

Proposal Tugas Akhir

NIM	Nama	Persentase kontribusi
11320035	Dewa Sembiring	33,3
11320026	Rafael Sihombing	33,3
11320015	Weny Ari Sinarni Purba	33,3

Usulan judul (sementara) : Implementasi Algoritma Regresi Logistik dalam Memprediksi Kemungkinan Terkena Penyakit Stroke

Pembimbing : Ike Fitriyaningsih, S.Si., M.Si

Program studi : D3 Teknologi Informasi

Jenis Tugas Akhir : Machine Learning

Matakuliah yang pernah diambil terkait dengan Tugas Akhir ini:

- **Teknologi kecerdasan Buatan**
- **Pengembangan Situs Web 1**

Latar belakang

Stroke merupakan penyakit yang sangat mematikan yang bisa menyerang orang-orang tanpa memandang umur. WSO mendefinisikan stroke sebagai suatu kondisi dimana suplai darah ke otak mengalami penghambatan, yang mengakibatkan pasokan oksigen ke otak menjadi terhalang sehingga sel-sel di otak akan rusak dan kehilangan fungsinya. Hal ini paling sering disebabkan oleh gumpalan di arteri yang memasok darah ke otak. Ini juga bisa disebabkan oleh pendarahan ketika pembuluh darah pecah menyebabkan darah bocor ke otak. Stroke dapat menyebabkan kerusakan permanen, termasuk kelumpuhan parsial dan gangguan dalam berbicara, pemahaman dan memori. Bagian otak yang terkena dan lamanya suplai darah terhenti mempengaruhi jenis dan tingkat keparahan kecacatan hingga kematian [1].

Berdasarkan data yang dikeluarkan oleh *World Stroke Organization* (WSO), penyakit stroke menjadi penyebab kematian tertinggi nomor dua di dunia dan nomor tiga dalam penyebab kematian akibat kecacatan. Perkiraan biaya stroke secara global sekitar US\$721 miliar (0,66% dari *Gross Domestic Product* (GDP) global). Dari tahun 1990 hingga 2019, dalam hal jumlah kasus absolut, stroke meningkat secara substansial (peningkatan 70,0% dalam insiden stroke, 43,0% kematian akibat stroke, 102,0% stroke yang lazim, dan 143,0% DALYs), dengan sebagian besar beban stroke global (86,0% kematian dan 89,0% DALYs) berada di negara berpenghasilan rendah dan negara berpenghasilan menengah ke bawah (LMIC) [2]. Saat ini, jumlah penderita penyakit stroke lebih banyak menyerang pada negara-negara berkembang daripada negara-negara maju. Cina merupakan salah satu negara maju dengan jumlah kasus stroke (115,61-219 per 100.000 per tahun) [3].

Penelitian sebelumnya yang ditulis oleh A.H. M. Rahmatullah Imon, berjudul *Prediction of Rainfall Using Logistic Regression* menggunakan beberapa variabel penting yang akan digunakan

dalam dalam memprediksi curah hujan seperti *evaporation*, *maximum temperature*, *minimum temperature*, *humidity at 8:30 am* dan *humidity at 4:30pm*. Dengan menggunakan algoritma regresi logistik, penulis dapat dengan benar memprediksi 562 dari 590 hari saat hujan (95.25%) dan 49 dari 58 hari saat tidak hujan (84.48%) dengan probabilitas kesalahan klasifikasi adalah 0.0571% [4]. Lalu penelitian yang dilakukan oleh *Institute of Biomedical and Clinical Science* yang berjudul *Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults*, menguji 7 algoritma *Gradient Boosting Machine*, *K-Nearest Neighbours*, *MARS (Multi Adaptive Regression Spline)*, *Support Vector Machine*, *Neural Network*, *Random Forest*, dan *Logistic Regression* didapatkan hasil bahwa rata-rata performa untuk validasi *internal* semua model adalah $ROC\ AUC \geq 0.94$ dan untuk validasi *external* mendapatkan $ROC\ AUC \geq 0.93$ untuk semua model kecuali regresi logistik, regresi logistik mendapatkan nilai tertinggi yaitu $ROC\ AUC\ 0.95$. Regresi logistik memiliki performa sangat baik dalam hal kalibrasi dan analisis kurva keputusan dibandingkan algoritma lainnya. Sedangkan *gradiend boosting machine* memiliki performa terbaik dalam hal kalibrasi dan *support vector machine* baik dalam hal analisis kurva keputusan sama halnya dengan *logistic regression* [5]. Kemudian penelitian yang dilakukan oleh *Department of Computer Science, Faculty of Computer Science, University of Pembangunan Nasional Veteran Jakarta* berjudul *A Web-Based Diabetes Prediction Application Using XGBoost Algorithm* menguji 768 sampel data yang terdiri dari 268 positif diabetes dan 500 negatif diabetes dengan jumlah 8 variabel, mendapatkan hasil *accuracy* 74,67%, *precision value* 57,40%, *recall value* 65,94%, dan *specificity value* 78,50% [6].

Berdasarkan hasil penelitian sebelumnya, algoritma *Logistic Regression* memiliki *accuracy* yang tinggi dan *classification error* yang rendah, sehingga pada penelitian kali ini akan menggunakan algoritma *Regresi Logistik* dalam memprediksi kemungkinan seseorang terkena penyakit stroke. Data yang akan digunakan pada penelitian ini memiliki 5110 dataset dan 11 variabel yaitu, jenis kelamin, umur, riwayat hipertensi, riwayat penyakit jantung, status pernikahan, jenis pekerjaan, jenis kependudukan, tingkat glukosa, *Body Mass Index (bmi)*, riwayat merokok, dan riwayat stroke. Nantinya data yang terkumpul akan diolah menggunakan metode dan algoritma untuk mendapatkan hasil yang sesuai dan akurat. Dengan data tersebut akan diterapkan pengimplementasian prediksi penyakit stroke. Segala data mengenai penyakit stroke akan dikelola agar dapat dideteksi dengan menggunakan *machine learning*. Perancangan *machine learning* tersebut menggunakan algoritma *Regresi Logistik*, guna untuk mendapatkan nilai performa dan *accuracy* yang tinggi.

Tujuan

Tujuan penelitian dalam Tugas Akhir ini adalah mengimplementasikan sebuah mesin pintar untuk mendeteksi prediksi kemungkinan penyakit stroke dengan menggunakan algoritma Regresi Logistik.

Research question(s)

Adapun rumusan masalah dalam penelitian Tugas Akhir ini adalah sebagai berikut.

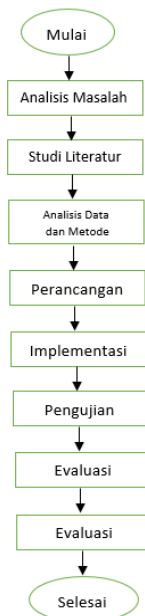
1. Bagaimana cara algoritma Regresi Logistik dalam memprediksi penyakit stroke?
2. Bagaimana tingkat akurasi implementasi algoritma Regresi Logistik dalam memprediksi penyakit stroke?

Hasil yang diharapkan

Hasil yang diharapkan pada Tugas Akhir ini adalah sebuah model dalam prediksi kemungkinan penyakit stroke.

Metodologi penelitian

Metodologi penelitian yang digunakan dalam pengerjaan Tugas Akhir terlampir pada Gambar 1.



Gambar 1 Metode Penelitian.

Berikut metodologi penelitian yang dilakukan dalam pengerjaan TA.

1. Analisis Masalah

Tahapan analisis masalah adalah tahapan yang bertujuan untuk mengidentifikasi masalah dalam penelitian untuk menentukan topik serta membuat poin-poin utama dalam penelitian.

2. Studi Literatur

Studi literatur mencakup beberapa hal seperti mengumpulkan, membaca dan mencatat informasi dari literatur yang berkaitan dengan topik TA. Tahap ini bertujuan sebagai pembelajaran dari penelitian atau proyek yang sudah pernah dilakukan serta dapat sebagai landasan dalam tugas akhir yang akan dibangun dalam teknik dan metode serta algoritma yang telah ditentukan.

3. Analisis Data dan Metode

proses tahap ini bertujuan untuk dapat memahami masalah serta menentukan metode yang tepat untuk memecahkan masalah tersebut dan untuk memastikan bahwa kebutuhan yang diperlukan oleh *user* tepat, dimana dalam penelitian tugas akhir ini akan menganalisis *dataset*.

4. Perancangan

Tahap perancangan yaitu tahap untuk merancang pola-pola yang digunakan untuk merancang dan membangun model serta *browser extension*.

5. Implementasi

Proses implementasi merupakan pembangunan model yang telah dirancang dengan menggunakan Algoritma Regresi Logistik. Melalui hasil perancangan yang telah dibuat akan digunakan bahasa pemrograman yang telah ditentukan pada tahap perancangan.

6. Pengujian

Tahap pengujian merupakan dimana penguji akan menguji sistem yang telah dibangun. Penguji akan melakukan pengujian terhadap sistem apakah sistem sudah dapat menghasilkan hasil yang sesuai dengan yang diharapkan.

7. Evaluasi

Pada tahap evaluasi yaitu dimana peneliti akan mengamati model serta aplikasi yang telah dibangun pada tahap pengujian. Pada tahap ini peneliti akan melakukan evaluasi terhadap hasil uji aplikasi yang sudah dilakukan pada tahap sebelumnya. peneliti akan melakukan analisis apakah aplikasi sudah menghasilkan tingkat akurasi sesuai yang diharapkan.

Resiko

Risiko yang mungkin terjadi dalam pengerjaan tugas akhir ini adalah hasil yang diperoleh saat melakukan implementasi tidak sesuai dengan hasil yang diharapkan.

Rencana kerja

Tabel 1 adalah rencana kerja dalam pengerjaan TA

Tabel 1 Rencana Kerja

Jenis Kegiatan	Minggu															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Konsultasi dengan pembimbing																
Penentuan topik dan judul Tugas Akhir																
Pengerjaan proposal Tugas Akhir																
Ujian Tengah Semester																
Seminar proposal																
Perbaikan proposal																
Pengerjaan laporan TA Bab I dan II																
Pengerjaan laporan TA Bab III																
Perbaikan dan melakukan finalisasi laporan TA Bab I-III																
Seminar TA I																

Referensi

- [1] M. E. A. Boehme, C. Esenwa, “Stroke: Risk factors and prevention,” *J. Pak. Med. Assoc.*, vol. 60, no. 3, p. 412, 2018, doi: 10.1161/CIRCRESAHA.116.308398.Stroke.
- [2] V. L. Feigin, “World Stroke Organization (WSO)- Global Stroke Fact Sheet 2022,” *Int. J. Stroke*, vol. 17(1) 18–, 2021, doi: 10.1177/17474930211065917.
- [3] R. C. Li *et al.*, “The risk of stroke and associated risk factors in a health examination population: A cross-sectional study,” *Med. (United States)*, vol. 98, no. 40, 2019, doi: 10.1097/MD.00000000000017218.
- [4] A. H. M. Rahmatullah Imon, M. C. Roy, and S. K. Bhattacharjee, “Prediction of rainfall using logistic regression,” *Pakistan J. Stat. Oper. Res.*, vol. 8, no. 3, pp. 655–667, 2012, doi: 10.18187/pjsor.v8i3.535.
- [5] A. L. Lynam *et al.*, “Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults,” *Diagnostic Progn. Res.*, vol. 4, no. 1, pp. 0–9, 2020, doi: 10.1186/s41512-020-00075-2.
- [6] Herlambang Dwi Prasetyo, Pandu Ananto Hogantara, and Ika Nurlaili Isnainiyah, “A Web-Based Diabetes Prediction Application Using XGBoost Algorithm,” *Data Sci. J. Comput. Appl. Informatics*, vol. 5, no. 2, pp. 49–59, 2021, doi: 10.32734/jocai.v5.i2-6290.