

Website Forum Diskusi Dengan Filter Spam Menggunakan Metode Clustering K-Medoids dan Metode Stemming Tala

Wilson

Program Studi Teknik Informatika, STMIK Time, Medan, Indonesia

Email: wu95.wilson@email.com

Email Penulis Korespondensi: wu95.wilson@email.com

Abstrak—Pada pembelajaran daring, dikenal ada istilah pembelajaran sinkron dan pembelajaran asinkron. Beberapa kegiatan pembelajaran asinkron yang umum adalah berinteraksi dengan Learning Management System (LMS), berkomunikasi menggunakan email, memposting di forum diskusi dan membaca artikel. Dalam penelitian ini, akan dirancang forum diskusi dalam mendukung proses pembelajaran daring. Dalam forum diskusi, sering terdapat thread yang dibuka mengandung pesan yang tidak berguna atau tidak berhubungan dengan topik yang dibahas. Pesan ini sering disebut sebagai spam. Untuk menyelesaikan permasalahan tersebut, maka dapat diterapkan algoritma clustering K-Medoids. Sementara itu, untuk melakukan proses stemming akan digunakan metode Tala. Hasil dari penelitian ini adalah sebuah aplikasi website forum diskusi yang menyediakan fitur filter spam sehingga forum diskusi dapat bersih dari spam. Website juga mendukung proses penambahan dan penyimpanan daftar data spam baru, sehingga user dapat memasukkan kata spam baru untuk meningkatkan akurasi pendeteksian spam.

Kata Kunci: Website; Forum Diskusi; Spam; Algoritma Clustering K-Medoids; Metode Tala

Abstract—In online learning, there are known terms of synchronous learning and asynchronous learning. Some common asynchronous learning activities are interacting with the Learning Management System (LMS), communicating using email, posting on discussion forums and reading articles. In this study, a discussion forum will be designed to support the online learning process. In discussion forums, there are often threads that are opened containing messages that are not useful or not related to the topic being discussed. These messages are often referred to as spam. To solve this problem, the K-Medoids clustering algorithm can be applied. Meanwhile, to carry out the stemming process, the Tala method will be used. The result of this research is a discussion forum website application that provides a spam filter feature so that discussion forums can be clean from spam. The website also supports the process of adding and storing new spam data lists, so that users can enter new spam words to improve spam detection accuracy.

Keywords: Websites; Discussion Forums; Spam; K-Medoids Clustering Algorithm; Tala Method

1. PENDAHULUAN

Pada awal tahun 2020, dunia tengah dilanda oleh penyebaran virus baru yang sangat pesat, tidak terkecuali bagi Indonesia. Virus Corona atau yang dikenal dengan COVID-19 saat ini diketahui telah menginfeksi jutaan orang di dunia. Sektor pendidikan di Indonesia tidak luput dari adanya COVID-19 dan ikut merasakan dampak dari penyebaran virus tersebut. Dalam surat edaran Kemendikbud Nomor 4 Tahun 2020 tentang Pelaksanaan Kebijakan Pendidikan dalam Masa Darurat Penyebaran Coronavirus Disease (COVID- 19) yang dikeluarkan oleh pemerintah sebagai tindakan solusi untuk tetap dapat mewujudkan pendidikan yang efektif, salah satu kebijakannya adalah proses belajar mengajar dilakukan dari rumah secara daring atau pembelajaran jarak jauh untuk mencegah penyebaran virus di lingkungan pendidikan [1].

Pada pembelajaran daring, dikenal ada istilah pembelajaran sinkron dan pembelajaran asinkron. Dalam pembelajaran sinkron, siswa dan guru berada di tempat yang sama pada waktu yang sama. Salah satu contoh pembelajaran sinkron adalah ketika siswa dan guru berpartisipasi dalam ruang kelas virtual melalui aplikasi web conference. Sedangkan, beberapa kegiatan pembelajaran asinkron yang umum adalah berinteraksi dengan Learning Management System (LMS), berkomunikasi menggunakan email, memposting di forum diskusi dan membaca artikel [2]. Dalam penelitian ini, akan dirancang forum diskusi dalam mendukung proses pembelajaran daring. Adapun beberapa kelebihan forum diskusi yaitu dapat menginspirasi peserta untuk lebih kreatif terutama dalam memberikan ide, dapat dilatih untuk beradaptasi dengan curah pendapat tentang berbagai masalah dan peserta dapat dilatih untuk mengungkapkan pendapat atau gagasan secara lisan [3].

Dalam forum diskusi, sering terdapat thread yang dibuka mengandung pesan yang tidak berguna atau tidak berhubungan dengan topik yang dibahas. Pesan ini sering disebut sebagai spam. Spam adalah pesan yang tidak diminta yang berisi promosi produk, pornografi, virus dan content-content yang tidak penting [4]. Munculnya spam sangat mengganggu pengguna karena dapat menambah penggunaan bandwidth koneksi internet, serta akan menjadi suatu sampah yang menumpuk sehingga mengurangi kapasitas penyimpanan [5]. Untuk menyelesaikan permasalahan tersebut, maka dapat diterapkan algoritma clustering K-Medoids.

Clustering atau klusterisasi adalah salah satu alat bantu pada data mining yang bertujuan mengelompokkan objek-objek ke dalam cluster-cluster [6]. Algoritma K-Medoids atau Partitioning Around Medoids (PAM) merupakan suatu metode partisi clustering untuk mengelompokkan sekumpulan (n) objek menjadi sejumlah (k) cluster [7]. Kelebihan dari Algoritma K-Medoids lebih mudah untuk mengatasi kelemahan yang sering terjadi pada algoritma K-Means seperti sensitive terhadap noise dan outlier, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data, dan hasil proses clustering dari algoritma K-Medoids tidak bergantung pada urutan masuk dari dataset [8].

Sementara itu, untuk melakukan proses stemming akan digunakan metode Tala. Proses stemming yaitu proses menghilangkan semua imbuhan dan akhiran yang melekat di beberapa kata. Semua hasil dari stemming, harus berupa kata dasar dan tidak ada lagi imbuhan maupun akhiran yang masih melekat [9]. Penelitian ini menggunakan metode Tala

karena metode ini lebih sederhana dan teknik aturan imbuhan sangat mudah untuk diimplementasikan dan tidak membutuhkan waktu komputasi yang tinggi. Selain itu penelitian juga menggunakan input berupa isi thread yang relatif pendek sehingga metode Tala ini tepat sekali digunakan sebagai alur program yang dapat meningkatkan kualitas indeks dan performa stemming berbasis aturan yang relatif stabil. Metode Tala ini digunakan sebagai proses stemming atau menghilangkan imbuhan dari sebuah kata dan metode ini adalah adopsi dari algoritma stemmer bahasa Inggris terkenal Porter [10]. Proses pengecekan spam akan dimulai dengan membuang stopword dari isi thread ataupun pesan balasan yang dimasukkan. Kemudian, proses dilanjutkan dengan melakukan proses stemming dengan metode Tala. Terakhir, akan digunakan algoritma K-Medoids untuk menentukan apakah isi thread ataupun pesan balasan yang dimasukkan tersebut termasuk dalam pesan spam atau bukan.

2. METODOLOGI PENELITIAN

2.1 Forum Diskusi

Forum online atau komunitas online merupakan konsep pengembangan yang bisa dimanfaatkan di dalam dunia pendidikan. Selain berguna untuk menjalin hubungan silaturahmi forum online berguna juga untuk menunjang dalam meningkatkan efektivitas belajar dengan cara membentuk komunitas dan diskusi sesuai dengan bidang dan minat bagi orang-orang yang terlibat di dalamnya. Melalui forum online seseorang dapat saling berbagi pengalaman dan informasi mengenai topik yang sedang dibahas tanpa harus saling berhadapan atau saling mengenal satu sama lain [11].

2.2 Filter Spam

SPAM merupakan akronim dari Stupid Pointless Annoying Message. Spam pesan yaitu pesan yang tidak diinginkan atau diminta oleh penerimanya. Spam muncul pertama kali pada bulan mei tahun 1978. Spam tersebut bersifat iklan yang dikirimkan oleh Digital Equipment Corporation (DEC). Spam juga dapat berupa pengiriman pesan secara berulang-ulang ke berbagai newsgroup atau server milis dengan pokok bahasan yang tidak berkaitan [12].

2.3 Metode Clustering K-Medoids

K-Medoids menggunakan k sebagai jumlah pusat cluster awal yang dihasilkan secara acak diawal proses clustering. Setiap obyek yang lebih dekat dengan pusat cluster akan dikelompokkan dan membentuk cluster baru. Algoritma kemudian secara acak menentukan cluster center baru dari setiap cluster yang terbentuk sebelumnya dan menghitung ulang jarak antara obyek dan pusat cluster baru yang dihasilkan. Jarak antar obyek i dan j dihitung dengan menggunakan dissimilarity measurement function, dimana salah satunya adalah Euclidean Distance Function yang ditunjukkan dalam persamaan berikut [13]:

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2}, i = 1, \dots, n; j = 1, \dots, n \quad (1)$$

dimana x_{ia} adalah variabel ke-a dari obyek i ($i=1, \dots, n; a=1, \dots, p$) dan d_{ij} adalah nilai Euclidean Distance. Algoritma juga menghitung probabilitas penukaran setiap obyek dengan pusat cluster yang lain menggunakan fungsi kriteria. Salah satu fungsi kriteria yang digunakan adalah absolute-error, seperti pada persamaan berikut [13]:

$$E = \sum_{j=1}^k \sum_{p \in C_j} |p - o_j| \quad (2)$$

di mana E adalah jumlah dari absolut error untuk semua objek dalam dataset; p adalah titik dalam ruang yang mewakili suatu objek dalam kluster C_j , dan o_j adalah obyek didalam cluster C_j [13].

2.4 Metode Stemming Tala

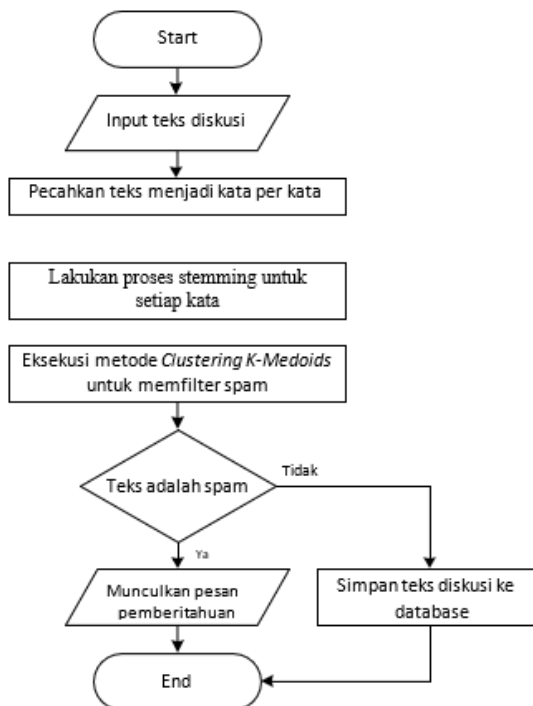
Pada stemmer tala terdapat 5 langkah utama dengan 3 langkah awal dan 2 langkah pilihan, langkah- langkah tersebut sebagai berikut:

1. Menghilangkan partikel.
2. Menghilangkan kata sandang dan kepunyaan.
3. Menghilangkan awalan 1.
4. Jika suatu aturan terpenuhi jalankan sebagai berikut:
 - a. Hilangkan akhiran.
 - b. Jika suatu aturan terpenuhi, hilangkan awalan 2. Jika tidak proses stemming selesai.
5. Jika tidak ada aturan yang terpenuhi jalankan sebagai berikut:
 - a. Hilangkan awalan 2.
 - b. Hilangkan akhiran.
 - c. Proses stemming selesai.

3. HASIL DAN PEMBAHASAN

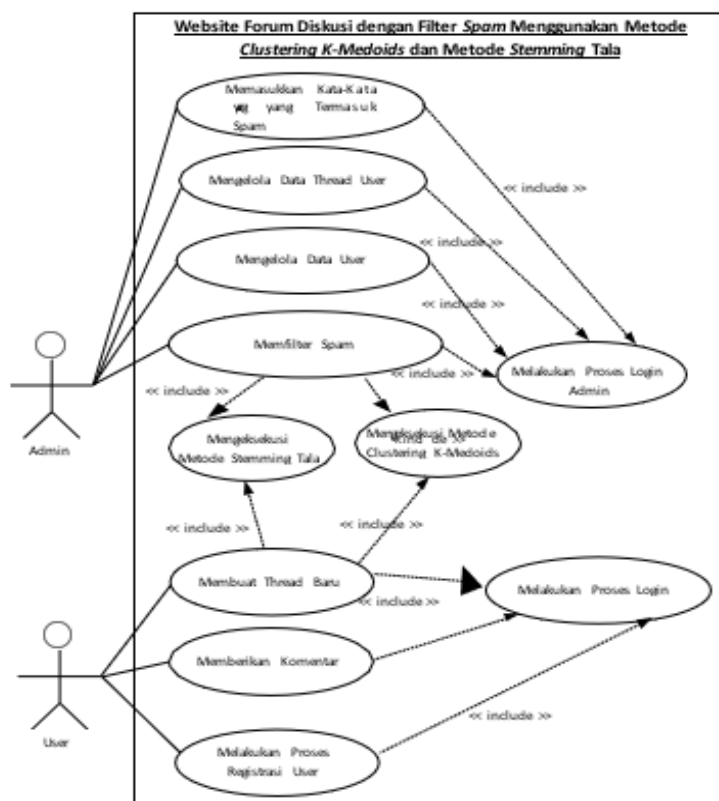
Proses kerja dari metode Clustering K-Medoids dan metode Stemming Tala dalam memfilter spam dapat digambarkan

dalam bentuk flowchart seperti terlihat pada gambar 1.



Gambar 1. Flowchart dari Website Forum Diskusi dengan Filter Spam Menggunakan Metode Clustering K-Medoids dan Metode Stemming Tala

Untuk melakukan pemodelan terhadap website forum diskusi yang mampu memfilter spam dengan metode Clustering K-Medoids dan metode Stemming Tala, akan digunakan bantuan use case diagram. Adapun rancangan use case diagram dari aplikasi dapat dilihat pada gambar 2.



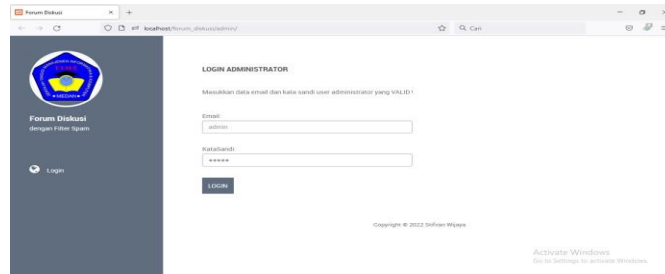
Gambar 2. Use Case dari Website Forum Diskusi dengan Filter Spam Menggunakan Metode Clustering K-Medoids dan Metode Stemming Tala

3. HASIL DAN PEMBAHASAN

Tampilan dari website forum diskusi dengan filter spam menggunakan metode Clustering K-Medoids dan metode Stemming Tala ini terbagi menjadi dua bagian yaitu bagian admin dan bagian user.

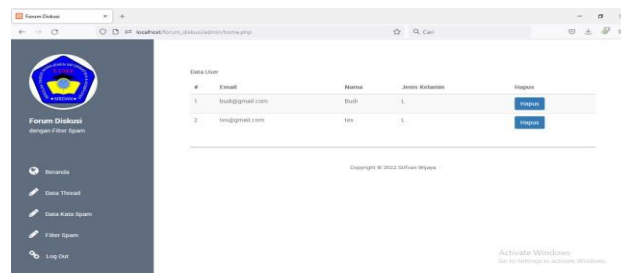
3.1 Tampilan Admin

Pada saat menjalankan bagian admin dari website forum diskusi dengan filter spam menggunakan metode Clustering K-Medoids dan metode Stemming Tala, maka akan muncul tampilan login seperti terlihat pada gambar berikut:



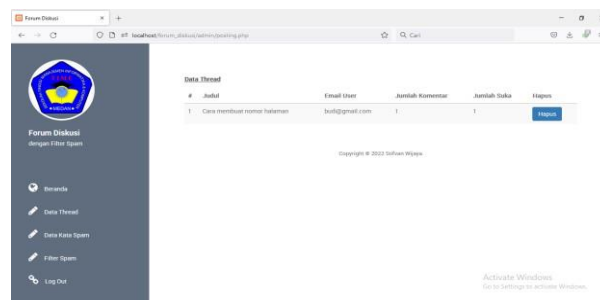
Gambar 3. Halaman Login Admin

User admin harus memasukkan data user id dan kata sandi admin yang valid agar dapat menggunakan sistem. Apabila proses login berhasil, maka sistem akan menampilkan halaman Home Admin seperti terlihat pada gambar berikut:



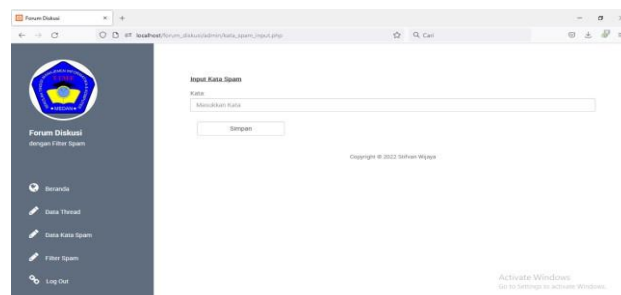
Gambar 4. Halaman Home Admin

Apabila user admin ingin menampilkan data posting yang telah tersimpan sebelumnya, maka user admin dapat mengklik menu Data Thread sehingga sistem akan menampilkan halaman Daftar Thread seperti terlihat pada gambar berikut:



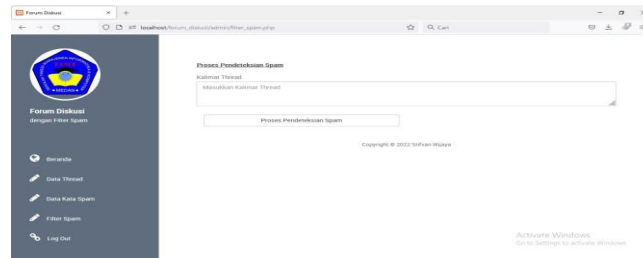
Gambar 5. Halaman Daftar Thread

Sementara itu, apabila user admin ingin menambah kata spam baru, maka dapat mengklik tombol Tambah Kata Spam Baru, sehingga sistem akan menampilkan halaman Tambah Kata Spam seperti terlihat pada gambar berikut:



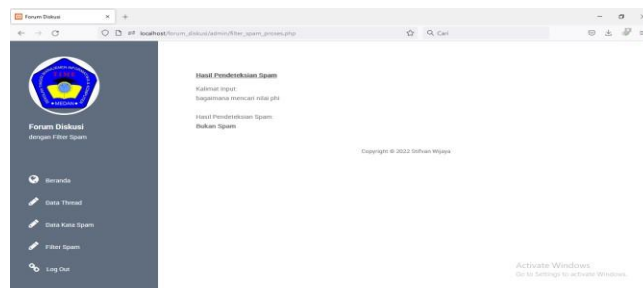
Gambar 6. Halaman Tambah Kata Spam

User admin dapat memasukkan kata spam. Setelah itu, klik tombol Simpan untuk menyimpan data kata spam yang dimasukkan ke dalam database. Terakhir, apabila user admin ingin mencoba mendeteksi spam dari thread yang dimasukkan, maka dapat klik menu Filter Spam, sehingga sistem akan menampilkan halaman Filter Spam seperti terlihat pada gambar berikut:



Gambar 7. Halaman Filter Spam

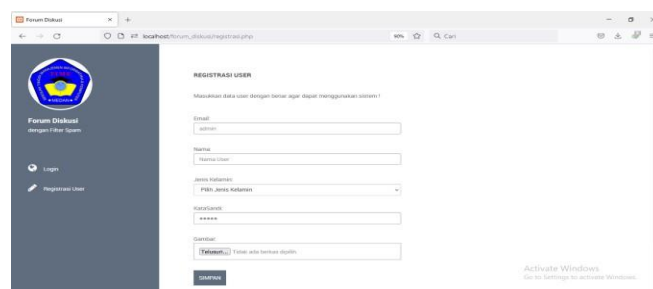
User admin dapat memasukkan kalimat thread yang ingin difilter spam dan klik tombol Proses Pendeteksian Spam, sehingga sistem akan menampilkan halaman Hasil Filter Spam seperti terlihat pada gambar berikut:



Gambar 8. Halaman Hasil Filter Spam

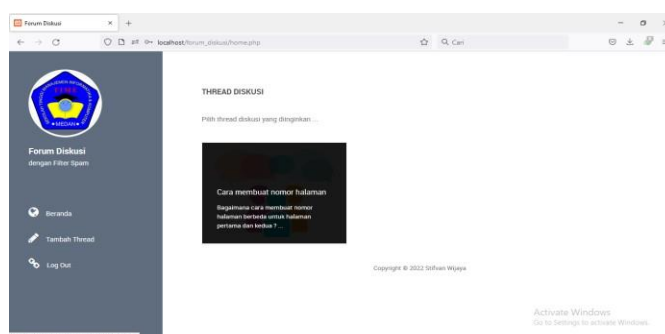
3.2 Tampilan User

Pada saat menjalankan bagian user dari website forum diskusi dengan filter spam menggunakan metode Clustering K-Medoids dan metode Stemming Tala, maka akan muncul tampilan login. Apabila user belum terdaftar, maka user dapat mengklik menu Registrasi User untuk melakukan proses pendaftaran user baru ke dalam sistem, sehingga sistem akan menampilkan halaman Registrasi User seperti terlihat pada gambar berikut:



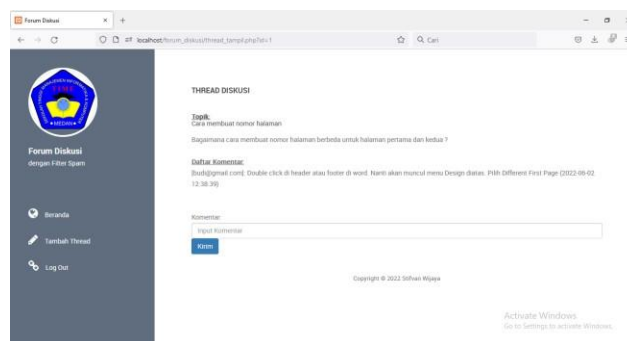
Gambar 9. Halaman Registrasi User

User dapat memasukkan data user baru dengan benar dan mengklik tombol Simpan untuk menyimpandata user ke dalam database. Setelah itu, maka user dapat melakukan proses login ke dalam sistem. Apabila user berhasil melakukan proses login ke dalam sistem, maka sistem akan menampilkan halaman Home User seperti terlihat pada gambar berikut:



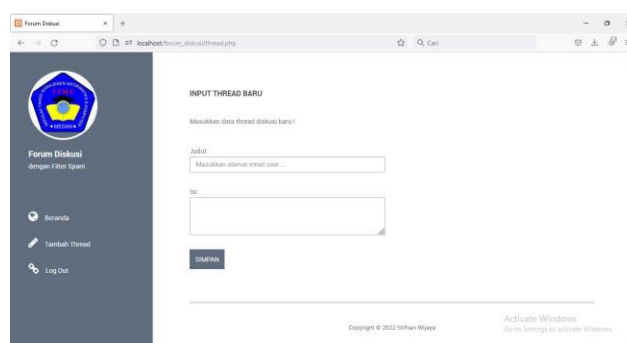
Gambar 10. Halaman Home User

Pada bagian tengah website akan terdapat berbagai thread diskusi yang dimasukkan oleh user. Apabila user ingin memberikan komentar ke dalam thread yang bersangkutan, maka user dapat klik link Topik Thread tersebut sehingga sistem akan menampilkan halaman Detail Thread seperti terlihat pada gambar berikut:



Gambar 11. Halaman Detail Posting

User dapat memasukkan komentar yang diinginkan dan klik tombol Kirim. Sementara itu, apabila user ingin menambah thread baru, maka dapat klik menu Tambah Thread, sehingga sistem akan menampilkan halaman Tambah Thread seperti terlihat pada gambar berikut:



Gambar 12. Halaman Tambah Thread

User dapat memasukkan data thread baru, yang mencakup judul posting dan isi thread. Setelah itu, user dapat klik tombol Simpan, untuk menyimpan data thread ke dalam database.

4. KESIMPULAN

Setelah menyelesaikan skripsi ini, penulis menarik beberapa kesimpulan sebagai berikut: Website dapat mendeteksi spam berdasarkan data thread yang dimasukkan oleh user. Perangkat lunak mendukung proses penambahan dan penyimpanan daftar data spam baru, sehingga user dapat memasukkan kata spam baru untuk meningkatkan akurasi pendeteksian spam.

REFERENCES

- [1] D. N. Baety dan D. R. Munandar, "Analisis Efektifitas Pembelajaran Daring Dalam Menghadapi Wabah Pandemi Covid-19," Edukatif : Jurnal Ilmu Pendidikan, vol. 3, no. 3, pp. 880-889, 2021.
- [2] A. Kurniasari, F. S. P. Pribowo dan D. A. Putra, "ANALISIS EFEKTIVITAS PELAKSANAAN BELAJAR DARI RUMAH (BDR) SELAMA PANDEMI COVID-19," Jurnal Review Pendidikan Dasar: Jurnal Kajian Pendidikan dan Hasil Penelitian, vol. 6, no. 3, 2020.
- [3] Sulandari, "ANALISIS TERHADAP METODA PEMBELAJARAN KLASIKAL DAN METODA PEMBELAJARAN E-LEARNING DI LINGKUNGAN BADIKLAT KEMHAN," Jurnal Pendidikan Indonesia, vol. 1, no. 2, pp. 176-187, 2020.
- [4] A. Wibisono, S. D. Rizkiono dan A. Wantoro, "FILTERING SPAM EMAIL MENGGUNAKAN METODE NAIVE BAYES," TELEFORTECH, vol. 1, no. 1, pp. 9-17, 2020.
- [5] F. Rozi dan R. Kartadie, "DETEKSI E-MAIL DAN SPAM MENGGUNAKAN FUZZY ASSOCIATION RULE MINING," JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 02, no. 02, pp. 94-98, 2017.
- [6] W. Kurniawan, A. Rifai, W. Gata dan D. Gunawan, "Analisis Algoritma K-Medoids Clustering Dalam Menentukan Pemesanan Hotel," JURNAL SWABUMI, vol. 8, no. 2, pp. 182-187, 2020.
- [7] S. Nurlaela, A. Primajaya dan T. N. Padilah, "ALGORITMA K-MEDOIDS UNTUK CLUSTERING PENYAKIT MAAG DI KABUPATEN KARAWANG," Jurnal Informatika, Manajemen dan Komputer, vol. 12, no. 2, pp. 56-62, 2020.
- [8] N. Hidayati, A. I. Rizmayanti, C. B. S. Dewi, R. Fatmasari dan W. Gata, "Penerapan Algoritma Klusterisasi dan Klasifikasi pada Tingkat Kepentingan Sistem Pembelajaran di Universitas Terbuka," JURNAL SWABUMI, vol. 8, no. 2, pp. 134-142, 2020.
- [10] P. F. Ariyani, A. Rahmala dan N. Juliasari, "Implementasi Metode Stemming Tala Dan Fungsi Jaccard Pada Aplikasi Katalog

- Perpustakaan,” Seminar Nasional Inovasi dan Aplikasi Teknologi di Industri, pp.128-133, 2019.
- [11] F. S. N. Ramadhan, N. Chasanah dan D. Iskandar, “APLIKASI FORUM DISKUSI HIMPUNAN MAHASISWA TEKNIK UNIVERSITAS JENDERAL SOEDIRMAN MENGGUNAKAN LARAVEL,” Jurnal Teknik Informatika (JUTIF), vol. 1, no. 1, pp. 1-6, 2020.
- [12] R. Y. Hayuningtyas, “Aplikasi Filtering of Spam Email Menggunakan Naïve Bayes,” IJCIT (Indonesian Journal on Computer and Information Technology), vol. 2, no. 1, pp. 53-60, 2017.
- [13] Z. Mustofa dan I. S. Suasana, “ALGORITMA CLUSTERING K-MEDOIDS PADA E-GOVERNMENT BIDANG INFORMATION AND COMMUNICATION TECHNOLOGY DALAM PENENTUAN STATUS EDGI,” Jurnal Teknologi Informasi dan Komunikasi, vol. 9, no. 1, pp. 1-10, 2018.