Pratiwi, B. Sutijo, and S. Ulama, “Klasifikasi Email Spam dengan Menggunakan Metode Support Vector Machine dan k- Nearest,” vol. 5, no. 2, pp. 344–349, 2016.
- [4] A. Wijaya, “Hybrid Decision Tree and Logistic Regression Classifier for Email Spam Detection,” pp. 5–8, 2016.
- [5] Y. D. Setiyaningrum, “Classification of Twitter Contents using Chi-Square and K-Nearest Neighbour Algorithm,” *2019 Int. Semin. Appl. Technol. Inf. Commun.*, pp. 1–4, 2019.
- [6] K. Spam, E. Menggunakan, C. Algoritma, P. Teknik, I. Universitas, and D. Nuswantoro, “Klasifikasi spam email menggunakan algoritma c4.5 dengan seleksi fitur,” vol. 10, no. April, pp. 19–30, 2014.
- [7] S. Singh and S. Kaur, “IMPROVED SPAMBASE DATASET PREDICTION USING SVM RBF KERNEL WITH ADAPTIVE BOOST,” pp. 2319–2322, 2015.
- [8] H. Kaur and A. Sharma, “Improved Email Spam Classification Method Using Integrated Particle Swarm Optimization and Decision Tree,” no. October, pp. 516–521, 2016.

- [9] H. Aliady *et al.*, “IMPLEMENTASI SUPPORT VECTOR MACHINE (SVM) DAN RANDOM FOREST,” vol. 2018, no. Sentika, pp. 23–24, 2018.
- [10] M. I. Halim, A. Dwi, A. N. S. T, and C. S. S. T, “Analisis Sentimen Menggunakan Metode Learning Vector Quantization Sentiment Analysis Using Learning Vector Quantization Method,” vol. 4, no. 2, pp. 2283–2292, 2017.
- [11] A. Subasi, S. Alzahrani, A. Aljuhani, and M. Aljedani, “Comparison of Decision Tree Algorithms for Spam E-mail Filtering,” *2018 1st Int. Conf. Comput. Appl. Inf. Secur.*, pp. 1–5, 2018.
- [12] T. Vyas, “A Survey and Evaluation of Supervised Machine Learning Techniques for Spam E-Mail Filtering.”
- [13] W. Li and W. Meng, “An Empirical Study on Email Classification Using Supervised Machine Learning in Real Environments,” pp. 7438–7443, 2015.
- [14] P. Tang, X. Tang, Z. Tao, J. Li, and C. Technology, “RESEARCH ON FEATURE SELECTION ALGORITHM BASED ON MUTUAL INFORMATION AND GENETIC ALGORITHM,” vol. 1, no. 1, pp. 403–406.
- [15] A. M. Al-zoubi, J. Alqatawna, and H. Faris, “Spam Profile Detection in Social Networks Based on Public Features,” no. May, 2017.
- [16] A. A. Naem, N. I. Ghali, and A. A. Saleh, “ScienceDirect Antlion optimization and boosting classifier for spam email detection,” *Futur. Comput. Informatics J.*, no. xxxx, 2018.
- [17] N. Rachburee, “A Comparison of Feature Selection Approach Between Greedy , IG-ratio , Chi-square , and mRMR in Educational Mining,” pp.

420–424, 2015.

- [18] E. L. S, “Klasifikasi Malware Trojan Ransomware Dengan Algoritma Support Vector Machine (SVM),” vol. 2, no. 1, pp. 137–142, 2016.
- [19] D. Z. Nathania and F. A. Bachtiar, “Klasifikasi Spam Pada Twitter Menggunakan Metode Improved K-Nearest Neighbor,” vol. 2, no. 10, pp. 3948–3956, 2018.