

TUGAS AKHIR

**APLIKASI DETEKSI DINI RISIKO PENYAKIT *STROKE*
MELALUI PENERAPAN METODE *OVERSAMPLING* DAN
UNDERSAMPLING MENGGUNAKAN METODE
*LOGISTIC REGRESSION***



Oleh:

Aldo Kadafi 1923250038

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER DAN REKAYASA
UNIVERSITAS MULTI DATA PALEMBANG
PALEMBANG
2024**

Fakultas Ilmu Komputer dan Rekayasa Universitas Multi Data Palembang

Program Studi Informatika
Tugas Akhir Sarjana Komputer
Semester Gasal Tahun 2023/2024

Aplikasi Deteksi Dini Risiko Penyakit *Stroke* Melalui Penerapan Metode *Oversampling* dan *Undersampling* Menggunakan Metode *Logistic regression*

Aldo Kadafi 1923250038

Abstrak

Stroke merupakan penyakit yang serius dan memiliki dampak signifikan pada kesehatan manusia. Prediksi risiko *Stroke* memiliki peran penting dalam upaya pencegahan dan pengobatan yang efektif. Penelitian ini bertujuan untuk menerapkan Metode *Logistic Regression* dalam meningkatkan akurasi prediksi *Stroke* serta membandingkan efektivitas teknik *oversampling* dan *undersampling* dalam mengatasi ketidakseimbangan data. Tiga teknik sampling yang dibandingkan adalah SMOTE (*Oversampling*), *Random Undersampling*, dan SMOTE-ENN (kombinasi *Oversampling* dan *Undersampling*). Pengujian dilakukan dengan menggunakan metode *Logistic Regression* pada *dataset* yang telah dilakukan sampling, menghasilkan evaluasi berupa *Confusion Matrix* yang menunjukkan akurasi, presisi, *recall*, dan *F1-score*. Hasil pengujian menunjukkan bahwa setiap metode sampling memberikan hasil akurasi yang berbeda. Teknik SMOTE dengan *Logistic Regression* menghasilkan akurasi tertinggi sebesar 76,77%. Model yang dihasilkan dapat diimplementasikan dalam aplikasi *web* menggunakan *framework Flask* dan dapat memprediksi risiko *Stroke* secara efektif. Penelitian ini menyimpulkan bahwa ketidakseimbangan *dataset* dapat diatasi dengan metode *sampling* tertentu, dan *Logistic Regression* efektif dalam melatih data yang telah *di-sampling*. Implementasi model ini dalam aplikasi *web* memiliki potensi besar dalam membantu identifikasi risiko *Stroke* pada pasien.

Kata kunci: *Flask, Logistic Regression, SMOTE, Stroke, Web*



BAB I

PENDAHULUAN

1.1 Latar Belakang

Stroke menurut *World Health Organization* adalah suatu keadaan dimana ditemukan tanda klinis yang berkembang cepat berupa defisit neurologik fokal dan global, yang dapat memberat dan berlangsung lama selama 24 jam atau lebih dan atau dapat menyebabkan kematian, tanpa adanya penyebab lain yang jelas selain vaskuler (Rachman, 2018). *Stroke* didefinisikan sebagai gangguan suplai darah ke otak yang biasanya disebabkan adanya sumbatan oleh gumpalan darah (Puspitasari, 2020). Hal ini menyebabkan gangguan pasokan oksigen dan nutrisi di otak sehingga terjadi kerusakan jaringan otak. Gejala *Stroke* umumnya mencakup kesulitan berbicara atau bicara tidak jelas, kelemahan atau mati rasa pada satu sisi tubuh, gangguan pada penglihatan, sakit kepala parah yang muncul secara mendadak, dan masalah keseimbangan atau kesadaran diri. *Stroke* terbagi dalam dua tipe, tipe pertama adalah *Stroke* iskemik disebabkan kurangnya suplai darah ke otak dikarenakan menyempitnya atau tersumbatnya pembuluh darah oleh deposit lemak yang disebut plak sehingga jaringan otak mengalami iskemik. Tipe yang kedua adalah *Stroke* hemoragik yang disebabkan pemecahan aneurisma pada *parenchyma* otak atau pada rongga antara otak dan tengkorak sehingga menyebabkan terjadinya iskemik dan desakan pada jaringan otak (Fadli, 2022).

Permasalahan *Stroke* Dalam 20 tahun terakhir terjadi peningkatan jumlah pasien *Stroke* di beberapa negara, WHO mengestimasi peningkatan jumlah pasien *Stroke* di beberapa negara Eropa sebesar 1,1 juta pertahun pada tahun 2000 menjadi 1,5 juta pertahun pada tahun 2025 *Prevalensi Stroke* di Indonesia

berdasarkan wawancara sebesar 8,3% pada tahun 2007 dan meningkat menjadi 12,1% pada tahun 2013 (Ghani dkk., 2016).

Jumlah penderita penyakit *Stroke* di Indonesia tahun 2013 berdasarkan diagnosis tenaga kesehatan diperkirakan sebanyak 1.236.825 orang (7,0%), sedangkan penderita yang menunjukkan gejala *Stroke* diperkirakan sebanyak 2.137.941 orang (12,1%) (Pusat Data dan Informasi Kementerian Kesehatan RI, 2014). Dalam tinjauan sistematis dari 36 studi epidemiologi berbasis populasi, tingkat kejadian *Stroke* hemoragik per 100.000 orang/tahun adalah 51,8% pada orang Asia, 24,2% di Whites, 22,9% di Blacks, dan 19,6% di Hispanik (Puspitasari, 2020).

Keterkaitan antara kesehatan mental dan fisik pada pasien stroke sudah lama dirasakan dan terbukti terjadi keselarasan antara fisiologis dan psikologis pada pasien Stroke (Permatasari, 2023). Stroke adalah penyakit neurologis yang jumlahnya meningkat setiap tahun dan juga menyebabkan rekor kasus dan kematian di seluruh dunia (Darma Perbasya, 2022). Pada hal ini *Stroke* merupakan penyakit gangguan fungsi otak berupa kelumpuhan pada saraf (defisit neurologis) akibat gangguan aliran darah pada salah satu bagian otak (Tamburion dkk., 2020). *Stroke* memiliki dampak serius dan potensi fatal karena menyerang bagian otak (Prakarsa, 2023) . Dalam dunia medis, tantangan utama yang dihadapi adalah dalam memprediksi risiko *Stroke*, yang sangat terkait dengan faktor-faktor seperti umur, tekanan darah, riwayat keluarga, pola makan, dan gaya hidup pasien (Suharno, 2017) . *Stroke* memiliki banyak gejala seperti gangguan fungsi syaraf pada *Stroke*, yang disebabkan oleh gangguan peredaran darah otak non traumatik, gangguan saraf tersebut menimbulkan gejala antara lain: kelumpuhan wajah atau anggota badan, bicara tidak lancar, bicara tidak jelas, perubahan kesadaran, gangguan penglihatan, dan lain-lain (Kemenkes RI, 2017). Dimana pasokan darah ke otak terganggu atau terputus. Kondisi ini dapat mengakibatkan kerusakan otak yang serius jika tidak segera diatasi. Oleh karena itu, deteksi dini dan tindakan medis cepat sangat penting untuk mengurangi risiko kerusakan otak yang permanen dan meminimalkan dampak *Stroke*.

Banyak faktor yang dapat mempengaruhi kejadian *Stroke*, diantaranya yaitu umur, jenis kelamin, keturunan, ras, hipertensi, hiperkolesterolemia, diabetes melitus, merokok, aterosklerosis, penyakit jantung, obesitas, konsumsi alkohol, stres, kondisi sosial ekonomi yang mendukung, diet yang tidak baik, aktivitas fisik yang kurang dan penggunaan obat anti hamil (Yonata dkk., 2016). Namun dari banyaknya faktor yang mempengaruhi kejadian *Stroke* hanya hipertensi yang secara signifikan mempengaruhi kejadian *Stroke* sedangkan kadar lipid dan kebiasaan merokok tidak secara signifikan berhubungan dengan kejadian *Stroke*. Status perkawinan bukan merupakan faktor risiko terjadinya penyakit *Stroke* namun, status pernikahan merupakan salah satu dukungan sosial terhadap pasien sehingga dengan adanya pasangan hidup dapat memberikan dukungan kepada pasangan untuk menjalankan perilaku yang sehat dan positif (Vihandayani dkk., 2019).

Status pekerjaan juga mempunyai hubungan dengan status ekonomi, sedangkan berbagai jenis penyakit yang timbul sering di dalam keluarga biasanya ada kaitannya dengan jenis pekerjaan yang bisa mempengaruhi pendapatan di dalam keluarga (Laily, 2017). Kejadian kematian yang disebabkan *stroke* sangat erat hubungannya dengan pekerjaan dan pendapatan di dalam keluarga, pada umumnya angka kematian *Stroke* meningkat pada pasien yang mempunyai status sosial ekonomi rendah. Risiko terjadinya *stroke* pada orang tidak bekerja karena adanya kecenderungan hidup yang santai, pola makan yang tidak teratur, malas berolahraga, dan tingkat stres yang lebih tinggi dibandingkan dengan orang yang mempunyai pekerjaan.

Peningkatan risiko *stroke* ini dapat diperantarai oleh beberapa mekanisme, yaitu: penurunan genetik faktor risiko *stroke*, penurunan kepekaan terhadap faktor risiko *stroke*, pengaruh keluarga pada pola hidup dan paparan lingkungan, interaksi antara faktor genetik dan lingkungan (Sary, 2016).

Dengan makin banyak kasus diabetes, sudah pasti angka kejadian *Stroke* juga akan bertambah (Alchuriyah & Wahjuni, 2016). Pasien *stroke* dengan

glukosa darah yang tinggi, apapun sebabnya, berpeluang lebih besar untuk makin memburuk dari pada pasien *stroke* yang kadar gula darah normal. Dasar timbulnya *stroke* adalah terjadinya aterosklerosis atau penyempitan pembuluh darah di otak.

Berat badan yang berlebihan sangat erat dengan terjadinya berbagai penyakit seperti jantung, diabetes dan *stroke*, merokok juga berarti menghirup zat yang dapat merupakan organ dalam tubuh seperti paru-paru dapat menyebabkan penyakit *Stroke* (Simbolon dkk., 2018) menurut *American Diabetes Association*, mengidap diabetes dapat meningkatkan risiko terkena *stroke* 1,5 kali lebih tinggi ketimbang orang yang tidak mengidap diabetes.

Berdasarkan hasil wawancara yang telah dilakukan bersama dr. Theresia Christin Sps.K, Penyakit *Stroke* awalnya diawali dengan gula darah yang tinggi, namun faktor utamanya biasanya umur yang sudah tua diatas 50 tahun dan ditandai dengan berat badan yang berlebihan maka akan memperkuat adanya gejala *Stroke* yang diterima pasien.

Berdasarkan hasil wawancara yang telah dilakukan bersama Dr. Jimmy Sabirin, Sp.S, selaku dokter spesialis penyakit saraf banyak pasien yang tidak mengetahui jika pasien itu sendiri mengalami *Stroke*, jadi ketika datang ke rumah sakit dalam keadaan terkena *Stroke* yang sangat berbahaya jika tidak segera ditangani, untuk itu dengan adanya aplikasi yang dibuat ini bisa membantu pasien untuk mencegah sebelum terjadi risiko *Stroke* yang lebih berbahaya.

Berdasarkan hasil wawancara yang telah dilakukan Bersama Dr. Billy Indra Gunawan, Sp., S(K) selaku dokter spesialis penyakit saraf *Stroke* ditandai dengan *fast*, yaitu *face*, *Arm*, *speech time*. *Face* merupakan gejala yang dilihat dari wajah seseorang yang tampak tidak normal seperti turun sebelah dan tidak simetris. *Arm* merupakan gejala *Stroke* yang dilihat dari lengan penderita menjadi lemah. Jika lengan itu diangkat maka tingginya tidak sama dengan lengan satunya, *Speech*, merupakan gejala *stroke* dilihat dari cara bicara penderita

yang menjadi sulit, tidak jelas, atau bahkan tidak bisa bicara. *Time*, metode terakhir setelah tiga metode sebelumnya terindikasi pada penderita, maka sudah waktunya membawa penderita ke rumah sakit.

Berdasarkan hasil wawancara yang telah dilakukan bersama Dr. Risayogi Wicaksana Asaf H SpBS, selaku dokter bedah saraf kebanyakan pasien *stroke* biasanya mengalami Hipertensi yang tinggi dan mengalami kenaikan gula darah, faktor risiko utama yang pasti karena umur dan pola makan dan pola hidup.

Berdasarkan hasil wawancara yang telah dilakukan bersama Dr. Sahat Edison Sitorus, SpBS(K), selaku dokter bedah Saraf biasanya ditandai dengan hipertensi yang tinggi dan gula darah yang tinggi tapi biasanya faktor umum dari gejala *Stroke* ini adalah faktor gen dan umur yang sudah rentan karena umur yang rentan biasanya pola gaya hidup yang tidak teratur menyebabkan adanya gejala *stroke*.

Berdasarkan hasil wawancara yang telah dilakukan dan dianalisis menurut para ahli. Secara umum *stroke* merupakan salah satu penyakit berbahaya yang menyerang banyak orang, khususnya orang tua banyak juga pasien yang tidak tau jika memiliki gejala *stroke*, *stroke* menjadi berbahaya apabila tidak ditindaklanjuti, menurut beberapa dokter yang telah diwawancarai penyebab yang paling umum pasien terkena *stroke* biasanya karena gula darah yang tinggi, faktor gen keluarga, dan juga hipertensi yang melebihi batas hipertensi pada umumnya, dan juga faktor BMI atau *body maks indeks* menjadi salah satu faktor yang memperkuat adanya gejala *stroke*.

Berdasarkan latar belakang yang sudah dijelaskan, maka penulis membuat suatu aplikasi pendeteksi gejala dini risiko penyakit *stroke* untuk digunakan oleh dokter agar lebih cepat mengetahui gejala pada pasien.

1.2 Rumusan Masalah

Dibutuhkan suatu aplikasi untuk mendeteksi gejala dini penyakit *Stroke* untuk Membantu dokter mengenali gejala *Stroke* pada pasien.

1.3 Analisis Terhadap Batasan

1.3.1 Analisis dari Aspek Ekonomi

Dari sudut pandang pengguna, aplikasi deteksi dini risiko *Stroke*. Meskipun pengguna mungkin dihadapkan pada biaya awal, pengguna yaitu dokter perlu dipertimbangkan sebagai investasi dalam kesehatan masa depan. Deteksi dini dapat mengurangi beban biaya pengobatan dan rehabilitasi, memberikan nilai ekonomis jangka panjang. Meski demikian, harga aplikasi perlu disesuaikan agar dapat diakses oleh berbagai lapisan masyarakat. Efisiensi waktu, ketika pengguna dapat dengan cepat memahami hasil prediksi dan mengambil tindakan, juga menjadi faktor penting.

Dari sudut pandang pengembang, biaya pengembangan awal dianggap sebagai investasi yang dapat meningkatkan reputasi dan mendatangkan manfaat ekonomis jangka panjang. Menetapkan model bisnis yang berkelanjutan, seperti berlangganan atau kerjasama dengan lembaga kesehatan, menjadi kunci untuk mendukung kelangsungan proyek. Meskipun biaya pengembangan dan pemeliharaan tinggi, keselarasan antara keamanan data, kepatuhan regulasi, dan manfaat ekonomis menjadi fokus utama. Dengan demikian, keberlanjutan dan keberhasilan ekonomis aplikasi ini bergantung pada keseimbangan yang cermat antara manfaat jangka panjang bagi pengguna dan biaya yang dihadapi oleh pengembang. Analisis ekonomis yang dilakukan dari hasil wawancara pada Tabel 1.1.

Tabel 1.1 Tabel Analisis dari Aspek Ekonomis

Aspek	Organisasi 1 Apotik Buana	Organisasi 2 Klinik Bedah Saraf	Organisasi 3 RSMH	Organisasi 4 Rs siloam	Organisasi 5 Klinik Apotik bersama
Menurut bapak/ibu, berapakah biaya yang dapat ditawarkan untuk harga aplikasi yang saya buat berdasarkan prototypenya?	Saya tidak memberikan harga yang signifikan karena ini bukan bidang saya yang pasti tidak lebih dari 5 juta	Dengan fitur yang ada, sebenarnya saya tidak berkompeten di bidang ini namun saya hanya bisa memberikan harga 5 juta saja	Fiturnya menarik namun harus ada tambahan fitur lebih lanjut terkait aplikasi ini , seperti <i>Stroke</i> ringan , sedang dan berat, saya hargai diharga 7 juta	Dengan parameter (umur, hipertensi, penyakit jantung, status menikah, pernah merokok, <i>BMI</i> , gula darah) . yang ada saya rasa untuk di bidang pendataan ini bisa dipakai di sini, jika ini dijual saya berani beli diharga 7 juta saja	Akan lebih baik tampilanya ditambah fitur baru biar makin banyak pilihannya, yang saya sarankan fitur saran, untuk harga yang kam bisa tawar jika kami beli di harga 7 juta saja

1.3.2 Analisis dari Aspek Manufakturabilitas

Dalam *menghadapi* tantangan pengembangan perangkat lunak, pengaturan waktu yang realistis menjadi kunci keberhasilan. analisis beberapa aspek manufakturabilitas dan sudut pandang pengguna, membawa kita pada pemahaman yang lebih mendalam tentang berapa lama batasan waktu yang harus dipenuhi untuk menyelesaikan perangkat lunak yang efektif. Tabel 1.2 untuk melihat sejauh mana setiap organisasi dapat memenuhi waktu penyelesaian yang ditetapkan.

Tabel 1.2 Tabel Analisis dari Aspek Manufakrurabilitas

Aspek	Organisasi 1 Apotik Buana	Organisasi 2 Klinik Bedah saraf	Organisasi 3 RSMH	Organisasi 4 Rs siloam	Organisasi 5 Klinik Apotik Bersama
Kemudahan dalam mengakses aplikasi (3 minggu)	OK	OK	OK	OK	OK
Keakuratan Deteksi (3 Minggu)	OK	OK	OK	OK	OK
Kemudahan pengoperasian dengan perangkat mobile/pc (4 Minggu)	OK	OK	OK	OK	OK
Total = 10 Minggu					

Dari hasil analisis yang tercantum dalam Tabel 1.2, dapat disimpulkan bahwa aspek-aspek kunci terkait rencana penggunaan aplikasi deteksi risiko *Stroke* telah memperoleh penilaian yang memuaskan dari setiap organisasi. Kemudahan dalam mengakses aplikasi, keakuratan deteksi, dan kemudahan pengoperasian dengan perangkat

mobile atau PC mendapatkan penilaian "OK" dari Organisasi 1 hingga Organisasi 5. Hasil ini menggambarkan bahwa aplikasi ini dapat diakses dengan mudah, memberikan deteksi risiko yang akurat, dan dapat dioperasikan dengan lancar di berbagai perangkat, menciptakan pengalaman pengguna yang konsisten dan dapat diandalkan. Dengan demikian, Aplikasi ini dapat dianggap berhasil memenuhi standar kualitas yang diperlukan untuk memberikan layanan deteksi risiko *Stroke* yang efektif kepada pengguna dari berbagai organisasi.

1.3.3 Analisis dari Aspek Sustainability

Dalam mengevaluasi kinerja perangkat lunak deteksi risiko *Stroke*, salah satu aspek yang penting untuk diperhatikan adalah kecepatan deteksi. Kecepatan ini tidak hanya berkaitan dengan efisiensi operasional, tetapi juga memberikan dampak langsung terhadap pengalaman pengguna dan keberlanjutan aplikasi. Tabel berikut menyajikan analisis terhadap kemampuan perangkat lunak dari setiap organisasi dalam memberikan hasil deteksi dalam waktu yang singkat, yaitu 1 detik. Hasil analisis untuk mendapatkan pemahaman yang lebih mendalam tentang kinerja setiap organisasi dalam hal kecepatan deteksi. Tabel 1.3 merupakan Tabel analisis dari aspek sustainability

Tabel 1.3 Tabel Analisis dari Aspek Sustainability

Aspek	Organisasi Apotik Buana	Organisasi Klinik bedah saraf	Organisasi RSMH	Organisasi RS Siloam	Organisasi Klinik Apotik bersama
Perangkat lunak dapat memberikan hasil deteksi dalam waktu singkat (1s)	OK	OK	OK	OK	OK

1.4 Analisis Terhadap Karakteristik Solusi

Dalam menjelajahi karakteristik solusi perangkat lunak deteksi risiko *Stroke*, penting untuk memahami fungsi-fungsi yang dirancang untuk secara konkret memecahkan masalah yang dihadapi. Fungsi-fungsi tersebut memiliki *traceability*, yang merupakan kejelasan mengenai alasan keberadaan setiap fungsi dalam perangkat lunak. Tabel 1.4 merupakan tabel analisis terhadap karakter solusi.

Tabel 1.4 Tabel Analisis Terhadap Karakter Solusi

No.	Masalah	Fungsi
1.	Pasien tidak tau dia mengalami <i>Stroke</i> atau tidak.	Perangkat lunak dapat digunakan untuk deteksi risiko dengan menginput data yang ada pada parameter yang disediakan
2.	Pasien hanya perlu mengecek <i>BMI</i> , gula darah dan hipertensi saja	Dengan perangkat lunak bisa mengetahuinya adanya gejala <i>Stroke</i>
3.	Pasien tidak tau harus apa Ketika dia mengalami <i>Stroke</i>	Dengan perangkat lunak dapat diberikan saran untuk pasien yang terkena gejala <i>Stroke</i> maupun tidak.

1.5 Pemilihan Solusi

Pada penelitian terdahulu yang pernah oleh (Yadav S. et dkk. 2020) yang berjudul *Handling Imbalanced Dataset Classification in Machine Learning*. Penelitian ini bertujuan untuk mencari strategi dalam menangani dan menyelesaikan masalah yang terkait dengan *dataset* yang tidak seimbang (*Imbalance Data*) terutama dalam penerapan data mining dan machine learning. Distribusi data yang tidak teratur menjadi masalah di balik menurunnya kinerja pengklasifikasian. Eksperimen dilakukan pada *dataset* churn prediction pada pelanggan telekomunikasi. Dengan menerapkan metode *oversampling* dan *undersampling* diperoleh performa terbaik adalah

metode Random Forest yang cukup baik dalam hal *Accuracy* sebesar 85,3%, *Precision* 82,6%, *recall* 88,2% dan nilai *ROC* 88,6%.

Penelitian lain yang serupa dilakukan oleh (Wang, Yao and Chen, 2021) yang berjudul *An Imbalanced-Data Processing Algorithm for the Prediction of Heart Attack in Stroke Patients*. Pada penelitian ini, diusulkan algoritma *Random Forest* dengan menggabungkan tiga metode, antara lain metode *undersampling*, metode *clustering*, dan metode *oversampling*. Algoritma tersebut dapat menangani data pasien *Stroke* yang tidak seimbang (*Imbalance*). Dari data asli, dipilih delapan indikator medis yang mempengaruhi serangan jantung. Kemudian membandingkan kinerja beberapa model pembelajaran mesin dalam memprediksi serangan jantung. Hasil eksperimen menunjukkan bahwa RF merupakan model terbaik untuk memprediksi kemungkinan serangan jantung pada *database* MIMIC-III *dataset* pasien *Stroke*. Akurasinya 70,29%, dan presisinya 70,05 %.

Dalam penelitian sebelumnya (Tomasouw & Rumlawang, 2023), SVM digunakan dalam prediksi dini risiko penyakit *Stroke* dengan pembagian proporsi data latih dan data uji sebesar 60% dan 40% secara berurutan. Tingkat akurasi yang dihasilkan adalah 81,25% dengan proporsi tersebut. Namun, ketika proporsi data latih ditambah menjadi 70%, akurasi menurun menjadi 76%. Oleh karena itu, proporsi data yang menghasilkan akurasi tertinggi adalah 60%. Penelitian lain yang memiliki judul yang sama (Puspitawuri, Santoso, & Dewi, 2019) menggunakan metode K-NN dengan proporsi data latih 45% dan data uji 30%. Akurasi tertinggi yang diperoleh adalah 96,67%.

Metode *klasifikasi* Naïve Bayes digunakan dalam penelitian lain (Paramitha et dkk., 2023), dengan proporsi data latih 80% dan data uji 20%, menghasilkan akurasi sebesar 80%. Penelitian lainnya (Iskandar, Ernawati, & Widiastiwi, 2022) menggunakan metode Random Forest dengan proporsi data latih 90% dan data uji 10%, menghasilkan tingkat akurasi sebesar 95,2%.

Dari analisis latar belakang dan solusi yang telah dijelaskan dalam penelitian sebelumnya, terlihat bahwa masalah ketidakseimbangan data dapat diatasi melalui penggunaan metode *oversampling* dan *undersampling*. Oleh karena itu, tujuan dari penelitian ini adalah untuk mengatasi masalah ketidakseimbangan data dalam prediksi risiko *Stroke* dengan menggunakan metode *Logistic Regression*, dengan membandingkan efektivitas kedua teknik tersebut. Dengan pemahaman yang lebih mendalam tentang manfaat dan keterbatasan masing-masing teknik, kami berharap penelitian ini akan memberikan kontribusi yang signifikan dalam meningkatkan efektivitas prediksi risiko *Stroke*. Selain itu, hasil penelitian ini juga diharapkan dapat memberikan panduan berharga bagi praktisi medis dan peneliti kesehatan dalam memanfaatkan teknologi ini untuk meningkatkan perawatan kesehatan masyarakat secara keseluruhan.

Berdasarkan penelitian terdahulu yang telah dijelaskan sebelum penjelasan ini, tiga metode yang dipilih untuk penelitian ini adalah *SVM*, *KNN*, dan *Logistic Regression*. Akurasi tertinggi diperoleh oleh *Logistic Regression*, diikuti oleh *KNN* dan *SVM*. Dalam penelitian ini, *Logistic Regression* menghasilkan akurasi sebesar 95,23%, sehingga dipilih sebagai metode yang tepat. Dalam tahap perancangan, *Logistic Regression* sesuai dengan fokus aplikasi yang akan dibuat, karena metode ini berfokus pada prediksi nilai dari satu faktor berdasarkan faktor lainnya, dengan prediksi umumnya berupa kategori biner seperti ya atau tidak.

1.6 Skenario Pemanfaatan Produk Oleh Pengguna

Perangkat lunak yang dibuat merupakan aplikasi untuk memprediksi adanya gejala dini risiko *Stroke*. Aplikasi ini dapat dijalankan melalui perangkat telepon pintar (*smartphone/pc*) dan dioperasikan oleh dokter untuk mengenali gejala *Stroke* secara cepat dan dapat mencegah adanya indikasi *Stroke*, namun sebelum menjalankan aplikasi ini harus dilakukan dulu pemeriksaan gula darah *BMI (Body mass indeks)*, dan hipertensi sebelum dokter bisa menggunakan aplikasi ini.

1.7 Tujuan

Tujuan penelitian dari rumusan masalah di atas adalah :

1. Menerapkan Metode *Logistic Regression* untuk meningkatkan akurasi prediksi *Stroke*.
2. Membandingkan efektivitas teknik *oversampling* dan *undersampling* dalam mengatasi ketidakseimbangan data pada prediksi *Stroke*.
3. Menyediakan solusi yang dapat digunakan dalam praktik medis untuk membantu identifikasi risiko *Stroke*.



DAFTAR PUSTAKA

- Alchuriyah, S., & Wahjuni, C. U. (2016). Faktor Risiko Kejadian Stroke Usia Muda pada Pasien Rumah Sakit Brawijaya Surabaya. *Jurnal Berkala Epidemiologi*, 4(1), 62. <https://doi.org/10.20473/jbe.v4i12016.62-73>
- Arafa, A., El-Fishawy, N., Badawy, M., & Radad, M. (2022). RN-SMOTE: Reduced Noise SMOTE based on DBSCAN for enhancing imbalanced data classification. *Journal of King Saud University - Computer and Information Sciences*, 34(8), 5059–5074. <https://doi.org/10.1016/j.jksuci.2022.06.005>
- Azhar, Y., Firdausy, A. K., & Amelia, P. J. (2022). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke. *SINTECH (Science and Information Technology) Journal*, 5(2), 191–197. <https://doi.org/10.31598/sintechjournal.v5i2.1222>
- Darma Perbasya, S. T. (2022). Hubungan Hipertensi Terhadap Stroke. *Jurnal Ilmu Kesehatan Indonesia (JIKSI)*, 2(2), 393–404. <https://doi.org/10.57084/jiksi.v2i2.775>
- Fadli, R. (2022). *STROKE*. <https://www.halodoc.com/kesehatan/stroke>
- Ghani, L., Mihardja, L. K., & Delima, D. (2016). Faktor Risiko Dominan Penderita Stroke di Indonesia. *Buletin Penelitian Kesehatan*, 44(1), 49–58. <https://doi.org/10.22435/bpk.v44i1.4949.49-58>
- Husin, T. F., Pribadi, M. R., & Yohannes. (2022). Implementation of LSSVM in Classification of Software Defect Prediction Data with Feature Selection. *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 126–131. <https://doi.org/10.23919/EECSI56542.2022.9946611>
- Kemenkes RI. (2017). *Stroke Dapat Dicegah, Kenali Faktor Risiko dan Gejalanya*. P2PTM Kemenkes RI. <http://p2ptm.kemkes.go.id/post/stroke-dapat-dicegah-kenali-faktor-risiko-dan-gejalanya>
- Laily, R. S. (2017). Hubungan Karakteristik Penderita dan Hipertensi dengan Kejadian Stroke Iskemik. *Jurnal Berkala Epidemiologi*, 5(1), 48–59. <https://doi.org/10.20473/jbe.v5i1>
- Permatasari, N. A. (2023). Penyebab Stroke Perspektif Kesehatan Mental. *Jurnal Impresi Indonesia*, 2(1), 1–9. <https://doi.org/10.58344/jii.v2i1.1790>

- Prakarsa, R. S. (2023). *Mengenal Gejala Awal dan Jenis Stroke*. Siloam Hospitals. <https://www.siloamhospitals.com/en/informasi-siloam/artikel/mengenal-gejala-awal-dan-jenis-stroke>
- Puspitasari, P. N. (2020). Hubungan Hipertensi Terhadap Kejadian Stroke. *Jurnal Ilmiah Kesehatan Sandi Husada*, 12(2), 922–926. <https://doi.org/10.35816/jiskh.v12i2.435>
- Rachman, T. (2018). Definsi Stroke dan Faktor terjadinya Stroke. *Angewandte Chemie International Edition*, 6(11), 951–952., 10–27.
- Sary, A. N. (2016). Faktor Risiko Kejadian Stroke pada Pasien Rawat Inap di Rumah Sakit Stroke Nasional Bukittinggi Tahun 2013. *Jurnal Medika Sainika*, 7(2), 45–54.
- Simbolon, P., Simbolon, N., & Siringo-ringo, M. (2018). Faktor merokok dengan kejadian stroke di Rumah Sakit Santa Elisabeth Medan. *Jurnal Kesehatan Manarang*, 4(1), 18–25. <http://www.jurnal.poltekkesmamuju.ac.id/index.php/m/article/view/53>
- Suharno, S. (2017). Faktor-Faktor Individu Yang Berhubungan Dengan Hipertensi Di Wilayah Kerja Uptd Puskesmas Cigasong Kabupaten Majalengka Tahun 2017. *Jurnal Kampus STIKES YPIB Majalengka*, 6(1), 1–10. <https://doi.org/10.51997/jk.v6i1.8>
- Tamburian, A. G., Ratag, T. B., & Nelwan, E. J. (2020). Hubungan antara hipertensi, diabetes melitus dan hiperkolesterolemia dengan kejadian stroke iskemik. *Journal of Public Health and Community Medicine*, 1(1), 27–33.
- Vihandayani, M., Wiratmo, P. A., & Hijriati, Y. (2019). Hubungan Dukungan Keluarga Sebagai Support Sistem dan Kualitas Hidup Pasien Stroke Infark. *Program Studi Ilmu Keperawatan, Universitas Binawan*, 1(2), 74–79.
- Yonata, A., Satria, A., & Pratama, P. (2016). Hipertensi sebagai Faktor Pencetus Terjadinya Stroke Majority. *Jurnal Majority*, 5(3), 17. <http://juke.kedokteran.unila.ac.id/index.php/majority/article/view/1030>