

PERANCANGAN SISTEM KLASIFIKASI PASIEN STROKE DENGAN METODE K-NEAREST NEIGHBOR

¹⁾ Rahmat Ardila Dwi Yulianto, ²⁾ Imam Riadi, ³⁾ Rusydi Umar

¹⁾Magister Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan

²⁾Sistem Informasi, Fakultas Sains dan Teknologi Terapan, Universitas Ahmad Dahlan

³⁾Teknik Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan

^{1,2,3)} Jl. Ahmad Yani (Ringroad Selatan) Tamanan Banguntapan Bantul Yogyakarta 55166, Indonesia

E-mail : 2108048016@webmail.ac.id, imam.riadi@is.uad.ac.id, rusydi@mti.uad.ac.id

ABSTRAK

Stroke merupakan penyakit yang ditandai gangguan fungsi otak yang di sebabkan kurangnya pasokan oksigen dan aliran darah ke otak sehingga mempengaruhi beberapa fungsi otak yang membuat penderita mengalami kesulitan dalam melakukan aktifitas. klasifikasi pasien stroke yang di temukan masih berupa catatan medis yang belum terintegrasi sehingga perlu waktu yang lebih lama untuk mendeteksi. Algoritma K-NN merupakan bagian dari algoritma machine learning yang dapat digunakan untuk mengklasifikasikan salah satu kasusnya yaitu klasifikasi pasien stroke. K-NN digunakan sebagai algoritma penentu kelas untuk memasukkan data baru yang diinputkan sesuai format. Berdasarkan hasil yang diperoleh penelitian ini mengarah pada perancangan sistem menggunakan Unified Model Language (UML) dan perancangan antarmuka pengguna sistem.

Kata Kunci: Klasifikasi, K-NN, Stroke, UML.

ABSTRACT

Stroke is a disease characterized by impaired brain function caused by a lack of oxygen supply and blood flow to the brain, affecting several brain functions that make sufferers experience difficulty in carrying out activities. the classification of stroke patients found is still in the form of medical records that have not been integrated so it takes longer time to detect. The K-NN algorithm is part of a machine learning algorithm that can be used to classify one of the cases, namely the classification of stroke patients. K-NN is used as a class determining algorithm to enter new data that is input according to the format. Based on the results obtained, this study leads to system design using the Unified Model Language (UML) and system user interface design.

Keyword: Classification, K-NN, Stroke, UML

PENDAHULUAN

Saat ini, stroke merupakan masalah utama kesehatan masyarakat. Stroke adalah penyakit dengan ditandai oleh terganggunya fungsi otak disebabkan kurangnya pasokan oksigen dan aliran darah ke otak sehingga mempengaruhi beberapa fungsi otak yang membuat penyintas mengalami kesulitan dalam melakukan aktifitas [1] Berdasarkan data orang yang menderita stroke pada penduduk berusia ≥ 15 tahun di Indonesia tahun 2018 mengalami peningkatan di bandingkan tahun 2013, dari 7% menjadi 10,9%, atau diperkirakan 2.120.362, menurut diagnosis dokter Stroke menjadi penyakit yang semakin serius yang dapat ditemui hampir di

seluruh negara karena stroke menyerang secara tiba-tiba dan dapat menyebabkan cacat fisik, cacat mental, bahkan sampai kematian. Stroke dapat menyerang pada usia produktif sampai dengan lanjut usia sehingga pencegahan dan penanganan harus sedini mungkin dilakukan [2].

Pencegahan stroke dapat dilakukan meminimalisir faktor risiko yang ada seperti hipertensi, peningkatan kadar lemak, diabetes, dan gaya hidup seperti merokok, dan diet tidak sehat [3]. Penanganan stroke dapat dilakukan dengan mendeteksi dini orang-orang yang memiliki risiko stroke dengan cara merekam catatan medis mereka dengan efektif dan efisien[4]. Namun, tidak semua catatan medis

yang ditemukan masih berupa catatan medis yang belum terintergrasi dengan sistem sehingga perlu waktu yang lebih lama untuk mendeteksi penderita Stroke [5]. Merancang sebuah sistem merupakan solusi yang dapat mempercepat dan mempermudah deteksi seseorang menderita penyakit stroke, dimana sistem tersebut mengimplementasikan machine learning untuk melakukan klasifikasi data dari hasil rekam medis.

Machine learning merupakan bagian dari kecerdasan buatan (AI) [6]. Machine learning bertujuan untuk merancang dan mengembangkan model matematika yang dapat dilatih tanpa pengetahuan lengkap tentang semua faktor eksternal yang mempengaruhi untuk memungkinkan keputusan dibuat secara cerdas [6].

Klasifikasi merupakan aktivitas menilai suatu objek untuk dimasukkan ke dalam kelas tertentu dari beberapa kelas yang tersedia. Ada 2 tugas utama dalam klasifikasi yaitu: membangun model dan menggunakan model untuk melakukan prediksi deteksi/klasifikasi/Ini memungkinkan Anda mengetahui di kelas pada suatu objek data baru [7]. K-Nearest Neighbor K-NN merupakan metode untuk klasifikasi antara objek yang terdekat dengan objek berdasarkan data latih. Data pembelajaran diproyeksikan ke dalam ruang multidimensi, dimana setiap dimensi mewakili fitur data [8]. Dalam penambangan data metode K-NN adalah algoritma yang telah banyak digunakan untuk membantu dalam pengambilan keputusan [9]. K-NN memiliki keunggulan berupa kumpulan data latih yang efektif untuk data set yang noise dan jumlah data training tinggi/besar. Namun kekurangan dari K-NN masih diperlukan untuk menentukan nilai K dan memilih atribut yang terbaik [10].

Pada penelitian sebelumnya K-NN

dimanfaatkan untuk melakukan Klasifikasi Kematangan Buah Mangga yang dilakukan oleh Husnul Khotimah, Nur Nafi'iyah dan Masruro. Menggunakan Data set yang sejumlah 129 data training, juga 40 data testing. Menghasilkan tingkat nilai akurasi tertinggi pada $k = 2$ sebesar 80% [11]. Nilai k mendefinisikan parameter pada algoritma K-NN. Parameter k pada pengujian ditentukan berdasarkan nilai k optimum selama pelatihan. Nilai k optimum diperoleh dengan mencoba-coba [12].

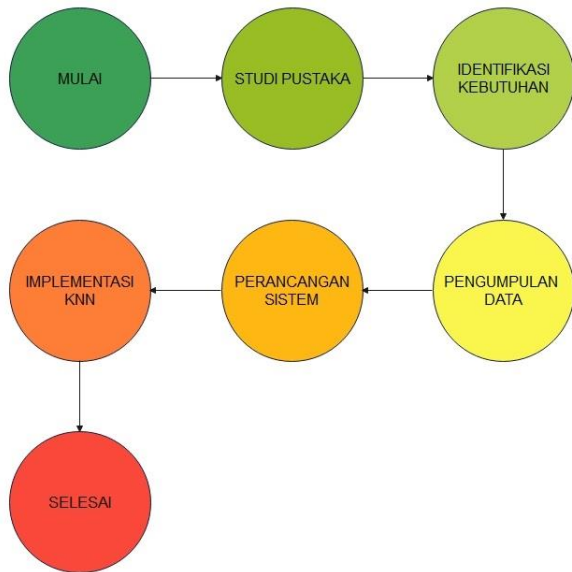
Pada penelitian lain digunakan metode K-NN untuk mengklasifikasi penyakit Stroke. Dari hasil pengujian yang dilakukan, performa algoritma KNN terbaik didapatkan dengan nilai $k=5$ dan akurasi 93.54% [13].

Penelitian lain menggunakan metode K-NN untuk menentukan hubungan antara klasifikasi stroke dengan gangguan fungsi kognitif pada pasien stroke. Jenis penelitian yang digunakan dalam penelitian ini adalah analisa korelasi, dengan rancangan cross-sectional, sampel 36 responden sesuai dengan kriteria, menggunakan Chi Square pada $\alpha = 0,05$. Ada hubungan antara klasifikasi stroke dengan gangguan fungsi kognitif pada pasien stroke dengan nilai $pvalue = 0,000$. [14].

Penelitian ini bertujuan untuk merancang sistem klasifikasi pasien stroke membantu peneliti membangun sistem yang mengacu pada rancangan yang ada.

METODE

Adapun proses penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur proses penelitian

Studi Pustaka

Studi pustaka adalah kegiatan penelitian yang dilakukan dengan mengumpulkan informasi kepustakaan sebagai alat penunjang penelitian yang bersumber dari buku, paper ataupun referensi artikel yang berkaitan dengan penelitian ini (Melinda & Zainil, 2020).

Identifikasi Kebutuhan

Identifikasi kebutuhan diperlukan dalam mendukung pelaksanaan penelitian yang muncul dari tinjauan literatur. Ada 2 kebutuhan yang harus dipetakan, yaitu kebutuhan perangkat keras dan kebutuhan perangkat lunak.

Pengumpulan Data

Data yang digunakan sebagai parameter perancangan sistem pada penelitian ini adalah data penyakit stroke dari kaggle. Ada 11 atribut yaitu jenis kelamin, umur, hipertensi, penyakit jantung, pernah menikah, jenis pekerjaan, jenis tempat tinggal, kadar glukosa rata-rat, indeks massa tubuh (BMI), merokok, dan kelas binary yaitu 0 = tidak stroke dan 1 = stroke.

Perancangan Sistem

Pada proses ini akan dibuat suatu perancangan dari sistem klasifikasi penyakit stroke dengan memanfaatkan Unified Modeling Language

(UML), mengembangkan sebuah perangkat lunak atau sistem membutuhkan sebuah bahasa standar untuk menyampaikan bentuk model oleh karena itulah sebabnya UML digunakan [15]. Use case diagram adalah diagram UML yang mendefinisikan fungsionalitas dan secara grafis dari suatu sistem yang digambarkan dalam istilah aktor, use case, dan relasi [16]. Activity Diagram merupakan diagram penting yang memodelkan aspek dinamis dari suatu sistem untuk alur kerja [17].

Implementasi K-Nearest Neighbor

Implementasi *K-Nearest Neighbor* adalah algoritma yang gampang diimplementasikan dan termasuk dalam *supervised learning* algoritma yang dapat mengatasi tugas yang cukup kompleks. Prinsip pengoperasian K-NN adalah menghitung jarak dari titik data baru ke titik paling dekat dengannya. Ada beberapa fungsi jarak yang dapat digunakan seperti Euclidean, Cosine, Manhattan dimana nilai k adalah bilangan bulat acak [18]. Perhitungan jarak menggunakan fungsi Euclidean yang ditunjukkan pada persamaan 1 sebagai berikut.

$$d(a,b) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (1)$$

Keterangan:

d(a,b) = jarak

i = variabel data

n = dimensi data

ai = data latih

bi = data uji

HASIL

Dalam perancangan Sistem klasifikasi pasien stroke dengan metode K-NN, akan dibuat analisis kebutuhan baik dari sisi perangkat keras dan perangkat lunak, UML digunakan untuk membuat use case diagram dan activity diagram untuk mendeskripsikan sistem yang

akan dibuat juga perancangan desain antarmuka.

Identifikasi Kebutuhan

a. Kebutuhan Perangkat Keras

Tabel 1. List Hardware.

No.	Hardware	Detail
1.	Processor	Intel Core i5 7200
2.	RAM	8 GB
3.	HDD	500 GB

Dari Tabel 1 dapat dilihat daftar perangkat keras yang digunakan untuk menunjang penelitian ini dengan laptop yang digunakan adalah Acer aspire E14 E5-475G.

b. Kebutuhan Perangkat Lunak

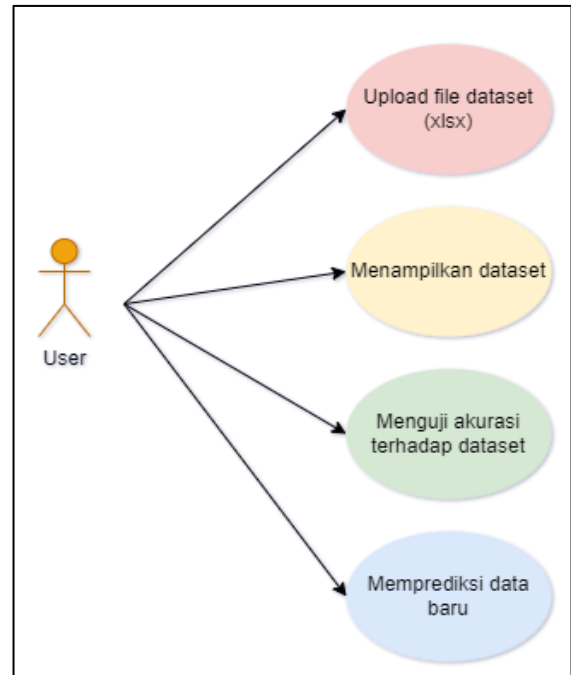
Tabel 2. List Software.

No.	Perangkat lunak
1.	Windows 10 Pro 64-bit
2.	Microsoft Office 2016
3.	Photoshop
4.	Edraw

Dari Tabel 2 terdapat beberapa perangkat lunak yang mendukung penelitian ini diantaranya Microsoft Office digunakan sebagai aplikasi pengolah kata, photoshop untuk melakukan desain antar muka sistem yang akan dibuat, dan draw io untuk pembuatan use case diagram, activity diagram.

Use Case Diagram

Pada tahapan ini akan dibuat use case diagram seperti yang ditunjukkan dalam Gambar 2 sebagai deskripsi functional requirements.



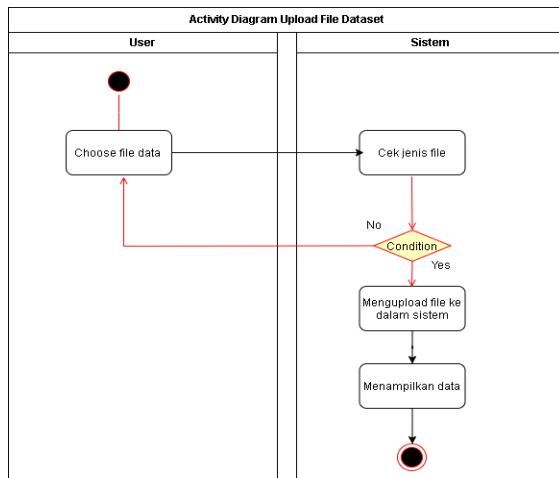
Gambar 2. Use Case Diagram.

Use case digunakan sebagai skenario untuk merepresentasikan fungsionalitas dari sistem yang akan dibangun dimana terdapat 4 proses diantaranya, Upload file yang berfungsi untuk mengunggah dataset yang akan digunakan dengan ekstensi file xlsx. Melihat Dataset proses ini menampilkan data yang sudah ada pada database berupa tabel. Menguji akurasi, dalam proses ini sistem akan melakukan split data menjadi data latih dan data uji untuk diklasifikasi menggunakan algoritma K-NN lalu menampilkan hasil accuracy, precision, recall, f1-score serta hasil validasi dengan cross validation. Dan yang terakhir adalah prediksi data, data baru akan diklasifikasikan berdasarkan jarak data satu dengan yang lainnya.

Activity Diagram

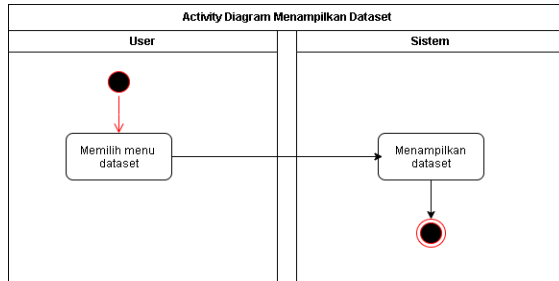
Activity diagram digunakan sebagai penggambaran aktivitas pada sistem yang akan dibuat seperti pada Gambar 3 menjelaskan skema diantara aktor dengan sistem saat melakukan proses pemilihan file dataset, jenis file yang dapat di unggah pada sistem ini hanya

file yang bereksistensi xlsx.



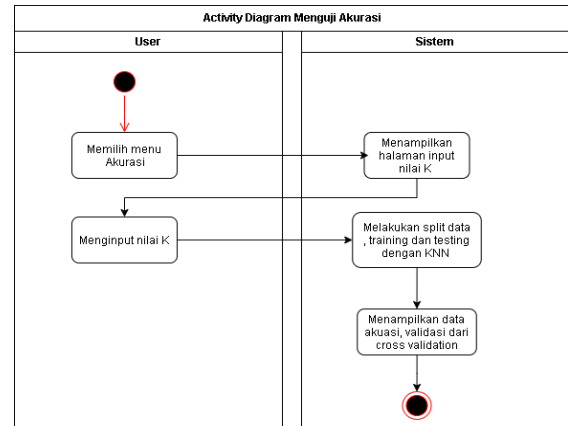
Gambar 3. Activity Diagram Upload File.

Gambar 4 dibawah mendeskripsikan proses jika user ingin menampilkan dataset yang ada dalam sistem dengan cara user memilih menu dataset dan sistem akan menampilkannya dalam bentuk tabel pada layar user.



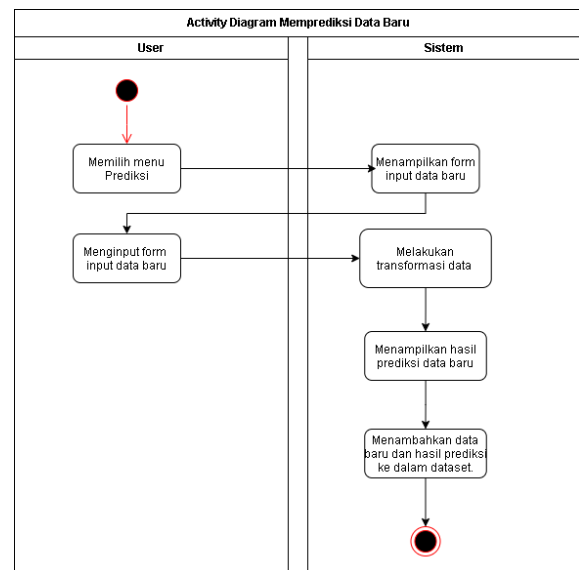
Gambar 4. Activity Diagram Menampilkan Dataset.

Pengujian akurasi dapat dilakukan user dengan memilih menu akurasi, lalu sistem akan meminta user memasukkan nilai bilangan bulat untuk menjadi nilai k seperti activity diagram pada Gambar 5 dibawah.



Gambar 5. Activity Diagram Menguji Akurasi.

User juga dapat melakukan prediksi kelas data baru seperti yang digambarkan pada activity diagram Gambar 6 dibawah dengan menginputkan nilai atribut yang ada. Jika berhasil sistem akan menampilkan prediksi kelas data baru yang ada.



Gambar 6. Activity Diagram Memprediksi Data Baru.

Desain User Interface

a. Halaman Dataset

Pada gambar 6. terdapat rancangan halaman yang akan menampilkan dataset serta klasifikasi dari data tersebut. Jika user melakukan prediksi data baru, maka data yang

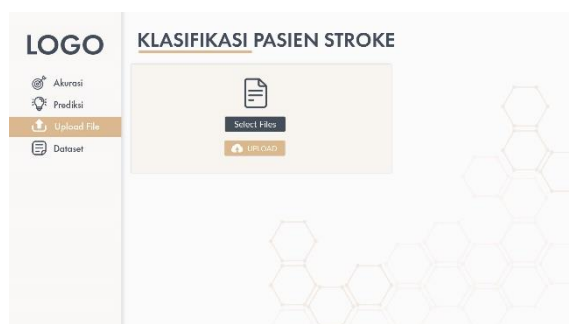
diinputkan tersimpan dalam database dan dapat dilihat pada halaman ini.



Gambar 7. Halaman Dataset.

b. Halaman Upload

Halaman upload ini berfungsi untuk mengunggah data ke dalam database yang akan digunakan untuk dataset sistem ini. File yang diunggah diharuskan bereksistensi .xlsx dengan nama dan jumlah atribut yang ditentukan.



Gambar 8. Halaman Upload Dataset.

c. Halaman Tingkat Akurasi

Halaman ini digunakan user untuk melakukan training dan evaluasi terhadap sistem klasifikasi menggunakan algoritma K-NN. *User* akan diminta untuk menginputkan nilai k bilangan bulat acak. Dan sistem akan menampilkan hasil akurasi, presisi, recall, f1-score.



Gambar 9. Halaman Tingkat Akurasi.

d. Halaman Pemrediksian

Gambar 10 adalah rancangan antarmuka halaman prediksi yang berfungsi untuk melakukan prediksi kelas data baru yang diinputkan oleh user sesuai variabel yang ada dan sesuai dengan format yang sudah ditentukan. Jika berhasil, sistem akan menampilkan hasil prediksi.



Gambar 10. Halaman Prediksi

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa penelitian ini membuat suatu rancangan sistem untuk mengklasifikasikan pasien stroke. Perancangan sistem didasarkan pada UML yang terdiri dari use case diagram dan activity diagram. Terdapat juga desain antarmuka pengguna untuk membantu dalam membangun sistem klasifikasi pasien stroke. Sistem ini dirancang untuk dibuat dalam bentuk platform web, sehingga tampilan antarmuka pengguna dapat di perbaharui dan dikembangkan lebih lanjut.

DAFTAR PUSTAKA

- [1] C. W. Yean *et al.*, “An emotion assessment of stroke patients by using bispectrum features of EEG signals,” *Brain Sci.*, vol. 10, no. 10, pp. 1–22, 2020, doi: 10.3390/brainsci10100672.
- [2] R. E. Pambudi, Sriyanto, and Firmansyah, “Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C.45 1,2,3,” vol. 16, no. x, pp. 221–226, 2022.
- [3] Kemenkes RI, “Stroke Dont Be The One,” p. 10, 2018.
- [4] R. S. Hutama, N. Hidayat, and E. Santoso, “Sistem Pakar Deteksi Dini Penyakit Stroke Menggunakan Metode Naïve Bayes-Certainty Factor,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 4333–4339, 2018.
- [5] P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, “Classification of stroke disease using machine learning algorithms,” *Neural Comput. Appl.*, vol. 32, no. 3, pp. 817–828, 2020, doi: 10.1007/s00521-019-04041-y.
- [6] A. H. Alamri, “Application of machine learning to stress corrosion cracking risk assessment,” *Egypt. J. Pet.*, vol. 31, no. 4, pp. 11–21, 2022, doi: 10.1016/j.ejpe.2022.09.001.
- [7] I. G. Bendesa Subawa, “Prediksi Kelulusan Mahasiswa Menggunakan Teorema Teorema Bayes,” *Janapati*, vol. 8, no. 3, pp. 227–236, 2019.
- [8] Isman, Andani Ahmad, and Abdul Latief, “Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 557–564, 2021, doi: 10.29207/resti.v5i3.3006.
- [9] N. N. Dzirkulloh and B. D. Setiawan, “Penerapan Metode K – Nearest Neighbor (KNN) dan Metode Weighted Product (WP) Dalam Penerimaan Calon Guru Dan Karyawan Tata Usaha Baru Berwawasan Teknologi (Studi Kasus : Sekolah Menengah Kejuruan Muhammadiyah 2 Kediri),” *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 5, pp. 378–385, 2017.
- [10] A. Bode, “K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika,” *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 188–195, 2017, doi: 10.33096/ilkom.v9i2.139.188-195.
- [11] N. Nafiah, “Klasifikasi Kematangan Buah Mangga Berdasarkan Citra HSV dengan KNN,” *J. Elektron. List. dan Teknol. Inf. Terap.*, vol. 1, no. 2, pp. 1–4, 2019, [Online]. Available: <https://ojs.politeknikjambi.ac.id/elti>
- [12] D. Marini Umi Atmaja, A. Rahman Hakim, D. Haryadi, and N. Suwaryo, “Penerapan Algoritma K-Nearest Neighbor Untuk Prediksi Pengelompokan Tingkat Risiko Penyebaran Covid-19 Jawa Barat,” *Dewi Mar. Umi Atmaja, SNTEM*, vol. 1, pp. 1218–1226, 2021.
- [13] Z. Zuriati and N. Qomariyah, “Klasifikasi Penyakit Stroke Menggunakan Algoritma K-Nearest Neighbor (KNN) Classification of Stroke Using the K-Nearest Neighbor (KNN) Algorithm,” vol. 1, no. 1, pp. 1–8, 2023.
- [14] D. Nopia and Z. Huzaifah, “Hubungan Antara Klasifikasi Stroke,” *J. Nurs. Invent.*, vol. 1, no. 1, pp. 16–22, 2020.
- [15] R. G. Tiwari, A. Pratap Srivastava, G. Bhardwaj, and V. Kumar, “Exploiting UML Diagrams for Test Case Generation: A Review,” *Proc. 2021 2nd Int. Conf. Intell. Eng. Manag. ICIEM 2021*, pp. 457–460, 2021, doi: 10.1109/ICIEM51511.2021.9445383.
- [16] M. N. Arifin and D. Siahaan, “Structural and Semantic Similarity Measurement of UML Use Case Diagram,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 2, p. 88, 2020, doi: 10.24843/lkjiti.2020.v11.i02.p03.
- [17] T. Ahmad, J. Iqbal, A. Ashraf, D. Truscan, and I. Porres, “Model-based testing using UML activity diagrams: A systematic mapping study,” *Comput. Sci. Rev.*, vol. 33, pp. 98–112, 2019, doi: 10.1016/j.cosrev.2019.07.001.
- [18] H. Abbad Ur Rehman, C. Y. Lin, and Z. Mushtaq, “Effective K-Nearest Neighbor Algorithms Performance Analysis of Thyroid Disease,” *J. Chinese Inst. Eng. Trans. Chinese Inst. Eng. A*, vol. 44, no. 1, pp. 77–87, 2021, doi: 10.1080/02533839.2020.1831967.