

MENINGKATKAN EFEKTIVITAS DAN EFISIENSI IMPLEMENTASI ELECTRONIC DATA PROCESSING PEMBOBOTAN KATA PADA KLASIFIKASI TEXT MINING

Nofiyani¹, Wulandari², Yesi Puspita Dewi³.

¹²³ Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur

¹²³ Ciledug Raya, RT.10/RW.2, Petungkana Utara, Kec. Pesanggrahan,
Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12260

¹ nofiyani@budiluhur.ac.id, ² wulandari@budiluhur.ac.id, ³ yesi.puspitadewi@budiluhur.ac.id

Abstrak : Kemajuan teknologi telah menyebabkan peningkatan jumlah data yang semakin besar dan terus bertambah setiap harinya. Perkembangan ini dapat dimanfaatkan untuk menambang data yang kemudian dapat diolah menjadi teks atau informasi yang dibutuhkan. Klasifikasi adalah bagian penting dari text mining yang digunakan untuk mengelompokkan objek berdasarkan karakteristik yang dimiliki oleh objek. Data yang diperoleh adalah data mentah dari file XML Google Alert yang kemudian dimasukkan ke dalam sistem yang dikembangkan menggunakan bahasa pemrograman PHP. Metode pengolahan data yang diterapkan dalam penelitian ini adalah Electronic Data Processing (EDP). Diharapkan bahwa penggunaan sistem ini dapat mempercepat proses klasifikasi data, terutama ketika jumlah sampel data yang diperlukan sangat banyak. Hasil penelitian menunjukkan bahwa sistem informasi yang dikembangkan dapat melakukan perhitungan probabilitas data untuk 60 dokumen hanya memerlukan waktu 0,093346003691355 menit dan proses klasifikasi data untuk 40 dokumen hanya memakan waktu 0,0069851199785868 menit. Dengan persentase 77,5% data dapat terklasifikasi dengan baik, sementara 22,5% tidak dapat terklasifikasi. Sehingga hasil penerapan Electronic Data Processing dapat meningkatkan efektivitas dan efisiensi dalam proses text mining.

Kata Kunci : Electronic Data Processing, Klasifikasi Data, Efektivitas, Efisiensi

Abstract: *Technological advancements have led to an increasing amount of data that grows continuously each day. This development can be utilized to mine data, which can then be processed into the needed text or information. Classification is a crucial part of text mining, used to group objects based on their characteristics. The data obtained is raw data from Google Alert XML files, which are then input into a system developed using the PHP programming language. The data processing method applied in this research is Electronic Data Processing (EDP). It is hoped that the use of this system can accelerate the data classification process, especially when a large number of data samples are required. The research results show that the developed information system can perform data probability calculations for 60 documents in just 0.093346003691355 minutes, and the data classification process for 40 documents takes only 0.0069851199785868 minutes. With 77.5% of the data successfully classified and 22.5% not classified. Thus, the implementation of Electronic Data Processing can enhance the effectiveness and efficiency of the text mining process.*

Keywords: *Electronic Data Processing, Data Classification, Effectiveness, Efficiency.*

PENDAHULUAN

Dalam era digital penggunaan text mining menjadi sangat penting untuk menggambarkan atau mengevaluasi pengetahuan dari sekumpulan besar dokumen teks (Deolika et al., 2019). Seiring dengan bertambahnya jumlah dokumen setiap harinya dapat dimanfaatkan untuk menggali data yang nantinya dapat diolah menjadi text kemudian dari text tersebut dapat diolah lagi menjadi informasi yang dibutuhkan. Saat ini sudah banyak organisasi yang telah mengimplementasi teknologi ini tidak hanya meningkatkan produktivitas, tetapi juga mempercepat proses pengambilan keputusan.

Teknik yang digunakan untuk menganalisis data dalam jumlah besar yaitu menggunakan text mining (Hermawan et al., 2023). Menurut (Kowsari et al., 2019), (Nota et al., 2022), (Tandel et al., 2019), (Hermawan et al., 2023) Text Mining adalah proses menemukan pengetahuan yang tidak diketahui melalui ekstraksi informasi otomatis dari sejumlah besar teks yang tidak terstruktur. Dalam proses text mining, terdapat pemecahan kata yang bertujuan untuk memberikan nilai atau bobot pada term yang ada dalam suatu dokumen (Deolika et al., 2019). Tantangan utama adalah bagaimana mengklasifikasikan data secara efektif dan efisien agar kategori tersebut sesuai dengan kebutuhan. Klasifikasi adalah bagian penting dari text mining yang digunakan untuk mengelompokkan objek berdasarkan karakteristik yang dimiliki oleh objek. Untuk mengklasifikasikan teks merupakan proses untuk membentuk golongan atau kelas dari sebuah dokumen yang pada dasarnya kategori atau kelompok sudah ditetapkan atau diketahui (Nurhadi, 2016), (Andriani & Wibowo, 2021). Kendala lain yaitu ketika melakukan klasifikasi data dalam jumlah besar, tentunya akan memakan waktu yang cukup lama. Sehingga dibutuhkan sebuah metode yaitu Electronic Data Processing (EDP) yang diharapkan yang sistem yang akan dibangun dapat

mempercepat proses klasifikasi data dalam jumlah banyak dan sistem yang dibangun lebih efektif dan efisien dalam menyajikan informasi.

KAJIAN PUSTAKA DAN LANDASAN TEORI

Efektivitas

Efektivitas adalah kemampuan untuk memilih tujuan-tujuan atau sasaran yang tepat dan mencapainya, hal ini dijelaskan dalam jurnal yang ditulis oleh (Pahmawati et al., 2020). Tujuan dalam jurnal tersebut untuk mengetahui pengaruh etos kerja dan konseling terhadap efektivitas kerja pegawai. Definisi lain Efektivitas adalah Sebuah kegiatan yang sukses dilaksanakan sesuai dengan target atau tujuan yang diharapkan dan dalam waktu yang telah dijadwalkan (Imam et al., 2022). Tujuan dalam jurnal tersebut Memahami seberapa efektif para pelaku ekonomi dalam mendukung pertumbuhan ekonomi di Indonesia.

Efisiensi

Efisiensi adalah Hubungan erat produktivitas yang mampu menunjukkan kehematan dari segi tenaga, waktu dan biaya. Hal ini dijelaskan dalam jurnal yang ditulis (Tri et al., 2018), (Febrianty et al., 2021). Tujuan dalam penelitian tersebut dapat meningkatkan kinerjanya agar lebih baik lagi.

Electronic Data Processing

Electronic Data Processing (EDP) merupakan Proses mengumpulkan dan mengevaluasi bukti apakah sistem TI dapat melindungi aset, menjaga integritas data, serta mencapai tujuan bisnis secara optimal dan efektif dengan menggunakan aset perusahaan secara hemat dan hati-hati. Hal ini dijelaskan dalam jurnal yang ditulis oleh (Lestari et al., 2022). Definisi lain Sebuah sistem pengolahan data di mana data-data tersebut dimanipulasi menggunakan peralatan elektronik, yaitu komputer (Septriawan, 2021). Tujuan dari penelitian yang dilakukan oleh Septriawan

Dengan penggunaan komputer di dalam sebuah perusahaan, diharapkan penerapan sistem informasi akuntansi dapat menghasilkan informasi yang berguna bagi para penggunanya.

Klasifikasi

Klasifikasi adalah metode pengelompokan objek berdasarkan karakteristik yang dimiliki oleh objek tersebut. Hal ini dijelaskan dalam jurnal yang ditulis oleh (Rini & Solecha, 2022). Tujuan dari penelitian yang dilakukan oleh rini dan solecha yaitu untuk menentukan algoritma yang paling sesuai digunakan guna memperoleh model algoritma terbaik dengan contoh data pada Aplikasi Halodoc. Definisi lain klasifikasi yaitu salah satu teknik dalam data mining yang mengelompokkan data ke dalam kategori atau kelas yang telah ditetapkan. Hal ini dijelaskan dalam jurnal yang ditulis oleh (Setio et al., 2020).

Text Mining

Text mining adalah proses mengeksplorasi dan menganalisis sejumlah besar data teks tidak terstruktur yang dibantu oleh perangkat lunak yang dapat mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lainnya dalam data. Hal ini dijelaskan dalam jurnal yang ditulis oleh (Fathonah & Herliana, 2021).

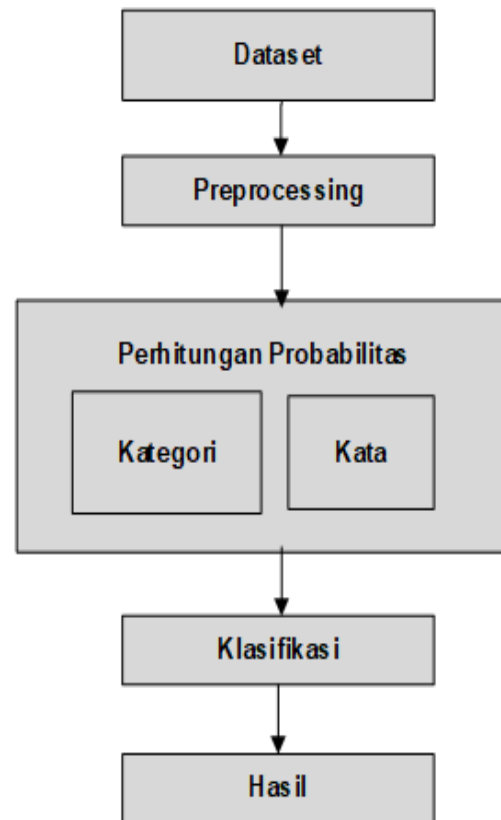
Skala Pengukuran

Dalam statistik untuk menentukan uji statistik ada empat skala yang umum digunakan (Asari et al., 2023):

1. Skala nominal yang merupakan skala terikat variabel yang hanya digunakan untuk membedakan satu jenis dari jenis lainnya.
2. Skala ordinal merupakan skala yang berhubungan dengan suatu variabel.
3. Skala interval merupakan gabungan semua karakteristik skala pengukuran nominal atau ordinal dengan satu fitur tambahan.
4. Skala rasio merupakan skala variabel yang dapat menunjukkan tingkat, rentang nilai, dan perbandingan.

METODE

Untuk mencapai hasil yang optimal dalam penelitian ini, diperlukan beberapa langkah kegiatan penelitian. Langkah-langkah penelitian ditunjukkan pada gambar 1.



Gambar 1 Langkah Penelitian

Keterangan:

1. DataSet

Tahapan pertama adalah mengumpulkan dataset yang akan digunakan dalam penelitian. Data yang tersebut didapatkan dari google alert berupa text dokumen yang di ubah menjadi file xml lalu dikirimkan ke feed rss kemudian di crawling agar mudah dibaca oleh sistem.

2. Preprocessing

Langkah berikutnya adalah preprocessing, yaitu setelah data diperoleh, kemudian data dipreprocessing menjadi beberapa tahapan yaitu case folding, penghapusan simbol, konversi

slangword, penghapusan stopword, stemming, dan tokenisasi.

3. Perhitungan Probabilitas

Dalam tahap ini, dilakukan perhitungan probabilitas untuk menentukan kemungkinan data masuk ke dalam kategori atau kelas tertentu. Ini melibatkan dua subproses utama:

- **Kategori:** Menentukan kategori atau kelas data berdasarkan fitur-fitur yang relevan.
- **Kata:** Menghitung frekuensi atau probabilitas kemunculan kata-kata tertentu dalam setiap kategori.

4. Klasifikasi

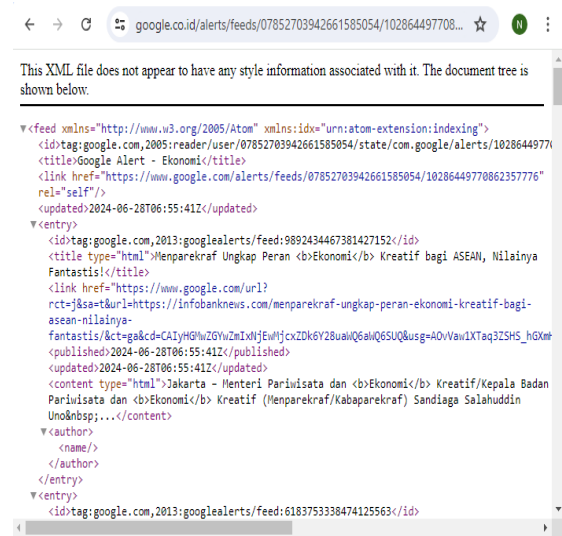
Setelah probabilitas dihitung, data diklasifikasikan ke dalam kategori atau kelas yang paling sesuai berdasarkan hasil perhitungan probabilitas.

5. Hasil

Tahapan terakhir adalah menampilkan hasil klasifikasi. Hasil ini menunjukkan kategori atau kelas akhir dari data yang telah diproses.

HASIL DAN PEMBAHASAN

Pada penelitian ini data yang digunakan diperoleh dari file XML Google Alert yang diperlukan dalam melakukan proses tranning dan testing yang ditunjukkan pada gambar 2.



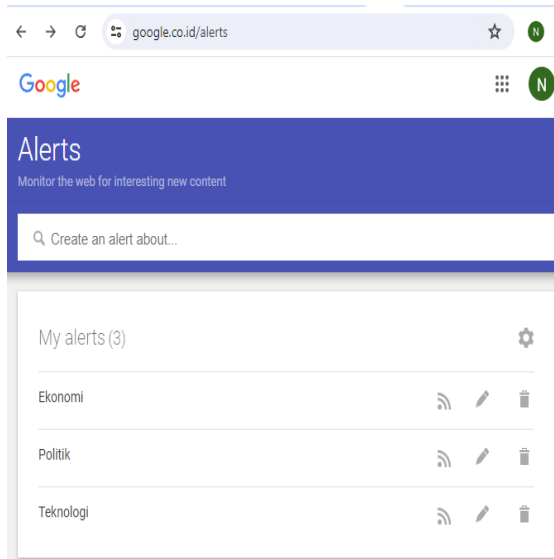
Gambar 2 file XML Google Alert

Data tranning dalam tahap klasifikasi akan digunakan sebagai data pembelajaran untuk menandai atau melabelkan data baru ke dalam sebuah kelas. Pada gambar 3 terlihat menu utama sistem yang sedang dikembangkan.



Gambar 3 Menu Utama

Data tranning yang digunakan dalam proses pembelajaran text mining berupa 60 dokumen data yang sudah melalui tahap preprocessing dan sudah memiliki label sesuai kata kunci pencarian pada Google Alert. Kategori berita yang digunakan terdiri dari Politik, Ekonomi dan Teknologi yang ditunjukkan pada Gambar 4.



Gambar 4 Kategori Berita Data Traning

Beberapa data traning dapat dilihat pada Gambar 5.

No.	Data Traning	Kategori
1	pekanbaru riauposco huawei device indonesia resmi luncur huawei matepad s pasar tablet tanah air tablet tawar alam tingkat kreasi layar tablet tekstur layak kertas berkat dukung teknologi	Teknologi
2	kompascom openai luncur aplikasi desktop chatgpt pertama perangkat macos tengah mei awal aplikasi unduh download guna bayar langgan chatgpt plus	Teknologi
3	sekjen pks abu bakar al habsy tanya dinamika ganti pasang anies baswedan sohibil iman buka quotmasih	Politik
4	kumpul contoh soal bekonomib ma ksm jawab jadi bahan latih serta seleksi ksm ikut info	Ekonomi
5	usaha nilai putus perintah pindah ribu asn ikn jadi magnet utama tarik investasi ibu kota baru	Ekonomi
56	catat tumbuh bekonomib gorontalo lambat tumbuh bekonomib kawasan sulampua besar persen yoy nasional besar	Ekonomi
57	lombokposthuawei device indonesia resmi luncur huawei matepad s pasar tablet tanah air tablet tawar alam tingkat kreasi layar tablet tekstur layak kertas berkat dukung teknologi papermatte	Teknologi
58	toba mistarid laku umkm sumondang tabita nainggolan murni asima sitanggang milik ide bekonomib kreatip buat shandal eceng	Ekonomi
59	teliti tahun beri judul the paradox of political dynasties of regeneration type and poverty in regional autonomy era tulis	Politik
60	kbrn palembang kabar gembira guna youtube android platform video populer hadir fitur sleep timer nanti fitur mungkin guna atur waktu putar video putar	Teknologi

Gambar 5 Data Traning

Dalam proses perhitungan probabilitas terdiri dari perhitungan nilai probabilitas pada setiap kategori berita dan nilai probabilitas setiap kata pada setiap

kategori. Implementasi sistem perhitungan probabilitas untuk 3 jenis kategori dan 2076 kata pada setiap kategori dapat dilihat pada Gambar 6 hanya membutuhkan waktu 0.093346003691355 menit.

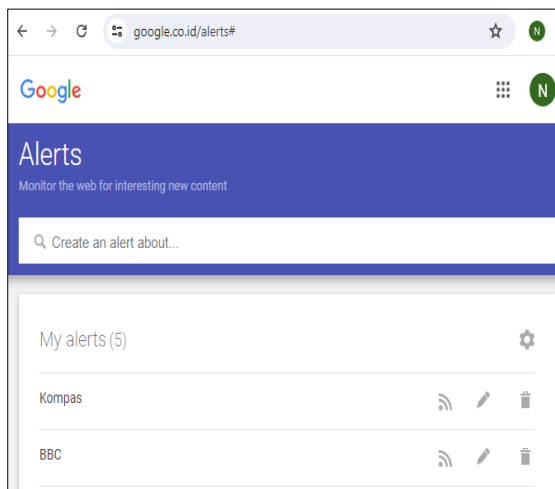
Nilai Probabilitas Pada Setiap Kategori			
No.	Kategori Berita	Frekuensi Dokumen	Nilai Probabilitas
1	Ekonomi	20	0.333333333333333
2	Politik	20	0.333333333333333
3	Teknologi	20	0.333333333333333

Nilai Probabilitas Setiap Kata Pada Setiap Kategori				
No.	Kategori	Kata	Frekuensi	Nilai Probabilitas
1	Ekonomi	acara	0	0.00072150072150072
2	Politik	acara	0	0.00072150072150072
3	Teknologi	acara	2	0.0021645021645022
4	Ekonomi	aceh	1	0.0014430014430014
5	Politik	aceh	0	0.00072150072150072
6	Teknologi	aceh	0	0.00072150072150072
7	Ekonomi	action	0	0.00072150072150072
8	Politik	action	0	0.00072150072150072
9	Teknologi	action	1	0.0014430014430014
2068	Ekonomi	ying	0	0.00072254335260116
2069	Politik	ying	0	0.00072254335260116
2070	Teknologi	ying	2	0.0021676300578035
2071	Ekonomi	youtube	0	0.00072254335260116
2072	Politik	youtube	0	0.00072254335260116
2073	Teknologi	youtube	1	0.0014450867052023
2074	Ekonomi	z	0	0.00072254335260116
2075	Politik	z	0	0.00072254335260116
2076	Teknologi	z	1	0.0014450867052023

Total Execution Time: 0.093346003691355 Mins

Gambar 6 Implementasi sistem perhitungan probabilitas

Setelah perhitungan nilai probabilitas, maka tahap selanjutnya dilakukan proses klasifikasi menggunakan algoritma Naïve Bayes Classifier. Data testing yang digunakan dalam proses klasifikasi berupa 40 dokumen data yang diperoleh dari file XML Google Alert yang sudah melalui tahap preprocessing. Kata kunci pencarian berdasarkan situs berita yang terdiri dari BBC dan Kompas yang ditunjukkan pada Gambar 7.



Gambar 7 Kategori Berita Data Testing

Implementasi sistem untuk menentukan nilai probabilitas tertinggi dari setiap dokumen ke dalam 3 jenis kategori dapat dilihat pada Gambar 8 yang hanya membutuhkan waktu 0.0069851199785868 menit dengan jumlah data terklasifikasi 31 dokumen dan jumlah tidak terklasifikasi 9 dokumen.

Proses Klasifikasi		
No.	Data Testing	Kategori
1	yakin jam akhir benarbenar jadi kata radio bbbcb pulau mariana utara pilih	Politik (0.0009633911)
2	adil pulau saipan kira izin diri wikileaks juli assange bebas kembali negara asal australia	Kategori Tidak Ditemukan
3	agatha christies miss marple vhs tape set bbbcb aampe home video lot new sealead buy it now or best offer freeday returns	Teknologi (0.0007225434)
4	bbcb selama aku data pasien bobol synnovis milik bukti qilin terbit data curi pihak	Teknologi (0.0000000008)
5	ne fifty x bbbcb billionaire boys club ice cream piping original authentic type fitted hat size cm tag	Politik (0.0000003480)

35	barusan petru clej berita bbbcb chisinau moldovapetru clejbbcbchisinau asa ibu kota eropa warga moldova	Politik (0.0004816956)
36	new keystudio smart home kit for bbbcb microbit kit tokopedia promo guna baru bebas ongkir cicil kurir instan	Teknologi (0.0000083531)
37	oran knowlson lansir bbbcb tahu derita sindrom lennoxgastaut bentuk epilepsi resistan obat	Teknologi (0.0000006961)
38	fitur compass produk elektronik ganggu alat listrik medan magnet komponen keras suara	Teknologi (0.0000000008)
39	perdana tri rishi sunak umum tanggal milu tri skotlandia alister jack kata bbbcb	Politik (0.0007225434)
40	belanja kaset pita led zeppelin bbbcb sessions harga murah lapak bebas toko kirim cepat bayar aman	Kategori Tidak Ditemukan

Data Terklasifikasi: 31 Dokumen
Data Tidak Terklasifikasi: 9 Dokumen
Total Execution Time: 0.0072022318840027 Mins

Gambar 8 Implementasi Sistem Klasifikasi Data Testing

Berdasarkan data klasifikasi diatas didapatkan presentase data dapat terklasifikasi sebesar 77,5% dan data tidak dapat terklasifikasi sebesar 22,5%. Dimana kategori hasil dalam skala pengukuran menggunakan skala ordinal dengan kategori(Nuryanita & Malika, 2021):
 a) Kategori baik jika nilainya $\geq 76-100\%$
 b) Kategori cukup jika nialinya $60 - 75\%$
 c) Kategori kurang jika nilainya $\leq 60\%$
 Sehingga hasil presentase implementasi sistem dinyatakan baik.

KESIMPULAN

Berdasarkan hasil penelitian penerapan Electronic Data Processing dapat meningkatkan efektivitas dan efisiensi dalam proses text mining. Sebagai contoh, pada gambar 2, tahap perhitungan probabilitas data untuk 60 dokumen hanya memerlukan waktu 0,093346003691355 menit. Sedangkan pada gambar 3, proses klasifikasi data untuk 40 dokumen hanya memakan waktu 0,0069851199785868 menit. Hasilnya menunjukkan bahwa 77,5% data dapat terklasifikasi dengan baik, sementara 22,5% tidak dapat terklasifikasi.

REFERENSI

- [1.] Andriani, N., & Wibowo, A. (2021). Implementasi Text Mining Klasifikasi Topik Tugas Akhir Mahasiswa Teknik Informatika Menggunakan Pembobotan TF-IDF dan Metode Cosine Similarity Berbasis Web. *Senamika*, September, 130–137.
- [2.] Asari, A., Zulkarnaini, Hartatik, Choirul, A. A., Suparto, Vonny, J. L., Mubina, F. D., Rini, D. P., Maswar, Alfa, W. S., Sari, N. M., & Sukwika, T. (2023). *Pengantar Statistika*.
- [3.] Deolika, A., Kusriani, K., & Luthfi, E. T. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. *Jurnal Teknologi Informasi*, 3(2), 179. <https://doi.org/10.36294/jurti.v3i2.1077>
- [4.] Fathonah, F., & Herliana, A. (2021). Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes. *Jurnal Sains Dan Informatika*, 7(2), 155–164. <https://doi.org/10.34128/jsi.v7i2.331>
- [5.] Febrianty, P., Novianty, I., & Ernita, E. S. (2021). Indonesian Accounting Research Journal Analisis Efektivitas Dan Efisiensi Anggaran Belanja Langsung Analysis Of The Effectiveness And Efficiency Of The Direct Expenditure Budget (Case Study On Bandung City Social Service) Ira Novianty Etti Ernita Sembirin. *Indonesian Accounting Research Journal*, 2(1), 74–82.
- [6.] Hermawan, A., Jowensen, I., Junaedi, J., & Edy. (2023). Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine. *JST (Jurnal Sains Dan Teknologi)*, 12(1), 129–137. <https://doi.org/10.23887/jstundiksha.v12i1.52358>
- [7.] Imam, D. S., Rahmadani, D., Rambe, M., Abdul, M. F., Fadillah, P. H., Siagian, S., & Wulandari, S. (2022). Efektivitas Para Pelaku Ekonomi Dalam Menunjang Pertumbuhan Ekonomi Indonesia. *Jurnal Inovasi Penelitian*, 2(10), 3201–3208.
- [8.] Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information (Switzerland)*, 10(4), 1–68. <https://doi.org/10.3390/info10040150>
- [9.] Lestari, V., Hariki, N. M., & Izzalqurny, T. R. (2022). Analisis Dampak Electronic Data Processing (EDP) Terhadap Pelaksanaan Audit Internal Bank Rakyat Indonesia Tbk Wilayah Bondowoso. *Prosiding National Seminar on Accounting, Finance, and Economics (NSAFE)*, 2(6), 98–105.
- [10.] Nota, G., Postiglione, A., & Carvello, R. (2022). Text mining techniques for the management of predictive maintenance. *Procedia Computer Science*, 200, 778–792. <https://doi.org/10.1016/j.procs.2022.01.276>
- [11.] Nurhadi, A. (2016). Implementasi Algoritma Naïve Bayes Classifier Berbasis Particle Swarm Optimization (PSO) Untuk Klasifikasi Konten Berita Digital Berbahasa Indonesia. *Journal Speed – Sentra Penelitian Engineering Dan Edukasi*, 8(3), 48–56. <https://core.ac.uk/download/pdf/228814999.pdf>
- [12.] Nuryanita, I., & Malika, R. (2021). Prosiding SEMNAS BIO 2021 Tingkat Pengetahuan Remaja terhadap Kesehatan Reproduksi. *Prosiding Semnas BIO*, 98–106.
- [13.] Pahmawati, R., Darna, N., & Herlina, E. (2020). Pengaruh etos kerja dan konseling terhadap efektivitas kerja pegawai (Suatu Studi pada Kantor Dinas Kependudukan dan Pencatatan Sipil Kabupaten Ciamis). *Jurnal Bussiness Management And Entrepreneurship*, 2(4), 1–12.
- [14.] Rini, E. Y., & Solecha, K. (2022). Pengujian Algoritma Teks Mining Untuk Klasifikasi Analisis Review Aplikasi Halodoc. *Jurnal Teknik Komputer AMIK BSI*, 8(2), 174–180. <https://doi.org/10.31294/jtk.v4i2>
- [15.] Septriawan, M. R. (2021). Pengolahan Data Akuntansi Dengan Sistem Electronic Data Processing Pada Pt. Mulia Tani Jaya. *Warta*

- Dharmawangsa*, 15(4), 487–492.
<https://doi.org/10.46576/wdw.v15i4.1523>
- [16.] Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- [17.] Tandel, S. S., Jamadar, A., & Dudugu, S. (2019). A Survey on Text Mining Techniques. *2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019, Icaccs*, 1022–1026.
<https://doi.org/10.1109/ICACCS.2019.8728547>
- [18.] Tri, D. Y., Sukarno, H., & Maria, S. W. (2018). Pengaruh Belanja Modal Dan Alokasi Dana Desa Terhadap Kemandirian Dan Kinerja Keuangan Desa Di Kabupaten Jember. *Bisma (Jurnal Bisnis Dan Manajemen)*, 12(1), 37–50.