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

**OLEH**

**SITI PEBSYA ROISATUN SHOLIHAH**

**09011281520102**

**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**

# LEMBAR PENGESAHAN

## 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**

**Indralaya, Agustus 2020**

| **Mengetahui,** | **Pembimbing** |
|---|---|
| **Ketua Jurusan Sistem Komputer** | |
| **Dr. Ir. H. Sukemi, M.T** | **Deris Stiawan, M.T., Ph.D.** |
| **NIP. 196612032006041001** | **NIP. 197806172006041002** |

# HALAMAN PERSETUJUAN

**Telah diuji dan lulus pada :**

**Hari** : **Senin**

**Tanggal** : **10 Agustus 2020**

**Tim Penguji :**

**1. Ketua** : **Ahmad Heryanto, S.Kom., M.T.**

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

**3. Anggota I** : **Ahmad Fali Oklilas, M.T.**

**4. Anggota II** : **Rahmat Fadli Isnanto, M.Sc.**

**Mengetahui,**
**Ketua Jurusan Sistem Komputer**

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

# HALAMAN PERNYATAAN

**Yang bertanda tangan dibawah ini :**

**Nama** : **Siti Pebsya Roisatun Sholihah**

**NIM** : **09011281520102**

**Judul** : **Klasifikasi Spam Email Menggunakan Algoritma *Principal Component Analysis* (PCA) dan *Decision Tree***

**Hasil Pengecekan *Software iThenticate/Turnitin* : 9%**

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**


**Siti Pebsya Roisatun Sholihah**
**NIM. 09011281520102**



iv

# HALAMAN PERSEMBAHAN

***"Boleh jadi kamu membenci sesuatu, padahal ia amat baik bagimu, dan boleh jadi (pula) kamu menyukai sesuatu, padahal ia amat buruk bagimu; Allah mengetahui, sedang kamu tidak mengetahui."***
***(Q.S. Al-Baqarah: 216)***

***"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)***

***"Dan apabila hamba-hamba-Ku bertanya kepadamu (Muhammad) tentang Aku, maka (jawablah), bahwasanya Aku adalah dekat."***
***(Q.S. Al-Baqarah: 186)***

***Kupersembahkan khusus untuk Mama dan Papa. Ma, Pa, terima kasih banyak karena telah dengan sabar selalu mendoakanku, menyemangatiku, menasehatiku, dan terus mencintaiku.Untuk keempat saudariku, terima kasih banyak sudah menemaniku, mengingatkanku, dan terus menyayangiku. Dan untuk adik bungsu kami yang paling tampan, terima kasih sudah menghibur dikala sedih dan letih. Tak akan pernah bosan untuk mengatakan bahwa aku sangat berterima kasih dan sangat mencintai kalian.***

# KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT, atas segala karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan Proposal Tugas Akhir ini dengan judul "**Klasifikasi Spam Email Menggunakan Algoritma *Principal Component Analysis* (PCA) Dan *Decision Tree***".

Dalam laporan ini penulis menjelaskan mengenai Penerapan Algoritma *Principal Component Analysis* dan *Decision Tree* pada klasifikassi spam *e-mail*. Penulis berharap tulisan ini dapat bermanfaat bagi orang banyak, dan menjadi tambahan bahan bacaan bagi yang tertarik meneliti tentang Spam Email serta penerapan *feature selection* dan klasifikasi spam dan non-spam.

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 terima kasih kepada yang terhormat :

1. Allah SWT, yang telah melimpahkan kesehatan, kesempatan, kekuatan, serta kemudahan dalam penulisan dan penyusunan Tugas Akhir ini.
2. Mama, Papa, Ayuk-Ayuk, serta Adikku tercinta yang telah memberikan nasehat-nasehat serta motivasi kepadaku selama ini. Terima kasih telah memberikan limpahan cinta dengan terus mendukungku baik secara moral, materi, dan juga spiritual.
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 dan Pembimbing Akademik di Jurusan Sistem Komputer. Terima Kasih karena telah meluangkan waktunya untuk membimbing penulis dalam menyelesaikan tugas akhir ini serta telah memberikan bimbingan dan nasehat selama perkuliahan.

6. Seluruh Dosen Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Sahabat-sahabatku tersayang Ulviyana, Ria Siti Juariah, Arfattustary Noorfizir, dan Dyah Citra Soraya. Terima kasih telah menemani selama kurang lebih empat tahun ini, baik itu suka, duka, ceria, canda, tawa, tangis, bahagia.
8. Semua teman-temanku yang banyak berjasa dalam masa perkuliahanku. Terima kasih banyak sudah banyak membantuku selama perkuliahan dan selama mengerjakan tugas akhir ini.
9. Teman-teman seperjuangan angkatan 2015 dan anak-anak eS KaCang khususnya yang selalu membersamai selama perkuliahan ini.
10. Almamater Universitas Sriwijaya yang telah memberikan kesempatan dan fasilitas selama saya menempuh pendidikan Sarjana disini.
11. Serta semua pihak yang telah membantu baik moril maupun material yang tidak dapat disebutkan satu persatu dalam penyelesaian tugas akhir ini. Terima kasih banyak semuanya.

Penulis menyadari bahwa masih terdapat banyak kekurangan dalam penulisan Tugas Akhir ini, baik dari materi maupun teknik penyajiannya, mengingat kurangnya pengetahuan dan pengalaman penulis. Untuk itu, penulis mengharapkan adanya kritik dan saran yang membangun agar dapat memperbaiki kekurangan-kekungan tersebut kedepannya nanti.

Akhir kata dengan segala keterbatasan, penulis berharap semoga penulisan Tugas Akhir ini dapat menjadi tambahan wawasan dan ilmu pengetahuan bagi mahasiswa yang memerlukan khususnya 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 PRINCIPAL COMPONENT ANALYSIS (PCA) AND DECISION TREE ALGORITHM*

**Siti Pebsya Roisatun Sholihah (09011281520102)**
*Departement of Computer Engineering, Faculty of Computer Science, Sriwijaya University*
Email: sitipebsya@gmail.com

## *ABSTRACT*

*Spam email is unsolicited email that is sent en masse to email users. Examples of spam email include advertisements, sweepstakes, false information, phishing, and so on. Classification is a method in data mining that can classify email as spam and non-spam. This research was conducted on two email spam datasets, namely the Spambase dataset obtained from UCI Machine Learning, and the Emails dataset obtained from Kaggle. Spam classification is done using the Decision Tree algorithm. The classification process is carried out after the pre-processing stage, namely by doing text mining (Email dataset only), separating data, scaling data, and applying the Principal Component Analysis (PCA) algorithm as a sign of the number of features in the dataset based on the value that is important the influence of each feature. The results of the classification using the Decision Tree Algorithm are 93.16% for the Spambase dataset and 94.24% for the Emails dataset. Meanwhile, the application of PCA to the Decision Tree resulted in a value of 90% for the Spambase dataset and 89.53% for the Emails dataset.*

***Keyword :*** *Spam Email, Classification, Decision Tree, Principal Component Analysis (PCA)*

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

**Siti Pebsya Roisatun Sholihah (09011281520102)**
Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya
Email: sitipebsya@gmail.com

## ABSTRAK

Spam email adalah email yang tidak diminta dan dikirim secara massal kepada para pengguna email. Contoh spam pada email yaitu berupa iklan, undian, informasi palsu, phishing, dan lain sebagainya. Klasifikasi adalah salah satu metode dalam data mining yang dapat mengklasfikasikan email sabagai spam dan non-spam. Penelitian ini dilakukan pada dua buah dataset spam email yaitu dataset *Spambase* yang diperoleh dari *UCI Machine Learning*, dan dataset *Emails* yang diperoleh dari *Kaggle*. Klasifikasi spam email dilakukan menggunakan algoritma *Decision Tree.* Proses klasifikasi dilakukan setelah tahap *pre-processing* yaitu dengan melakukan *text mining* (hanya dataset *Emails*), *split data*, *scaling data*, dan menerapkan algoritma *Principal Component Analysis* (PCA) sebagai pengurangan jumlah fitur pada dataset berdasarkan nilai seberapa penting pengaruh dari masing-masing fitur. Hasil akurasi dari klasifikasi yang dilakukan menggunakan Algoritma *Decision Tree* yaitu sebesar 93,16% untuk dataset *Spambase* dan 94,24% untuk dataset *Emails.* Sedangkan untuk penerapan PCA pada *Decision Tree* menghasilkan nilai akurasi sebesar 90% untuk dataset *Spambase* dan sebesar 89,53% untuk dataset *Emails*.

**Kata Kunci :** Spam Email, Klasifikasi, *Decision Tree, Principal Component Analysis* (PCA)

# DAFTAR ISI

# DAFTAR GAMBAR

# DAFTAR TABEL

## DAFTAR LAMPIRAN

# BAB I

# PENDAHULUAN

## 1.1 Latar Belakang

Email merupakan salah satu alat untuk komunikasi yang tercepat, termudah, termurah, dan populer saat ini. Email saat ini telah menjadi bagian dalam kehidupan sehari-hari manusia. Sehingga dapat dipercaya apabila dalam setiap harinya terdapat lebih dari 205 miliar email yang dikirim. Menurut survei yang dilakukan oleh Radicati Group, sebuah perusahaan riset berbasis di California, dari 205 miliar email yang dikirim setiap hari, sekitar 18,5% tidak relevan bagi penerima dan 22,8% email yang tidak perlu dikirim. Email spam menyebabkan kerugian sekitar 20 juta dolar per tahun, yang sebagian merupakan akibat dari karyawan yang membuang-buang waktu perusahaan yang kritis untuk membaca dan menghapusnya[1].

Spam dapat dikategorikan ke dalam berbagai kelas seperti *web spam*, *mobile phone messaging spam*, *Voice over Internet Protocol (VoIP) spam*, *instant messaging (IM) spam social networking spam*, , *Usenet newsgroup spam*, tetapa spam email adalah spam yang paling populer[2] dan juga menjadi fokus tulisan ini. Spam email merupakan masalah yang akan terus meningkat karena mudah dan murahnya mengirim email, sehingga dapat mengganggu dan menghabiskan waktu bagi para pengguna.

Sampai saat ini, klasifikasi spam email masih menantang karena spam email masih banyak terjadi dan nilai akurasi klasifikasi masih perlu ditingkatkan.

Klasifikasi spam email dapat dilakukan dengan klasifikasi biner dengan pembelajaran mesin sebagai alat klasifikasi. Beberapa metode dalam pembelajaran mesin berhasil meningkatkan kinerja dalam klasifikasi seperti *Decision Tree* (DT). DT adalah salah satu metode klasifikasi terkenal karena DT mampu menangani atribut nominal dan numerik dan meningkatkan efisiensi komputasi[3].

Pada penelitian sebelumnya[3], telah dilakukan sebuah riset mengenai klasifikasi spam email menggunakan metode *Decision Tree* dan *Logistic Regression* (LR). Dari penelitian tersebut, hasil menunjukkan bahwa metode yang diterapkan menghasilkan hasil yang mengesankan dan menjanjikan dengan nilai akurasi adalah 91,67%. Sementara nilai akurasi yang didapatkan dari penggunakan *Decision Tree* tanpa diterapkan algoritma LR adalah 90,65%. Dari hasil tersebut, disimpulkan bahwa LR dapat meningkatkan kinerja *Decision Tree* dengan mengurangi *noisy data*.

Selain algoritma LR, beberapa algoritma lainnya dapat diterapkan bersama-sama dengan metode *Decision Tree*. Salah satu algoritma tersebut adalah algoritma *Principal Component Analysis* (PCA). PCA digunakan sebagai *dimensionality reduction* yang berfungsi untuk mengurangi dimensi/fitur data sambil tetap menjaga karakteristik penting data. Dengan menerapkan metode ini diharapkan dapat meningkatkan nilai akurasi secara keseluruhan. Pada penelitian sebelumnya[4], PCA telah berhasil diterapkan pada metode *Random Forest.*

Dari pembahasan diatas, penulis mengusulkan untuk melakukan penelitian dengan menerapkan klasifikasi *Decision Tree* untuk menentukan data spam dan non spam pada dataset yang akan digunakan. Dimana dataset yang akan

digunakan yaitu dataset *Spambase* yang diperoleh dari UCI *Machine Learning Repository* dan dataset *Emails* yang diperoleh dari UCI Kaggle. Dataset *Spambase* memiliki 4601 data dengan 58 atribut. Dataset *Emails* memiliki 6728 data dimana semua data ini masih berbentuk teks sehingga untuk memproses data tersebut akan melalui tahap *text mining* terlebih dahulu. Penelitian ini juga akan menerapkan algoritma PCA sebagai perbandingan nilai akurasi yang lebih baik ketika sebelum dan sesudah digunakan.

## 1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah :

1. Menerapkan tahapan *text mining* dalam mengubah dataset berbentuk teks ke dalam bentuk angka.
2. Menerapkan algoritma *Principal Component Analysis* (PCA) sebagai *dimensionality reduction* untuk mengurangi dimensi atau atribut pada dataset yang digunakan.
3. Melakukan klasifikasi data spam dan non-spam menggunakan metode *Decision Tree.*
4. Menganalisa dan membandingkan hasil klasifikasi menggunakan metode *Decision Tree* tanpa menerapkan PCA dan metode *Decision Tree* setelah menerapkan algoritma PCA.

## 1.3 Manfaat Penelitian

Manfaat dari penelitian ini antara lain :

1. Dapat menerapkan algoritma *text mining* untuk mengubah data berbentuk teks menjadi bentuk angka.
2. Dapat mengetahui apakah nilai akurasi dari klasifikasi menggunakan PCA pada algoritma *Decision Tree* lebih baik dibandingkan dengan tidak menggunakan PCA.
3. Dapat mempelajari proses klasifikasi spam dan non-spam.
4. Dapat menjadi referensi untuk penelitian selanjutnya.

## 1.4 Rumusan dan Batasan Masalah

Rumusan masalah yang akan diambil dari penelitian ini adalah bagaimana hasil dari penerapan algoritma *Principal Component Analysis* dan *Decision Tree*. Sedangkan batasan masalah pada penelitian ini terdapat pada nilai yang diukur, yaitu *Accuracy*, *F-Measure*, *Precision*, *Recall*, dan *Error*. Dataset yang digunakan berbatas pada spam email yang terdapat pada dataset *Spambase* yang diperoleh dari UCI *Machine Learning Respository* dan dataset *Emails* yang diperoleh dari UCI Kaggle.

## 1.5 Metodologi Penelitian

Penelitian tugas akhir ini akan melewati serangkaian metodologi (tahap-tahap) sebagai berikut:

1. Studi Pustaka

   Pada tahap ini penulis akan mencari referensi atau literature pada *keyword* yang diangkat dari judul yang bertujuan menjadi penunjang untuk penelitian ini.

2. Konsultasi

   Pada tahap ini penulis akan melakukan beberapa konsultasi kepada orang-orang yang dianggap mempunyai pengetahuan dan wawasan mengenai permasalah yang ditemui saat pembuatan tugas akhir.

3. Pengumpulan Data

   Selanjutnya, penulis mengumpulkan data dengan cara mencari dataset yang akan digunakan pada beberapa website yang menyediakan dataset. Sehingga pada akhirnya dataset yang akan digunakan dalam penelitian ini adalah dataset *Spambase* yang diperoleh dari UCI *Machine Learning Respository*[5] dan dataset *Emails* yang diperoleh dari UCI Kaggle.

4. Pengolahan Data

   Tahap selanjutnya yaitu melakukan pengolahan data pada dataset *Emails* dengan mengubah data teks menjadi angka menggunakan algoritma *text mining.* Setelah itu, penulis menerapkan algoritma *Principal Component Analysis* sebagai metode pra-proses dan *Decision Tree* sebagai metode klasifikasi pada kedua dataset yang digunakan.

5. Analisa

   Pada tahap ini, penulis melakukan pengambilan data yang telah dilakukan pengolahan sebelumnya dan selanjutnya dilakukan analisa terhadap data tersebut.

6. Kesimpulan dan Saran

Tahap terakhir ialah tahap menarik kesimpulan dari analisa sebelumnya serta memberikan saran untuk menjadi bahan referensi bagi penelitian selanjutnya.

## 1.6 Sistematika Penulisan

Berikut ini adalah sistematika penulisan dalam penelitian:

**BAB I PENDAHULUAN**

Pada bab ini akan dipaparkan mengenai latar belakang dari penelitian, manfaat penelitian, tujuan penelitian, rumusan dan batasan masalah dalam penelitian, metodologi penelitian, serta sistematika penulisan yang dipakai dalam laporan penelitian ini.

**BAB II TINJAUAN PUSTAKA**

Pada bab ini akan dijelaskan mengenai teori-teori dasar dari beberapa materi yang terkait dalam penelitian ini. Teori-teori dasar tersebut antara lain mengenai *machine learning*, spam email, *text mining*, *Principal Component Analysis*, dan *Decision Tree*.

**BAB III METODOLOGI**

Dalam bab ini akan dijelaskan mengenai *framework*, perancangan system, serta langkah-langkah atau metodologi yang akan dilakukan dalam proses penelitian ini.

**BAB IV HASIL DAN ANALISA**

Bab ini akan menjelaskan tentang proses dari penelitian, serta analisa dari hasil penerapan algoritma *Principal Component Analysis* dan metode *Decision Tree* pada kedua dataset.

**BAB V KESIMPULAN DAN SARAN**

Bab terakhir ini akan ditarik beberapa kesimpulan dari hasil penjelasan pada bab-bab sebelumnya dan memberikan beberapa saran yang diharapkan dapat digunakan pada penelitian selanjutnya.

# DAFTAR PUSTAKA


[1] A. Pacific, “Email Statistics Report , 2015-2019,” vol. 44, no. 0, 2019.

[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] A. Wijaya, “Hybrid Decision Tree and Logistic Regression Classifier for Email Spam Detection,” pp. 5–8, 2016.

[4] Y. Li, C. Yan, W. Liu, and M. Li, “A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification,” *Appl. Soft Comput. J.*, vol. 70, pp. 1000–1009, 2018.

[5] A. Subasi, S. Alzahrani, A. Aljuhani, and M. Aljedani, “Comparison of Decision Tree Algorithms for Spam E-mail Filtering,” *1st Int. Conf. Comput. Appl. Inf. Secur. ICCAIS 2018*, pp. 1–5, 2018.

[6] I. C. Freeman, A. J. Haigler, S. E. Schmeelk, L. R. Ellrodt, T. L. Fields, and W. Discovery, “What are they researching ? Examining Industry-based Doctoral Dissertation Research through the Lens of Machine Learning,” *2018 17th IEEE Int. Conf. Mach. Learn. Appl.*, pp. 1338–1340, 2018.

[7] D. Ucci, L. Aniello, and R. Baldoni, “Survey of Machine Learning Techniques for Malware Analysis,” *Comput. Secur.*, 2018.

[8] A. A. Alurkar *et al.*, “A Proposed Data Science Approach for Email Spam b Classification using Machine Learning Techniques b,” 2017.

[9] M. Habib, H. Faris, M. A. Hassonah, J. Alqatawna, A. F. Sheta, and A. M.

Al-Zoubi, “Automatic Email Spam Detection using Genetic Programming with SMOTE,” *ITT 2018 - Inf. Technol. Trends Emerg. Technol. Artif. Intell.*, no. Itt, pp. 185–190, 2019.

[10] K. Agarwal, “approach of Naïve Bayes and Particle Swarm Optimization,” *2018 Second Int. Conf. Intell. Comput. Control Syst.*, no. Iciccs, pp. 685–690, 2018.

[11] R. Shao, W. Hu, Y. Wang, and X. Qi, “The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform,” *MEASUREMENT*, vol. 54, pp. 118–132, 2014.

[12] I. T. Jolliffe, J. Cadima, and J. Cadima, “Principal component analysis : a review and recent developments Subject Areas : Author for correspondence :,” 2016.

[13] J. Li *et al.*, “Fault Diagnosis for Constant Deceleration Braking System of Mine Hoist based on Principal Component Analysis and SVM 2 Principal Component Analysis 3 Support Vector Machine,” vol. 05015, pp. 1–5, 2017.

[14] S. Singh and S. Kaur, “Improved Spambase Dataset Prediction Using SVM RBF Kernel with Adaptive Boost,” pp. 2319–2322, 2015.

[15] M. Morchid, R. Dufour, P. Bousquet, and G. Linarès, “Feature selection using Principal Component Analysis for massive retweet detection q,” *PATTERN Recognit. Lett.*, vol. 49, pp. 33–39, 2014.

[16] H. Kaur and A. Sharma, “Improved email spam classification method using integrated particle swarm optimization and decision tree,” *Proc. 2016 2nd*

*Int. Conf. Next Gener. Comput. Technol. NGCT 2016*, no. October, pp. 516–521, 2017.

[17] A. Trabelsi, Z. Elouedi, and E. Lefevre, "CO," *Fuzzy Sets Syst.*, vol. 1, pp. 1–17, 2018.

[18] W. Li and W. Meng, "An Empirical Study on Email Classification Using Supervised Machine Learning in Real Environments," pp. 7438–7443, 2015.

[19] S. Shrivastava, "Spam Mail Detection through Data Mining Techniques," 2017.

[20] D. Farid, L. Zhang, C. Mofizur, M. A. Hossain, and R. Strachan, "Expert Systems with Applications Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1937–1946, 2014.

[21] N. Daud, N. L. Mohd Noor, S. A. Aljunid, N. Noordin, and N. I. M. Fahmi Teng, "Predictive Analytics: The Application of J48 Algorithm on Grocery Data to Predict Obesity," *2018 IEEE Conf. Big Data Anal. ICBDA 2018*, pp. 1–6, 2019.

[22] A. M. Al-Zoubi, J. Alqatawna, and H. Faris, "Spam profile detection in social networks based on public features," *2017 8th Int. Conf. Inf. Commun. Syst. ICICS 2017*, no. May, pp. 130–135, 2017.

[23] D. K. Thara, B. G. Premasudha, and F. Xiong, "Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques," *Pattern Recognit. Lett.*, vol. 128, pp. 544–550, 2019.