

Analisis Part of Tagging Bahasa Indonesia tentang Swamedikasi Pada Dialog Interactive Qestion dengan Brill TAGGER

Maharani^{1*)}, Lili Andraini²⁾
¹Informatika
²Teknik Komputer
*) Maharani33@gmail.com

Abstrak

Part of Speech Tagging merupakan kegiatan pemberian label kelas kata pada suatu kata sehingga akan diketahui keterangan dari masing-masing kata. Dalam membangun sistem interactive question answering (IQA), part of speech tagging dibutuhkan pada tahapan awal sistem IQA. Sistem IQA memberikan kesempatan kepada pengguna untuk bertanya dan memperjelas pertanyaan sampai didapatkan sebuah informasi yang diinginkan berdasarkan pada feedback yang diberikan oleh sistem. Swamedikasi adalah kegiatan pemilihan dan penggunaan obat baik itu obat modern, herbal, maupun obat tradisional oleh seorang individu untuk mengatasi penyakit atau gejala penyakit.

Hasil yang diperoleh berdasarkan pengujian yang dilakukan adalah sebesar 70,37037 %. Pengujian ini dilakukan dengan menggunakan cross validation dan recall. Pada cross validation didapatkan hasil bahwa terdapat 86 kata yang tidak terlabel (tertagging) dengan benar dan 269 kata tertagging dengan benar termasuk tanda baca, sedangkan pada pengujian recall dilakukan dengan melakukan pengujian sebanyak 10 kali pengujian yang terdiri dari masing-masing 5 kalimat dan menghasilkan akurasi yang berbeda-beda sehingga diperoleh akurasi rata-ratanya sebesar 70,37037%.

Kata Kunci: tokenisasi, normalisasi kata, part of speech tagging, brill tagger.

PENDAHULUAN

Question Answering (Tanya Jawab) merupakan suatu task dalam Natural Language Processing (NLP) yang secara otomatis akan memberikan jawaban dari pertanyaan yang diberikan dalam bahasa alami. Sistem Question Answering dibangun melalui beberapa tahapan dimulai dari tahap Pre-Processing sampai Answer Generator (Indrayuni, 2019);(Parnabhakti & Ulfa, 2020). Tahap Pre-Processing meliputi Tokenizer, normalisasi sampai dengan Part of speech tagging. Part-of-Speech Tagging (Postagging) adalah proses pemberian label setiap kata dalam urutan kata-kata dengan tag yang menunjukkan kategori lexical syntactic dengan mengasumsikan urutan yang diberikan. Part-of-Speech Tagging merupakan suatu proses untuk menandai kata-kata dalam teks .

Pendekatan yang kami gunakan dalam part of speech tagging ini adalah pendekatan berbasis rule(aturan) dan transformasi yaitu dengan menggunakan metode brill tagger

(Suaidah & Sidni, 2018);(Abidin, Wijaya, et al., 2021). proses tokenisasi digunakan untuk memecah kalimat menjadi kata agar dapat dikenali oleh system, tokenizer dalam penelitian ini yaitu tokenisasi yang dapat mengelompokkan kata frase dan bukan kata frase. Frasa adalah kata yang terdiri dari lebih dari satu kata tetapi memiliki arti (Abidin & Permata, 2021);(Abidin, Permata, et al., 2021). Kemudian ada langkah normalisasi kata yang digunakan untuk menormalkan kata-kata yang salah ketik. Dalam dialog, pengguna sering menggunakan kata yang disingkat, atau pengguna sering salah ketik, sehingga harus dilakukan proses normalisasi (Ambarwari et al., 2020);(Aziz & Fauzi, 2022);(Pajar & Putra, 2021). Normalisasi kata dilakukan dengan menggunakan metode jarak Levenshtein untuk menghitung jarak antar kata dari kata yang dimasukkan oleh pengguna dan kata dalam kamus normalisasi (Styawati & Mustofa, 2019);(Aldino, Saputra, et al., 2021).

KAJIAN PUSTAKA

Data

Data adalah kumpulan informasi atau deskripsi fakta yang diperoleh melalui pengamatan atau pencarian sumber-sumber tertentu (Darwis et al., 2021);(Mahmuda et al., 2021);(Darwis & Pauristina, 2020). Data yang diterima tetapi tidak diproses lebih lanjut mungkin faktual atau hipotetis. Data kualitatif didefinisikan, atau kadang-kadang disebut data yang secara kasar mencirikan sesuatu. Data ini dapat diamati dan dicatat (*Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis*, 2021);(Satria & Haryadi, 2018);(Al-Ayyubi et al., 2021). Tipe data ini bukan numerik. Jenis data ini dikumpulkan melalui observasi, wawancara satu lawan satu, melakukan kelompok fokus, dan metode serupa (Oktaviani, 2021);(Aminatun & Oktaviani, 2019);(rusliyawati et al., 2020).

Tokenisasi

Tokenisasi adalah proses memecah teks panjang menjadi bagian-bagian yang lebih kecil, memecah urutan kata dalam kalimat, paragraf, atau halaman menjadi token atau fragmen kata, atau kata-kata yang muncul dengan sendirinya (Rahman Isnain et al., 2021). Lanjutkan pemrosesan saat kalimat adalah token. Tokenisasi menghilangkan huruf dan simbol selain a-z, dan jeda kalimat dan kata dilakukan berdasarkan spasi dalam kalimat

(Borman & Putra, 2018). Fase ini juga menghapus karakter tertentu seperti tanda baca dan mengubah semua token menjadi huruf kecil (lower case).

Levenshtein Distance

Dalam teori informasi dan ilmu komputer, jarak Levenshtein adalah matriks untuk mengukur perbedaan antara dua string (Aldino, Hendra, et al., 2021);(Ahdan et al., 2020). Jarak Levenshtein dari dua string adalah jumlah minimum operasi yang diperlukan untuk mengubah string (source string) menjadi string yang lain (target string), dimana suatu operasi melibatkan penyisipan (insertion), penghapusan(deletion), dan penggantian (substitution) dari suatu karakter tunggal. Levenshtein Distance sering digunakan pada aplikasi untuk menentukan seberapa mirip, atau perbedaan dua buah string, seperti aplikasi pengecekan suatu ejaan, atau yang biasa dikenal dengan spell checkers (Ahmad et al., 2021);(Sarasvananda et al., 2021).

MWE Tokenizer

MWE Tokenizer merupakan proses pemotongan kata berdasarkan spasi tetapi juga mempertimbangkan karakter spesial seperti tanda seru, dan tanda tanya untuk kalimat pembatas (Riskiono & Pasha, 2020);(Styawati et al., 2021). Selain karakter diatas, baris baru dianggap sebagai pembatas kalimat. Kata multi ekspresi biasanya terdiri dari dua kata atau lebih (Aldino, Darwis, et al., 2021);(Sulistiani, Wardani, et al., 2019).

Brill Tagger

Brill Tagger merupakan program tagging yang diperkenalkan oleh Eric Brill pada tahun 1992. Secara umum Brill Tagger disebut juga Transformation-based Error-driven Learning (TEL). Brill Tagger ini didasarkan pada aturan atau transformasi, yaitu tata bahasa yang diinduksi langsung dari pelatihan corpus tanpa adanya campur tangan manusia atau ahli pengetahuan (Fakhrurozi & Puspita, 2021);(Puspaningrum & Susanto, 2021). Brill Tagger melakukan pemberian anotasi (tag) pada corpus dengan dua aturan, yaitu aturan leksikal untuk memberi tag pada unknownword (kata yang tidak diketahui) dan aturan kontekstual untuk memperoleh aturan yang meningkatkan akurasi dari perbaikan skor dan kata ambigu (Rusliyawati et al., 2021).

Cross Validation

Metode statistik yang dapat digunakan untuk mengevaluasi kinerja model atau algoritma. Pisahkan data menjadi dua bagian: data proses pembelajaran dan data validasi/evaluasi (Sulistiani & Aldino, 2020);(Sulistiani, Muludi, et al., 2019);(Samsugi et al., 2021). Model atau algoritma dilatih oleh subset pelatihan dan divalidasi oleh subset validasi (Sulistiani & Tjahyanto, 2016);(Alita et al., 2021). Validasi silang ini mendefinisikan kumpulan data untuk "menguji" model selama fase pelatihan, membatasi masalah seperti overfitting yang terjadi, dan bagaimana model digeneralisasikan secara independen dari kumpulan data (Abidin, 2021);(Alita et al., 2020).

METODE

Observasi

Paper ini menggunakan corpus bahasa Indonesia mengenai kesehatan. Pada tahapan normalisasi kata menggunakan kamus normalisasi yang didapat dari <http://Indodic.com/indoWordList.zip> yang berjumlah 41057 kata yang kemudian dilakukan penyortiran kata untuk bidang kesehatan sehingga didapatkan 2578 kata yang digunakan untuk sebagai kamus kata. Selanjutnya pada tahapan MWE Tokenizer menggunakan kamus multiword bahasa Indonesia yang terdiri dari 20677 kata yang kemudian dilakukan penyortiran sehingga menghasilkan 250 kata frase selanjutnya ditambahkan kata frase sebanyak 264 kata yang diperoleh dari internet serta pedoman penggunaan obat bebas dan bebas terbatas tahun 2008 dan sudah divalidasi menggunakan kamus besar bahasa Indonesia. Pada tahapan part of speech tagging menggunakan corpus yang berasal dari penelitian Pisceldo dkk dan buku pedoman obat bebas dan obat bebas terbatas tahun 2008 dengan dilakukan penyortiran terlebih dahulu sehingga didapatkan jumlah kata sebanyak 1583 yang berhubungan dengan kesehatan.

Part of speech tagging

Post tagging kami diawali dengan proses tokenisasi dilanjutkan dengan normalisasi kata. Pada paper ini proses tokenizer dilakukan dengan menggunakan MWE Tokenizer dan Normalisasi kata menggunakan Levenshtein Distance sedangkan untuk proses tagging itu sendiri menggunakan metode Brill Tagger.



Gambar 1 Part Of Speech Tagging

Tokenisasi

Sistem akan memisahkan pertanyaan berdasarkan spasi dan mengubah semua token ke bentuk huruf kecil.

Normalisasi Kata

Normalisasi kata yaitu proses dimana semua input kalimat yang sudah dilakukan tokenisasi, dinormalisasikan terlebih dahulu sesuai dengan bahasa Indonesia yang benar dengan tujuan agar dapat disesuaikan dengan korpus yang digunakan untuk melakukan tahapan Pre-processing selanjutnya yaitu post tagging. Dalam proses normalisasi pada penelitian ini menggunakan algoritma Levenshtein Distance dengan menghitung jarak kedekatan antar kata. Perhitungan atau pengecekan dilakukan perkata yang ada didalam kamuskata. Contoh pertanyaan user yang mengandung kata singkatan atau tidak baku “Apa efek samping dri paracetamol?” Pada pertanyaan tersebut kata “dri” akan dirubah kebentuk kata normal yaitu kata “dari”.

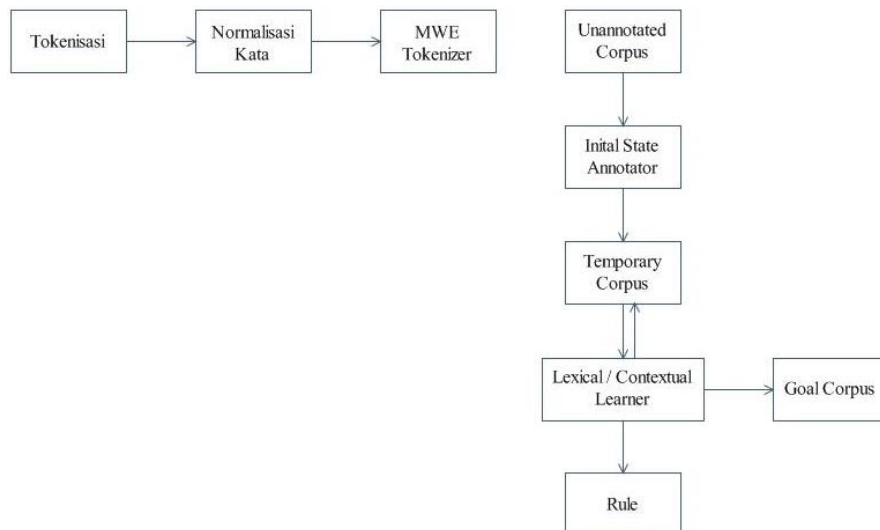
MWE tokenizer

MWE Tokenizer Proses pemotongan kata yang terdiri dari satu atau lebih kata namun memiliki arti yang sama. Setiap dialog pertanyaan akan dipisahkan berdasarkan spasi mempertimbangkan tanda baca seperti tanda tanya, tanda seru sebagai pembatas dialog dan dengan memperhatikan apakah kata tersebut termasuk kata frase atau kata bukan frase.

Aturan yang dibuat untuk proses pemotongan kata berdasarkan struktur pohon yang telah dibuat pada corpus. Dimana aturan tersebut adalah sistem akan mengecek setiap token pada corpus, jika kata tersebut terdapat pada corpus maka sistem akan memeriksa kata selanjutnya sampai kata selanjutnya yang tidak cocok dalam membentuk kata frase dan jika kata cocok dalam membentuk kata frase maka sistem akan memberikan token kedalam satu token. Jika token tidak terdapat pada corpus, maka sistem akan memberikan token terpisah berdasarkan spasi.

Tagging

Berikut ini adalah proses tagging yang digunakan dengan mengimplementasikan metode brill tagger. Proses tersebut adalah sebagai berikut :



Gambar 2 Proses Tagging

1. Unannotated corpus adalah data yang nantinya akan di tag (training).
2. Initial state anotator adalah membaca token atau posisi frekuensi kata yang muncul atau dengan kata lain mempersiapkan untuk ditagging.
3. Temporary corpus adalah penyimpanan sementara dari hasil initial state annotator.
4. Lexical/Contextual Learner adalah proses rule melakukan pembelajaran lexical dan contextual dari goal corpus.
5. Goal corpus adalah bertugas sebagai bahan atau contoh untuk data training sesuai dengan contoh temporary corpus yang akan dibandingkan dengan lexical dan contextual learner. Setelah diproses maka rule mengambil temporary dari corpus yang sudah dipersiapkan. Aturan lexical dan contextual yang berasal dari goal corpus akan diterapkan ke dalam training corpus, maka akan terbentuk rule-rule untuk membuat tagging.

HASIL DAN PEMBAHASAN

Pengujian dilakukan dengan cara membandingkan setiap kata pada data training dengan data testing, dalam hal ini menggunakan recall dan cross validation.

A. Cross Validation Berdasarkan pengujian cross validation dari 300 kata dan 55 tanda baca yang diuji dengan 18 tagset maka didapatkan hasil bahwa terdapat 86 kata yang tidak terlabel (tertagging) dengan benar dan 269 kata tertagging dengan benar termasuk tanda baca.

B. Pengujian dengan menggunakan recall digunakan untuk mendapatkan akurasi. Pengujian dengan recall.

Pada tabel dibawah ini menunjukkan kesepuluh hasil pengujian recall, masing-masing berjumlah 5 kalimat. Pengujian kalimat diuji secara acak berdasarkan kata-kata pada data training. Berdasarkan pengujian tersebut didapatkan rata-rata pengujian recall sebesar 70,37037%.

Pengujian	Jumlah Kalimat	Jumlah Kata	Jumlah Tanda Baca	Total Benar	Total Salah	Akurasi Recall
1	5	16	5	20	1	95,2380%
2	5	24	5	21	8	72,4138%
3	5	29	5	19	15	55,8823%
4	5	29	5	21	13	61,7647%
5	5	40	7	34	13	72,3404%
6	5	19	3	18	4	81,8181%
7	5	44	4	37	11	77,0833%
8	5	34	5	28	11	71,7949%
9	5	28	5	20	13	60,6060%
10	5	37	5	23	19	54,7619%
Akurasi Rata-rata Recall						70,37034%

SIMPULAN

Kesimpulan yang dihasilkan dari penelitian ini yaitu:

1. Normalisasi kata dapat memudahkan proses tagging jika terjadi kesalahan penulisan pada pertanyaan yang diinputkan oleh pengguna
2. MWE tokenizer berhasil mengelompokkan kata yang termasuk frase dan bukan frase sehingga akan diperoleh 1 tag untuk kata frase.
3. Proses tagging menggunakan metode brill tagger dan kata uji sebanyak 269 kata menghasilkan akurasi sebesar 70,37%.

REFERENSI

- Abidin, Z. (2021). Pelatihan Dasar-Dasar Algoritma Dan Pemograman Untuk Membangkitkan Minat Siswa-Siswi Smk Pada Dunia Pemograman. *Journal of Social Sciences and Technology for Community Service (JSSTCS)*, 2(2), 54. <https://doi.org/10.33365/jsstcs.v2i2.1326>
- Abidin, Z., Permata, Ahmad, I., & Rusliyawati. (2021). Effect of mono corpus quantity on statistical machine translation Indonesian-Lampung dialect of nyo. *Journal of Physics: Conference Series*, 1751(1). <https://doi.org/10.1088/1742-6596/1751/1/012036>
- Abidin, Z., & Permata, P. (2021). Pengaruh Penambahan Korpus Paralel Pada Mesin Penerjemah Statistik Bahasa Indonesia Ke Bahasa Lampung Dialek Nyo. *Jurnal Teknoinfo*, 15(1), 13. <https://doi.org/10.33365/jti.v15i1.889>
- Abidin, Z., Wijaya, A., & Pasha, D. (2021). Aplikasi Stemming Kata Bahasa Lampung Dialek Api Menggunakan Pendekatan Brute-Force dan Pemograman C. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(1), 1–8.
- Ahdan, S., Putri, A. R., & Sucipto, A. (2020). Aplikasi M-Learning Sebagai Media Pembelajaran Conversation Pada Homey English. *Sistemasi*, 9(3), 493. <https://doi.org/10.32520/stmsi.v9i3.884>
- Ahmad, I., Borman, R. I., Caksana, G. G., & Fakhrurozi, J. (2021). IMPLEMENTASI STRING MATCHING DENGAN ALGORITMA BOYER-MOORE UNTUK MENENTUKAN TINGKAT KEMIRIPAN PADA PENGAJUAN JUDUL SKRIPSI/TA MAHASISWA (STUDI KASUS: UNIVERSITAS XYZ). *SINTECH (Science and Information Technology) Journal*, 4(1), 53–58.
- Al-Ayyubi, M. S., Sulistiani, H., Muhaqiqin, M., Dewantoro, F., & Isnain, A. R. (2021). Implementasi E-Government untuk Pengelolaan Data Administratif pada Desa Banjar Negeri, Lampung Selatan. *E-Dimas: Jurnal Pengabdian Kepada Masyarakat*, 12(3), 491–497. <https://doi.org/10.26877/e-dimas.v12i3.6704>
- Aldino, A. A., Darwis, D., Prastowo, A. T., & Sujana, C. (2021). Implementation of K-

means algorithm for clustering corn planting feasibility area in south lampung regency. *Journal of Physics: Conference Series*, 1751(1), 12038.

- Aldino, A. A., Hendra, V., & Darwis, D. (2021). Pelatihan Spada Sebagai Optimalisasi Lms Pada Pembelajaran Di Masa Pandemi Covid 19. *Journal of Social Sciences and Technology for Community Service (JSSTCS)*, 2(2), 72. <https://doi.org/10.33365/jsstcs.v2i2.1330>
- Aldino, A. A., Saputra, A., & Nurkholis, A. (2021). *Application of Support Vector Machine (SVM) Algorithm in Classification of Low-Cape Communities in Lampung Timur*. 3(3), 325–330. <https://doi.org/10.47065/bits.v3i3.1041>
- Alita, D., Fernando, Y., & Sulistiani, H. (2020). Implementasi Algoritma Multiclass SVM pada Opini Publik Berbahasa Indonesia di Twitter. *Jurnal Tekno Kompak*, 14(2), 86–91.
- Alita, D., Putra, A. D., & Darwis, D. (2021). Analysis of classic assumption test and multiple linear regression coefficient test for employee structural office recommendation. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(3), 1–5.
- Ambarwari, A., Adria, Q. J., Herdiyeni, Y., & Hermadi, I. (2020). Plant species identification based on leaf venation features using SVM. *Telkomnika*, 18(2), 726–732.
- Aminatun, D., & Oktaviani, L. (2019). Memrise: Promoting Students' Autonomous Learning Skill through Language Learning Application. *Metathesis: Journal of English Language, Literature, and Teaching*, 3(2), 214. <https://doi.org/10.31002/metathesis.v3i2.1982>
- Aziz, M., & Fauzi, A. (2022). *CNN UNTUK DETEKSI BOLA MULTI POLA STUDI KASUS : LIGA HUMANOID ROBOCUP CNN For Multi Pattern Ball Detection Case Study : RoboCup Humanoid League*. 5(1), 23–34.
- Borman, R. I., & Putra, A. S. (2018). Game Pengenalan Huruf Hijaiyah Untuk Anak Autis Dengan Penerapan Pendekatan Edukasi Multisensori. *SEMNASTEKNOMEDIA ONLINE*, 6(1), 1–6.
- Darwis, D., & Pauristina, D. M. (2020). AUDIT SISTEM INFORMASI MENGGUNAKAN FRAMEWORK COBIT 4.1 SEBAGAI UPAYA EVALUASI PENGOLAHAN DATA PADA SMKK BPK PENABUR BANDAR LAMPUNG. *Jurnal Ilmiah Infrastruktur Teknologi Informasi*, 1(1), 1–6.
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131–145.
- Fakhrurozi, J., & Puspita, D. (2021). KONSEP PIIL PESENGGIRI DALAM SASTRA LISAN WAWANCAN LAMPUNG SAIBATIN. *JURNAL PESONA*, 7(1), 1–13.
- Indrayuni, E. (2019). Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks

- Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Khatulistiwa Informatika*, 7(1), 29–36. <https://doi.org/10.31294/jki.v7i1.1>
- Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis, (2021).
- Mahmuda, S., Sucipto, A., & Setiawansyah, S. (2021). Pengembangan Sistem Informasi Pengolahan Data Tunjangan Karyawan Bulog (TKB)(Studi Kasus: Perum Bulog Divisi Regional Lampung). *Jurnal Ilmiah Sistem Informasi Akuntansi*, 1(1), 14–23.
- Oktaviani, L. (2021). Penerapan Sistem Pembelajaran Dalam Jaringan Berbasis Web Pada Madrasah Aliyah Negeri 1 Pesawaran. *Jurnal WIDYA LAKSMI (Jurnal Pengabdian Kepada Masyarakat)*, 1(2), 68–75.
- Pajar, M., & Putra, K. (2021). A Novel Method for Handling Partial Occlusion on Person Re-identification using Partial Siamese Network. 12(7), 313–321.
- Parnabhakti, L., & Ulfa, M. (2020). Perkembangan Matematika Dalam Filsafat. *Jurnal Ilmiah Matematika Realistik*, 1(1), 11–14.
- Puspaningrum, A. S., & Susanto, E. R. (2021). Penerapan Dan Pelatihan e-Learning Pada SMA Tunas Mekar Indonesia. *Jurnal Pengabdian Kepada Masyarakat (JPKM) TABIKPUN*, 2(2), 91–100.
- Rahman Isnain, A., Indra Sakti, A., Alita, D., & Satya Marga, N. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm. *Jdmsi*, 2(1), 31–37. <https://t.co/NfhnfMjtXw>
- Riskiono, S. D., & Pasha, D. (2020). Analisis Metode Load Balancing Dalam Meningkatkan Kinerja Website E-Learning. *Jurnal TeknoInfo*, 14(1), 22–26.
- rusliyawati, rusliyawati, Suryani, A. D., & Ardian, Q. J. (2020). Rancang Bangun Identifikasi Kebutuhan Kalori Dengan Aplikasi Go Healthy Life. *Jurnal Teknologi Dan Sistem Informasi*, 1(1), 47–56. <http://jim.teknokrat.ac.id/index.php/sisteminformasi/article/view/51>
- Rusliyawati, R., Muludi, K., Wantoro, A., & Saputra, D. A. (2021). Implementasi Metode International Prostate Symptom Score (IPSS) Untuk E-Screening Penentuan Gejala Benign Prostate Hyperplasia (BPH). *Jurnal Sains Dan Informatika*, 7(1), 28–37.
- Samsugi, S., Nurkholis, A., Permatasari, B., Candra, A., & Prasetyo, A. B. (2021). Internet of Things Untuk Peningkatan Pengetahuan Teknologi Bagi Siswa. *Journal of Technology and Social for Community Service (JTSCS)*, 2(2), 174.
- Sarasvananda, I. B. G., Anwar, C., Donaya, P., & Styawati. (2021). ANALISIS SURVEI KEPUASAN MASYARAKAT MENGGUNAKAN PENDEKATAN E-CRM (Studi Kasus: BP3TKI Lampung). ... *Dan Sistem Informasi*, 2(1), 1–9. <https://ejurnal.teknokrat.ac.id/index.php/JDMSI/article/view/1026>
- Satria, M. N. D., & Haryadi, S. (2018). Effect of the content store size to the performance of named data networking: Case study on Palapa Ring topology. *Proceeding of 2017*

11th International Conference on Telecommunication Systems Services and Applications, TSSA 2017, 2018-Janua, 1–5.
<https://doi.org/10.1109/TSSA.2017.8272911>

- Styawati, Andi Nurkholis, Zaenal Abidin, & Heni Sulistiani. (2021). Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(5), 904–910. <https://doi.org/10.29207/resti.v5i5.3380>
- Styawati, S., & Mustofa, K. (2019). A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(3), 219–230.
- Suaidah, S., & Sidni, I. (2018). Perancangan Monitoring Prestasi Akademik dan Aktivitas Siswa Menggunakan Pendekatan Key Performance Indicator (Studi Kasus SMA N 1 Kalirejo). *Jurnal Tekno Kompak*, 12(2), 62–67.
- Sulistiani, H., & Aldino, A. A. (2020). Decision Tree C4.5 Algorithm for Tuition Aid Grant Program Classification (Case Study: Department of Information System, Universitas Teknokrat Indonesia). *EduTic - Scientific Journal of Informatics Education*, 7(1), 40–50. <https://doi.org/10.21107/edutic.v7i1.8849>
- Sulistiani, H., Muludi, K., & Syarif, A. (2019). Implementation of Dynamic Mutual Information and Support Vector Machine for Customer Loyalty Classification. *Journal of Physics: Conference Series*, 1338(1). <https://doi.org/10.1088/1742-6596/1338/1/012050>
- Sulistiani, H., & Tjahyanto, A. (2016). Heterogeneous feature selection for classification of customer loyalty fast moving consumer goods (Case study: Instant noodle). *Journal of Theoretical and Applied Information Technology*, 94(1), 77–83. <https://doi.org/10.5281/zenodo.579836>
- Sulistiani, H., Wardani, F., & Sulistyawati, A. (2019). Application of Best First Search Method to Search Nearest Business Partner Location (Case Study: PT Coca Cola Amatil Indonesia, Bandar Lampung). *2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)*, 102–106.