

LAPORAN TUGAS AKHIR

METODE *NEAR MISS UNDERSAMPLING* UNTUK OPTIMALISASI *DATASET* PADA MODEL PREDIKSI PENYAKIT STROKE

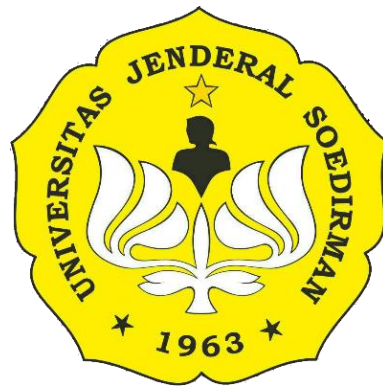
Disusun untuk memenuhi kebutuhan akan laporan tugas akhir
di Jurusan Teknik Elektro Universitas Jenderal Soedirman



Disusun oleh:

Yusuf Muhammad Nur Zaman
H1A018043

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI
UNIVERSITAS JENDERAL SOEDIRMAN
FAKULTAS TEKNIK
JURUSAN/PROGRAM STUDI TEKNIK ELEKTRO
PURBALINGGA
2021**



LAPORAN TUGAS AKHIR

METODE *NEAR MISS UNDERSAMPLING* UNTUK OPTIMALISASI *DATASET* PADA MODEL PREDIKSI PENYAKIT STROKE

Disusun untuk memenuhi kebutuhan akan laporan tugas akhir
di Jurusan Teknik Elektro Universitas Jenderal Soedirman



Disusun oleh:

Yusuf Muhammad Nur Zaman
H1A018043

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI
UNIVERSITAS JENDERAL SOEDIRMAN
FAKULTAS TEKNIK
JURUSAN/PROGRAM STUDI TEKNIK ELEKTRO
PURBALINGGA
2021**

HALAMAN PENGESAHAN

Laporan Tugas Akhir dengan Judul:

METODE *NEAR MISS UNDERSAMPLING* UNTUK OPTIMALISASI DATASET PADA MODEL PREDIKSI PENYAKIT STROKE

Disusun oleh:

Yusuf Muhammad Nur Zaman

H1A018043

Diajukan untuk memenuhi salah satu persyaratan
memperoleh gelar Sarjana Teknik pada
Jurusan/Program Studi Teknik Elektro
Fakultas Teknik
Universitas Jenderal Soedirman

Diterima dan disetujui

Pada Tanggal : _____

Pembimbing I

Pembimbing II

Farida Asriani, S.Si., M.T.
(NIP : 197502012000032005)

Ari Fadli, S.T., M.Eng.
(NIP : 198407312019031007)

Mengetahui:
Dekan Fakultas Teknik

Prof. Dr. Eng Suroso S.T., M.Eng.
NIP. 197812142001121002

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Laporan Tugas Akhir dengan judul “**Metode *Near Miss Undersampling* Untuk Optimalisasi *Dataset* Pada Model Prediksi Penyakit Stroke**” ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Purbalingga, 5 Juli 2021

[materai sesuai ketentuan uu]

Ttd.

Yusuf Muhammad Nur Zaman

NIM. H1A018043

HALAMAN MOTTO DAN PERSEMBAHAN

MOTTO

The goal is not to be perfect by the end. The goal is to be better tomorrow.

PERSEMBAHAN

Tugas akhir dan juga laporan tugas akhir ini ditulis dan dipersembahkan untuk:

1. Tuhan yang Maha Esa,
2. Nabi Muhammad SAW,
3. Bapak, Ibu, saudara-saudara, serta teman-teman,
4. Semua pihak yang tidak dapat disebutkan satu persatu yang telah memberi bantuan penulis baik dalam pelaksanaan.

RINGKASAN

METODE *NEAR MISS UNDERSAMPLING* UNTUK OPTIMALISASI *DATASET* PADA MODEL PREDIKSI PENYAKIT STROKE

Yusuf Muhammad Nur Zaman

Stroke merupakan kondisi yang terjadi ketika suplai darah ke otak terganggu atau berkurang akibat adanya penyumbatan atau pecahnya pembuluh darah. Pada tahun 2018 penyakit stroke menjadi penyebab kematian nomor satu di Indonesia, salah satu penyakit dengan tingkat prevalensi yang tinggi, dan juga biaya *pre-diagnosis* dan pengobatan yang tidak murah.

Beberapa data set medis yang memiliki dua kelas atau *binomial class* mengalami ketidakseimbangan kelas, hal ini disebut dengan *imbalanced dataset*, yaitu suatu kondisi ketika kelas tujuan yang akan diklasifikasi tidak memiliki rasio yang seimbang. Sehingga menyebabkan hasil klasifikasi yang bias karena *classifier* akan lebih condong mendeteksi kelas mayoritas dibanding dengan kelas minoritas. Kasus *imbalanced dataset* ini dapat diatasi dengan menerapkan metode *Near-Miss undersampling*.

Berdasarkan hasil pelatihan dan penelitian, model dengan *dataset* orisinal memiliki nilai skor *f1*, *precision*, dan *recall* bernilai 0.00, sedangkan model yang menggunakan *dataset* hasil dari *near-miss undersampling* versi 1 memiliki skor *f1* sebesar 0.731, skor *precision* sebesar 0.717, dan skor *recall* sebesar 0.745, lalu model yang menggunakan *dataset* hasil dari *near-miss undersampling* versi 2 memiliki skor *f1* sebesar 0.660, skor *precision* sebesar 0.673, dan skor *recall* sebesar 0.647, dan model yang menggunakan *dataset* hasil dari *near-miss undersampling* versi 3 memiliki skor *f1* sebesar 0.747, skor *precision* sebesar 0.673, dan skor *recall* sebesar 0.725. Sehingga model dengan menggunakan *dataset* hasil *near-miss undersampling* memiliki performa yang jauh lebih baik dalam klasifikasi.

Kata kunci : *imbalanced dataset*, stroke, metode *near-miss undersampling*

SUMMARY

NEAR MISS UNDERSAMPLING METHOD FOR DATASET OPTIMIZATION IN STROKE DISEASE PREDICTION MODEL

Yusuf Muhammad Nur Zaman

Stroke is a condition that occurs when the blood supply to the brain is interrupted or reduced due to a blockage or rupture of a blood vessel. In 2018 stroke became the number one cause of death in Indonesia, one of the diseases with a high prevalence rate, and the cost of pre-diagnosis and treatment is not cheap.

Some medical data sets that have two classes or binomial classes experience a class imbalance, this is called an imbalanced data set, which is a condition when the destination class to be classified does not have a balanced ratio. This causes biased classification results because the classifier will be more inclined to detect the majority class than the minority class. This imbalanced dataset case can be overcome by applying the Near Miss undersampling method.

Based on the results of training and testing, the model with original dataset has a f1 score, precision score, and recall score are 0.00, while the model with dataset from near-miss undersampling version 1 has a f1 score of 0.731, precision score of 0.717, and recall of 0.745, the the model with dataset from near-miss undersampling version 2 has a f1 score of 0.660, precision score of 0.673, and recall score of 0.647, and the model with dataset from near-miss undersampling version 3 has a f1 score of 0.747, precision score of 0.673, dan recall score of 0.725. So the model with undersampled using near-miss has a mush better performance in classification.

Keywords : imbalanced dataset, stroke, near miss undersampling method

PRAKATA

Puji syukur kehadiran Allah S.W.T. yang telah melimpahkan berkah dan rahmat-Nya dokumen “**Metode *Near Miss Undersampling* Untuk Optimalisasi *Dataset* Pada Model Prediksi Penyakit Stroke**” ini dapat disusun. Terimakasih kami sampaikan kepada seluruh pihak yang telah membantu terwujudnya dokumen ini, diantaranya: Dekan FT Unsoed, Wakil Dekan Akademik FT Unsoed, Kajur Teknik Elektro Unsoed, Sekretaris Jurusan Teknik Elektro Unsoed, bapak-ibu dosen Teknik Elektro Unsoed, Teman-teman Teknik Elektro Unsoed dan pihak-pihak lain yang tidak dapat kami sebutkan satu persatu.

Purbalingga, 5 Juli 2021

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN	v
HALAMAN MOTTO DAN PERSEMBAHAN.....	vi
RINGKASAN	vii
<i>SUMMARY</i>	viii
PRAKATA.....	ix
DAFTAR ISI	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
DAFTAR ISTILAH DAN SINGKATAN	xv
DAFTAR SIMBOL	xvi
BAB 1 PENDAHULUAN	17
1.1 Latar Belakang	17
1.2 Rumusan Masalah	18
1.3 Batasan Masalah.....	18
1.4 Tujuan dan Manfaat.....	19
1.4.1 Tujuan	19
1.4.2 Manfaat.....	19
1.5 Sistematika Penulisan.....	19
BAB 2 TINJAUAN PUSTAKA.....	21
2.1 Penelitian Terdahulu	21
2.2 Penyakit Stroke.....	22
2.3 <i>Overfitting</i>	24
2.4 <i>Imbalanced Dataset</i>	25
2.5 <i>Google Colaboratory</i>	28
2.6 <i>Framework Tensorflow dan Keras</i>	30
2.7 Metode <i>Undersampling</i>	31
2.8 Metode <i>Near Miss Undersampling</i>	31
BAB 3 METODE PENELITIAN.....	33

3.1 Waktu dan Tempat Penelitian	33
3.2 Alat dan Bahan	33
3.3 Alur dan Tahap Penelitian	34
3.3.1 Tahap Persiapan	35
3.3.2 Tahap Persiapan Model dan Pre-Proses Dataset.....	36
3.3.3 Tahap Penerapan Metode <i>Near Miss Undersampling</i>	36
3.3.4 Tahap Perbandingan Dataset Sebelum dan Sesudah Dioptimalisasi ...	37
3.3.5 Tahap Perancangan Arsitektur Model.....	37
3.3.6 Tahap Pengujian.....	38
3.3.7 Tahap Analisis dan Evaluasi	38
3.4 Waktu dan Jadwal Pelaksanaan Tugas Akhir	39
BAB 4 HASIL DAN PEMBAHASAN.....	40
4.1 Pengolahan <i>Dataset</i>	40
4.1.1 <i>Preprocessing Dataset</i>	41
4.1.2 Penerapan Metode <i>Near-Miss Undersampling</i>	42
4.2 Hasil Pelatihan dan Pengujian	44
4.2.1 <i>Dataset</i> tanpa <i>Near-Miss Undersampling</i>	44
4.2.2 <i>Dataset</i> dengan <i>Near-Miss Undersampling</i>	75
4.3 Perbandingan Evaluasi Performa Model	96
BAB 5 PENUTUP.....	100
5.1 Kesimpulan.....	100
5.2 Saran.....	101
DAFTAR PUSTAKA	102
LAMPIRAN	103
Lampiran 1. Tampilan Aplikasi MedikalTech Capstone Project 2021	103
Lampiran 2. <i>Preprocessing Dataset</i>	104
Lampiran 3. Model dengan <i>dataset</i> orisinil	107
Lampiran 4. Model dengan <i>dataset</i> hasil <i>near-miss</i>	109
BIODATA PENULIS	115

DAFTAR TABEL

Tabel 2.1 Rincian penelitian terdahulu.....	21
Tabel 3.1 Jadwal rinci penelitian.....	39
Tabel 4.1 Data Uji Menggunakan Model dengan Dataset Orisinil	46
Tabel 4.2 Nilai evaluasi model dengan dataset orisinil	75
Tabel 4.3 Data Uji Menggunakan Model dengan Dataset Hasil Near-Miss 1	77
Tabel 4.4 Nilai evaluasi model dengan dataset near-miss 1	82
Tabel 4.5 Data Uji Menggunakan Model dengan Dataset Hasil Near-Miss 2	84
Tabel 4.6 Nilai evaluasi model dengan dataset near-miss 2.....	89
Tabel 4.7 Data Uji Menggunakan Model dengan Dataset Hasil Near-Miss 3	91
Tabel 4.8 Nilai evaluasi model dengan dataset near-miss 3.....	95
Tabel 4.9 Perbandingan evaluasi pada akurasi model.....	96
Tabel 4.10 Perbandingan evaluasi pada nilai loss model	96
Tabel 4.11 Perbandingan hasil confusion matrix model	97

DAFTAR GAMBAR

Gambar 2.1 Distribusi kelas pada Dataset	22
Gambar 2.2 Ilustrasi performa model	24
Gambar 2.3 Ilustrasi kondisi dataset	26
Gambar 2.4 Distribusi dari kasus imbalanced dataset	27
Gambar 2.5 Tampilan Google Colaboratory	29
Gambar 2.6 Peringkat Framework Deep Learning pada pasar	30
Gambar 3.1 Diagram alur tahap penelitian	35
Gambar 3.2 Arsitektur model yang digunakan	37
Gambar 3.3 Parameter optimizer, loss, dan metrics yang digunakan pada kompilasi model.....	37
Gambar 4.1 Rincian data dan jumlah pada dataset	40
Gambar 4.2 Distribusi kelas pada Dataset Stroke.....	41
Gambar 4.3 Hasil pengolahan data kategori objek dengan Ordinal Encoder	42
Gambar 4.4 Distribusi kelas dalam dataset pada tiap hasil near-miss undersampling.....	43
Gambar 4.5 Jumlah data hasil penerapan metode near-miss	43
Gambar 4.6 Grafik kurva akurasi training dan test pada dataset orisinal	45
Gambar 4.7 Grafik kurva loss training dan test pada dataset orisinal	46
Gambar 4.8 Confusion Matrix pada model dengan dataset orisinal.....	74
Gambar 4.9 Grafik kurva akurasi training dan test pada dataset near-miss 1	76
Gambar 4.10 Grafik kurva loss training dan test pada dataset near-miss 1	77
Gambar 4.11 Confusion Matrix pada model dengan dataset near-miss 1.....	82
Gambar 4.12 Grafik kurva akurasi training dan test pada dataset near-miss 2.....	83
Gambar 4.13 Grafik kurva loss training dan test pada dataset near-miss 2	84
Gambar 4.14 Confusion Matrix pada model dengan dataset near-miss 2.....	88
Gambar 4.15 Grafik kurva akurasi training dan test pada dataset near-miss 3.....	89
Gambar 4.16 Grafik kurva loss training dan test pada dataset near-miss 3	90
Gambar 4.17 Confusion Matrix pada model dengan dataset near-miss 3.....	95

DAFTAR LAMPIRAN

Lampiran 1. Tampilan Aplikasi MedikalTech Capstone Project 2021	Error! Bookmark not defined.
Lampiran 2. <i>Preprocessing Dataset</i>	Error! Bookmark not defined.
Lampiran 3. Model dengan <i>dataset</i> orisinil	Error! Bookmark not defined.
Lampiran 4. Model dengan <i>dataset</i> hasil <i>near-miss</i>	Error! Bookmark not defined.

DAFTAR ISTILAH DAN SINGKATAN

Dataset : Obyek yang merepresentasikan data dan relasinya di memori

Python : Bahasa pemrograman tingkat tinggi yang bersifat *open source*

Imbalanced dataset : Dataset yang jumlah data pada tiap kelasnya tidak seimbang

Imbalanced class : ketidakseimbangan jumlah kelas dalam *dataset*

Library : kumpulan kode yang memiliki fungsi tertentu yang dapat dipanggil

Majority class : kelas dengan data terbanyak dari keseluruhan dataset

Minority class : kelas dengan data tersedikit dari keseluruhan dataset

Classifier : Pengklasifikasi objek ke dalam kategori

KNN : *K-nearest neighbors*

Precision score : nilai/skor presisi

Recall score : nilai/skor pengingatan

DAFTAR SIMBOL

$f1$: nilai rata-rata harmonis dari *precision* dan *recall*

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Stroke merupakan suatu keadaan ketika terjadi tanda-tanda klinis berupa defisit neurologik dan global yang berkembang dengan cepat. Sehingga terjadi gangguan pada fungsi otak yang berkaitan dengan kurangnya pasokan darah ke otak atau terjadi adanya penyakit pembuluh darah. Penyakit stroke menjadi salah satu penyebab kematian dan kecacatan kronik yang paling tinggi pada usia di atas 45 tahun terbanyak di Indonesia.

Beberapa dataset medis yang memiliki dua kelas atau binomial class cenderung akan mengalami ketidakseimbangan kelas, hal ini disebut dengan *imbalanced dataset*, yaitu suatu kondisi ketika kelas tujuan yang akan diklasifikasi tidak memiliki rasio yang seimbang. Sehingga menyebabkan hasil klasifikasi yang bias karena *classifier* akan lebih condong mendeteksi kelas mayoritas dibanding dengan kelas minoritas [1].

Overfitting adalah suatu permasalahan mendasar dalam *machine learning* yang dapat menyebabkan tercegahnya generalisasi model yang sempurna agar sesuai dengan data yang diamati pada *train data*, serta data yang selain dari *train data*. Dikarenakan oleh adanya derau, terbatasnya ukuran dari *training set*, dan kompleksitas *classifier*, maka terjadilah *overfitting* [2].

Permasalahan *class imbalanced* mendapat perhatian yang signifikan dalam bidang *Machine Learning* dan *Pattern Recognition* dalam beberapa tahun ini. Dataset yang memiliki dua buah kelas, secara implisit akan mengalami

imbalanced ketika kelas minoritas sangat kontras jumlahnya dengan kelas mayoritas [3]. Kasus *imbalanced dataset* ini dapat diatasi dengan menerapkan metode *Near Miss undersampling*.

Near Miss undersampling merupakan salah satu metode *undersampling*. Ide utama dari metode ini adalah memilih himpunan data dari kelas mayoritas yang dekat dengan data kelas minoritas untuk mendapat merepresentasikan batas antar kelas lebih baik [4].

Oleh karena itu penulis akan melakukan penelitian dengan judul “**Metode *Near Miss Undersampling* Untuk Optimalisasi *Dataset* Pada Model Prediksi Penyakit Stroke**”.

1.2 Rumusan Masalah

Sesuai dengan latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Apakah metode *Near Miss Undersampling* dapat diterapkan dalam optimalisasi dataset?
2. Bagaimana pengaruh metode *Near Miss Undersampling* terhadap akurasi model?
3. Bagaimana hasil analisis dan evaluasi dari model?

1.3 Batasan Masalah

Agar dalam penelitian ini dapat diselesaikan dengan hasil yang optimal, maka berikut adalah batasan masalah penelitian ini.

1. Menggunakan model prediksi penyakit stroke pada aplikasi MedikalTech Capstone Project Bangkit 2021.

2. Dataset yang digunakan adalah dataset “*Stroke Prediction Dataset - Fedesoriano*” pada situs *kaggle*.
3. Metode *undersampling* yang digunakan adalah metode *Near Miss undersampling*.
4. Penerapan metode *Near Miss undersampling* ini menggunakan bahasa pemrograman *python* dengan menggunakan antarmuka dan infrastruktur *Google Colaboratory*.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut.

1. Penyempurnaan model prediksi penyakit stroke.
2. Penerapan metode Near Miss Undersampling pada dataset model prediksi penyakit stroke.

1.4.2 Manfaat

Manfaat yang diharapkan dalam penelitian ini adalah sebagai berikut.

1. Mampu menerapkan ilmu pengetahuan yang didapat pada mata kuliah yang bersangkutan dalam menyelesaikan tugas akhir.
2. Pengembangan model ke dalam aplikasi diagnosis stroke dapat memudahkan akses dan menekan tingginya biaya *pre-diagnosis* penyakit stroke.
3. Pengembangan model ke dalam aplikasi diagnosis stroke dapat menjadi referensi untuk mengembangkan sistem prediksi stroke pada Rumah Sakit.

1.5 Sistematika Penulisan

Adapun sistematika penulisan dalam penelitian ini adalah sebagai berikut.

BAB 1 Pendahuluan

Bab Pendahuluan ini berisikan mengenai judul penelitian, latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian.

BAB 2 Tinjauan Pustaka

Bab Tinjauan Pustaka ini berisikan berbagai gagasan yang menjadi landasan ide tentang penelitian terdahulu serta berbagai teori pendukung dalam proses penelitian.

BAB 3 Metode Penelitian

Bab Metode Penelitian berisikan mengenai metode penelitian yang dilakukan oleh penulis, seperti waktu dan tempat penelitian, alat dan bahan yang digunakan dalam penelitian, metode penelitian, sumber data, alur penelitian, dan *timeline* penelitian.

BAB 4 Hasil dan Pembahasan

Bab Pembahasan berisikan mengenai hasil dan analisis penelitian mengatasi masalah *overfitting* pada model prediksi penyakit stroke yang diakibatkan oleh terjadinya *imbalanced dataset* dengan menerapkan metode *near miss undersampling*. Hasil dari proses ini adalah meningkatnya perfoma dari model prediksi penyakit stroke yang dilihat dari tidak terjadinya masalah *overfitting*.

BAB 5 Penutup

Bab Penutup berisikan mengenai kesimpulan dan saran dari hasil yang didapatkan dari penelitian yang dilaksanakan oleh penulis.

BAB 2 TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

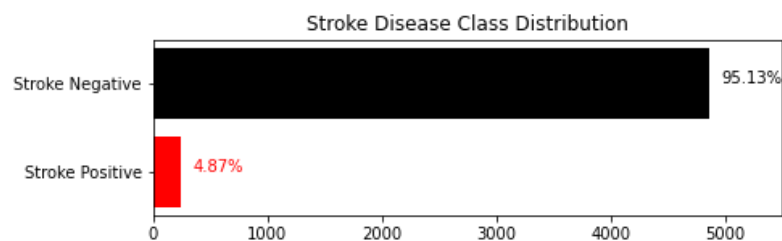
Dalam proses penelitian tugas akhir dengan judul “Metode *Near Miss Undersampling* Untuk Optimalisasi *Dataset* Pada Model Prediksi Penyakit Stroke”, penulis melakukan studi literatur dari beberapa penelitian terdahulu yang memiliki keterkaitan. Hal ini dilakukan demi mendalami dan memperdalam teori serta gagasan yang digunakan. Penelitian tersebut diantaranya adalah berikut.

Tabel 2.1 Rincian penelitian terdahulu

No	Nama Penulis	Judul	Metode	Dataset	Akurasi
1.	Nhlakanipho M. Mqadi	<i>Solving Misclassification of the Credit Card Imbalance Problem Using Near Miss</i>	<i>The Data-Point Approach</i> , Pemilihan fitur, <i>Near Miss-Based Undersampling</i> , Studi desain, <i>Datasets</i> (30% data <i>test</i> dan 70% data <i>training</i>), Algoritma klasifikasi, metrik performansi.	<i>Europeans Credit Card Fraud Detection</i> , 284807 sampel (492 data penipuan), presentase <i>minority class</i> adalah 0.173%. <i>UCI Credit Card Dataset</i> dari April 2005 sampai September 2005, 30000 sampel dan 6636 kasus penipuan, presentase <i>minority class</i> adalah 22.12%. Dengan parameter V1-V28, amount, dan class.	100% ~ 90% 78% ~ 73%
	Nalindren Naicker,				
	Timothy Adeliyi				
2.	Ajinkya More	<i>Survey of Resampling Techniques for Improving Classification Performance in Unbalanced Datasets</i>	Membagi dataset menjadi 70% data <i>training</i> dan 30% data <i>test</i> . Melakukan validasi silang 5 kali lipat pada set <i>training</i> untuk memilih parameter terbaik dan melaporkan hasil set <i>test</i> . Hasilnya diperoleh menggunakan <i>python library scikit-learn</i> dan <i>imbalanced-learn</i> .	<i>Make_classification</i> , 10000 sampel, 2 kelas, dengan rasio antar kelas adalah 1:9 sehingga presentase <i>minority class</i> sebesar 10%.	91%
3.	Timothy Oladunni	<i>COVID-19 County Level</i>	Data COVID-19, Normalisasi,	<i>John Hopkins</i>	47.73%

	Sourou Tossou	<i>Severity Classification with Imbalanced Dataset: A</i>	NearMiss Undersampling, Split data (90% Data <i>testing</i> dan 10% data <i>training</i>), membuat dan melatih model, evaluasi performa model, perbandingan performa model.	University COVID-19 Repository, 3006 sampel	
	Yayehyrad Haile				
	Adonias Kidane	<i>NearMiss Undersampling Approach</i>			

Dapat disimpulkan dari data penelitian terdahulu di atas, bahwa penelitian yang akan penulis lakukan menggunakan dataset dengan jumlah kelas dan parameter yang berbeda, juga menggunakan dataset yang berbeda pula, di mana menggunakan dataset *Stroke Prediction - Fedesoriano* dengan total data berjumlah 5110 data dengan distribusi kelas positif stroke 249 data dan kelas negatif stroke 4861 data, yang mana distribusi *minority class* sebesar 4.87% dan *majority class* sebesar 95.13%, distribusi kelas ini dapat dilihat pada gambar 2.1 di bawah. Serta *dataset* terbagi menjadi dua bagian dengan komposisi 80% data *training* dan 20% data *testing*.



Gambar 2.1 Distribusi kelas pada Dataset

2.2 Penyakit Stroke

Stroke merupakan suatu keadaan ketika terjadi tanda-tanda klinis berupa defisit neurologik dan global yang berkembang dengan cepat. Sehingga terjadi gangguan pada fungsi otak yang berkaitan dengan kurangnya pasokan darah ke otak

atau terjadi adanya penyakit pembuluh darah. Aliran darah yang berhenti dapat mengakibatkan pasokan zat makanan dan oksigen ke otak berhenti, sehingga terjadi tidak berfungsinya sebagian otak (Nabyl, 2012). WHO (2010) mendefinisikan bahwa stroke merupakan manifestasi klinis dari gangguan fungsi otak, baik secara fokal maupun secara global, yang berlangsung lebih dari 24 jam bahkan sampai menyebabkan kematian tanpa terdapat penyebab lain selain gangguan vaskuler

Penyakit stroke menjadi salah satu penyebab kematian dan kecacatan kronik yang paling tinggi pada usia di atas 45 tahun terbanyak di Indonesia. Berdasarkan Riset Kesehatan Dasar (Riskesdas) Nasional pada tahun 2013, prevalensi penyakit stroke di Indonesia berdasar atas diagnosis tenaga kesehatan mencapai angka tujuh per mil serta yang terdiagnosis oleh tenaga kesehatan atau gejala mencapai 12,1 per mil. Sebanyak 57.9% penyakit stroke sudah terdiagnosis oleh tenaga kesehatan.

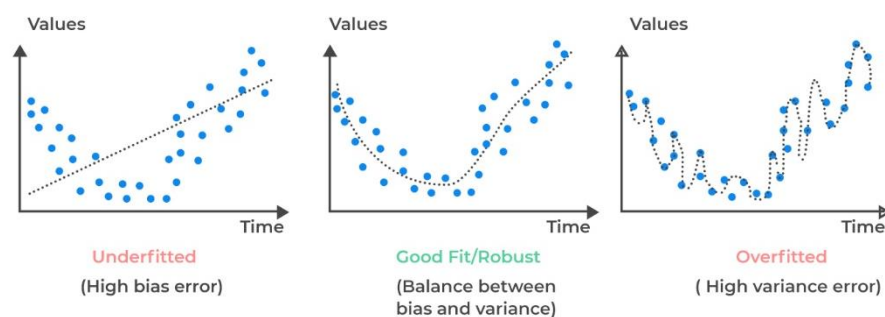
Prevalensi stroke berdasarkan diagnosis nakes tertinggi di Sulawesi Utara (10,8%), diikuti DI Yogyakarta (10,3%), Bangka Belitung dan DKI Jakarta masing-masing 9,7 per mil sedangkan Sumatera Barat 7,4 per mil. Prevalensi stroke berdasarkan diagnosis nakes dan gejala tertinggi terdapat di Sulawesi Selatan (17,9%), DI Yogyakarta (16,9%), Sulawesi Tengah (16,6%), diikuti Jawa Timur sebesar 16 per mil sedangkan Sumatera Barat sebesar 12,2 per mil [5].

Menurut Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan tahun 2016 Stroke menghabiskan biaya pelayanan kesehatan sebesar 1,43 Trilyun, tahun 2017 naik menjadi 2,18 Trilyun dan tahun 2018 mencapai 2,56 Trilyun rupiah.

2.3 Overfitting

Overfitting adalah suatu permasalahan mendasar dalam *machine learning* yang dapat menyebabkan tercegahnya generalisasi model yang sempurna agar sesuai dengan data yang diamati pada *train data*, serta data yang selain dari *train data*. Dikarenakan oleh adanya derau, terbatasnya ukuran dari *training set*, dan kompleksitas *classifier*, maka terjadilah *overfitting* [2].

Mendeteksi *overfitting* hampir tidak dapat dilakukan sebelum dilakukannya pengujian data. Hal ini dapat membantu dalam mengatasi karakteristik dari *overfitting*, yaitu ketidakmampuan dalam menggeneralisasi dataset. Performa model dapat diukur dengan mengamati presentase akurasi ketika tahap *train* dan validasi. Apabila performa model lebih baik pada saat tahap *train* daripada saat tahap validasi, kemungkinan besar model tersebut mengalami *overfitting*.



Gambar 2.2 Ilustrasi performa model

Secara visual performa dari model yang dihasilkan dapat dilihat pada gambar 2.1 di atas. Pada grafik paling kiri, merupakan grafik *underfitting*, yang menunjukkan bahwa model memiliki nilai galat yang sangat tinggi karena garis regresi yang dihasilkan memiliki jarak yang jauh terhadap data-data latihan.

Underfitting terjadi ketika masih ada ruang untuk perbaikan pada data latihan. Hal ini dapat terjadi karena beberapa alasan: Jika model tidak cukup kuat, terlalu diatur, atau tidak cukup lama dilatih. Ini berarti jaringan belum mempelajari pola yang relevan dalam data latihan .

Pada grafik paling kanan, merupakan grafik yang mempresentasikan bahwa model mengalami *overfitting*, menunjukkan bahwa garis regresi yang dihasilkan sangat akurat dan menyentuh setiap data latihan, meskipun menghasilkan nilai akurasi yang tinggi, namun *overfitting* ini sangat dihindari karena model tidak mampu dalam menggeneralisasi *dataset*. Untuk mencegah *overfitting*, solusi terbaik adalah menggunakan data pelatihan yang lebih lengkap. Dataset harus mencakup berbagai input yang diharapkan dapat ditangani oleh model. Data tambahan mungkin hanya berguna jika mencakup kasus-kasus baru dan menarik.

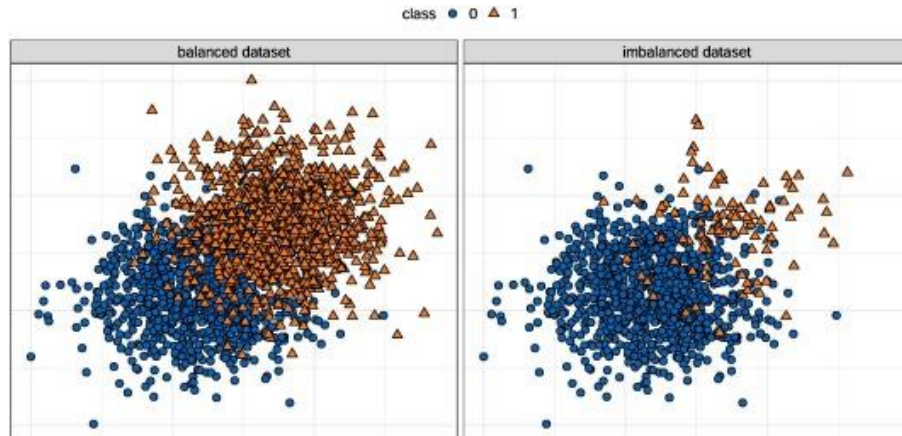
Lalu grafik yang berada ditengah, menunjukan performa model yang mengalami *good fit* suatu kondisi yang tidak mengalami *overfitting* ataupun *underfitted*, kondisi yang seimbang antara biasnya dan variansinya.

2.4 Imbalanced Dataset

Imbalanced dataset adalah suatu data set yang biasa digunakan dalam klasifikasi memiliki perbedaan jumlah yang sangat kontras dari tiap kelasnya. Pada *imbalanced dataset* terdapat kelas dengan data yang sedikit, disebut dengan kelas minoritas, dan terdapat pula kelas dengan jumlah yang sangat banyak, disebut dengan kelas mayoritas.

Permasalahan *class imbalanced* ini mendapat perhatian yang signifikan dalam bidang *Machine Learning* dan *Pattern Recognition* dalam beberapa tahun ini. Dataset yang memiliki dua buah kelas, secara implisit akan mengalami *imbalanced* ketika kelas minoritas sangat kontras jumlahnya dengan kelas mayoritas [3].

Adanya ketidakseimbangan distribusi kelas dapat mempengaruhi performa dari suatu model *machine learning*, karena model klasifikasi akan bekerja dengan mengasumsikan bahwa distribusi kelas pada dataset relatif seimbang dan memiliki biaya galat klasifikasi yang sama. Hal ini tentu menimbulkan resiko terjadinya kesalahan klasifikasi terhadap dataset, yang akan mengakibatkan tidak maksimalnya performa dari suatu model.



Gambar 2.3 Ilustrasi kondisi dataset

Secara visual kondisi distribusi kelas pada dataset dapat dilihat pada Gambar 2.2 di atas ini. Dapat dilihat bahwa dataset dengan kondisi “*balanced dataset*” atau dataset seimbang, memiliki distribusi data yang sama banyak antara data biru bulat dengan data segitiga jingga, sebaliknya pada kondisi dataset tidak

seimbang atau “*imbalanced dataset*” terlihat bahwa data segitiga jingga sangat sedikit jika dibandingkan dengan data biru bulat.

Sebagai contoh dari kasus yang sering terjadi adanya *imbalanced dataset* yaitu seperti *credit scoring*, *fraud*, data penyakit, dan lain-lain. Dari kasus tersebut biasanya pada *imbalanced dataset* terdapat *minority class* dengan proporsi kurang dari 1% dari keseluruhan data.



Gambar 2.4 Distribusi dari kasus *imbalanced dataset*

Misal *minority class* 0.2% dari data keseluruhan. Dengan asumsi model yang dimiliki tidak cukup baik untuk membedakan antara *minority class* dengan *majority class*. Model memprediksi seluruh data menjadi kategori *majority class*. Maka nilai akurasi yang didapat adalah sebesar 99.8%, terlihat bahwa model memiliki nilai akurasi yang sangat tinggi, namun model tersebut tidak dapat membedakan antara *minority class* dengan *majority class* [6]. Situasi ini membuat *classifier* kesulitan untuk menentukan *minority class* dan kemungkinan besar faktanya adalah *minority class* akan diprediksi sebagai *majority class* [7]. Sehingga dapat menyebabkan kekeliruan dalam melakukan klasifikasi. Data yang mengalami *imbalanced dataset* antar *class* akan mengandalkan *class* mayoritas dalam pengklasifikasi. Konsekuensi dari diagnosis yang dilakukan adalah mendapatkan hasil klasifikasi yang salah [8].

Kasus *imbalanced dataset* pula akan lebih sulit untuk dilakukan berbagai pengolahan dan analisis data seperti klasifikasi, pengklasteran, prediksi, dan sebagainya. Hal tersebut terjadi karena model analisis data tidak dirancang untuk mempertimbangkan distribusi kelas dalam meningkatkan nilai akurasi dari model. Banyak penelitian yang melaporkan hasil analisis data dengan menggunakan *imbalanced dataset* seringkali memberi hasil yang keliru [9].

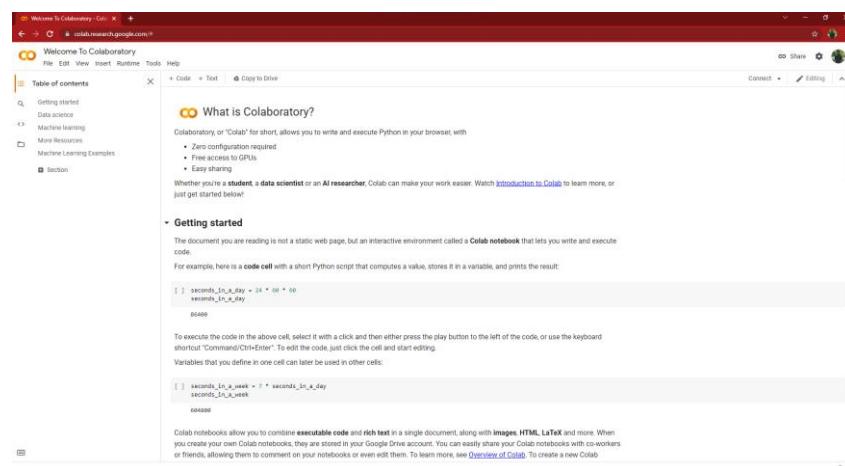
Pada data yang mengalami *imbalanced dataset* yang digunakan akan menghasilkan model dengan hasil klasifikasi yang tidak baik. Model yang tidak baik tersebut sehubungan dengan adanya peristiwa *overfitting* dan tidak dapat melakukan klasifikasi data dengan baik. Yang mana *overfitting* pada kasus *imbalanced dataset* ini dapat terjadi karena kurva logistik yang dibentuk akan cenderung untuk mengarah pada salah satu kelas saja [10].

Seperti pada penelitian Zeju Li, Konstantinos Kamnitsas, dan Ben Glocker [11], temuan mereka mengungkapkan bahwa model yang *overfitting* dengan data *train* yang *imbalanced* memiliki bias untuk mengelompokkan kelas yang kurang terwakili pada data validasi. Pelatihan dengan lebih sedikit data mengarah pada *overfitting*.

2.5 Google Colaboratory

Google Colaboratory atau *Colab* adalah sebuah *executable document* yang dapat digunakan untuk menyimpan, menulis, serta membagikan program yang telah ditulis melalui *Google Drive*. Software ini pada dasarnya serupa dengan *Jupyter Notebook* gratis berbentuk cloud yang dijalankan menggunakan browser, seperti *Mozilla Firefox* dan *Google Chrome* [12].

Memungkinkan pengguna untuk menjalankan kode Python tanpa perlu melakukan proses instalasi dan setup lainnya. Justru, semua keperluan setting dan adjustment akan diserahkan ke cloud. Keuntungan terbesar dari *Google Colaboratory* adalah bahwa ia memiliki kumpulan *built-in-library machine learning* paling populer yang dapat dimuat dengan mudah dalam *notebook*.



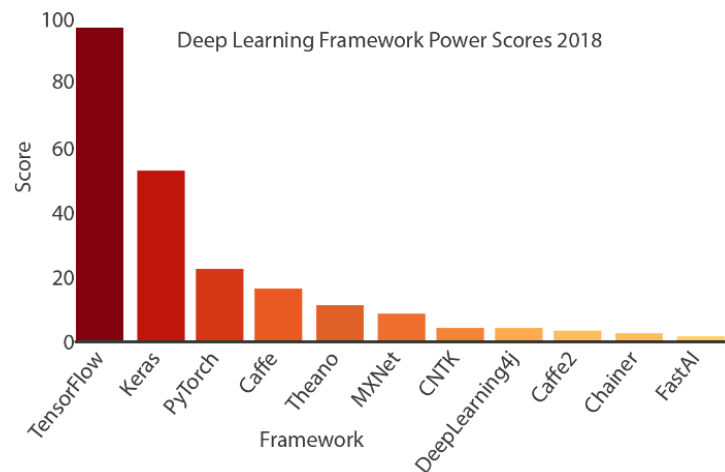
Gambar 2.5 Tampilan Google Colaboratory

Tidak sedikit kelebihan yang ditawarkan oleh *Google Colaboratory*, antara lainnya adalah sebagai berikut.

1. *Built-in-library machine learning* yang lengkap.
2. Berbasis cloud, sehingga tidak memakan space dalam memori komputer.
3. Data dalam *Google Colaboratory* dapat diakses dan diedit dengan mudah.
4. Mempermudah proses kolaborasi antar tim.
5. Memiliki fitur GPU dan TPU yang dapat dimanfaatkan secara gratis.

2.6 Framework Tensorflow dan Keras

TensorFlow adalah *end-to-end platform* sumber terbuka untuk *machine learning*. Merupakan alat, pustaka, dan sumber daya lain yang komprehensif dan fleksibel yang menyediakan framework dengan API tingkat tinggi [13].



Gambar 2.6 Peringkat Framework Deep Learning pada pasar

Keras adalah API *deep learning* yang ditulis dengan *Python*, berjalan di atas *platform machine learning* TensorFlow. Menggunakan Keras dalam *deep learning* memungkinkan pembuatan prototipe yang mudah dan cepat serta berjalan mulus di CPU dan GPU [14]. *Framework* ini ditulis dalam kode *Python* yang mudah didebug dan memungkinkan kemudahan untuk diperpanjang. Keuntungan utama Keras dijelaskan di bawah ini:

- a. *User-Friendly*: Keras memiliki antarmuka yang sederhana dan konsisten yang dioptimalkan untuk kasus penggunaan umum yang memberikan umpan balik yang jelas dan dapat ditindaklanjuti untuk kesalahan pengguna.
- b. Modular dan Dapat Disusun: Model Keras dibuat dengan menghubungkan blok yang dapat dikonfigurasi bersama-sama, dengan sedikit batasan.

- c. Mudah Diperpanjang: Dengan bantuan Keras, dapat dengan mudah menulis blok bangunan khusus untuk ide dan penelitian baru.
- d. Mudah Digunakan: Keras menawarkan API yang konsisten dan sederhana yang membantu meminimalkan jumlah tindakan pengguna yang diperlukan untuk kasus penggunaan umum, juga memberikan umpan balik yang jelas dan dapat ditindaklanjuti atas kesalahan pengguna.

2.7 Metode *Undersampling*

Metode *Undersampling heuristic* digunakan untuk dua dasar pendekatan: suatu metode kelas diidentifikasi dan data yang tidak penting untuk *learning* akan dihapus. Metode kelas yang kedua lebih berfungsi sebagai metode *data cleaning*, mengidentifikasi derau dari data. Pada umumnya, metode *data cleaning* digunakan sebagai metode *undersampling*, hanya data pada kelas mayoritas yang dipangkas, meskipun data pada kelas minoritas dicurigai sebagai derau. Argumen yang menyatakan memangkas sebagian dari data pada kelas mayoritas, karena data pada kelas minoritas terbilang langka. Berikut merupakan yang termasuk metode *undersampling*: *NearMiss*; *Condensed Nearest Neighbor Rule*; *Tomek Links*; *One-sided selection*; *Neighborhood Cleaning Rule*.

2.8 Metode *Near Miss Undersampling*

Near Miss Undersampling merupakan salah satu metode *undersampling*. Ide utama dari metode ini adalah memilih himpunan data dari kelas mayoritas yang dekat dengan data kelas minoritas untuk mendapat merepresentasikan batas antar kelas lebih baik.

Near Miss Undersampling mengacu pada kumpulan metode *undersampling* yang memilih contoh data berdasarkan pada jarak contoh data kelas mayoritas terhadap kelas minoritas. Metode pendekatan tersebut diusulkan oleh Jianping Zhang dan Inderjeet Mani dalam makalah mereka pada tahun 2003 yang berjudul “*KNN Approach to Unbalanced Data Distributions: A Case Study Involving Information Extraction*”.

Terdapat tiga buah versi dari metode *near miss* ini, yaitu *NearMiss-1*, *NearMiss-2*, dan *NearMiss-3*. Berdasarkan penentuan jarak ruang fitur menggunakan jarak Euclidean, berikut merupakan perbedaan dari ketiga versi tersebut.

1. *NearMiss-1* : contoh data pada kelas mayoritas dengan jarak rata-rata minimum ketiga contoh data kelas minoritas terdekat.
2. *NearMiss-2* : contoh data pada kelas mayoritas dengan jarak rata-rata minimum ketiga contoh data kelas minoritas terjauh.
3. *NearMiss-3* : contoh data pada kelas mayoritas dengan jarak minimum ke setiap contoh data pada kelas minoritas.

BAB 3 METODE PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan dalam waktu 3 bulan dimulai dari bulan Juli 2021 hingga akhir bulan Desember 2021 bertempat di Kampus Fakultas Teknik Unsoed, Kabupaten Purbalingga.

3.2 Alat dan Bahan

Selama penelitian, berikut adalah daftar alat dan bahan penunjang proses penelitian.

1. *Hardware*:

- a. Satu unit *laptop* Asus VivoBook X412FL.307 dengan spesifikasi *processor Intel core i5*, kapasitas RAM 8Gb, dan sistem operasi *Windows 10 64-bit*.
- b. Komputer virtual *Google Colaboratory* dengan spesifikasi GPU: 1x Tesla K80, dengan inti CUDA 2496, nilai komputasi 3.7, VRAM sebesar 12 Gb GDDR5, CPU: 1x *single core hyper threaded Xeon Processor @2.3 GHz*, RAM: ~12.6 Gb, *Disk*: ~33Gb.

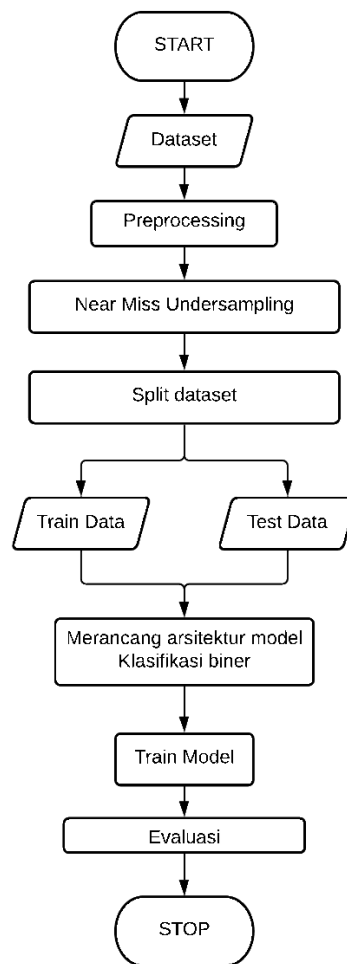
2. *Software*:

- a. Sistem operasi *Windows 10 64-bit*
- b. Peramban internet *Google Chrome* versi 92.0.4515.107 64-bit
- c. *Google Colaboratory* atau *Jupyter Notebook*
- d. *Version Control System* dan layanan repositori pada *GitHub*

3. Dataset *train* dan *validation* berupa data numerik yang terdiri dari parameter-parameter kesehatan tubuh seperti id, jenis kelamin, usia, hipertensi, sakit jantung, status menikah, tipe pekerjaan, tipe tempat tinggal, kadar glukosa, bmi, status merokok, dan stroke untuk mendeteksi penyakit stroke yang didapatkan dari situs *kaggle. Stroke Prediction Dataset - Fedesoriano* dengan total data berjumlah 5110 data dengan distribusi kelas positif stroke 249 data dan kelas negatif stroke 4861 data, dapat dilihat bahwa dataset ini mengalami *imbalanced dataset* dengan presentase *minority class* sebesar 4.87% dari jumlah keseluruhan data.

3.3 Alur dan Tahap Penelitian

Penelitian dilakukan untuk meningkatkan performa pada model prediksi penyakit stroke yang mengalami ketidakseimbangan kelas pada dataset dengan menerapkan metode *near-miss undersampling*. Terdapat beberapa tahapan yang dilakukan dalam penelitian ini. Berikut ini adalah tahap-tahap yang dilakukan.



Gambar 3.1 Diagram alur tahap penelitian

3.3.1 Tahap Persiapan

Tahap persiapan penelitian merupakan hal yang pertama dilakukan. Pada tahap ini penulis terlebih dahulu menentukan topik yang akan dikembangkan, yakni optimalisasi model yang mengalami *imbalanced dataset* pada datesetnya. Selanjutnya merumuskan berbagai masalah yang terjadi, menentukan tujuan serta manfaat dari penelitian, dan menentukan batasan masalah. Dilanjutkan dengan studi literatur mengenai kondisi penyakit stroke di Indonesia.

Pada tahap studi literatur pula mencari serta mengumpulkan referensi berupa jurnal, buku, artikel, maupun referensi lainnya yang memiliki keterkaitan dengan penelitian ini.

3.3.2 Tahap Persiapan Model dan Pre-Proses Dataset

Pada tahap persiapan model dan dataset, penulis menggunakan model prediksi stroke yang sudah dirancang pada *Capstone Project* MedikalTech, lalu mencari dataset penyakit stroke yang didapatkan dari situs *kaggle*.

Dataset stroke mula-mula melalui tahap pre-proses, penulis mengolah *dataset* membuang kolom yang tidak perlu, mengisi data yang kosong atau “*NaN*”, serta melakukan *encoding* dengan mengubah data pada dataset yang bertipe kageteri menjadi data dengan tipe vektor biner.

3.3.3 Tahap Penerapan Metode *Near Miss Undersampling*

Pada tahap ini, penulis menulis *source code* algoritma metode *near miss undersampling* dalam bahasa pemrograman *python* pada platform *Google Colaboratory* dengan mengimplementasikan *library imbalanced-learn* dan disimpan pada *Github* dengan format file “*ipynb*”.

Selanjutnya *source code* metode ini digunakan untuk mengolah dataset stroke yang mengalami *imbalanced dataset*. Proses pengolahan dataset ini pula dilakukan pada platform *Google Colaboratory*. Dari penerapan metode *near-miss undersampling* akan didapat dataset stroke baru yang tidak mengalami ketidakseimbangan yang siap untuk dilatih kepada model.

3.3.4 Tahap Perbandingan Dataset Sebelum dan Sesudah Dioptimalisasi

Pada tahap perbandingan dataset ini, penulis melakukan perbandingan antara dataset yang belum dioptimalisasi dengan dataset yang sudah dioptimalisasi menggunakan metode *NearMiss Undersampling*, hal yang diamati adalah jumlah data pada masing-masing kelas pada dataset, yang mana akan terlihat pada dataset yang sudah optimalisasi jumlah data masing-masing kelas berjumlah sama.

3.3.5 Tahap Perancangan Arsitektur Model

Model yang disusun menggunakan *framework tensorflow keras* memiliki arsitektur sebagai berikut ini.

```
model = tf.keras.Sequential([
    tf.keras.layers.Dense(16, activation='relu', input_shape=[10]),
    tf.keras.layers.Dense(16, activation='relu'),
    tf.keras.layers.Dense(1, activation='sigmoid')
])
```

Gambar 3.2 Arsitektur model yang digunakan

Model memiliki sepuluh buah node input yang dapat dilihat pada *input shape* dengan nilai 10. Enam belas buah node pada *dense layer* 1 dengan aktivasi ReLU, enam belas buah node pada *dense layer* 2 dengan aktivasi ReLU, serta satu buah node output dengan aktivasi *sigmoid*.

```
model.compile(optimizer='adam',
              loss='BinaryCrossentropy',
              metrics=['accuracy'])
```

Gambar 3.3 Parameter *optimizer*, *loss*, dan *metrics* yang digunakan pada kompilasi model

Sebelum model memasuki tahap latihan, model perlu dikompilasi terlebih dahulu dengan parameter yang ditunjukkan pada gambar 3.3 di atas. *Optimizer* yang digunakan adalah “adam” dan *loss function* yang digunakan adalah *binary*

crossentropy karena model ditujukan untuk melakukan klasifikasi biner yang mana untuk menentukan terkenatidaknya seseorang pada penyakit stroke, serta akurasi di sini menjadi besaran yang ingin ditampilkan pada proses latihan.

3.3.6 Tahap Pengujian

Tahap pengujian dilakukan dengan melatih model prediksi penyakit stroke menggunakan *dataset Stroke Prediction Dataset - Fedesoriano* yang belum dilakukan *sampling* dan *dataset* yang sudah dilakukan *sampling*. Setiap *dataset* dibagi dengan proporsi yang sama, yakni 80% data *train* dan 20% data validasi. *Dataset* pertama yang digunakan adalah *dataset* orisinil dengan rincian total data berjumlah 5110 data dengan distribusi kelas positif stroke 249 data dan kelas negatif stroke 4861 data, dan presentase *minority class* sebesar 4.87% terhadap keseluruhan data. Lalu *dataset* kedua yang digunakan adalah *dataset* yang sudah melalui *near-miss undersampling* yang perbandingan antar kelasnya adalah 1:1, dengan masing-masing kelas memiliki jumlah data yang sama yakni 249, sehingga *dataset* tersebut sudah tidak mengalami *imbalanced*.

3.3.7 Tahap Analisis dan Evaluasi

Selanjutnya tahap analisis dan evaluasi yang merupakan tahap terakhir dari penelitian ini. Pada tahap ini, penulis melakukan pencatatan terhadap hasil prediksi yang dikeluarkan oleh model pada tahap pengujian, lalu penulis melakukan pengamatan serta analisis terhadap data yang sudah diperoleh dan parameter akurasi dari model yang sudah dilatih. Lalu penulis melakukan evaluasi model berdasarkan hasil pengamatan dan analisis, guna melihat apakah penerapan metode

near-miss undersampling memberikan hasil yang signifikan dalam mengatasi *overfitting* akibat dataset yang tidak seimbang.

3.4 Waktu dan Jadwal Pelaksanaan Tugas Akhir

Penelitian ini dilaksanakan dalam waktu 5 bulan dimulai dari bulan Agustus 2021 hingga akhir bulan Desember 2021, berikut merupakan jadwal penelitian yang disusun secara rinci.

Tabel 3.1 Jadwal rinci penelitian

[illegible]

BAB 4 HASIL DAN PEMBAHASAN

4.1 Pengolahan *Dataset*

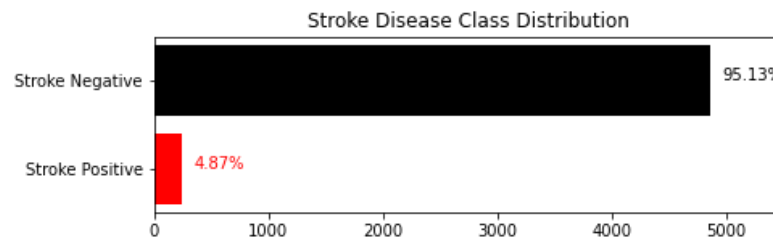
Dataset yang digunakan pada klasifikasi penyakit stroke terdiri dari dua buah *dataset*, yakni *dataset* yang mengalami *imbalanced class* dan *dataset* yang *balanced class* dengan menerapkan metode *near-miss undersampling*. Masing-masing *dataset* ini terbagi ke dalam dua bagian, yakni data *train* dan data *test*. Data *train* dan data *test* dari masing-masing *dataset* akan digunakan untuk melakukan pelatihan untuk memperoleh dua buah model yang dihasilkan dari dua buah *dataset* tersebut. Seluruh *dataset* tersebut diperoleh dari <https://www.kaggle.com/fedesoriano/stroke-prediction-dataset> milik Fedesoriano.

Adapun rincian serta jumlah data dari *dataset* orisinal sebelum melalui tahap *preprocessing* dan penerapan metode *near-miss undersampling* yang dapat dilihat pada gambar 4.1 dan 4.2 berikut ini.

	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	9046	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	51676	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	NaN	never smoked	1
2	31112	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
3	60182	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
4	1665	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1
...
5105	18234	Female	80.0	1	0	Yes	Private	Urban	83.75	NaN	never smoked	0
5106	44873	Female	81.0	0	0	Yes	Self-employed	Urban	125.20	40.0	never smoked	0
5107	19723	Female	35.0	0	0	Yes	Self-employed	Rural	82.99	30.6	never smoked	0
5108	37544	Male	51.0	0	0	Yes	Private	Rural	166.29	25.6	formerly smoked	0
5109	44679	Female	44.0	0	0	Yes	Govt_job	Urban	85.28	26.2	Unknown	0

5110 rows x 12 columns

Gambar 4.1 Rincian data dan jumlah pada dataset



Gambar 4.2 Distribusi kelas pada Dataset Stroke

Rincian jumlah data antara *minority class* yakni kelas yang menyatakan bahwa seseorang positif stroke dan *majority class* yang menyatakan seseorang negatif stroke, dapat dilihat persentasenya pada gambar 4.2 di atas, dengan jumlah 4861 data pada *majority class* dan 249 data pada *minority class*.

4.1.1 Preprocessing Dataset

Dataset yang akan digunakan untuk melatih model prediksi penyakit *stroke* mula-mula melalui tahap *preprocessing*. Tahap *preprocessing* pada penelitian penulis ini dilakukan melalui *Google Colab*, untuk melihat distribusi kelas pada *dataset* tersebut yang dapat dilihat pada gambar 4.2 di atas, jumlah data *minority class* dan *majority class* dalam persentase masing-masing adalah 4,87% dan 95,13%, karena terdapat perbedaan jumlah data yang sangat jauh maka *dataset* penyakit stroke ini mengalami *imbalanced class*.

Serta ditemui bahwa terdapat data yang kosong atau “NaN” pada kolom *BMI*, sehingga untuk menanggulangi kekosongan data ini yaitu diisi dengan menggunakan nilai rata-rata dari keseluruhan data pada kolom *BMI*.

Data yang bersifat kategori objek pada *dataset* diolah menjadi data kategori dalam bentuk *integer* dengan menggunakan metode *Ordinal Encoder*, kolom dengan data kategori objek tersebut ialah *gender*, *ever-married*, *work type*,

residence type, dan *smoking status*. Berikut ini hasil dari *Ordinal Encoder* yang dilakukan pada *dataset*.

```

Before Encoding Categorical Data

[30] for col in categorical_col:
      print(df[col].unique())

['Male' 'Female' 'Other']
['Yes' 'No']
['Private' 'Self-employed' 'Govt_job' 'children' 'Never_worked']
['Urban' 'Rural']
['formerly smoked' 'never smoked' 'smokes' 'Unknown']

[31] #create encoder using OrdinalEncoder
ord = OrdinalEncoder()
df_encoded = df.copy()

#encoding categorical data
for col in categorical_col:
    df_encoded[col] = ord.fit_transform(df_encoded[[col]]).astype(np.int64)

After Encoding Categorical Data

[32] for col in categorical_col:
      print(df_encoded[col].unique())

[1 0 2]
[1 0]
[2 3 0 4 1]
[1 0]
[1 2 3 0]

```

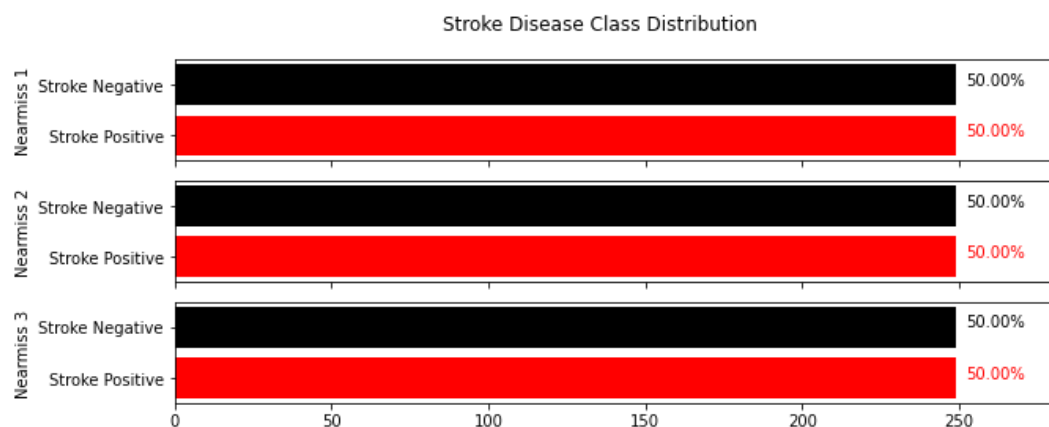
Gambar 4.3 Hasil pengolahan data kategori objek dengan *Ordinal Encoder*

4.1.2 Penerapan Metode *Near-Miss Undersampling*

Pada tahap *preprocessing* ditemui bahwa *dataset* penyakit stroke yang akan digunakan mengalami *imbalanced class*, hal ini perlu dihindari karena berdasarkan dari hasil proyek penulis pada *Capstone Project* Program Bangkit, model yang dilatih dengan menggunakan *dataset* yang mengalami *imbalanced class* akan menghasilkan model yang *overfitting* terhadap kelas yang memiliki jumlah distribusi lebih banyak atau *majority class*.

Oleh karena itu, untuk mengatasi permasalahan *imbalanced class* pada *dataset* ini penulis menerapkan metode *undersampling* terhadap *dataset* yakni, metode *Near-miss undersampling*. Metode *near-miss undersampling* ini memiliki

tiga jenis yang masing-masing memiliki pendekatan terhadap data yang berbeda. Hasil dari proses *near-miss undersampling* dapat dilihat pada gambar 4.4 di bawah ini, terlihat bahwa distribusi antar kelas pada masing-masing *dataset* hasil dari setiap tipe metode *near-miss undersampling* menghasilkan pebandingan yang sama, dengan distribusi antar kelasnya adalah 1:1.



Gambar 4.4 Distribusi kelas dalam dataset pada tiap hasil *near-miss undersampling*

Serta pada gambar 4.5, menunjukan bahwa setelah melalui metode *near-miss undersampling*, *majority class* mengalami pemangkasan data yang pada mulanya sebanyak 4861 data menjadi seimbang dengan jumlah data pada *minority class* yakni sebanyak 249 data.

```
[24] print('Nearmiss Undersampling Result')
      print('Nearmiss 1 : ', Counter(df_near1['stroke']).items())
      print('Nearmiss 2 : ', Counter(df_near2['stroke']).items())
      print('Nearmiss 3 : ', Counter(df_near3['stroke']).items())

Nearmiss Undersampling Result
Nearmiss 1 : dict_items([(0, 249), (1, 249)])
Nearmiss 2 : dict_items([(0, 249), (1, 249)])
Nearmiss 3 : dict_items([(0, 249), (1, 249)])
```

Gambar 4.5 Jumlah data hasil penerapan metode *near-miss*

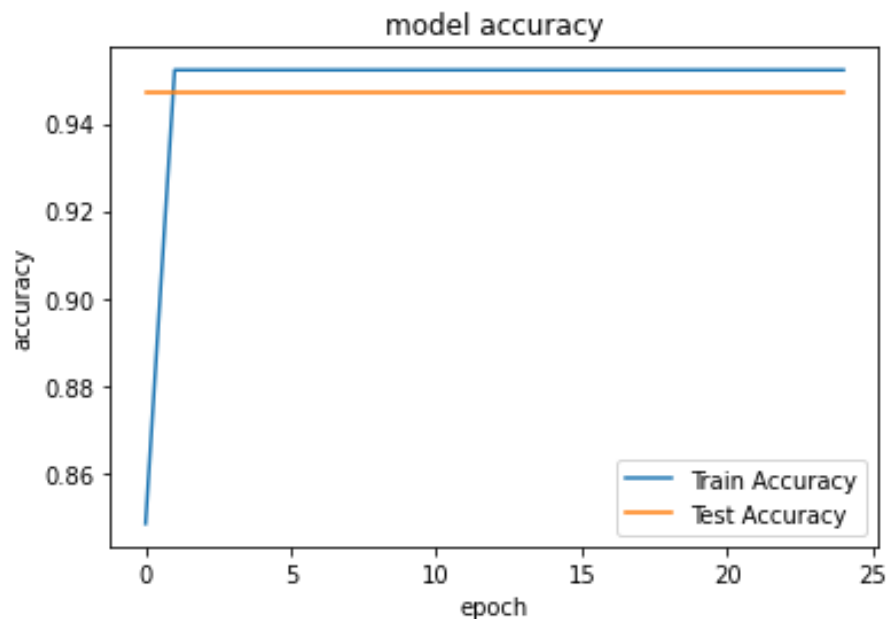
4.2 Hasil Pelatihan dan Pengujian

Pelatihan dan pengujian ini dilakukan dengan menggunakan satu arsitektur model yang sama serta empat buah *dataset* yang berbeda, yakni *dataset* dengan kondisi *imbalanced class*, *dataset* hasil dari *near-miss* 1, *dataset* hasil dari *near-miss* 2, dan *dataset* hasil dari *near-miss* 3. Perbedaan perilaku terhadap model yang dilakukan melalui *dataset* yang variatif adalah untuk melihat perbedaan hasil performa model yang diperoleh, apakah dengan menerapkan metode *near-miss undersampling* ini dapat menghasilkan model dengan performa yang lebih baik dibandingkan dengan model yang dilatih dengan *dataset* yang mengalami *imbalanced class*.

4.2.1 Dataset tanpa Near-Miss Undersampling

Proses *train* model dan pengujian menggunakan *dataset* orisinal, tanpa melalui *near-miss undersampling*, dengan 10 buah data sebagai input. *Dataset* terbagi menjadi dua bagian dengan komposisi 80% data *train* dan 20% data *test* yang masing-masing berjumlah 4088 dan 1022 data. Serta pada tahap pengujian *dataset* yang digunakan adalah data *test*.

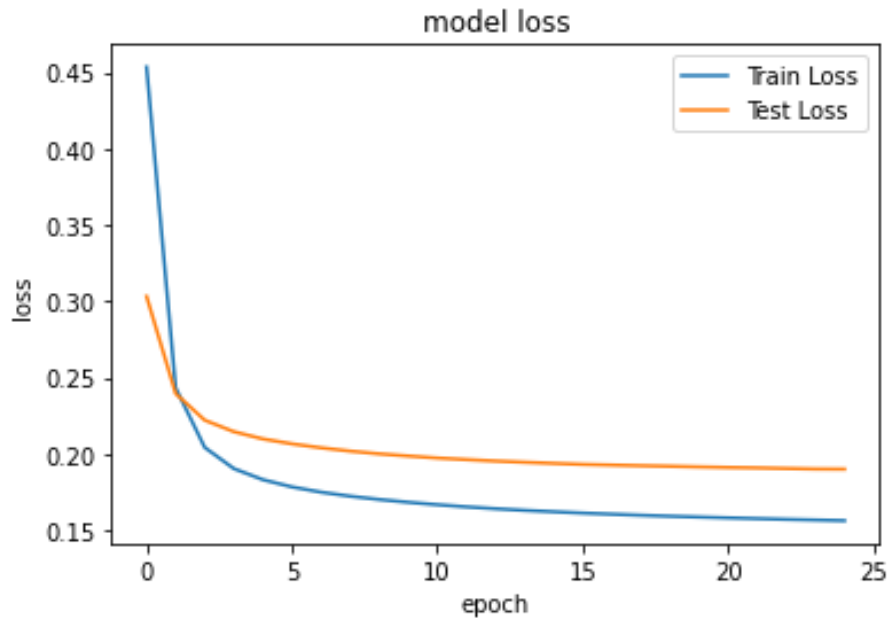
Setelah itu model dengan arsitektur *deep neural network keras* dengan *input shape* sebanyak 10. Delapan buah node pada *dense layer* 1 dengan aktivasi ReLU, empat buah node pada *dense layer* 2 dengan aktivasi ReLU, serta satu buah node output dengan aktivasi sigmoid. Hasil dari proses *train* dan *test* dapat dilihat pada grafik dalam gambar 4.6.



Gambar 4.6 Grafik kurva akurasi training dan test pada dataset orisinal

Dapat dilihat pada gambar 4.6 di atas, terdapat dua buah grafik yang masing-masing merepresentasikan nilai akurasi seiring bertambahnya *epoch*. Grafik berwarna biru merupakan grafik yang menunjukkan nilai akurasi *train* dan grafik berwarna jingga menunjukkan nilai akurasi *test*. Dari kedua grafik tersebut menunjukkan kurva akurasi *train* yang konstan mendekati nilai akurasi 100%, yakni sebesar 95,53% sejak *epoch* ke 2, namun pada grafik akurasi *test* nilai akurasi sudah konstan pada nilai 94,72% sejak *epoch* pertama. Dengan hasil yang ditampilkan oleh grafik di atas, bahwa model yang dihasilkan memiliki nilai akurasi yang tinggi serta dapat dinyatakan berhasil dalam melakukan klasifikasi.

Adapun pada gambar 4.7 menampilkan nilai *loss* yang diperoleh dari hasil proses *train* dan *test*.



Gambar 4.7 Grafik kurva loss training dan test pada dataset orisinal

Pada gambar 4.7 di atas terdapat grafik berwarna biru merupakan grafik yang menunjukkan nilai *train loss* dan grafik berwarna jingga menunjukkan nilai *test loss*. Nilai *train loss* dan *test loss* pada akhir *epoch* masing-masing sebesar 0.1588 dan 0.1830. Dengan hasil yang ditampilkan oleh grafik di atas, bahwa model yang dihasilkan memiliki nilai *loss* yang rendah sehingga dapat dinyatakan bahwa model memiliki kemampuan klasifikasi yang baik.

Tabel 4.1 Data Uji Menggunakan Model dengan Dataset Orisinal

Uji ke-	Data Uji	Hasil Prediksi	Benar/Salah
1	Positif Stroke	Negatif Stroke	Salah
2	Negatif Stroke	Negatif Stroke	Benar
3	Negatif Stroke	Negatif Stroke	Benar
4	Negatif Stroke	Negatif Stroke	Benar
5	Negatif Stroke	Negatif Stroke	Benar
6	Negatif Stroke	Negatif Stroke	Benar
7	Negatif Stroke	Negatif Stroke	Benar
8	Negatif Stroke	Negatif Stroke	Benar
9	Negatif Stroke	Negatif Stroke	Benar

10	Negatif Stroke	Negatif Stroke	Benar
11	Negatif Stroke	Negatif Stroke	Benar
12	Negatif Stroke	Negatif Stroke	Benar
13	Negatif Stroke	Negatif Stroke	Benar
14	Negatif Stroke	Negatif Stroke	Benar
15	Negatif Stroke	Negatif Stroke	Benar
16	Negatif Stroke	Negatif Stroke	Benar
17	Negatif Stroke	Negatif Stroke	Benar
18	Negatif Stroke	Negatif Stroke	Benar
19	Negatif Stroke	Negatif Stroke	Benar
20	Negatif Stroke	Negatif Stroke	Benar
21	Negatif Stroke	Negatif Stroke	Benar
22	Negatif Stroke	Negatif Stroke	Benar
23	Negatif Stroke	Negatif Stroke	Benar
24	Negatif Stroke	Negatif Stroke	Benar
25	Positif Stroke	Negatif Stroke	Salah
26	Negatif Stroke	Negatif Stroke	Benar
27	Negatif Stroke	Negatif Stroke	Benar
28	Negatif Stroke	Negatif Stroke	Benar
29	Negatif Stroke	Negatif Stroke	Benar
30	Negatif Stroke	Negatif Stroke	Benar
31	Negatif Stroke	Negatif Stroke	Benar
32	Negatif Stroke	Negatif Stroke	Benar
33	Positif Stroke	Negatif Stroke	Salah
34	Negatif Stroke	Negatif Stroke	Benar
35	Negatif Stroke	Negatif Stroke	Benar
36	Negatif Stroke	Negatif Stroke	Benar
37	Negatif Stroke	Negatif Stroke	Benar
38	Negatif Stroke	Negatif Stroke	Benar
39	Negatif Stroke	Negatif Stroke	Benar
40	Negatif Stroke	Negatif Stroke	Benar
41	Negatif Stroke	Negatif Stroke	Benar
42	Negatif Stroke	Negatif Stroke	Benar
43	Negatif Stroke	Negatif Stroke	Benar
44	Negatif Stroke	Negatif Stroke	Benar
45	Negatif Stroke	Negatif Stroke	Benar
46	Negatif Stroke	Negatif Stroke	Benar
47	Negatif Stroke	Negatif Stroke	Benar
48	Negatif Stroke	Negatif Stroke	Benar

49	Negatif Stroke	Negatif Stroke	Benar
50	Negatif Stroke	Negatif Stroke	Benar
51	Negatif Stroke	Negatif Stroke	Benar
52	Negatif Stroke	Negatif Stroke	Benar
53	Negatif Stroke	Negatif Stroke	Benar
54	Positif Stroke	Negatif Stroke	Salah
55	Negatif Stroke	Negatif Stroke	Benar
56	Negatif Stroke	Negatif Stroke	Benar
57	Positif Stroke	Negatif Stroke	Salah
58	Negatif Stroke	Negatif Stroke	Benar
59	Negatif Stroke	Negatif Stroke	Benar
60	Negatif Stroke	Negatif Stroke	Benar
61	Negatif Stroke	Negatif Stroke	Benar
62	Negatif Stroke	Negatif Stroke	Benar
63	Negatif Stroke	Negatif Stroke	Benar
64	Negatif Stroke	Negatif Stroke	Benar
65	Negatif Stroke	Negatif Stroke	Benar
66	Negatif Stroke	Negatif Stroke	Benar
67	Negatif Stroke	Negatif Stroke	Benar
68	Negatif Stroke	Negatif Stroke	Benar
69	Negatif Stroke	Negatif Stroke	Benar
70	Negatif Stroke	Negatif Stroke	Benar
71	Negatif Stroke	Negatif Stroke	Benar
72	Negatif Stroke	Negatif Stroke	Benar
73	Positif Stroke	Negatif Stroke	Salah
74	Negatif Stroke	Negatif Stroke	Benar
75	Negatif Stroke	Negatif Stroke	Benar
76	Negatif Stroke	Negatif Stroke	Benar
77	Negatif Stroke	Negatif Stroke	Benar
78	Negatif Stroke	Negatif Stroke	Benar
79	Negatif Stroke	Negatif Stroke	Benar
80	Negatif Stroke	Negatif Stroke	Benar
81	Negatif Stroke	Negatif Stroke	Benar
82	Negatif Stroke	Negatif Stroke	Benar
83	Negatif Stroke	Negatif Stroke	Benar
84	Negatif Stroke	Negatif Stroke	Benar
85	Negatif Stroke	Negatif Stroke	Benar
86	Negatif Stroke	Negatif Stroke	Benar
87	Negatif Stroke	Negatif Stroke	Benar

88	Negatif Stroke	Negatif Stroke	Benar
89	Negatif Stroke	Negatif Stroke	Benar
90	Negatif Stroke	Negatif Stroke	Benar
91	Positif Stroke	Negatif Stroke	Salah
92	Negatif Stroke	Negatif Stroke	Benar
93	Negatif Stroke	Negatif Stroke	Benar
94	Negatif Stroke	Negatif Stroke	Benar
95	Negatif Stroke	Negatif Stroke	Benar
96	Positif Stroke	Negatif Stroke	Salah
97	Negatif Stroke	Negatif Stroke	Benar
98	Negatif Stroke	Negatif Stroke	Benar
99	Negatif Stroke	Negatif Stroke	Benar
100	Negatif Stroke	Negatif Stroke	Benar
101	Negatif Stroke	Negatif Stroke	Benar
102	Negatif Stroke	Negatif Stroke	Benar
103	Negatif Stroke	Negatif Stroke	Benar
104	Negatif Stroke	Negatif Stroke	Benar
105	Negatif Stroke	Negatif Stroke	Benar
106	Negatif Stroke	Negatif Stroke	Benar
107	Negatif Stroke	Negatif Stroke	Benar
108	Negatif Stroke	Negatif Stroke	Benar
109	Negatif Stroke	Negatif Stroke	Benar
110	Negatif Stroke	Negatif Stroke	Benar
111	Negatif Stroke	Negatif Stroke	Benar
112	Negatif Stroke	Negatif Stroke	Benar
113	Negatif Stroke	Negatif Stroke	Benar
114	Negatif Stroke	Negatif Stroke	Benar
115	Negatif Stroke	Negatif Stroke	Benar
116	Negatif Stroke	Negatif Stroke	Benar
117	Negatif Stroke	Negatif Stroke	Benar
118	Negatif Stroke	Negatif Stroke	Benar
119	Negatif Stroke	Negatif Stroke	Benar
120	Negatif Stroke	Negatif Stroke	Benar
121	Positif Stroke	Negatif Stroke	Salah
122	Negatif Stroke	Negatif Stroke	Benar
123	Negatif Stroke	Negatif Stroke	Benar
124	Negatif Stroke	Negatif Stroke	Benar
125	Negatif Stroke	Negatif Stroke	Benar
126	Negatif Stroke	Negatif Stroke	Benar

127	Negatif Stroke	Negatif Stroke	Benar
128	Negatif Stroke	Negatif Stroke	Benar
129	Negatif Stroke	Negatif Stroke	Benar
130	Negatif Stroke	Negatif Stroke	Benar
131	Negatif Stroke	Negatif Stroke	Benar
132	Negatif Stroke	Negatif Stroke	Benar
133	Negatif Stroke	Negatif Stroke	Benar
134	Negatif Stroke	Negatif Stroke	Benar
135	Negatif Stroke	Negatif Stroke	Benar
136	Positif Stroke	Negatif Stroke	Salah
137	Negatif Stroke	Negatif Stroke	Benar
138	Negatif Stroke	Negatif Stroke	Benar
139	Negatif Stroke	Negatif Stroke	Benar
140	Negatif Stroke	Negatif Stroke	Benar
141	Negatif Stroke	Negatif Stroke	Benar
142	Negatif Stroke	Negatif Stroke	Benar
143	Positif Stroke	Negatif Stroke	Salah
144	Negatif Stroke	Negatif Stroke	Benar
145	Negatif Stroke	Negatif Stroke	Benar
146	Positif Stroke	Negatif Stroke	Salah
147	Negatif Stroke	Negatif Stroke	Benar
148	Negatif Stroke	Negatif Stroke	Benar
149	Negatif Stroke	Negatif Stroke	Benar
150	Negatif Stroke	Negatif Stroke	Benar
151	Positif Stroke	Negatif Stroke	Salah
152	Negatif Stroke	Negatif Stroke	Benar
153	Negatif Stroke	Negatif Stroke	Benar
154	Negatif Stroke	Negatif Stroke	Benar
155	Negatif Stroke	Negatif Stroke	Benar
156	Negatif Stroke	Negatif Stroke	Benar
157	Negatif Stroke	Negatif Stroke	Benar
158	Negatif Stroke	Negatif Stroke	Benar
159	Negatif Stroke	Negatif Stroke	Benar
160	Negatif Stroke	Negatif Stroke	Benar
161	Negatif Stroke	Negatif Stroke	Benar
162	Negatif Stroke	Negatif Stroke	Benar
163	Negatif Stroke	Negatif Stroke	Benar
164	Negatif Stroke	Negatif Stroke	Benar
165	Negatif Stroke	Negatif Stroke	Benar

166	Negatif Stroke	Negatif Stroke	Benar
167	Negatif Stroke	Negatif Stroke	Benar
168	Negatif Stroke	Negatif Stroke	Benar
169	Negatif Stroke	Negatif Stroke	Benar
170	Positif Stroke	Negatif Stroke	Salah
171	Negatif Stroke	Negatif Stroke	Benar
172	Negatif Stroke	Negatif Stroke	Benar
173	Negatif Stroke	Negatif Stroke	Benar
174	Negatif Stroke	Negatif Stroke	Benar
175	Negatif Stroke	Negatif Stroke	Benar
176	Negatif Stroke	Negatif Stroke	Benar
177	Negatif Stroke	Negatif Stroke	Benar
178	Negatif Stroke	Negatif Stroke	Benar
179	Negatif Stroke	Negatif Stroke	Benar
180	Negatif Stroke	Negatif Stroke	Benar
181	Negatif Stroke	Negatif Stroke	Benar
182	Negatif Stroke	Negatif Stroke	Benar
183	Negatif Stroke	Negatif Stroke	Benar
184	Negatif Stroke	Negatif Stroke	Benar
185	Negatif Stroke	Negatif Stroke	Benar
186	Negatif Stroke	Negatif Stroke	Benar
187	Negatif Stroke	Negatif Stroke	Benar
188	Negatif Stroke	Negatif Stroke	Benar
189	Negatif Stroke	Negatif Stroke	Benar
190	Negatif Stroke	Negatif Stroke	Benar
191	Negatif Stroke	Negatif Stroke	Benar
192	Negatif Stroke	Negatif Stroke	Benar
193	Negatif Stroke	Negatif Stroke	Benar
194	Positif Stroke	Negatif Stroke	Salah
195	Negatif Stroke	Negatif Stroke	Benar
196	Negatif Stroke	Negatif Stroke	Benar
197	Positif Stroke	Negatif Stroke	Salah
198	Negatif Stroke	Negatif Stroke	Benar
199	Negatif Stroke	Negatif Stroke	Benar
200	Negatif Stroke	Negatif Stroke	Benar
201	Negatif Stroke	Negatif Stroke	Benar
202	Negatif Stroke	Negatif Stroke	Benar
203	Negatif Stroke	Negatif Stroke	Benar
204	Negatif Stroke	Negatif Stroke	Benar

205	Negatif Stroke	Negatif Stroke	Benar
206	Negatif Stroke	Negatif Stroke	Benar
207	Negatif Stroke	Negatif Stroke	Benar
208	Negatif Stroke	Negatif Stroke	Benar
209	Negatif Stroke	Negatif Stroke	Benar
210	Negatif Stroke	Negatif Stroke	Benar
211	Negatif Stroke	Negatif Stroke	Benar
212	Negatif Stroke	Negatif Stroke	Benar
213	Negatif Stroke	Negatif Stroke	Benar
214	Negatif Stroke	Negatif Stroke	Benar
215	Negatif Stroke	Negatif Stroke	Benar
216	Negatif Stroke	Negatif Stroke	Benar
217	Negatif Stroke	Negatif Stroke	Benar
218	Negatif Stroke	Negatif Stroke	Benar
219	Negatif Stroke	Negatif Stroke	Benar
220	Negatif Stroke	Negatif Stroke	Benar
221	Negatif Stroke	Negatif Stroke	Benar
222	Negatif Stroke	Negatif Stroke	Benar
223	Negatif Stroke	Negatif Stroke	Benar
224	Negatif Stroke	Negatif Stroke	Benar
225	Negatif Stroke	Negatif Stroke	Benar
226	Negatif Stroke	Negatif Stroke	Benar
227	Positif Stroke	Negatif Stroke	Salah
228	Negatif Stroke	Negatif Stroke	Benar
229	Negatif Stroke	Negatif Stroke	Benar
230	Negatif Stroke	Negatif Stroke	Benar
231	Negatif Stroke	Negatif Stroke	Benar
232	Negatif Stroke	Negatif Stroke	Benar
233	Negatif Stroke	Negatif Stroke	Benar
234	Negatif Stroke	Negatif Stroke	Benar
235	Negatif Stroke	Negatif Stroke	Benar
236	Negatif Stroke	Negatif Stroke	Benar
237	Negatif Stroke	Negatif Stroke	Benar
238	Negatif Stroke	Negatif Stroke	Benar
239	Negatif Stroke	Negatif Stroke	Benar
240	Negatif Stroke	Negatif Stroke	Benar
241	Negatif Stroke	Negatif Stroke	Benar
242	Negatif Stroke	Negatif Stroke	Benar
243	Negatif Stroke	Negatif Stroke	Benar

244	Negatif Stroke	Negatif Stroke	Benar
245	Negatif Stroke	Negatif Stroke	Benar
246	Negatif Stroke	Negatif Stroke	Benar
247	Negatif Stroke	Negatif Stroke	Benar
248	Negatif Stroke	Negatif Stroke	Benar
249	Negatif Stroke	Negatif Stroke	Benar
250	Negatif Stroke	Negatif Stroke	Benar
251	Negatif Stroke	Negatif Stroke	Benar
252	Negatif Stroke	Negatif Stroke	Benar
253	Negatif Stroke	Negatif Stroke	Benar
254	Negatif Stroke	Negatif Stroke	Benar
255	Negatif Stroke	Negatif Stroke	Benar
256	Negatif Stroke	Negatif Stroke	Benar
257	Negatif Stroke	Negatif Stroke	Benar
258	Negatif Stroke	Negatif Stroke	Benar
259	Negatif Stroke	Negatif Stroke	Benar
260	Negatif Stroke	Negatif Stroke	Benar
261	Negatif Stroke	Negatif Stroke	Benar
262	Negatif Stroke	Negatif Stroke	Benar
263	Negatif Stroke	Negatif Stroke	Benar
264	Negatif Stroke	Negatif Stroke	Benar
265	Negatif Stroke	Negatif Stroke	Benar
266	Negatif Stroke	Negatif Stroke	Benar
267	Negatif Stroke	Negatif Stroke	Benar
268	Negatif Stroke	Negatif Stroke	Benar
269	Negatif Stroke	Negatif Stroke	Benar
270	Negatif Stroke	Negatif Stroke	Benar
271	Negatif Stroke	Negatif Stroke	Benar
272	Negatif Stroke	Negatif Stroke	Benar
273	Negatif Stroke	Negatif Stroke	Benar
274	Negatif Stroke	Negatif Stroke	Benar
275	Negatif Stroke	Negatif Stroke	Benar
276	Positif Stroke	Negatif Stroke	Salah
277	Negatif Stroke	Negatif Stroke	Benar
278	Positif Stroke	Negatif Stroke	Salah
279	Negatif Stroke	Negatif Stroke	Benar
280	Negatif Stroke	Negatif Stroke	Benar
281	Negatif Stroke	Negatif Stroke	Benar
282	Negatif Stroke	Negatif Stroke	Benar

283	Negatif Stroke	Negatif Stroke	Benar
284	Negatif Stroke	Negatif Stroke	Benar
285	Positif Stroke	Negatif Stroke	Salah
286	Negatif Stroke	Negatif Stroke	Benar
287	Negatif Stroke	Negatif Stroke	Benar
288	Negatif Stroke	Negatif Stroke	Benar
289	Negatif Stroke	Negatif Stroke	Benar
290	Negatif Stroke	Negatif Stroke	Benar
291	Negatif Stroke	Negatif Stroke	Benar
292	Negatif Stroke	Negatif Stroke	Benar
293	Negatif Stroke	Negatif Stroke	Benar
294	Negatif Stroke	Negatif Stroke	Benar
295	Negatif Stroke	Negatif Stroke	Benar
296	Negatif Stroke	Negatif Stroke	Benar
297	Negatif Stroke	Negatif Stroke	Benar
298	Negatif Stroke	Negatif Stroke	Benar
299	Negatif Stroke	Negatif Stroke	Benar
300	Negatif Stroke	Negatif Stroke	Benar
301	Negatif Stroke	Negatif Stroke	Benar
302	Positif Stroke	Negatif Stroke	Salah
303	Negatif Stroke	Negatif Stroke	Benar
304	Positif Stroke	Negatif Stroke	Salah
305	Negatif Stroke	Negatif Stroke	Benar
306	Negatif Stroke	Negatif Stroke	Benar
307	Negatif Stroke	Negatif Stroke	Benar
308	Negatif Stroke	Negatif Stroke	Benar
309	Negatif Stroke	Negatif Stroke	Benar
310	Positif Stroke	Negatif Stroke	Salah
311	Negatif Stroke	Negatif Stroke	Benar
312	Negatif Stroke	Negatif Stroke	Benar
313	Negatif Stroke	Negatif Stroke	Benar
314	Negatif Stroke	Negatif Stroke	Benar
315	Negatif Stroke	Negatif Stroke	Benar
316	Negatif Stroke	Negatif Stroke	Benar
317	Negatif Stroke	Negatif Stroke	Benar
318	Negatif Stroke	Negatif Stroke	Benar
319	Negatif Stroke	Negatif Stroke	Benar
320	Negatif Stroke	Negatif Stroke	Benar
321	Negatif Stroke	Negatif Stroke	Benar

322	Negatif Stroke	Negatif Stroke	Benar
323	Negatif Stroke	Negatif Stroke	Benar
324	Negatif Stroke	Negatif Stroke	Benar
325	Negatif Stroke	Negatif Stroke	Benar
326	Negatif Stroke	Negatif Stroke	Benar
327	Negatif Stroke	Negatif Stroke	Benar
328	Negatif Stroke	Negatif Stroke	Benar
329	Negatif Stroke	Negatif Stroke	Benar
330	Negatif Stroke	Negatif Stroke	Benar
331	Negatif Stroke	Negatif Stroke	Benar
332	Negatif Stroke	Negatif Stroke	Benar
333	Negatif Stroke	Negatif Stroke	Benar
334	Negatif Stroke	Negatif Stroke	Benar
335	Negatif Stroke	Negatif Stroke	Benar
336	Negatif Stroke	Negatif Stroke	Benar
337	Negatif Stroke	Negatif Stroke	Benar
338	Negatif Stroke	Negatif Stroke	Benar
339	Negatif Stroke	Negatif Stroke	Benar
340	Negatif Stroke	Negatif Stroke	Benar
341	Negatif Stroke	Negatif Stroke	Benar
342	Negatif Stroke	Negatif Stroke	Benar
343	Negatif Stroke	Negatif Stroke	Benar
344	Negatif Stroke	Negatif Stroke	Benar
345	Negatif Stroke	Negatif Stroke	Benar
346	Negatif Stroke	Negatif Stroke	Benar
347	Negatif Stroke	Negatif Stroke	Benar
348	Negatif Stroke	Negatif Stroke	Benar
349	Negatif Stroke	Negatif Stroke	Benar
350	Negatif Stroke	Negatif Stroke	Benar
351	Negatif Stroke	Negatif Stroke	Benar
352	Negatif Stroke	Negatif Stroke	Benar
353	Negatif Stroke	Negatif Stroke	Benar
354	Negatif Stroke	Negatif Stroke	Benar
355	Negatif Stroke	Negatif Stroke	Benar
356	Negatif Stroke	Negatif Stroke	Benar
357	Negatif Stroke	Negatif Stroke	Benar
358	Negatif Stroke	Negatif Stroke	Benar
359	Negatif Stroke	Negatif Stroke	Benar
360	Negatif Stroke	Negatif Stroke	Benar

361	Negatif Stroke	Negatif Stroke	Benar
362	Negatif Stroke	Negatif Stroke	Benar
363	Negatif Stroke	Negatif Stroke	Benar
364	Negatif Stroke	Negatif Stroke	Benar
365	Positif Stroke	Negatif Stroke	Salah
366	Negatif Stroke	Negatif Stroke	Benar
367	Negatif Stroke	Negatif Stroke	Benar
368	Positif Stroke	Negatif Stroke	Salah
369	Negatif Stroke	Negatif Stroke	Benar
370	Negatif Stroke	Negatif Stroke	Benar
371	Negatif Stroke	Negatif Stroke	Benar
372	Negatif Stroke	Negatif Stroke	Benar
373	Negatif Stroke	Negatif Stroke	Benar
374	Negatif Stroke	Negatif Stroke	Benar
375	Negatif Stroke	Negatif Stroke	Benar
376	Negatif Stroke	Negatif Stroke	Benar
377	Negatif Stroke	Negatif Stroke	Benar
378	Negatif Stroke	Negatif Stroke	Benar
379	Negatif Stroke	Negatif Stroke	Benar
380	Negatif Stroke	Negatif Stroke	Benar
381	Negatif Stroke	Negatif Stroke	Benar
382	Negatif Stroke	Negatif Stroke	Benar
383	Negatif Stroke	Negatif Stroke	Benar
384	Negatif Stroke	Negatif Stroke	Benar
385	Negatif Stroke	Negatif Stroke	Benar
386	Negatif Stroke	Negatif Stroke	Benar
387	Negatif Stroke	Negatif Stroke	Benar
388	Negatif Stroke	Negatif Stroke	Benar
389	Negatif Stroke	Negatif Stroke	Benar
390	Negatif Stroke	Negatif Stroke	Benar
391	Negatif Stroke	Negatif Stroke	Benar
392	Negatif Stroke	Negatif Stroke	Benar
393	Negatif Stroke	Negatif Stroke	Benar
394	Negatif Stroke	Negatif Stroke	Benar
395	Negatif Stroke	Negatif Stroke	Benar
396	Negatif Stroke	Negatif Stroke	Benar
397	Negatif Stroke	Negatif Stroke	Benar
398	Negatif Stroke	Negatif Stroke	Benar
399	Negatif Stroke	Negatif Stroke	Benar

400	Negatif Stroke	Negatif Stroke	Benar
401	Negatif Stroke	Negatif Stroke	Benar
402	Negatif Stroke	Negatif Stroke	Benar
403	Negatif Stroke	Negatif Stroke	Benar
404	Negatif Stroke	Negatif Stroke	Benar
405	Negatif Stroke	Negatif Stroke	Benar
406	Negatif Stroke	Negatif Stroke	Benar
407	Negatif Stroke	Negatif Stroke	Benar
408	Negatif Stroke	Negatif Stroke	Benar
409	Negatif Stroke	Negatif Stroke	Benar
410	Negatif Stroke	Negatif Stroke	Benar
411	Positif Stroke	Negatif Stroke	Salah
412	Negatif Stroke	Negatif Stroke	Benar
413	Negatif Stroke	Negatif Stroke	Benar
414	Negatif Stroke	Negatif Stroke	Benar
415	Negatif Stroke	Negatif Stroke	Benar
416	Negatif Stroke	Negatif Stroke	Benar
417	Negatif Stroke	Negatif Stroke	Benar
418	Negatif Stroke	Negatif Stroke	Benar
419	Negatif Stroke	Negatif Stroke	Benar
420	Negatif Stroke	Negatif Stroke	Benar
421	Negatif Stroke	Negatif Stroke	Benar
422	Negatif Stroke	Negatif Stroke	Benar
423	Negatif Stroke	Negatif Stroke	Benar
424	Positif Stroke	Negatif Stroke	Salah
425	Negatif Stroke	Negatif Stroke	Benar
426	Negatif Stroke	Negatif Stroke	Benar
427	Negatif Stroke	Negatif Stroke	Benar
428	Negatif Stroke	Negatif Stroke	Benar
429	Negatif Stroke	Negatif Stroke	Benar
430	Negatif Stroke	Negatif Stroke	Benar
431	Negatif Stroke	Negatif Stroke	Benar
432	Negatif Stroke	Negatif Stroke	Benar
433	Negatif Stroke	Negatif Stroke	Benar
434	Negatif Stroke	Negatif Stroke	Benar
435	Negatif Stroke	Negatif Stroke	Benar
436	Negatif Stroke	Negatif Stroke	Benar
437	Positif Stroke	Negatif Stroke	Salah
438	Negatif Stroke	Negatif Stroke	Benar

439	Negatif Stroke	Negatif Stroke	Benar
440	Negatif Stroke	Negatif Stroke	Benar
441	Negatif Stroke	Negatif Stroke	Benar
442	Negatif Stroke	Negatif Stroke	Benar
443	Negatif Stroke	Negatif Stroke	Benar
444	Negatif Stroke	Negatif Stroke	Benar
445	Negatif Stroke	Negatif Stroke	Benar
446	Negatif Stroke	Negatif Stroke	Benar
447	Negatif Stroke	Negatif Stroke	Benar
448	Negatif Stroke	Negatif Stroke	Benar
449	Positif Stroke	Negatif Stroke	Salah
450	Negatif Stroke	Negatif Stroke	Benar
451	Negatif Stroke	Negatif Stroke	Benar
452	Negatif Stroke	Negatif Stroke	Benar
453	Negatif Stroke	Negatif Stroke	Benar
454	Negatif Stroke	Negatif Stroke	Benar
455	Negatif Stroke	Negatif Stroke	Benar
456	Negatif Stroke	Negatif Stroke	Benar
457	Negatif Stroke	Negatif Stroke	Benar
458	Negatif Stroke	Negatif Stroke	Benar
459	Negatif Stroke	Negatif Stroke	Benar
460	Negatif Stroke	Negatif Stroke	Benar
461	Negatif Stroke	Negatif Stroke	Benar
462	Negatif Stroke	Negatif Stroke	Benar
463	Negatif Stroke	Negatif Stroke	Benar
464	Negatif Stroke	Negatif Stroke	Benar
465	Negatif Stroke	Negatif Stroke	Benar
466	Negatif Stroke	Negatif Stroke	Benar
467	Negatif Stroke	Negatif Stroke	Benar
468	Negatif Stroke	Negatif Stroke	Benar
469	Negatif Stroke	Negatif Stroke	Benar
470	Negatif Stroke	Negatif Stroke	Benar
471	Negatif Stroke	Negatif Stroke	Benar
472	Negatif Stroke	Negatif Stroke	Benar
473	Negatif Stroke	Negatif Stroke	Benar
474	Negatif Stroke	Negatif Stroke	Benar
475	Negatif Stroke	Negatif Stroke	Benar
476	Negatif Stroke	Negatif Stroke	Benar
477	Negatif Stroke	Negatif Stroke	Benar

478	Negatif Stroke	Negatif Stroke	Benar
479	Positif Stroke	Negatif Stroke	Salah
480	Negatif Stroke	Negatif Stroke	Benar
481	Negatif Stroke	Negatif Stroke	Benar
482	Negatif Stroke	Negatif Stroke	Benar
483	Negatif Stroke	Negatif Stroke	Benar
484	Negatif Stroke	Negatif Stroke	Benar
485	Negatif Stroke	Negatif Stroke	Benar
486	Negatif Stroke	Negatif Stroke	Benar
487	Negatif Stroke	Negatif Stroke	Benar
488	Negatif Stroke	Negatif Stroke	Benar
489	Negatif Stroke	Negatif Stroke	Benar
490	Negatif Stroke	Negatif Stroke	Benar
491	Negatif Stroke	Negatif Stroke	Benar
492	Negatif Stroke	Negatif Stroke	Benar
493	Negatif Stroke	Negatif Stroke	Benar
494	Negatif Stroke	Negatif Stroke	Benar
495	Negatif Stroke	Negatif Stroke	Benar
496	Negatif Stroke	Negatif Stroke	Benar
497	Negatif Stroke	Negatif Stroke	Benar
498	Negatif Stroke	Negatif Stroke	Benar
499	Negatif Stroke	Negatif Stroke	Benar
500	Negatif Stroke	Negatif Stroke	Benar
501	Negatif Stroke	Negatif Stroke	Benar
502	Negatif Stroke	Negatif Stroke	Benar
503	Negatif Stroke	Negatif Stroke	Benar
504	Negatif Stroke	Negatif Stroke	Benar
505	Negatif Stroke	Negatif Stroke	Benar
506	Negatif Stroke	Negatif Stroke	Benar
507	Negatif Stroke	Negatif Stroke	Benar
508	Negatif Stroke	Negatif Stroke	Benar
509	Negatif Stroke	Negatif Stroke	Benar
510	Negatif Stroke	Negatif Stroke	Benar
511	Negatif Stroke	Negatif Stroke	Benar
512	Negatif Stroke	Negatif Stroke	Benar
513	Negatif Stroke	Negatif Stroke	Benar
514	Negatif Stroke	Negatif Stroke	Benar
515	Negatif Stroke	Negatif Stroke	Benar
516	Negatif Stroke	Negatif Stroke	Benar

517	Negatif Stroke	Negatif Stroke	Benar
518	Negatif Stroke	Negatif Stroke	Benar
519	Negatif Stroke	Negatif Stroke	Benar
520	Negatif Stroke	Negatif Stroke	Benar
521	Negatif Stroke	Negatif Stroke	Benar
522	Negatif Stroke	Negatif Stroke	Benar
523	Negatif Stroke	Negatif Stroke	Benar
524	Negatif Stroke	Negatif Stroke	Benar
525	Negatif Stroke	Negatif Stroke	Benar
526	Negatif Stroke	Negatif Stroke	Benar
527	Negatif Stroke	Negatif Stroke	Benar
528	Negatif Stroke	Negatif Stroke	Benar
529	Negatif Stroke	Negatif Stroke	Benar
530	Negatif Stroke	Negatif Stroke	Benar
531	Negatif Stroke	Negatif Stroke	Benar
532	Negatif Stroke	Negatif Stroke	Benar
533	Negatif Stroke	Negatif Stroke	Benar
534	Negatif Stroke	Negatif Stroke	Benar
535	Negatif Stroke	Negatif Stroke	Benar
536	Negatif Stroke	Negatif Stroke	Benar
537	Negatif Stroke	Negatif Stroke	Benar
538	Negatif Stroke	Negatif Stroke	Benar
539	Negatif Stroke	Negatif Stroke	Benar
540	Negatif Stroke	Negatif Stroke	Benar
541	Negatif Stroke	Negatif Stroke	Benar
542	Negatif Stroke	Negatif Stroke	Benar
543	Negatif Stroke	Negatif Stroke	Benar
544	Negatif Stroke	Negatif Stroke	Benar
545	Negatif Stroke	Negatif Stroke	Benar
546	Negatif Stroke	Negatif Stroke	Benar
547	Negatif Stroke	Negatif Stroke	Benar
548	Negatif Stroke	Negatif Stroke	Benar
549	Negatif Stroke	Negatif Stroke	Benar
550	Negatif Stroke	Negatif Stroke	Benar
551	Negatif Stroke	Negatif Stroke	Benar
552	Negatif Stroke	Negatif Stroke	Benar
553	Negatif Stroke	Negatif Stroke	Benar
554	Negatif Stroke	Negatif Stroke	Benar
555	Negatif Stroke	Negatif Stroke	Benar

556	Negatif Stroke	Negatif Stroke	Benar
557	Negatif Stroke	Negatif Stroke	Benar
558	Negatif Stroke	Negatif Stroke	Benar
559	Negatif Stroke	Negatif Stroke	Benar
560	Negatif Stroke	Negatif Stroke	Benar
561	Negatif Stroke	Negatif Stroke	Benar
562	Negatif Stroke	Negatif Stroke	Benar
563	Negatif Stroke	Negatif Stroke	Benar
564	Negatif Stroke	Negatif Stroke	Benar
565	Negatif Stroke	Negatif Stroke	Benar
566	Negatif Stroke	Negatif Stroke	Benar
567	Negatif Stroke	Negatif Stroke	Benar
568	Negatif Stroke	Negatif Stroke	Benar
569	Negatif Stroke	Negatif Stroke	Benar
570	Negatif Stroke	Negatif Stroke	Benar
571	Negatif Stroke	Negatif Stroke	Benar
572	Negatif Stroke	Negatif Stroke	Benar
573	Negatif Stroke	Negatif Stroke	Benar
574	Negatif Stroke	Negatif Stroke	Benar
575	Negatif Stroke	Negatif Stroke	Benar
576	Negatif Stroke	Negatif Stroke	Benar
577	Negatif Stroke	Negatif Stroke	Benar
578	Negatif Stroke	Negatif Stroke	Benar
579	Negatif Stroke	Negatif Stroke	Benar
580	Negatif Stroke	Negatif Stroke	Benar
581	Negatif Stroke	Negatif Stroke	Benar
582	Negatif Stroke	Negatif Stroke	Benar
583	Negatif Stroke	Negatif Stroke	Benar
584	Negatif Stroke	Negatif Stroke	Benar
585	Negatif Stroke	Negatif Stroke	Benar
586	Negatif Stroke	Negatif Stroke	Benar
587	Negatif Stroke	Negatif Stroke	Benar
588	Negatif Stroke	Negatif Stroke	Benar
589	Negatif Stroke	Negatif Stroke	Benar
590	Negatif Stroke	Negatif Stroke	Benar
591	Negatif Stroke	Negatif Stroke	Benar
592	Negatif Stroke	Negatif Stroke	Benar
593	Negatif Stroke	Negatif Stroke	Benar
594	Negatif Stroke	Negatif Stroke	Benar

595	Negatif Stroke	Negatif Stroke	Benar
596	Negatif Stroke	Negatif Stroke	Benar
597	Negatif Stroke	Negatif Stroke	Benar
598	Negatif Stroke	Negatif Stroke	Benar
599	Negatif Stroke	Negatif Stroke	Benar
600	Negatif Stroke	Negatif Stroke	Benar
601	Negatif Stroke	Negatif Stroke	Benar
602	Negatif Stroke	Negatif Stroke	Benar
603	Negatif Stroke	Negatif Stroke	Benar
604	Positif Stroke	Negatif Stroke	Salah
605	Negatif Stroke	Negatif Stroke	Benar
606	Negatif Stroke	Negatif Stroke	Benar
607	Negatif Stroke	Negatif Stroke	Benar
608	Negatif Stroke	Negatif Stroke	Benar
609	Negatif Stroke	Negatif Stroke	Benar
610	Negatif Stroke	Negatif Stroke	Benar
611	Negatif Stroke	Negatif Stroke	Benar
612	Negatif Stroke	Negatif Stroke	Benar
613	Negatif Stroke	Negatif Stroke	Benar
614	Negatif Stroke	Negatif Stroke	Benar
615	Negatif Stroke	Negatif Stroke	Benar
616	Negatif Stroke	Negatif Stroke	Benar
617	Negatif Stroke	Negatif Stroke	Benar
618	Negatif Stroke	Negatif Stroke	Benar
619	Negatif Stroke	Negatif Stroke	Benar
620	Negatif Stroke	Negatif Stroke	Benar
621	Negatif Stroke	Negatif Stroke	Benar
622	Negatif Stroke	Negatif Stroke	Benar
623	Negatif Stroke	Negatif Stroke	Benar
624	Negatif Stroke	Negatif Stroke	Benar
625	Positif Stroke	Negatif Stroke	Salah
626	Negatif Stroke	Negatif Stroke	Benar
627	Negatif Stroke	Negatif Stroke	Benar
628	Negatif Stroke	Negatif Stroke	Benar
629	Negatif Stroke	Negatif Stroke	Benar
630	Negatif Stroke	Negatif Stroke	Benar
631	Negatif Stroke	Negatif Stroke	Benar
632	Negatif Stroke	Negatif Stroke	Benar
633	Negatif Stroke	Negatif Stroke	Benar

634	Negatif Stroke	Negatif Stroke	Benar
635	Negatif Stroke	Negatif Stroke	Benar
636	Negatif Stroke	Negatif Stroke	Benar
637	Negatif Stroke	Negatif Stroke	Benar
638	Negatif Stroke	Negatif Stroke	Benar
639	Negatif Stroke	Negatif Stroke	Benar
640	Positif Stroke	Negatif Stroke	Salah
641	Negatif Stroke	Negatif Stroke	Benar
642	Negatif Stroke	Negatif Stroke	Benar
643	Negatif Stroke	Negatif Stroke	Benar
644	Negatif Stroke	Negatif Stroke	Benar
645	Positif Stroke	Negatif Stroke	Salah
646	Negatif Stroke	Negatif Stroke	Benar
647	Negatif Stroke	Negatif Stroke	Benar
648	Negatif Stroke	Negatif Stroke	Benar
649	Negatif Stroke	Negatif Stroke	Benar
650	Negatif Stroke	Negatif Stroke	Benar
651	Negatif Stroke	Negatif Stroke	Benar
652	Negatif Stroke	Negatif Stroke	Benar
653	Negatif Stroke	Negatif Stroke	Benar
654	Negatif Stroke	Negatif Stroke	Benar
655	Negatif Stroke	Negatif Stroke	Benar
656	Negatif Stroke	Negatif Stroke	Benar
657	Negatif Stroke	Negatif Stroke	Benar
658	Negatif Stroke	Negatif Stroke	Benar
659	Positif Stroke	Negatif Stroke	Salah
660	Negatif Stroke	Negatif Stroke	Benar
661	Positif Stroke	Negatif Stroke	Salah
662	Negatif Stroke	Negatif Stroke	Benar
663	Negatif Stroke	Negatif Stroke	Benar
664	Negatif Stroke	Negatif Stroke	Benar
665	Negatif Stroke	Negatif Stroke	Benar
666	Negatif Stroke	Negatif Stroke	Benar
667	Negatif Stroke	Negatif Stroke	Benar
668	Negatif Stroke	Negatif Stroke	Benar
669	Negatif Stroke	Negatif Stroke	Benar
670	Negatif Stroke	Negatif Stroke	Benar
671	Negatif Stroke	Negatif Stroke	Benar
672	Negatif Stroke	Negatif Stroke	Benar

673	Negatif Stroke	Negatif Stroke	Benar
674	Negatif Stroke	Negatif Stroke	Benar
675	Negatif Stroke	Negatif Stroke	Benar
676	Negatif Stroke	Negatif Stroke	Benar
677	Negatif Stroke	Negatif Stroke	Benar
678	Negatif Stroke	Negatif Stroke	Benar
679	Negatif Stroke	Negatif Stroke	Benar
680	Positif Stroke	Negatif Stroke	Salah
681	Negatif Stroke	Negatif Stroke	Benar
682	Negatif Stroke	Negatif Stroke	Benar
683	Negatif Stroke	Negatif Stroke	Benar
684	Negatif Stroke	Negatif Stroke	Benar
685	Negatif Stroke	Negatif Stroke	Benar
686	Negatif Stroke	Negatif Stroke	Benar
687	Negatif Stroke	Negatif Stroke	Benar
688	Negatif Stroke	Negatif Stroke	Benar
689	Negatif Stroke	Negatif Stroke	Benar
690	Negatif Stroke	Negatif Stroke	Benar
691	Negatif Stroke	Negatif Stroke	Benar
692	Negatif Stroke	Negatif Stroke	Benar
693	Negatif Stroke	Negatif Stroke	Benar
694	Negatif Stroke	Negatif Stroke	Benar
695	Positif Stroke	Negatif Stroke	Salah
696	Negatif Stroke	Negatif Stroke	Benar
697	Negatif Stroke	Negatif Stroke	Benar
698	Negatif Stroke	Negatif Stroke	Benar
699	Negatif Stroke	Negatif Stroke	Benar
700	Negatif Stroke	Negatif Stroke	Benar
701	Negatif Stroke	Negatif Stroke	Benar
702	Negatif Stroke	Negatif Stroke	Benar
703	Negatif Stroke	Negatif Stroke	Benar
704	Negatif Stroke	Negatif Stroke	Benar
705	Negatif Stroke	Negatif Stroke	Benar
706	Negatif Stroke	Negatif Stroke	Benar
707	Negatif Stroke	Negatif Stroke	Benar
708	Negatif Stroke	Negatif Stroke	Benar
709	Negatif Stroke	Negatif Stroke	Benar
710	Negatif Stroke	Negatif Stroke	Benar
711	Negatif Stroke	Negatif Stroke	Benar

712	Negatif Stroke	Negatif Stroke	Benar
713	Negatif Stroke	Negatif Stroke	Benar
714	Negatif Stroke	Negatif Stroke	Benar
715	Negatif Stroke	Negatif Stroke	Benar
716	Negatif Stroke	Negatif Stroke	Benar
717	Negatif Stroke	Negatif Stroke	Benar
718	Negatif Stroke	Negatif Stroke	Benar
719	Negatif Stroke	Negatif Stroke	Benar
720	Negatif Stroke	Negatif Stroke	Benar
721	Negatif Stroke	Negatif Stroke	Benar
722	Negatif Stroke	Negatif Stroke	Benar
723	Positif Stroke	Negatif Stroke	Salah
724	Negatif Stroke	Negatif Stroke	Benar
725	Negatif Stroke	Negatif Stroke	Benar
726	Negatif Stroke	Negatif Stroke	Benar
727	Negatif Stroke	Negatif Stroke	Benar
728	Negatif Stroke	Negatif Stroke	Benar
729	Positif Stroke	Negatif Stroke	Salah
730	Negatif Stroke	Negatif Stroke	Benar
731	Negatif Stroke	Negatif Stroke	Benar
732	Negatif Stroke	Negatif Stroke	Benar
733	Negatif Stroke	Negatif Stroke	Benar
734	Negatif Stroke	Negatif Stroke	Benar
735	Negatif Stroke	Negatif Stroke	Benar
736	Negatif Stroke	Negatif Stroke	Benar
737	Negatif Stroke	Negatif Stroke	Benar
738	Negatif Stroke	Negatif Stroke	Benar
739	Negatif Stroke	Negatif Stroke	Benar
740	Negatif Stroke	Negatif Stroke	Benar
741	Negatif Stroke	Negatif Stroke	Benar
742	Negatif Stroke	Negatif Stroke	Benar
743	Negatif Stroke	Negatif Stroke	Benar
744	Negatif Stroke	Negatif Stroke	Benar
745	Negatif Stroke	Negatif Stroke	Benar
746	Negatif Stroke	Negatif Stroke	Benar
747	Negatif Stroke	Negatif Stroke	Benar
748	Negatif Stroke	Negatif Stroke	Benar
749	Negatif Stroke	Negatif Stroke	Benar
750	Negatif Stroke	Negatif Stroke	Benar

751	Negatif Stroke	Negatif Stroke	Benar
752	Negatif Stroke	Negatif Stroke	Benar
753	Negatif Stroke	Negatif Stroke	Benar
754	Negatif Stroke	Negatif Stroke	Benar
755	Negatif Stroke	Negatif Stroke	Benar
756	Negatif Stroke	Negatif Stroke	Benar
757	Negatif Stroke	Negatif Stroke	Benar
758	Negatif Stroke	Negatif Stroke	Benar
759	Negatif Stroke	Negatif Stroke	Benar
760	Negatif Stroke	Negatif Stroke	Benar
761	Positif Stroke	Negatif Stroke	Salah
762	Negatif Stroke	Negatif Stroke	Benar
763	Negatif Stroke	Negatif Stroke	Benar
764	Negatif Stroke	Negatif Stroke	Benar
765	Negatif Stroke	Negatif Stroke	Benar
766	Negatif Stroke	Negatif Stroke	Benar
767	Negatif Stroke	Negatif Stroke	Benar
768	Negatif Stroke	Negatif Stroke	Benar
769	Negatif Stroke	Negatif Stroke	Benar
770	Negatif Stroke	Negatif Stroke	Benar
771	Negatif Stroke	Negatif Stroke	Benar
772	Negatif Stroke	Negatif Stroke	Benar
773	Negatif Stroke	Negatif Stroke	Benar
774	Negatif Stroke	Negatif Stroke	Benar
775	Positif Stroke	Negatif Stroke	Salah
776	Positif Stroke	Negatif Stroke	Salah
777	Negatif Stroke	Negatif Stroke	Benar
778	Negatif Stroke	Negatif Stroke	Benar
779	Negatif Stroke	Negatif Stroke	Benar
780	Negatif Stroke	Negatif Stroke	Benar
781	Negatif Stroke	Negatif Stroke	Benar
782	Negatif Stroke	Negatif Stroke	Benar
783	Negatif Stroke	Negatif Stroke	Benar
784	Negatif Stroke	Negatif Stroke	Benar
785	Negatif Stroke	Negatif Stroke	Benar
786	Negatif Stroke	Negatif Stroke	Benar
787	Negatif Stroke	Negatif Stroke	Benar
788	Negatif Stroke	Negatif Stroke	Benar
789	Negatif Stroke	Negatif Stroke	Benar

790	Negatif Stroke	Negatif Stroke	Benar
791	Negatif Stroke	Negatif Stroke	Benar
792	Negatif Stroke	Negatif Stroke	Benar
793	Negatif Stroke	Negatif Stroke	Benar
794	Negatif Stroke	Negatif Stroke	Benar
795	Negatif Stroke	Negatif Stroke	Benar
796	Negatif Stroke	Negatif Stroke	Benar
797	Negatif Stroke	Negatif Stroke	Benar
798	Negatif Stroke	Negatif Stroke	Benar
799	Negatif Stroke	Negatif Stroke	Benar
800	Negatif Stroke	Negatif Stroke	Benar
801	Negatif Stroke	Negatif Stroke	Benar
802	Negatif Stroke	Negatif Stroke	Benar
803	Negatif Stroke	Negatif Stroke	Benar
804	Negatif Stroke	Negatif Stroke	Benar
805	Negatif Stroke	Negatif Stroke	Benar
806	Negatif Stroke	Negatif Stroke	Benar
807	Negatif Stroke	Negatif Stroke	Benar
808	Negatif Stroke	Negatif Stroke	Benar
809	Negatif Stroke	Negatif Stroke	Benar
810	Negatif Stroke	Negatif Stroke	Benar
811	Negatif Stroke	Negatif Stroke	Benar
812	Negatif Stroke	Negatif Stroke	Benar
813	Negatif Stroke	Negatif Stroke	Benar
814	Negatif Stroke	Negatif Stroke	Benar
815	Negatif Stroke	Negatif Stroke	Benar
816	Negatif Stroke	Negatif Stroke	Benar
817	Negatif Stroke	Negatif Stroke	Benar
818	Negatif Stroke	Negatif Stroke	Benar
819	Negatif Stroke	Negatif Stroke	Benar
820	Negatif Stroke	Negatif Stroke	Benar
821	Positif Stroke	Negatif Stroke	Salah
822	Negatif Stroke	Negatif Stroke	Benar
823	Negatif Stroke	Negatif Stroke	Benar
824	Negatif Stroke	Negatif Stroke	Benar
825	Negatif Stroke	Negatif Stroke	Benar
826	Negatif Stroke	Negatif Stroke	Benar
827	Negatif Stroke	Negatif Stroke	Benar
828	Negatif Stroke	Negatif Stroke	Benar

829	Negatif Stroke	Negatif Stroke	Benar
830	Negatif Stroke	Negatif Stroke	Benar
831	Negatif Stroke	Negatif Stroke	Benar
832	Negatif Stroke	Negatif Stroke	Benar
833	Negatif Stroke	Negatif Stroke	Benar
834	Positif Stroke	Negatif Stroke	Salah
835	Negatif Stroke	Negatif Stroke	Benar
836	Negatif Stroke	Negatif Stroke	Benar
837	Negatif Stroke	Negatif Stroke	Benar
838	Negatif Stroke	Negatif Stroke	Benar
839	Negatif Stroke	Negatif Stroke	Benar
840	Negatif Stroke	Negatif Stroke	Benar
841	Negatif Stroke	Negatif Stroke	Benar
842	Negatif Stroke	Negatif Stroke	Benar
843	Positif Stroke	Negatif Stroke	Salah
844	Negatif Stroke	Negatif Stroke	Benar
845	Negatif Stroke	Negatif Stroke	Benar
846	Negatif Stroke	Negatif Stroke	Benar
847	Negatif Stroke	Negatif Stroke	Benar
848	Negatif Stroke	Negatif Stroke	Benar
849	Negatif Stroke	Negatif Stroke	Benar
850	Negatif Stroke	Negatif Stroke	Benar
851	Negatif Stroke	Negatif Stroke	Benar
852	Negatif Stroke	Negatif Stroke	Benar
853	Negatif Stroke	Negatif Stroke	Benar
854	Negatif Stroke	Negatif Stroke	Benar
855	Negatif Stroke	Negatif Stroke	Benar
856	Negatif Stroke	Negatif Stroke	Benar
857	Negatif Stroke	Negatif Stroke	Benar
858	Negatif Stroke	Negatif Stroke	Benar
859	Negatif Stroke	Negatif Stroke	Benar
860	Positif Stroke	Negatif Stroke	Salah
861	Negatif Stroke	Negatif Stroke	Benar
862	Negatif Stroke	Negatif Stroke	Benar
863	Negatif Stroke	Negatif Stroke	Benar
864	Positif Stroke	Negatif Stroke	Salah
865	Negatif Stroke	Negatif Stroke	Benar
866	Negatif Stroke	Negatif Stroke	Benar
867	Negatif Stroke	Negatif Stroke	Benar

868	Negatif Stroke	Negatif Stroke	Benar
869	Negatif Stroke	Negatif Stroke	Benar
870	Negatif Stroke	Negatif Stroke	Benar
871	Negatif Stroke	Negatif Stroke	Benar
872	Negatif Stroke	Negatif Stroke	Benar
873	Negatif Stroke	Negatif Stroke	Benar
874	Negatif Stroke	Negatif Stroke	Benar
875	Negatif Stroke	Negatif Stroke	Benar
876	Negatif Stroke	Negatif Stroke	Benar
877	Negatif Stroke	Negatif Stroke	Benar
878	Negatif Stroke	Negatif Stroke	Benar
879	Negatif Stroke	Negatif Stroke	Benar
880	Negatif Stroke	Negatif Stroke	Benar
881	Negatif Stroke	Negatif Stroke	Benar
882	Negatif Stroke	Negatif Stroke	Benar
883	Negatif Stroke	Negatif Stroke	Benar
884	Negatif Stroke	Negatif Stroke	Benar
885	Negatif Stroke	Negatif Stroke	Benar
886	Negatif Stroke	Negatif Stroke	Benar
887	Negatif Stroke	Negatif Stroke	Benar
888	Negatif Stroke	Negatif Stroke	Benar
889	Negatif Stroke	Negatif Stroke	Benar
890	Negatif Stroke	Negatif Stroke	Benar
891	Negatif Stroke	Negatif Stroke	Benar
892	Negatif Stroke	Negatif Stroke	Benar
893	Negatif Stroke	Negatif Stroke	Benar
894	Negatif Stroke	Negatif Stroke	Benar
895	Negatif Stroke	Negatif Stroke	Benar
896	Negatif Stroke	Negatif Stroke	Benar
897	Negatif Stroke	Negatif Stroke	Benar
898	Negatif Stroke	Negatif Stroke	Benar
899	Negatif Stroke	Negatif Stroke	Benar
900	Negatif Stroke	Negatif Stroke	Benar
901	Negatif Stroke	Negatif Stroke	Benar
902	Negatif Stroke	Negatif Stroke	Benar
903	Negatif Stroke	Negatif Stroke	Benar
904	Negatif Stroke	Negatif Stroke	Benar
905	Negatif Stroke	Negatif Stroke	Benar
906	Negatif Stroke	Negatif Stroke	Benar

907	Negatif Stroke	Negatif Stroke	Benar
908	Negatif Stroke	Negatif Stroke	Benar
909	Negatif Stroke	Negatif Stroke	Benar
910	Negatif Stroke	Negatif Stroke	Benar
911	Negatif Stroke	Negatif Stroke	Benar
912	Negatif Stroke	Negatif Stroke	Benar
913	Negatif Stroke	Negatif Stroke	Benar
914	Negatif Stroke	Negatif Stroke	Benar
915	Negatif Stroke	Negatif Stroke	Benar
916	Negatif Stroke	Negatif Stroke	Benar
917	Negatif Stroke	Negatif Stroke	Benar
918	Negatif Stroke	Negatif Stroke	Benar
919	Negatif Stroke	Negatif Stroke	Benar
920	Negatif Stroke	Negatif Stroke	Benar
921	Negatif Stroke	Negatif Stroke	Benar
922	Negatif Stroke	Negatif Stroke	Benar
923	Negatif Stroke	Negatif Stroke	Benar
924	Negatif Stroke	Negatif Stroke	Benar
925	Negatif Stroke	Negatif Stroke	Benar
926	Negatif Stroke	Negatif Stroke	Benar
927	Negatif Stroke	Negatif Stroke	Benar
928	Negatif Stroke	Negatif Stroke	Benar
929	Negatif Stroke	Negatif Stroke	Benar
930	Negatif Stroke	Negatif Stroke	Benar
931	Negatif Stroke	Negatif Stroke	Benar
932	Negatif Stroke	Negatif Stroke	Benar
933	Negatif Stroke	Negatif Stroke	Benar
934	Negatif Stroke	Negatif Stroke	Benar
935	Negatif Stroke	Negatif Stroke	Benar
936	Negatif Stroke	Negatif Stroke	Benar
937	Negatif Stroke	Negatif Stroke	Benar
938	Negatif Stroke	Negatif Stroke	Benar
939	Negatif Stroke	Negatif Stroke	Benar
940	Negatif Stroke	Negatif Stroke	Benar
941	Negatif Stroke	Negatif Stroke	Benar
942	Negatif Stroke	Negatif Stroke	Benar
943	Negatif Stroke	Negatif Stroke	Benar
944	Negatif Stroke	Negatif Stroke	Benar
945	Negatif Stroke	Negatif Stroke	Benar

946	Negatif Stroke	Negatif Stroke	Benar
947	Negatif Stroke	Negatif Stroke	Benar
948	Negatif Stroke	Negatif Stroke	Benar
949	Negatif Stroke	Negatif Stroke	Benar
950	Negatif Stroke	Negatif Stroke	Benar
951	Negatif Stroke	Negatif Stroke	Benar
952	Negatif Stroke	Negatif Stroke	Benar
953	Negatif Stroke	Negatif Stroke	Benar
954	Negatif Stroke	Negatif Stroke	Benar
955	Negatif Stroke	Negatif Stroke	Benar
956	Positif Stroke	Negatif Stroke	Salah
957	Negatif Stroke	Negatif Stroke	Benar
958	Negatif Stroke	Negatif Stroke	Benar
959	Negatif Stroke	Negatif Stroke	Benar
960	Negatif Stroke	Negatif Stroke	Benar
961	Negatif Stroke	Negatif Stroke	Benar
962	Negatif Stroke	Negatif Stroke	Benar
963	Negatif Stroke	Negatif Stroke	Benar
964	Negatif Stroke	Negatif Stroke	Benar
965	Negatif Stroke	Negatif Stroke	Benar
966	Negatif Stroke	Negatif Stroke	Benar
967	Negatif Stroke	Negatif Stroke	Benar
968	Negatif Stroke	Negatif Stroke	Benar
969	Negatif Stroke	Negatif Stroke	Benar
970	Negatif Stroke	Negatif Stroke	Benar
971	Negatif Stroke	Negatif Stroke	Benar
972	Negatif Stroke	Negatif Stroke	Benar
973	Negatif Stroke	Negatif Stroke	Benar
974	Negatif Stroke	Negatif Stroke	Benar
975	Positif Stroke	Negatif Stroke	Salah
976	Negatif Stroke	Negatif Stroke	Benar
977	Negatif Stroke	Negatif Stroke	Benar
978	Negatif Stroke	Negatif Stroke	Benar
979	Negatif Stroke	Negatif Stroke	Benar
980	Negatif Stroke	Negatif Stroke	Benar
981	Negatif Stroke	Negatif Stroke	Benar
982	Positif Stroke	Negatif Stroke	Salah
983	Negatif Stroke	Negatif Stroke	Benar
984	Negatif Stroke	Negatif Stroke	Benar

985	Negatif Stroke	Negatif Stroke	Benar
986	Negatif Stroke	Negatif Stroke	Benar
987	Negatif Stroke	Negatif Stroke	Benar
988	Negatif Stroke	Negatif Stroke	Benar
989	Negatif Stroke	Negatif Stroke	Benar
990	Negatif Stroke	Negatif Stroke	Benar
991	Negatif Stroke	Negatif Stroke	Benar
992	Negatif Stroke	Negatif Stroke	Benar
993	Positif Stroke	Negatif Stroke	Salah
994	Negatif Stroke	Negatif Stroke	Benar
995	Negatif Stroke	Negatif Stroke	Benar
996	Negatif Stroke	Negatif Stroke	Benar
997	Negatif Stroke	Negatif Stroke	Benar
998	Negatif Stroke	Negatif Stroke	Benar
999	Negatif Stroke	Negatif Stroke	Benar
1000	Negatif Stroke	Negatif Stroke	Benar
1001	Negatif Stroke	Negatif Stroke	Benar
1002	Negatif Stroke	Negatif Stroke	Benar
1003	Negatif Stroke	Negatif Stroke	Benar
1004	Negatif Stroke	Negatif Stroke	Benar
1005	Negatif Stroke	Negatif Stroke	Benar
1006	Negatif Stroke	Negatif Stroke	Benar
1007	Negatif Stroke	Negatif Stroke	Benar
1008	Negatif Stroke	Negatif Stroke	Benar
1009	Positif Stroke	Negatif Stroke	Salah
1010	Negatif Stroke	Negatif Stroke	Benar
1011	Negatif Stroke	Negatif Stroke	Benar
1012	Negatif Stroke	Negatif Stroke	Benar
1013	Negatif Stroke	Negatif Stroke	Benar
1014	Negatif Stroke	Negatif Stroke	Benar
1015	Negatif Stroke	Negatif Stroke	Benar
1016	Negatif Stroke	Negatif Stroke	Benar
1017	Negatif Stroke	Negatif Stroke	Benar
1018	Negatif Stroke	Negatif Stroke	Benar
1019	Negatif Stroke	Negatif Stroke	Benar
1020	Negatif Stroke	Negatif Stroke	Benar
1021	Positif Stroke	Negatif Stroke	Salah
1022	Negatif Stroke	Negatif Stroke	Benar

Dari hasil pada tabel 4.1 di atas, menunjukkan bahwa hasil prediksi oleh model dengan *dataset* orisinil memiliki tingkat akurasi yang tinggi. Berdasarkan tabel tersebut dapat dihitung nilai akurasi dan nilai *loss* dengan persamaan di bawah ini.

$$\%akurasi = \frac{Jumlah\ prediksi\ benar}{Total\ jumlah\ prediksi} \times 100\%$$

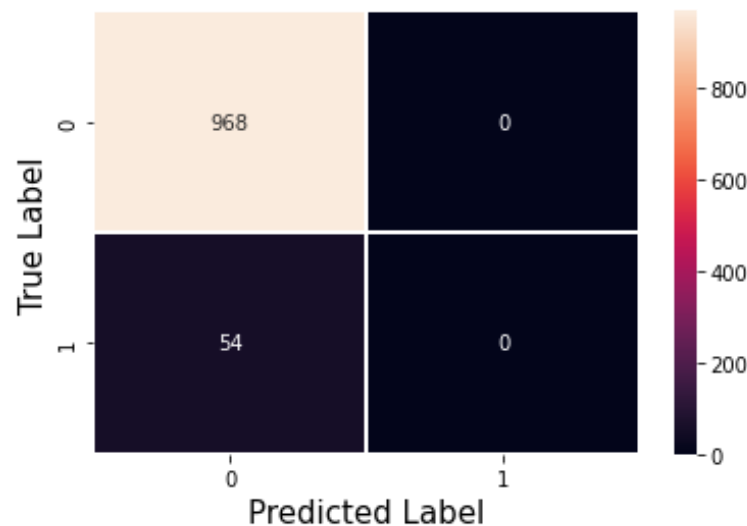
$$\%akurasi = \frac{968}{1022} \times 100\% = 94.72\%$$

$$\%loss = \frac{Jumlah\ prediksi\ benar}{Total\ jumlah\ prediksi} \times 100\%$$

$$\%loss = \frac{54}{1022} \times 100\% = 0.053\%$$

Berdasarkan perhitungan tersebut, terlihat bahwa nilai akurasi yang didapati oleh model dengan *dataset* orisinil sebesar 94.72% dan nilai *loss* sebesar 0.053%. Sehingga tergambar bahwa model ini memiliki kemampuan klasifikasi yang baik.

Selanjutnya untuk mengetahui tingkat kebenaran dan kesalahan hasil prediksi pada tahap pengujian yang dilakukan adalah dengan menggunakan *confusion matrix*. Dalam *confusion matrix* terdapat label *True label* dan *Predicted label*, serta terdapat label angka “0” dan “1” yang masing-masing merepresentasikan “negatif stroke” dan “positif stroke”. Dapat diperhatikan pada gambar 4.8 yang merupakan *confusion matrix* hasil dari pengujian pada model dengan *dataset* orisinil, menunjukkan bahwa model selalu menghasilkan prediksi negatif stroke pada semua data yang dimasukkan meskipun terdapat data yang seharusnya menunjukkan bahwa seseorang positif stroke.



Gambar 4.8 Confusion Matrix pada model dengan dataset orisinil

Meskipun hasil proses *train* mendapati nilai akurasi yang tinggi serta nilai *loss* yang rendah sehingga menunjukkan bahwa model memiliki kemampuan yang baik dalam klasifikasi, namun *confusion matrix* menunjukkan hal yang sebaliknya. Hal ini disebabkan karena *dataset* orisinil yang digunakan untuk melatih model mengalami *imbalanced class* serta selisih jumlah data yang sangat besar, model yang dihasilkan akan mengalami *overfitting*, yang mana *overfitting* pada kasus *imbalanced* ini dapat terjadi karena kurva logistik yang dibentuk akan cenderung untuk mengarah pada salah satu kelas saja [10], yakni cenderung mengarah ke arah *majority class*.

Terjadinya *overfitting* pada model ini juga didukung dengan didapatnya nilai *f1 score*, *precision score*, dan *recall score* yang dapat dilihat pada tabel 4.2 di bawah ini.

Tabel 4.2 Nilai evaluasi model dengan dataset orisinal

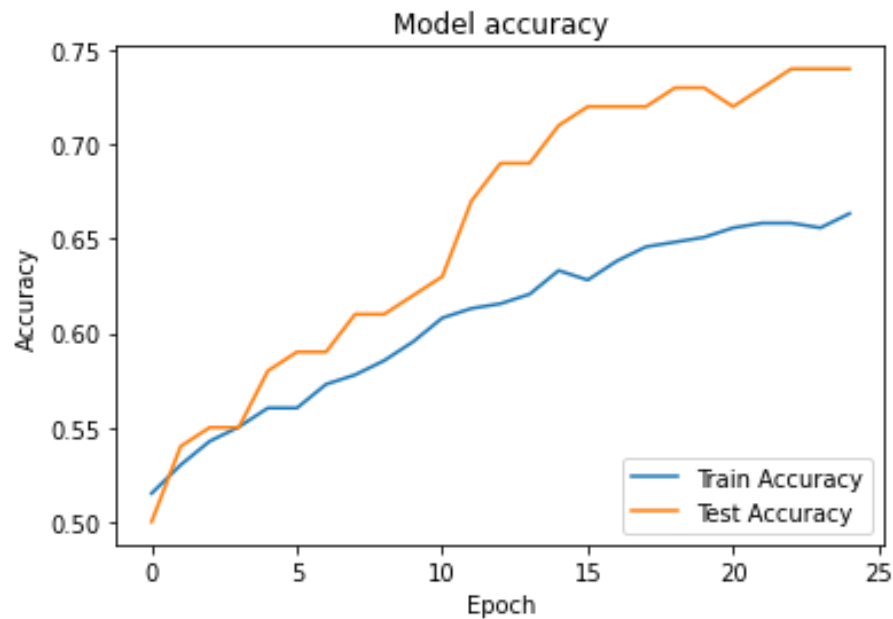
<i>F1 Score</i>	0.00
<i>Precision Score</i>	0.00
<i>Recall Score</i>	0.00

4.2.2 Dataset dengan Near-Miss Undersampling

Proses *train* model dan pengujian menggunakan *dataset* yang sudah dilakukan *near-miss undersampling*, dengan 10 buah data sebagai input. Berbeda dari *dataset* sebelumnya yang mengalami *imbalanced dataset*, pada *dataset* yang sudah melalui *near-miss undersampling* perbandingan antara kelas seimbang. *Dataset* terbagi ke dalam dua bagian yakni data *train* dan data *test* yang masing-masing berjumlah 398 dan 100 data dengan komposisi 80% data *train* dan 20% data *test*. Serta pada tahap pengujian *dataset* yang digunakan adalah data *test*.

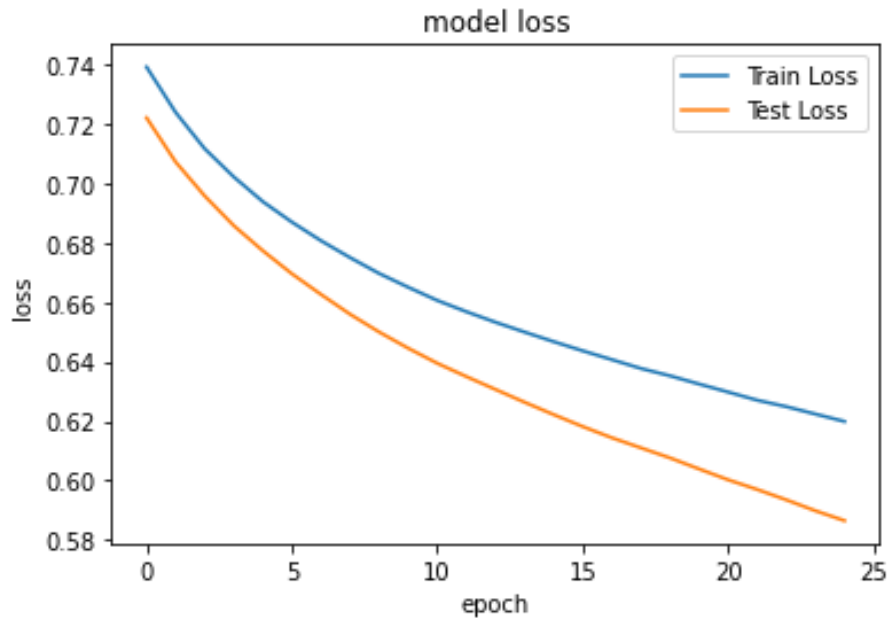
Menggunakan model yang sama seperti pada percobaan sebelumnya, dengan arsitektur *deep neural network keras* dengan *input shape* sebanyak 10. Enam belas buah node pada *dense layer* 1 dengan aktivasi ReLU, enam belas buah node pada *dense layer* 2 dengan aktivasi ReLU, serta satu buah node output dengan aktivasi sigmoid. Hasil dari proses *train* dan *test* terhadap model dari setiap *dataset* yang melalui *near-miss undersampling* tipe 1, 2, dan 3 dibahas pada subbab berikut.

Near-Miss Undersampling versi 1



Gambar 4.9 Grafik kurva akurasi training dan test pada dataset near-miss 1

Pada gambar 4.9 di atas, menampilkan dua buah grafik berwarna jingga dan biru yang menggambarkan nilai akurasi seiring bertambahnya *epoch*. Grafik berwarna biru merupakan grafik yang menunjukkan nilai akurasi dari hasil proses *train* dan grafik berwarna jingga menunjukkan nilai akurasi dari hasil proses *test*. Dapat dipahami bahwa kurva akurasi *train* berakhir pada nilai akurasi sebesar 66.33% pada *epoch* ke-25, serta pada grafik akurasi *test* nilai akurasi pada *epoch* ke-25 berada pada nilai 74%. Dari hasil proses *train* dan *test* yang ditampilkan oleh kurva pada grafik di atas, bahwa model yang dihasilkan memiliki nilai akurasi yang cukup tinggi, namun nilai akurasi yang didapati tidak setinggi dengan nilai akurasi pada proses *train* model ketika menggunakan *dataset* orisinil.



Gambar 4.10 Grafik kurva loss training dan test pada dataset near-miss 1

Dapat dilihat pada gambar 4.10 di atas terdapat grafik berwarna biru merupakan grafik yang menunjukkan nilai *train loss* dan grafik berwarna jingga menunjukkan nilai *test loss*. Nilai *train loss* dan *test loss* pada akhir *epoch* masing-masing sebesar 0.6199 dan 0.5865. Dengan hasil yang ditampilkan oleh kurva pada grafik di atas, bahwa model yang dihasilkan memiliki nilai *loss* yang cukup rendah, meskipun nilai *loss* yang didapati tersebut lebih besar jika dibandingkan dengan nilai *loss* pada model ketika menggunakan *dataset* orisinal.

Tabel 4.3 Data Uji Menggunakan Model dengan Dataset Hasil Near-Miss 1

Uji ke-	Data Uji	Hasil Prediksi	Benar/Salah
1	Negatif Stroke	Positif Stroke	Salah
2	Positif Stroke	Positif Stroke	Benar
3	Positif Stroke	Negatif Stroke	Salah
4	Positif Stroke	Positif Stroke	Benar
5	Positif Stroke	Positif Stroke	Benar
6	Negatif Stroke	Positif Stroke	Salah
7	Positif Stroke	Positif Stroke	Benar

8	Positif Stroke	Negatif Stroke	Salah
9	Negatif Stroke	Negatif Stroke	Benar
10	Negatif Stroke	Negatif Stroke	Benar
11	Negatif Stroke	Positif Stroke	Salah
12	Positif Stroke	Positif Stroke	Benar
13	Positif Stroke	Negatif Stroke	Salah
14	Positif Stroke	Positif Stroke	Benar
15	Negatif Stroke	Negatif Stroke	Benar
16	Positif Stroke	Positif Stroke	Benar
17	Positif Stroke	Positif Stroke	Benar
18	Positif Stroke	Negatif Stroke	Salah
19	Negatif Stroke	Positif Stroke	Salah
20	Negatif Stroke	Negatif Stroke	Benar
21	Negatif Stroke	Negatif Stroke	Benar
22	Positif Stroke	Negatif Stroke	Salah
23	Negatif Stroke	Negatif Stroke	Benar
24	Negatif Stroke	Negatif Stroke	Benar
25	Negatif Stroke	Negatif Stroke	Benar
26	Positif Stroke	Positif Stroke	Benar
27	Negatif Stroke	Negatif Stroke	Benar
28	Positif Stroke	Negatif Stroke	Salah
29	Negatif Stroke	Negatif Stroke	Benar
30	Negatif Stroke	Negatif Stroke	Benar
31	Positif Stroke	Positif Stroke	Benar
32	Negatif Stroke	Negatif Stroke	Benar
33	Negatif Stroke	Negatif Stroke	Benar
34	Negatif Stroke	Negatif Stroke	Benar
35	Positif Stroke	Positif Stroke	Benar
36	Positif Stroke	Positif Stroke	Benar
37	Positif Stroke	Positif Stroke	Benar
38	Positif Stroke	Positif Stroke	Benar
39	Positif Stroke	Positif Stroke	Benar
40	Negatif Stroke	Positif Stroke	Salah
41	Positif Stroke	Positif Stroke	Benar
42	Negatif Stroke	Positif Stroke	Salah
43	Positif Stroke	Negatif Stroke	Salah
44	Negatif Stroke	Negatif Stroke	Benar
45	Negatif Stroke	Negatif Stroke	Benar
46	Negatif Stroke	Positif Stroke	Salah

47	Negatif Stroke	Positif Stroke	Salah
48	Positif Stroke	Positif Stroke	Benar
49	Positif Stroke	Negatif Stroke	Salah
50	Negatif Stroke	Positif Stroke	Salah
51	Positif Stroke	Positif Stroke	Benar
52	Negatif Stroke	Negatif Stroke	Benar
53	Positif Stroke	Positif Stroke	Benar
54	Negatif Stroke	Negatif Stroke	Benar
55	Negatif Stroke	Negatif Stroke	Benar
56	Negatif Stroke	Positif Stroke	Salah
57	Positif Stroke	Positif Stroke	Benar
58	Negatif Stroke	Negatif Stroke	Benar
59	Negatif Stroke	Negatif Stroke	Benar
60	Positif Stroke	Positif Stroke	Benar
61	Negatif Stroke	Negatif Stroke	Benar
62	Positif Stroke	Positif Stroke	Benar
63	Positif Stroke	Positif Stroke	Benar
64	Positif Stroke	Positif Stroke	Benar
65	Positif Stroke	Positif Stroke	Benar
66	Negatif Stroke	Negatif Stroke	Benar
67	Negatif Stroke	Negatif Stroke	Benar
68	Positif Stroke	Positif Stroke	Benar
69	Positif Stroke	Positif Stroke	Benar
70	Negatif Stroke	Negatif Stroke	Benar
71	Negatif Stroke	Negatif Stroke	Benar
72	Negatif Stroke	Negatif Stroke	Benar
73	Positif Stroke	Negatif Stroke	Salah
74	Positif Stroke	Positif Stroke	Benar
75	Negatif Stroke	Positif Stroke	Salah
76	Positif Stroke	Positif Stroke	Benar
77	Positif Stroke	Positif Stroke	Benar
78	Negatif Stroke	Negatif Stroke	Benar
79	Positif Stroke	Positif Stroke	Benar
80	Positif Stroke	Negatif Stroke	Salah
81	Positif Stroke	Negatif Stroke	Salah
82	Negatif Stroke	Positif Stroke	Salah
83	Positif Stroke	Positif Stroke	Benar
84	Negatif Stroke	Negatif Stroke	Benar
85	Positif Stroke	Negatif Stroke	Salah

86	Positif Stroke	Positif Stroke	Benar
87	Positif Stroke	Positif Stroke	Benar
88	Negatif Stroke	Positif Stroke	Salah
89	Positif Stroke	Positif Stroke	Benar
90	Negatif Stroke	Negatif Stroke	Benar
91	Negatif Stroke	Negatif Stroke	Benar
92	Positif Stroke	Positif Stroke	Benar
93	Positif Stroke	Positif Stroke	Benar
94	Negatif Stroke	Negatif Stroke	Benar
95	Negatif Stroke	Negatif Stroke	Benar
96	Positif Stroke	Negatif Stroke	Salah
97	Negatif Stroke	Positif Stroke	Salah
98	Positif Stroke	Positif Stroke	Benar
99	Negatif Stroke	Negatif Stroke	Benar
100	Negatif Stroke	Positif Stroke	Salah

Dari hasil pada tabel 4.3 di atas, menunjukkan bahwa hasil prediksi oleh model dengan *dataset* orisinil memiliki tingkat akurasi yang cukup tinggi. Berdasarkan tabel tersebut dapat dihitung nilai akurasi dan nilai *loss* dengan persamaan di bawah ini.

$$\%akurasi = \frac{Jumlah\ prediksi\ benar}{Total\ jumlah\ prediksi} \times 100\%$$

$$\%akurasi = \frac{72}{100} \times 100\% = 72.00\%$$

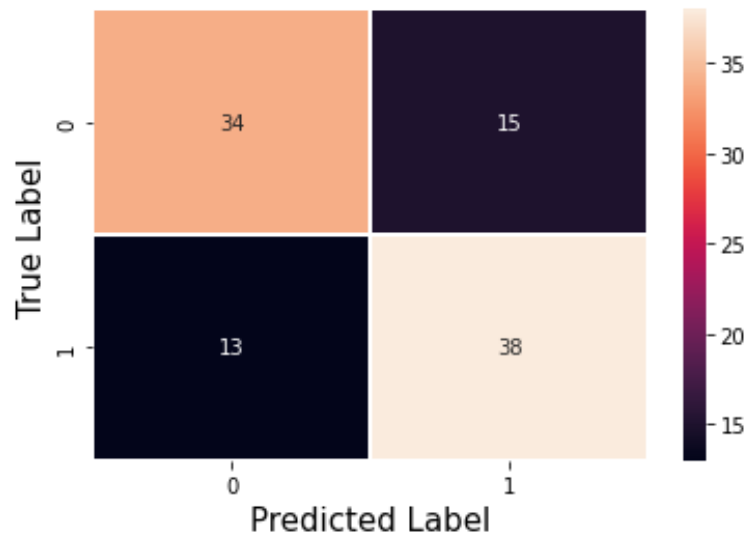
$$\%loss = \frac{Jumlah\ prediksi\ benar}{Total\ jumlah\ prediksi} \times 100\%$$

$$\%loss = \frac{28}{100} \times 100\% = 28.00\%$$

Berdasarkan perhitungan tersebut, terlihat bahwa nilai akurasi yang didapati oleh model dengan *dataset* hasil *near-miss* 1 sebesar 72.00% dan nilai *loss*

sebesar 28.00%. Sehingga dapat dikatakan bahwa model ini memiliki kemampuan klasifikasi yang cukup baik.

Kemudian terdapat *confusion matrix* yang digunakan untuk memastikan tingkat kebenaran dan kesalahan dari hasil prediksi pada tahap pengujian yang dilakukan. Dalam *confusion matrix* pada sumbu y diberi label *True label* lalu pada sumbu x diberi label *Predicted label*, serta terdapat label angka “0” dan “1” yang masing-masing mewakili kondisi “negatif stroke” dan “positif stroke”. Dapat diperhatikan pada gambar 4.11 yang merupakan *confusion matrix* hasil dari pengujian pada model dengan *dataset* yang dilakukan *near-miss undersampling*, menampilkan bahwa jumlah data yang diprediksi benar lebih tinggi dibandingkan dengan jumlah data yang salah diprediksi, hal ini ditunjukkan pada *confusion matrix*, di mana hasil prediksi yang menyatakan negatif stroke dan sesuai dengan data yang seharusnya mencapai angka 34, serta hasil prediksi yang menyatakan positif stroke dan sesuai dengan data yang seharusnya mencapai angka 38. Sehingga diperoleh nilai *True negative rate* sebesar 69.4% dan nilai *True positive rate* adalah sebesar 74.5%



Gambar 4.11 Confusion Matrix pada model dengan dataset near-miss 1

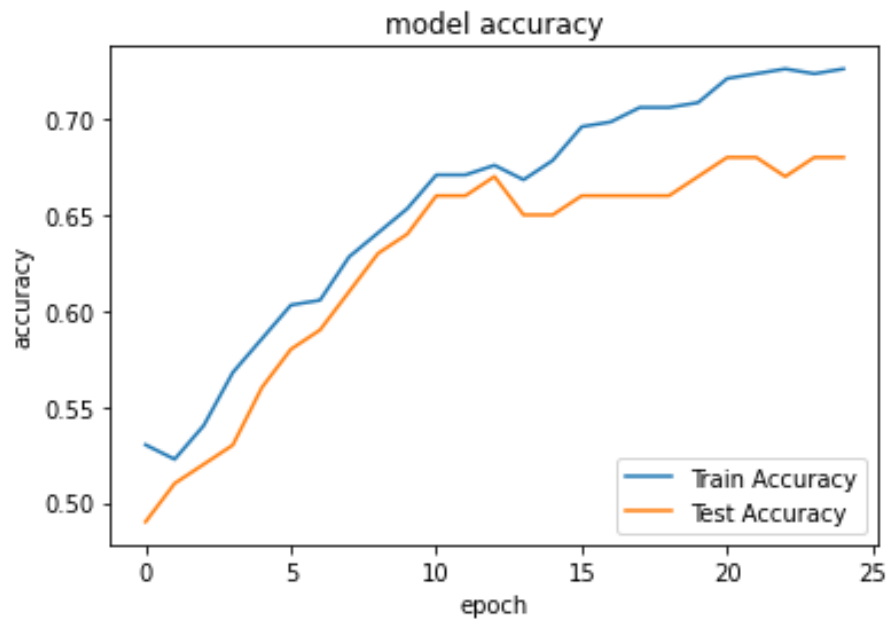
Meskipun dari hasil proses *train* diperoleh nilai akurasi yang lebih rendah dan nilai *loss* yang lebih tinggi, namun pada *confusion matrix* menunjukkan bahwa model yang dibentuk memiliki performa yang baik, serta tidak mengalami *overfitting* sebagaimana yang terjadi pada model sebelumnya ketika menggunakan *dataset* orisinal.

Performa dari model ini juga dapat dilihat dari berbagai parameter seperti nilai *f1 score*, *precision score*, dan *recall score* yang dapat dilihat pada tabel 4.4 di bawah ini.

Tabel 4.4 Nilai evaluasi model dengan dataset near-miss 1

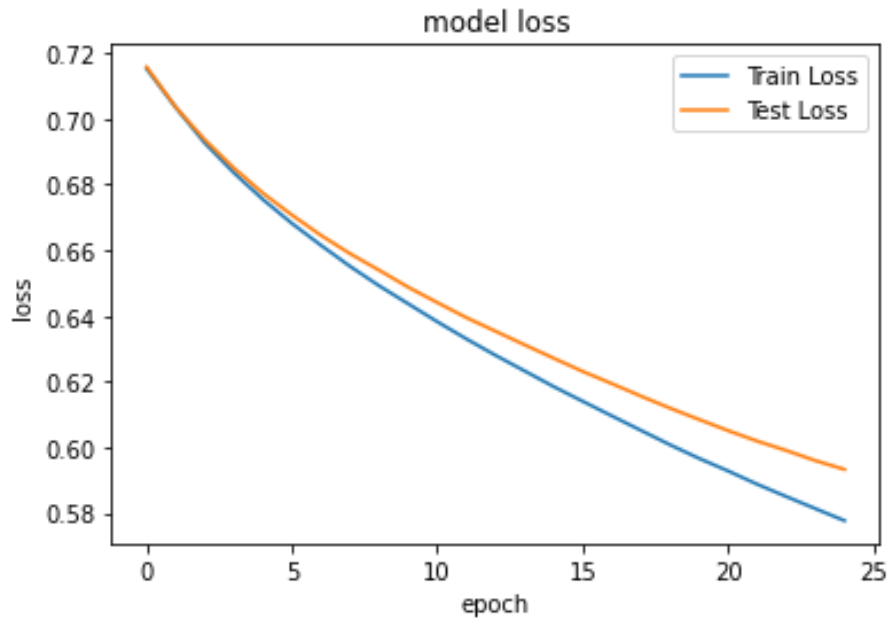
<i>F1 Score</i>	0.731
<i>Precision Score</i>	0.717
<i>Recall Score</i>	0.745

Near-Miss Undersampling versi 2



Gambar 4.12 Grafik kurva akurasi training dan test pada dataset near-miss 2

Gambar 4.12 di atas menampilkan grafik dua buah kurva berwarna jingga dan biru yang masing-masing merepresentasikan nilai akurasi *train* dan *test* seiring bertambahnya *epoch*. Dapat dilihat bahwa kurva akurasi *train* berakhir pada nilai akurasi sebesar 72.61% pada *epoch* ke-25, serta pada grafik akurasi *test* nilai akurasi pada *epoch* ke-25 berada pada nilai 68%. Berdasarkan hasil yang ditampilkan oleh kurva pada grafik di atas, bahwa model memiliki nilai akurasi yang cukup tinggi, namun nilai akurasi tersebut lebih rendah dibandingkan dengan nilai akurasi pada proses *train* model ketika menggunakan *dataset* orisinal.



Gambar 4.13 Grafik kurva loss training dan test pada dataset near-miss 2

Pada gambar 4.13 di atas menampilkan grafik dua buah kurva berwarna jingga dan biru yang masing-masing merepresentasikan nilai *train loss* dan *test loss* bertambahnya epoch. Nilai *train loss* dan *test loss* pada akhir *epoch* masing-masing sebesar 0.5776 dan 0.5932. Berdasarkan hasil yang ditampilkan oleh kurva pada grafik di atas, bahwa model yang dihasilkan memiliki nilai *loss* yang cukup rendah, meskipun nilai *loss* tersebut lebih besar daripada nilai *loss* pada model ketika menggunakan *dataset* orisinal.

Tabel 4.5 Data Uji Menggunakan Model dengan Dataset Hasil Near-Miss 2

Uji ke-	Data Uji	Hasil Prediksi	Benar/Salah
1	Negatif Stroke	Positif Stroke	Salah
2	Positif Stroke	Positif Stroke	Benar
3	Positif Stroke	Negatif Stroke	Salah
4	Positif Stroke	Negatif Stroke	Salah
5	Positif Stroke	Positif Stroke	Benar
6	Negatif Stroke	Negatif Stroke	Benar
7	Positif Stroke	Positif Stroke	Benar
8	Positif Stroke	Negatif Stroke	Salah

9	Negatif Stroke	Negatif Stroke	Benar
10	Negatif Stroke	Negatif Stroke	Benar
11	Negatif Stroke	Positif Stroke	Salah
12	Positif Stroke	Positif Stroke	Benar
13	Positif Stroke	Negatif Stroke	Salah
14	Positif Stroke	Negatif Stroke	Salah
15	Negatif Stroke	Negatif Stroke	Benar
16	Positif Stroke	Positif Stroke	Benar
17	Positif Stroke	Positif Stroke	Benar
18	Positif Stroke	Positif Stroke	Benar
19	Negatif Stroke	Negatif Stroke	Benar
20	Negatif Stroke	Negatif Stroke	Benar
21	Negatif Stroke	Negatif Stroke	Benar
22	Positif Stroke	Positif Stroke	Benar
23	Negatif Stroke	Negatif Stroke	Benar
24	Negatif Stroke	Positif Stroke	Salah
25	Negatif Stroke	Negatif Stroke	Benar
26	Positif Stroke	Positif Stroke	Benar
27	Negatif Stroke	Positif Stroke	Salah
28	Positif Stroke	Negatif Stroke	Salah
29	Negatif Stroke	Positif Stroke	Salah
30	Negatif Stroke	Positif Stroke	Salah
31	Positif Stroke	Positif Stroke	Benar
32	Negatif Stroke	Negatif Stroke	Benar
33	Negatif Stroke	Negatif Stroke	Benar
34	Negatif Stroke	Negatif Stroke	Benar
35	Positif Stroke	Positif Stroke	Benar
36	Positif Stroke	Positif Stroke	Benar
37	Positif Stroke	Negatif Stroke	Salah
38	Positif Stroke	Negatif Stroke	Salah
39	Positif Stroke	Positif Stroke	Benar
40	Negatif Stroke	Negatif Stroke	Benar
41	Positif Stroke	Positif Stroke	Benar
42	Negatif Stroke	Negatif Stroke	Benar
43	Positif Stroke	Positif Stroke	Benar
44	Negatif Stroke	Positif Stroke	Salah
45	Negatif Stroke	Negatif Stroke	Benar
46	Negatif Stroke	Negatif Stroke	Benar
47	Negatif Stroke	Negatif Stroke	Benar
48	Positif Stroke	Positif Stroke	Benar
49	Positif Stroke	Negatif Stroke	Salah
50	Negatif Stroke	Positif Stroke	Salah

51	Positif Stroke	Positif Stroke	Benar
52	Negatif Stroke	Negatif Stroke	Benar
53	Positif Stroke	Negatif Stroke	Salah
54	Negatif Stroke	Negatif Stroke	Benar
55	Negatif Stroke	Negatif Stroke	Benar
56	Negatif Stroke	Positif Stroke	Salah
57	Positif Stroke	Positif Stroke	Benar
58	Negatif Stroke	Positif Stroke	Salah
59	Negatif Stroke	Negatif Stroke	Benar
60	Positif Stroke	Positif Stroke	Benar
61	Negatif Stroke	Negatif Stroke	Benar
62	Positif Stroke	Positif Stroke	Benar
63	Positif Stroke	Positif Stroke	Benar
64	Positif Stroke	Negatif Stroke	Salah
65	Positif Stroke	Positif Stroke	Benar
66	Negatif Stroke	Negatif Stroke	Benar
67	Negatif Stroke	Positif Stroke	Salah
68	Positif Stroke	Negatif Stroke	Salah
69	Positif Stroke	Positif Stroke	Benar
70	Negatif Stroke	Negatif Stroke	Benar
71	Negatif Stroke	Negatif Stroke	Benar
72	Negatif Stroke	Positif Stroke	Salah
73	Positif Stroke	Negatif Stroke	Salah
74	Positif Stroke	Negatif Stroke	Salah
75	Negatif Stroke	Negatif Stroke	Benar
76	Positif Stroke	Positif Stroke	Benar
77	Positif Stroke	Negatif Stroke	Salah
78	Negatif Stroke	Negatif Stroke	Benar
79	Positif Stroke	Negatif Stroke	Salah
80	Positif Stroke	Positif Stroke	Benar
81	Positif Stroke	Positif Stroke	Benar
82	Negatif Stroke	Positif Stroke	Salah
83	Positif Stroke	Positif Stroke	Benar
84	Negatif Stroke	Negatif Stroke	Benar
85	Positif Stroke	Negatif Stroke	Salah
86	Positif Stroke	Positif Stroke	Benar
87	Positif Stroke	Negatif Stroke	Salah
88	Negatif Stroke	Positif Stroke	Salah
89	Positif Stroke	Positif Stroke	Benar
90	Negatif Stroke	Negatif Stroke	Benar
91	Negatif Stroke	Negatif Stroke	Benar
92	Positif Stroke	Positif Stroke	Benar

93	Positif Stroke	Positif Stroke	Benar
94	Negatif Stroke	Negatif Stroke	Benar
95	Negatif Stroke	Negatif Stroke	Benar
96	Positif Stroke	Positif Stroke	Benar
97	Negatif Stroke	Negatif Stroke	Benar
98	Positif Stroke	Positif Stroke	Benar
99	Negatif Stroke	Positif Stroke	Salah
100	Negatif Stroke	Positif Stroke	Salah

Dari hasil pada tabel 4.5 di atas, menunjukkan bahwa hasil prediksi oleh model dengan *dataset* orisinil memiliki tingkat akurasi yang cukup tinggi. Berdasarkan tabel tersebut dapat dihitung nilai akurasi dan nilai *loss* dengan persamaan di bawah ini.

$$\%akurasi = \frac{Jumlah\ prediksi\ benar}{Total\ jumlah\ prediksi} \times 100\%$$

$$\%akurasi = \frac{66}{100} \times 100\% = 66.00\%$$

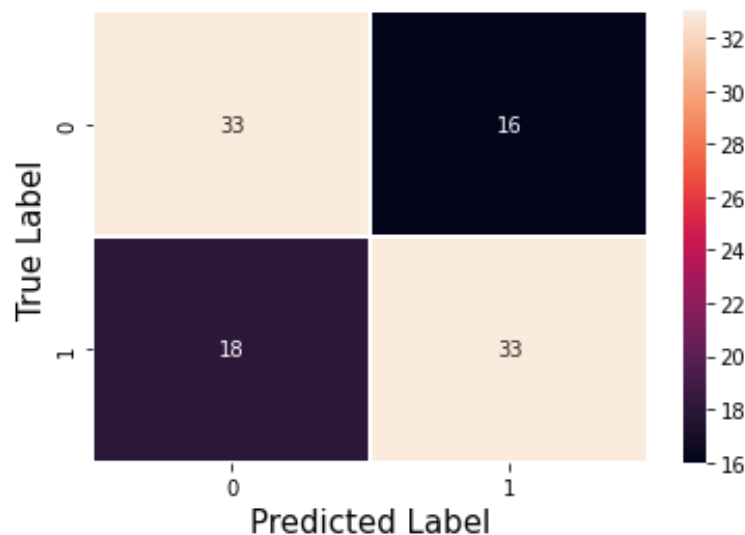
$$\%loss = \frac{Jumlah\ prediksi\ benar}{Total\ jumlah\ prediksi} \times 100\%$$

$$\%loss = \frac{34}{100} \times 100\% = 34.00\%$$

Berdasarkan perhitungan tersebut, terlihat bahwa nilai akurasi yang didapati oleh model dengan *dataset* hasil *near-miss* 2 sebesar 66.00% dan nilai *loss* sebesar 34.00%. Sehingga dapat dikatakan bahwa model ini memiliki kemampuan klasifikasi yang cukup baik.

Kemudian terdapat *confusion matrix* untuk memastikan tingkat kebenaran dan kesalahan dari hasil prediksi pada tahap pengujian yang dilakukan. Dalam *confusion matrix* terdapat sumbu y dan sumbu x yang masing-masing diberi label

True label dan *Predicted label*, serta terdapat label angka “0” dan “1” yang masing-masing merepresentasikan kondisi “negatif stroke” dan “positif stroke”. Dapat dilihat pada gambar 4.14 di bawah merupakan *confusion matrix* hasil dari pengujian pada model dengan *dataset* yang dilakukan *near-miss undersampling* versi 2, menampilkan jumlah data yang diprediksi benar lebih tinggi dibandingkan dengan jumlah data yang salah diprediksi, di mana hasil prediksi yang menyatakan negatif stroke dan sesuai dengan data yang seharusnya mencapai angka 33, begitu pula dengan hasil prediksi yang menyatakan positif stroke. Sehingga diperoleh nilai *True negative rate* sebesar 67.3% dan nilai *True positive rate* adalah sebesar 64.7%



Gambar 4.14 Confusion Matrix pada model dengan dataset near-miss 2

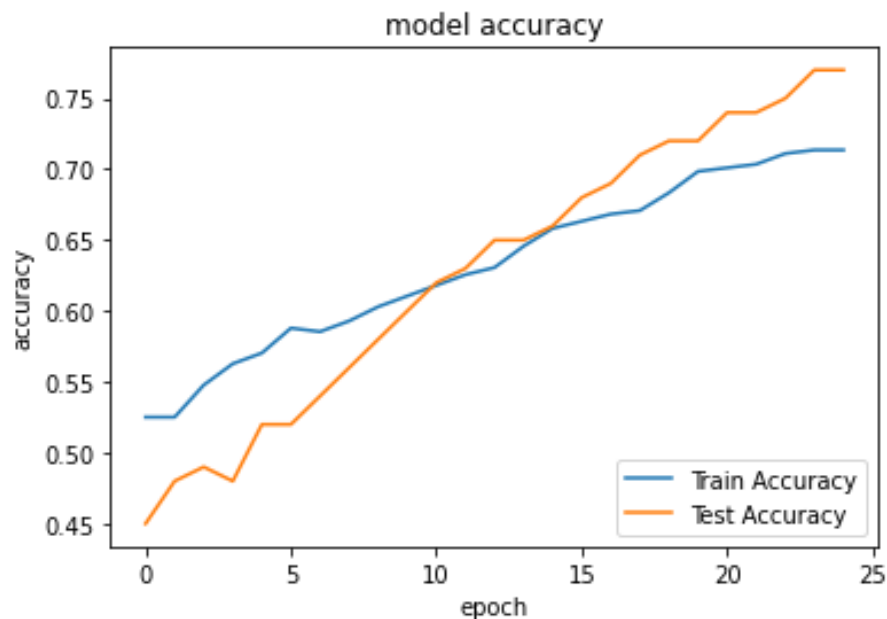
Meskipun dari hasil proses *train* diperoleh nilai akurasi yang lebih rendah dan nilai *loss* yang lebih tinggi, namun pada *confusion matrix* menunjukkan bahwa model yang dibentuk memiliki performa yang baik, serta tidak mengalami *overfitting* sebagaimana yang terjadi pada model sebelumnya ketika menggunakan *dataset* orisinal.

Adapun performa dari model dapat dilihat dari berbagai parameter seperti nilai *f1 score*, *precision score*, dan *recall score* yang dapat dilihat pada tabel 4.6 di bawah ini.

Tabel 4.6 Nilai evaluasi model dengan dataset near-miss 2

<i>F1 Score</i>	0.660
<i>Precision Score</i>	0.673
<i>Recall Score</i>	0.647

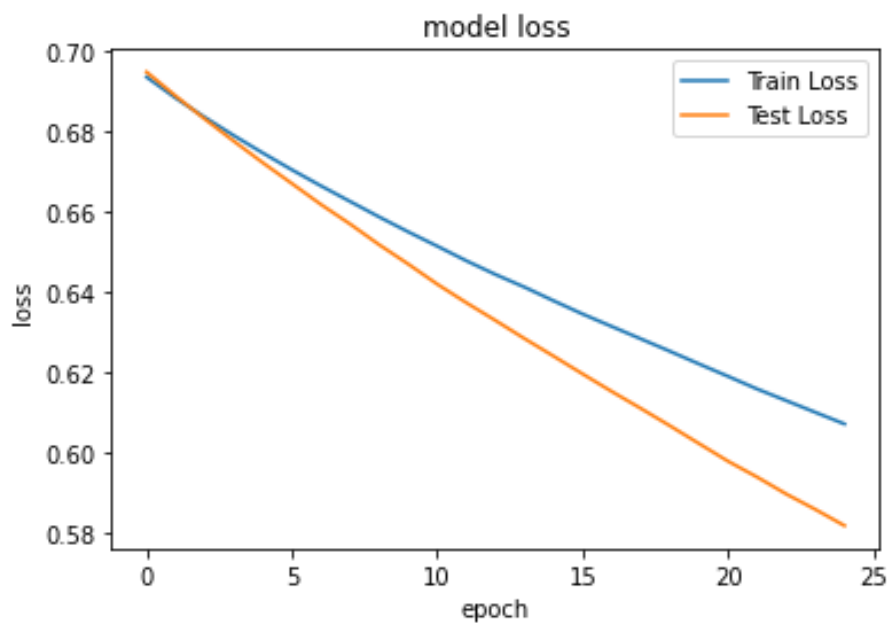
Near-Miss Undersampling versi 3



Gambar 4.15 Grafik kurva akurasi training dan test pada dataset near-miss 3

Pada gambar 4.15 di atas, menunjukkan dua buah grafik berwarna jingga dan biru yang masing-masing menggambarkan nilai akurasi *test* dan *train* seiring bertambahnya *epoch*. Gambar 4.15 juga menunjukkan bahwa kurva akurasi *train* berakhir pada nilai akurasi sebesar 71.36% pada *epoch* ke-25, serta pada grafik akurasi *test* nilai akurasi pada *epoch* ke-25 berada pada nilai 77%. Dari hasil proses

train dan *test* yang ditampilkan oleh kurva pada grafik di atas, bahwa model yang dihasilkan memiliki nilai akurasi yang cukup tinggi, namun nilai akurasi yang didapati tidak setinggi dengan nilai akurasi pada proses *train* model ketika menggunakan *dataset* orisinal.



Gambar 4.16 Grafik kurva loss training dan test pada dataset near-miss 3

Berdasarkan gambar 4.16 di atas terdapat grafik berwarna biru merupakan grafik yang menunjukkan nilai *train loss* dan grafik berwarna jingga menunjukkan nilai *test loss*. Nilai *train loss* dan *test loss* pada akhir *epoch* masing-masing sebesar 0.6072 dan 0.5819. Dengan hasil yang ditampilkan oleh kurva pada grafik di atas, bahwa model yang dihasilkan memiliki nilai *loss* yang cukup rendah, meskipun nilai *loss* tersebut lebih besar jika dibandingkan dengan nilai *loss* pada model ketika menggunakan *dataset* orisinal.

Tabel 4.7 Data Uji Menggunakan Model dengan Dataset Hasil Near-Miss 3

Uji ke-	Data Uji	Hasil Prediksi	Benar/Salah
1	Negatif Stroke	Negatif Stroke	Benar
2	Positif Stroke	Positif Stroke	Benar
3	Positif Stroke	Negatif Stroke	Salah
4	Positif Stroke	Positif Stroke	Benar
5	Positif Stroke	Negatif Stroke	Salah
6	Negatif Stroke	Positif Stroke	Salah
7	Positif Stroke	Positif Stroke	Benar
8	Positif Stroke	Positif Stroke	Benar
9	Negatif Stroke	Positif Stroke	Salah
10	Negatif Stroke	Negatif Stroke	Benar
11	Negatif Stroke	Negatif Stroke	Benar
12	Positif Stroke	Positif Stroke	Benar
13	Positif Stroke	Positif Stroke	Benar
14	Positif Stroke	Negatif Stroke	Salah
15	Negatif Stroke	Negatif Stroke	Benar
16	Positif Stroke	Negatif Stroke	Salah
17	Positif Stroke	Positif Stroke	Benar
18	Positif Stroke	Negatif Stroke	Salah
19	Negatif Stroke	Positif Stroke	Salah
20	Negatif Stroke	Negatif Stroke	Benar
21	Negatif Stroke	Negatif Stroke	Benar
22	Positif Stroke	Positif Stroke	Benar
23	Negatif Stroke	Negatif Stroke	Benar
24	Negatif Stroke	Positif Stroke	Salah
25	Negatif Stroke	Negatif Stroke	Benar
26	Positif Stroke	Positif Stroke	Benar
27	Negatif Stroke	Negatif Stroke	Benar
28	Positif Stroke	Positif Stroke	Benar
29	Negatif Stroke	Negatif Stroke	Benar
30	Negatif Stroke	Negatif Stroke	Benar
31	Positif Stroke	Positif Stroke	Benar
32	Negatif Stroke	Negatif Stroke	Benar
33	Negatif Stroke	Negatif Stroke	Benar
34	Negatif Stroke	Negatif Stroke	Benar
35	Positif Stroke	Positif Stroke	Benar
36	Positif Stroke	Positif Stroke	Benar
37	Positif Stroke	Positif Stroke	Benar
38	Positif Stroke	Negatif Stroke	Salah
39	Positif Stroke	Positif Stroke	Benar

40	Negatif Stroke	Negatif Stroke	Benar
41	Positif Stroke	Positif Stroke	Benar
42	Negatif Stroke	Positif Stroke	Salah
43	Positif Stroke	Negatif Stroke	Salah
44	Negatif Stroke	Negatif Stroke	Benar
45	Negatif Stroke	Negatif Stroke	Benar
46	Negatif Stroke	Negatif Stroke	Benar
47	Negatif Stroke	Positif Stroke	Salah
48	Positif Stroke	Positif Stroke	Benar
49	Positif Stroke	Positif Stroke	Benar
50	Negatif Stroke	Positif Stroke	Salah
51	Positif Stroke	Positif Stroke	Benar
52	Negatif Stroke	Negatif Stroke	Benar
53	Positif Stroke	Negatif Stroke	Salah
54	Negatif Stroke	Negatif Stroke	Benar
55	Negatif Stroke	Negatif Stroke	Benar
56	Negatif Stroke	Negatif Stroke	Benar
57	Positif Stroke	Positif Stroke	Benar
58	Negatif Stroke	Negatif Stroke	Benar
59	Negatif Stroke	Negatif Stroke	Benar
60	Positif Stroke	Positif Stroke	Benar
61	Negatif Stroke	Negatif Stroke	Benar
62	Positif Stroke	Negatif Stroke	Salah
63	Positif Stroke	Positif Stroke	Benar
64	Positif Stroke	Positif Stroke	Benar
65	Positif Stroke	Positif Stroke	Benar
66	Negatif Stroke	Positif Stroke	Salah
67	Negatif Stroke	Negatif Stroke	Benar
68	Positif Stroke	Positif Stroke	Benar
69	Positif Stroke	Positif Stroke	Benar
70	Negatif Stroke	Negatif Stroke	Benar
71	Negatif Stroke	Negatif Stroke	Benar
72	Negatif Stroke	Negatif Stroke	Benar
73	Positif Stroke	Negatif Stroke	Salah
74	Positif Stroke	Positif Stroke	Benar
75	Negatif Stroke	Negatif Stroke	Benar
76	Positif Stroke	Positif Stroke	Benar
77	Positif Stroke	Positif Stroke	Benar
78	Negatif Stroke	Negatif Stroke	Benar
79	Positif Stroke	Negatif Stroke	Salah
80	Positif Stroke	Positif Stroke	Benar
81	Positif Stroke	Positif Stroke	Benar

82	Negatif Stroke	Positif Stroke	Salah
83	Positif Stroke	Negatif Stroke	Salah
84	Negatif Stroke	Negatif Stroke	Benar
85	Positif Stroke	Negatif Stroke	Salah
86	Positif Stroke	Positif Stroke	Benar
87	Positif Stroke	Negatif Stroke	Salah
88	Negatif Stroke	Negatif Stroke	Benar
89	Positif Stroke	Positif Stroke	Benar
90	Negatif Stroke	Negatif Stroke	Benar
91	Negatif Stroke	Negatif Stroke	Benar
92	Positif Stroke	Positif Stroke	Benar
93	Positif Stroke	Positif Stroke	Benar
94	Negatif Stroke	Positif Stroke	Salah
95	Negatif Stroke	Negatif Stroke	Benar
96	Positif Stroke	Positif Stroke	Benar
97	Negatif Stroke	Negatif Stroke	Benar
98	Positif Stroke	Positif Stroke	Benar
99	Negatif Stroke	Negatif Stroke	Benar
100	Negatif Stroke	Positif Stroke	Salah

Dari hasil pada tabel 4.7 di atas, menunjukkan bahwa hasil prediksi oleh model dengan *dataset* orisinil memiliki tingkat akurasi yang cukup tinggi. Berdasarkan tabel tersebut dapat dihitung nilai akurasi dan nilai *loss* dengan persamaan di bawah ini.

$$\%akurasi = \frac{Jumlah\ prediksi\ benar}{Total\ jumlah\ prediksi} \times 100\%$$

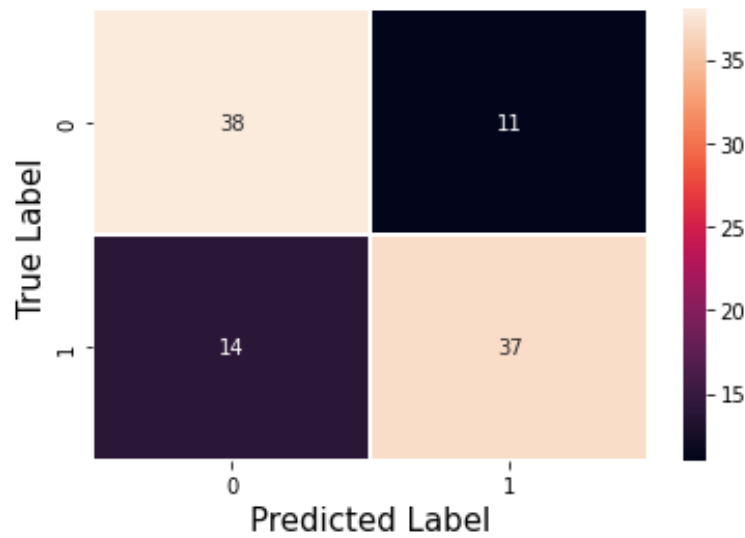
$$\%akurasi = \frac{75}{100} \times 100\% = 75.00\%$$

$$\%loss = \frac{Jumlah\ prediksi\ benar}{Total\ jumlah\ prediksi} \times 100\%$$

$$\%loss = \frac{25}{100} \times 100\% = 25.00\%$$

Berdasarkan perhitungan tersebut, terlihat bahwa nilai akurasi yang didapati oleh model dengan *dataset* hasil *near-miss* 3 sebesar 75.00% dan nilai *loss* sebesar 25.00%. Sehingga dapat dikatakan bahwa model ini memiliki kemampuan klasifikasi yang cukup baik.

Selanjutnya untuk melihat tingkat kebenaran dan kesalahan dari hasil prediksi pada tahap pengujian yang dilakukan terdapat *confusion matrix*. Dalam *confusion matrix* ini memiliki sumbu y dan sumbu x diberi label *True label* dan *Predicted label*, serta label angka “0” dan “1” yang masing-masing mewakili kondisi “negatif stroke” dan “positif stroke”. Berdasarkan pada gambar 4.14 yang merupakan *confusion matrix* hasil dari pengujian pada model dengan *dataset* yang dilakukan *near-miss undersampling* versi 3, menunjukkan bahwa jumlah data yang diprediksi benar lebih tinggi daripada jumlah data yang salah diprediksi, hasil prediksi yang menyatakan negatif stroke dan sesuai dengan data yang seharusnya mencapai angka 38, serta hasil prediksi yang menyatakan positif stroke dan sesuai dengan data yang seharusnya mencapai angka 37. Sehingga diperoleh nilai *True negative rate* sebesar 77.6% dan nilai *True positive rate* adalah sebesar 72.5%



Gambar 4.17 Confusion Matrix pada model dengan dataset near-miss 3

Walaupun hasil proses *train* diperoleh nilai akurasi yang lebih rendah dan nilai *loss* yang lebih tinggi, dapat dilihat pada *confusion matrix* yang menunjukkan model memiliki performa yang baik, serta tidak mengalami *overfitting* sebagaimana yang terjadi pada model sebelumnya ketika menggunakan *dataset* orisinal.

Berikut adalah performa dari model dari berbagai parameter seperti nilai *f1 score*, *precision score*, dan *recall score* yang dapat dilihat pada tabel 4.8 di bawah ini.

Tabel 4.8 Nilai evaluasi model dengan dataset near-miss 3

<i>F1 Score</i>	0.747
<i>Precision Score</i>	0.673
<i>Recall Score</i>	0.725

4.3 Perbandingan Evaluasi Performa Model

Berdasarkan dari hasil penelitian keempat buah percobaan terhadap model dengan *imbalanced dataset* dan tiga *balanced dataset* memiliki perbedaan. Berikut adalah perbedaan dari hasil evaluasi pada penelitian.

Tabel 4.9 Perbandingan evaluasi pada akurasi model

<i>Dataset</i>	<i>Train Accuracy</i>	<i>Test Accuracy</i>
<i>Dataset</i> orisinal	95.23%	94.72%
Hasil <i>near-miss</i> versi 1	66.58%	72%
Hasil <i>near-miss</i> versi 2	72.36%	69%
Hasil <i>near-miss</i> versi 3	71.36%	75%

Tabel 4.9 di atas menunjukkan perbandingan nilai akurasi model dari evaluasi empat model yang dihasilkan, yaitu model dengan *dataset* orisinal yang merupakan *imbalanced dataset*, model dengan *dataset* hasil *near-miss undersampling* versi 1, model dengan *dataset* hasil *near-miss undersampling* versi 2, dan model dengan *dataset* hasil *near-miss undersampling* versi 3. Berdasarkan tabel evaluasi akurasi tersebut model dengan *imbalanced dataset* memiliki akurasi yang lebih tinggi daripada model dengan *dataset* yang dilakukan *near-miss undersampling*, yakni akurasinya sebesar 95.23%, sedangkan untuk model dengan *dataset* hasil *near-miss undersampling* secara berurut memiliki nilai akurasi 66.58%, 72.36%, dan 71.36%.

Tabel 4.10 Perbandingan evaluasi pada nilai loss model

<i>Dataset</i>	<i>Train Loss</i>	<i>Test Loss</i>
<i>Dataset</i> orisinal	0.1558	0.1910
Hasil <i>near-miss</i> versi 1	0.6181	0.5972

Hasil <i>near-miss</i> versi 2	0.5754	0.5863
Hasil <i>near-miss</i> versi 3	0.6051	0.5808

Pada tabel 4.10 menunjukkan perbandingan nilai *loss* model dari evaluasi empat model yang dihasilkan, yakni model dengan *dataset* orisinil yang merupakan *imbalanced dataset*, model dengan *dataset* hasil *near-miss undersampling* versi 1, model dengan *dataset* hasil *near-miss undersampling* versi 2, dan model dengan *dataset* hasil *near-miss undersampling* versi 3. Pada hasil evaluasi tersebut nilai *loss* terkecil dimiliki oleh model dengan *imbalanced dataset*, yaitu dengan nilai *loss* sebesar 0.1558, selanjutnya untuk model dengan *dataset* hasil *near-miss undersampling* secara berurut memiliki nilai *loss* 0.6181, 0.5754, dan 0.6051.

Nilai akurasi dan nilai *loss* dari keempat model tersebut berdasarkan hasil proses *train* dan *test* memiliki nilai yang relatif sama. Nilai akurasi *train* dan *test* dari masing-masing model secara berurut adalah 95,53% dan 94,72%, 66.33% dan 74%, 72.61% dan 68%, serta 71.36% dan 77%. Lalu nilai *loss train* dan *test* dari masing-masing model secara berurut adalah 0.1588 dan 0.1830, 0.6199 dan 0.6865, 0.5776 dan 0.5932, serta 0.6072 dan 0.5819.

Tabel 4.11 Perbandingan hasil *confusion matrix* model

<i>Dataset</i>	<i>F1</i>	<i>Precision</i>	<i>Recall</i>
<i>Dataset</i> orisinil	0.00	0.00	0.00
Hasil <i>near-miss</i> versi 1	0.731	0.717	0.745
Hasil <i>near-miss</i> versi 2	0.660	0.673	0.647
Hasil <i>near-miss</i> versi 3	0.747	0.673	0.725

Dapat dilihat pada tabel 4.11 di atas, menunjukkan perbandingan nilai skor *f1*, *precision*, dan *recall* model dari empat model yang dihasilkan, yakni model dengan *dataset* orisinil yang merupakan *imbalanced dataset*, model dengan *dataset* hasil *near-miss undersampling* versi 1, model dengan *dataset* hasil *near-miss undersampling* versi 2, dan model dengan *dataset* hasil *near-miss undersampling* versi 3. Untuk model dengan menggunakan *dataset* orisinil memiliki nilai skor *f1*, *precision*, dan *recall* yang sama, yaitu bernilai 0.00, sedangkan untuk model yang menggunakan *dataset* hasil dari *near-miss undersampling* versi 1 memiliki skor *f1* sebesar 0.731, skor *precision* sebesar 0.717, dan skor *recall* sebesar 0.745, lalu untuk model yang menggunakan *dataset* hasil dari *near-miss undersampling* versi 2 memiliki skor *f1* sebesar 0.660, skor *precision* sebesar 0.673, dan skor *recall* sebesar 0.647, dan yang terakhir untuk model yang menggunakan *dataset* hasil dari *near-miss undersampling* versi 3 memiliki skor *f1* sebesar 0.747, skor *precision* sebesar 0.673, dan skor *recall* sebesar 0.725.

Berdasarkan hasil tersebut dapat dikatakan bahwa meskipun model dengan *dataset* orisinil memiliki tingkat akurasi yang sangat tinggi, sehingga dapat dikatakan bahwa model memiliki performa yang baik dalam klasifikasi, namun dari skor *f1*, *precision*, dan *recall* yang didapat menunjukkan bahwa model memiliki performa yang buruk dalam klasifikasi. Yang mana hal ini terjadi karena *dataset* orisinil yang digunakan merupakan *imbalanced dataset*, sehingga model memiliki kurva logistik yang cenderung mengarah ke arah kelas dengan distribusi terbanyak atau dengan kata lain model mengalami *overfitting*.

Model yang dilatih dengan menggunakan *dataset* yang sudah dioptimalisasi dengan diseimbangkan jumlah antar kelasnya dengan *near-miss undersampling* versi 1, 2, maupun 3 memiliki tingkat akurasi yang lebih rendah, namun dari skor *f1*, *precision*, dan *recall* yang didapat menunjukkan bahwa model memiliki performa yang jauh lebih baik dalam klasifikasi dibandingkan dengan model yang dilatih dengan *dataset* orisinal.

BAB 5 PENUTUP

5.1 Kesimpulan

Berdasarkan dari hasil penelitian yang sudah dilakukan penulis, maka dapat diambil kesimpulan sebagai berikut.

1. Optimalisasi *dataset* dengan menerapkan metode *near-miss undersampling* terhadap *imbalanced dataset* berpengaruh dengan signifikan terhadap performa model yang dibentuk.
2. Model dengan *dataset* orisinal memiliki nilai akurasi pelatihan sebesar 95.23%. Tiga model lainnya dengan *dataset* yang melalui *near-miss undersampling* masing-masing memiliki nilai akurasi pelatihan 66.58%, 72.36%, 71.36%, meskipun lebih rendah, namun performa secara keseluruhan dari model tersebut jauh lebih baik daripada model dengan *dataset* orisinal.
3. Model dengan *dataset* orisinal memiliki nilai skor f1, *precision*, dan *recall* bernilai 0.00. Model dengan *dataset* hasil dari *near-miss undersampling* versi 1 memiliki skor f1, skor *precision*, dan skor *recall* sebesar 0.731, 0.717, dan 0.745. Model dengan *dataset* hasil dari *near-miss undersampling* versi 2 memiliki skor f1, skor *precision*, dan skor *recall* sebesar 0.660, 0.673, dan 0.647, dan model dengan *dataset* hasil dari *near-miss undersampling* versi 3 memiliki skor f1, skor *precision*, dan skor *recall* sebesar 0.747, 0.673, dan 0.725.
4. Meskipun model yang dilatih dengan *dataset* yang melalui *near-miss undersampling* memiliki nilai akurasi yang rendah, namun performa secara

keseluruhan dari model tersebut jauh lebih baik daripada model yang dilatih dengan *dataset* orisinal.

5.2 Saran

Adapun beberapa saran yang penulis sampaikan untuk pengembangan penelitian ini adalah sebagai berikut.

1. Menambahkan variasi metode *sampling* yang digunakan sehingga dapat dilihat metode mana yang lebih signifikan dalam mengoptimisasi *dataset* dan performa model.
2. Melakukan modifikasi pada arsitektur model, sehingga dapat menghasilkan model dengan performa yang lebih baik.

DAFTAR PUSTAKA

- [1] M. Ahadi, "UNDERSAMPLING MAJORITY CLASS PADA KASUS IMBALANCED DATASET DAN APLIKASINYA PADA DETEKSI ANOMALI TRANSAKSI," 2019.
- [2] X. Ying, "An Overview of Overfitting and its Solutions," *J. Phys. Conf. Ser.*, vol. 1168, no. 2, 2019, doi: 10.1088/1742-6596/1168/2/022022.
- [3] D. Ramyachitra and P. Manikandan, "Imbalanced dataset classification and solutions: a review," *Int. J. Comput. Bus. Res.*, vol. 5, no. 4, 2014.
- [4] R. C. Prati, G. E. A. P. A. Batista, and M. C. Monard, "Data mining with unbalanced class distributions: Concepts and methods," *Proc. 4th Indian Int. Conf. Artif. Intell. IICAI 2009*, pp. 359–376, 2009.
- [5] E. E. Dooley, "American Heart Association," *Environ. Health Perspect.*, vol. 112, no. 15, pp. 1–8, 2004, doi: 10.1289/ehp.112-a873.
- [6] "Handling Imbalanced Dataset. Pada proses klasifikasi biasanya kita... | by Rusnandi Fikri | Medium."
<https://medium.com/@rusnandifikri96/handling-imbalanced-dataset-260378b2a21b> (accessed Dec. 15, 2021).
- [7] "Imbalanced Class Effect | M Reza Faisal."
<https://staf.ulm.ac.id/rezafaisal/2016/09/11/imbalance-class-effect/> (accessed Dec. 15, 2021).
- [8] S. Mutmainah, "Penanganan Imbalance Data Pada Klasifikasi Kemungkinan Penyakit Stroke," *J. SNATi*, vol. 1, pp. 10–16, 2021, [Online]. Available: <https://library.uui.ac.id/osr>.
- [9] A. Indrawati, L. Ilmu, P. Indonesia, and P. I. Diabetes, "Penerapan Teknik Kombinasi Oversampling Dan Undersampling Hybrid Oversampling and Undersampling Techniques To Handling Imbalanced Dataset," vol. 4, no. 1, pp. 38–43, 2021, doi: 10.33387/jiko.
- [10] A. Mujiit WS and R. Nooraeni, "Penerapan Metode Resampling Dalam Mengatasi Imbalanced Data Pada Determinan Kasus Diare Pada Balita Di Indonesia (Analisis Data Sdki 2017)," *J. MSA (Mat. dan Stat. serta Apl.)*, vol. 8, no. 1, p. 19, 2020, doi: 10.24252/msa.v8i1.13452.
- [11] Z. Li, K. Kamnitsas, and B. Glocker, "Analyzing Overfitting under Class Imbalance in Neural Networks for Image Segmentation," pp. 1–18.
- [12] "Google Colab: Arti, Manfaat, dan Cara Menggunakannya - Glints Blog."
<https://glints.com/id/lowongan/google-colab-adalah/#.YXgFMxpByHt> (accessed Oct. 26, 2021).
- [13] "TensorFlow vs Keras: Which One Should You Choose."
<https://analyticsindiamag.com/tensorflow-vs-keras-which-one-should-you-choose/> (accessed Oct. 26, 2021).
- [14] "About Keras." <https://keras.io/about/> (accessed Oct. 26, 2021).


LAMPIRAN

Lampiran 1. Tampilan Aplikasi MedikalTech Capstone Project 2021

The image shows a mobile application interface for a stroke-related app. The top status bar shows the time as 13:32 and battery level at 73%. The app's header is purple with the title "Stroke". On the left, there is a circular logo with a green leaf-like design and the text "MedikalTech" below it. The main content area is light blue and contains a registration form. The form starts with a greeting "Hello, Name" and a prompt "Please fill out this form!". It then lists several fields: "Name" (a text input field), "Age" (a text input field), "Gender" (radio buttons for "Male" and "Female"), "Marriage Status" (radio buttons for "Married", "Single", and "Divorce"), "Work Status" (radio buttons for "Employed", "Unemployed", and "Fired"), "Residency" (radio buttons for "City" and "Village"), and "Smoking Status" (radio buttons for "Never", "Smoking", and "Not again"). At the bottom of the form, there are two purple rounded rectangular buttons, both labeled "Next". The bottom of the screen shows the standard Android navigation bar with back, home, and recent apps icons.

13:32 73%

Stroke


MedikalTech

Hello, Name

Please fill out this form!

Marriage Status

☐ Married ☐ Single ☐ Divorce

Work Status

☐ Employed ☐ Unemployed ☐ Fired

Residency

☐ City ☐ Village

Smoking Status

☐ Never ☐ Smoking ☐ Not again

Name

Age

Gender

☐ Male ☐ Female

Next

Next


```
[ ] #fill empty bmi data with median bmi data
df['bmi'].fillna(df['bmi'].median(), inplace=True)
```

```
[ ] df.isna().any()
```

```
id          False
gender      False
age         False
hypertension False
heart_disease False
ever_married False
work_type   False
Residence_type False
avg_glucose_level False
bmi         False
smoking_status False
stroke      False
dtype: bool
```

Rename Column's names

```
1 #Preparing Stroke
df = df.rename(columns={'heart_disease':'HD',
                        'ever_married':'EH',
                        'work_type':'WT',
                        'Residence_type':'RT',
                        'avg_glucose_level':'AG',
                        'bmi':'BMI',
                        'smoking_status':'SS'})
```

```
[ ] df
```

	id	gender	age	hypertension	HD	EH	WT	RT	AG	BMI	SS	stroke
0	9046	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	51676	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	28.1	never smoked	1
2	31112	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
3	60182	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
4	1665	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1
...
5105	18234	Female	80.0	1	0	Yes	Private	Urban	83.75	28.1	never smoked	0
5106	44873	Female	81.0	0	0	Yes	Self-employed	Urban	125.20	40.0	never smoked	0
5107	19723	Female	35.0	0	0	Yes	Self-employed	Rural	82.99	30.6	never smoked	0
5108	37544	Male	51.0	0	0	Yes	Private	Rural	166.29	25.6	formerly smoked	0
5109	44670	Female	44.0	0	0	Yes	Govt_job	Urban	85.28	26.2	Unknown	0

5110 rows x 12 columns

```
[ ] categorical_col = []
for col in df:
    if (df[col].dtypes == object):
        categorical_col.append(col)

categorical_col

['gender', 'EH', 'WT', 'RT', 'SS']
```

```
[ ] df['gender'].unique()
```

```
array(['Male', 'Female'], dtype=object)
```

```
age gender age hypertension HD EH WT RT AG BMI SS stroke
0 9046 1 67.0 0 1 1 2 1 228.69 36.6 1 1
1 51676 0 61.0 0 0 1 3 0 202.21 28.1 2 1
2 31112 1 80.0 0 1 1 2 0 105.92 32.5 2 1
3 60182 0 49.0 0 0 1 2 1 171.23 34.4 3 1
4 1665 0 79.0 1 0 1 3 0 174.12 24.0 2 1
...
5105 18234 0 80.0 1 0 1 2 1 83.75 28.1 2 0
5106 44873 0 81.0 0 0 1 3 1 125.20 40.0 2 0
5107 19723 0 35.0 0 0 1 3 0 82.99 30.6 2 0
5108 37544 1 51.0 0 0 1 2 0 166.29 25.6 1 0
5109 44670 0 44.0 0 0 1 0 1 85.28 26.2 0 0
5110 rows x 12 columns
```

```
1 df_encoded['stroke'].value_counts()
```

```
0 4861
1 249
Name: stroke, dtype: int64
```

```
[ ] pos_stroke = df_encoded['stroke'].value_counts()[1]
neg_stroke = df_encoded['stroke'].value_counts()[0]

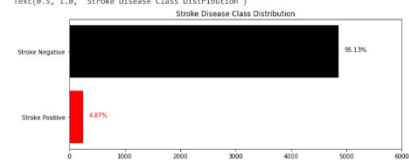
maj_class_rate = (neg_stroke/(pos_stroke + neg_stroke))*100
min_class_rate = (pos_stroke/(pos_stroke + neg_stroke))*100

labels = ['Stroke Positive', 'Stroke Negative']
values = [pos_stroke, neg_stroke]

plt.figure(figsize=(10,4))

plt.xlim(0,6000)
plt.bar(labels, values, color=['red', 'black'])
plt.text(pos_stroke + 100, labels[0], '{0:.2f}%'.format(maj_class_rate), color = 'red')
plt.text(neg_stroke + 100, labels[1], '{0:.2f}%'.format(min_class_rate), color = 'black')
plt.title('Stroke Disease Class Distribution')
```

Text(8.5, 1.8, 'Stroke Disease Class Distribution')



- Apply Near Miss Undersampling

```
[ ] #Importing NearMiss
from collections import Counter
from imblearn.under_sampling import NearMiss
```

```

#define the undersampling method
near1 = NearMiss(version=1)
near2 = NearMiss(version=2)
near3 = NearMiss(version=3)

[ ] #apply nearmiss 1
x_near1, y_near1 = near1.fit_resample(x,y)
x_near1 = pd.DataFrame(x_near1)
y_near1 = pd.DataFrame(y_near1)
df_near1 = x_near1.join(y_near1)

#apply nearmiss 2
x_near2, y_near2 = near2.fit_resample(x,y)
x_near2 = pd.DataFrame(x_near2)
y_near2 = pd.DataFrame(y_near2)
df_near2 = x_near2.join(y_near2)

#apply nearmiss 3
x_near3, y_near3 = near3.fit_resample(x,y)
x_near3 = pd.DataFrame(x_near3)
y_near3 = pd.DataFrame(y_near3)
df_near3 = x_near3.join(y_near3)

# print('Nearmiss Undersampling Result')
print('Nearmiss 1 : ', Counter(df_near1['stroke']).items())
print('Nearmiss 2 : ', Counter(df_near2['stroke']).items())
print('Nearmiss 3 : ', Counter(df_near3['stroke']).items())

Nearmiss Undersampling Result
Nearmiss 1 : dict_items([(0, 249), (1, 249)])
Nearmiss 2 : dict_items([(0, 249), (1, 249)])
Nearmiss 3 : dict_items([(0, 249), (1, 249)])

[ ] x = df_encoded.drop(['stroke'], axis = 1)
y = df_encoded['stroke']

[ ] #Plot Class Distribution
fig, ax = plt.subplots(1, sharex=True, figsize=(10,4))

fig.suptitle('Stroke Disease Class Distribution')
for x in range(len(ax)):
    ax[x].set_xlim([0, 280])

#Plot Class Distribution Nearmiss 1
pos_stroke1 = df_near1['stroke'].value_counts()[1]
neg_stroke1 = df_near1['stroke'].value_counts()[0]

maj_class_rate1 = (neg_stroke1/(pos_stroke1 + neg_stroke1))*100
min_class_rate1 = (pos_stroke1/(pos_stroke1 + neg_stroke1))*100

labels1 = ['Stroke Positive', 'Stroke Negative']
values1 = [pos_stroke1, neg_stroke1]
ax[0].set_ylabel('Nearmiss 1')
ax[0].barh(labels1, values1, color=['red', 'black'])
ax[0].text(pos_stroke1 + 3, labels1[0], '{0:.2f}%'.format(min_class_rate1), color = 'red')
ax[0].text(neg_stroke1 + 3, labels1[1], '{0:.2f}%'.format(maj_class_rate1), color = 'black')

#Plot Class Distribution Nearmiss 2
pos_stroke2 = df_near2['stroke'].value_counts()[1]
neg_stroke2 = df_near2['stroke'].value_counts()[0]

maj_class_rate2 = (neg_stroke2/(pos_stroke2 + neg_stroke2))*100
min_class_rate2 = (pos_stroke2/(pos_stroke2 + neg_stroke2))*100

labels2 = ['Stroke Positive', 'Stroke Negative']
values2 = [pos_stroke2, neg_stroke2]
ax[1].set_ylabel('Nearmiss 2')

```

```

#fig=plt.gcf()
#fig.savefig('nearmiss_undersampling.png')

```

Before Encoding Categorical Data

```

[ ] for col in categorical_col:
    print(df[col].unique())

['Male' 'female' 'Other']
['Yes' 'No']
['Private' 'Self-employed' 'Govt_Job' 'children' 'Never_worked']
['Urban' 'Rural']
['formerly smoked' 'never smoked' 'smokes' 'Unknown']

[ ] #create encoder using OrdinalEncoder
ord = OrdinalEncoder()
df_encoded = df.copy()

#encoding categorical data
for col in categorical_col:
    df_encoded[col] = ord.fit_transform(df_encoded[col]).astype(np.int64)

```

After Encoding Categorical Data

```

[ ] for col in categorical_col:
    print(df_encoded[col].unique())

[ 0 2]
[ 0]
[ 2 3 4 1]
[ 0]
[ 2 3 0]

[ ] df_encoded

   id  gender  age  hypertension  HD  EH  HT  RT  AG  BPL  SS  stroke
0   9046     1  67.0           0    1  1  2  1  228.69  36.6  1    1
1   51676    0  61.0           0    0  1  3  0  202.21  28.1  2    1
2   31112     1  80.0           0    1  1  2  0  105.92  32.5  2    1
3   60182    0  49.0           0    0  1  2  1  171.23  34.4  3    1
4   1665     0  79.0           1    0  1  3  0  174.12  24.0  2    1
...
5105 18234     0  80.0           1    0  1  2  1  83.75  28.1  2    0
5106 44873    0  81.0           0    0  1  3  1  125.20  40.0  2    0
5107 19723    0  35.0           0    0  1  3  0  82.99  30.6  2    0
5108 37544     1  51.0           0    0  1  2  0  166.20  25.6  1    0
5109 44679    0  44.0           0    0  1  0  1  85.28  26.2  0    0
5110 rows x 12 columns

```

Dealing with Imbalanced Dataset

Class Distribution

```

[ ] df_encoded

   id  gender  age  hypertension  HD  EH  HT  RT  AG  BPL  SS  stroke
0   9046     1  67.0           0    1  1  2  1  228.69  36.6  1    1
1   51676    0  61.0           0    0  1  3  0  202.21  28.1  2    1
2   31112     1  80.0           0    1  1  2  0  105.92  32.5  2    1
3   60182    0  49.0           0    0  1  2  1  171.23  34.4  3    1
4   1665     0  79.0           1    0  1  3  0  174.12  24.0  2    1
...
5105 18234     0  80.0           1    0  1  2  1  83.75  28.1  2    0
5106 44873    0  81.0           0    0  1  3  1  125.20  40.0  2    0
5107 19723    0  35.0           0    0  1  3  0  82.99  30.6  2    0
5108 37544     1  51.0           0    0  1  2  0  166.20  25.6  1    0
5109 44679    0  44.0           0    0  1  0  1  85.28  26.2  0    0
5110 rows x 12 columns

```



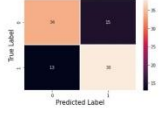
```

#Plot confusion matrix
import seaborn as sns

sns.heatmap(cm, annot=True, fmt='d', linewidths=1)
plt.xlabel('Predicted Label', fontsize=15)
plt.ylabel('True Label', fontsize=15)

text(35, 0.5, 'True Label')

```



```

1) true_negative_rate = tp / (tp + fn)
true_positive_rate = tp / (tp + fp)
print('TPR : %.2f' % true_positive_rate, 'FPR : %.2f' % true_positive_rate)

precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1_score = f1_score(y_test, y_pred)

print('Precision score : %.2f' % precision)
y_pred = pd.DataFrame(y_pred, columns=['stroke'])
f1_score = f1_score(y_test, y_pred)

```

```

1) y_pred

stroke
0    1
1    1
2    0
3    0
4    1
...
95    1
96    0
97    1
98    1
99    1
100 rows x 1 columns

```

```

1) y_test = pd.DataFrame(y_test, columns=['stroke'])

stroke
90    0
254    1
283    1
100 rows x 1 columns

```

```

1) y_pred["stroke"].value_counts()

0    51
1    49
Name: stroke, dtype: int64

```

```

1) #Plot confusion matrix values
cm = confusion_matrix(y_test, y_pred)
text(15, 0.5, 'cm = cm')

#Plot confusion matrix
import seaborn as sns

sns.heatmap(cm, annot=True, fmt='d', linewidths=1)
plt.xlabel('Predicted Label', fontsize=15)
plt.ylabel('True Label', fontsize=15)

```



```

# Confusion matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)

# True Positive Rate (TPR) and False Positive Rate (FPR)
fpr, tpr, thresholds = roc_curve(y_test, y_scores)
print(fpr, tpr, thresholds)

# Area Under the Curve (AUC)
auc = roc_auc_score(y_test, y_scores)
print(auc)

# Model Metrics
print(f"Model Metrics")
print(f"Accuracy: {accuracy_score(y_test, y_pred)}")
print(f"Precision: {precision_score(y_test, y_pred)}")
print(f"Recall: {recall_score(y_test, y_pred)}")
print(f"F1 Score: {f1_score(y_test, y_pred)}")

# Confusion Matrix
print(f"Confusion Matrix")
print(cm)

# ROC Curve
plt.figure()
plt.plot(fpr, tpr, label='ROC Curve')
plt.plot([0, 1], [0, 1], label='Random Guess')
plt.legend()
plt.show()

# Precision-Recall Curve
plt.figure()
plt.plot(precision, recall, label='Precision-Recall Curve')
plt.legend()
plt.show()

# Feature Importance
print(f"Feature Importance")
print(importance_scores)

# Model Summary
print(f"Model Summary")
print(model.summary())

```

BIODATA PENULIS



Biodata penulis berisi terkait dengan identitas penulis (nama, kontak email), riwayat akademis (pendidikan) penulis ditulis dari yang paling , skill, serta prestasi penulis.

A. Identitas

Nama : Yusuf Muhammad Nur Zaman
NIM : H1A018043
Tempat, tanggal lahir : Tasikmalaya, 1 Maret 2021
Alamat : Jl. EmpangSari no.53 Kec. Tawang, Kota Tasikmalaya, Jawa Barat
No. Telp. : 087823130029
Alamat e-mail : yusuf.zaman@mhs.unsoed.ac.id

B. Riwayat Pendidikan Akademik

Periode	Jenjang	Institusi
2018 – sekarang	S1	Teknik Elektro Universitas Jenderal Soedirman
2015 – 2018	SMA	SMAN 1 Tasikmalaya
2013 – 2015	SMP	SMPN 1 Tasikmalaya
2007 – 2013	SD	SDN Citapen 1 Tasikmalaya

C. Prestasi

Tahun	Tingkat	Prestasi
2021	Jurusan	Juara 1 PIM Jurusan Teknik Elektro Unsoed, Purbalingga
2021	Fakultas	Juara 1 PIM Fakultas Teknik Unsoed, Purbalingga
2021	Internasional	Medali Emas kategori <i>Power and Electricity</i> , Indonesia Invention Day 2021, Denpasar-Bali