

BAB 3

METODE PENELITIAN

3.1 SUMBER DATA

Dataset [Heart failure prediction / Kaggle](#) [7] merupakan kumpulan data yang berisi catatan medis dari 299 pasien gagal jantung yang dikumpulkan di Institut Kardiologi Faisalabad dan di Rumah Sakit Sekutu di Faisalabad (Punjab, Pakistan), selama April–Desember 2015. Kumpulan data tersebut berisi 13 fitur, yang melaporkan informasi klinis, tubuh, dan gaya hidup. Berikut ini adalah deskripsi dari 13 fitur tersebut.

Tabel 3.1 Variabel dan Deskripsi

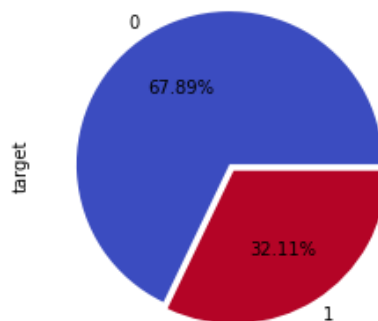
| Fitur | Deskripsi | Keterangan |
|---------------------------------|---|--------------------|
| <i>Age</i> | Umur pasien dalam tahun | Variabel Prediktor |
| <i>Anaemia</i> | Penurunan sel darah merah atau hemoglobin (dinyatakan dalam <i>Boolean</i>) | |
| <i>Creatinine_phosphokinase</i> | Tingkat enzim CPK di dalam darah (mcg/L) | |
| <i>Diabetes</i> | Jika pasien memiliki penyakit diabetes (dinyatakan dalam <i>Boolean</i>) | |
| <i>Ejection_fraction</i> | Persentase darah yang meninggalkan hati di setiap kontraksi | |
| <i>High_blood_pressure</i> | Jika pasien memiliki tekanan darah tinggi/hipertensi (dinyatakan dalam <i>Boolean</i>) | |
| <i>Platelets</i> | Trombosit dalam darah (kiloplatelet/mL) | |
| <i>Serum_creatinine</i> | Tingkat kreatinin serum dalam darah (mg/dL) | |

| | | |
|---------------------|---|-----------------|
| <i>Serum_sodium</i> | Tingkat sodium serum dalam darah (mEq/L) | |
| <i>Sex</i> | Gender: <i>Female</i> /Wanita atau <i>Male</i> /Pria | |
| <i>Smoking</i> | Pasien merokok atau tidak | |
| <i>Time</i> | Periode masa tindak lanjut/ <i>follow-up</i> (hari) | |
| <i>Death_event</i> | Jika pasien meninggal selama masa tindak lanjut/ <i>follow-up</i> | Variabel Target |

Tabel 3.1 merupakan deskripsi dari tiap fitur dan juga penggolongan fitur terhadap variabel target dan variabel prediktor. Berikut ini adalah penjelasan lebih lanjut mengenai fitur - fitur yang terdapat di dalam dataset.

A. Peristiwa Kematian (*Death event*)

Fitur peristiwa kematian atau *death_event* merupakan variabel target dalam penelitian ini. Variabel target ini sangat bergantung kepada variabel prediktor, karena *output* yang akan dihasilkan sangat bergantung terhadap *inputan* dari variabel prediktor.



Gambar 3.1 Persentase Peristiwa Kematian Pasien Gagal Jantung

Berdasarkan gambar 3.1 diketahui sebanyak 67,89% pasien selamat dari gagal jantung dan sisanya 32,11% pasien meninggal akibat gagal jantung. Berdasarkan persentase dapat diketahui bahwa pasien yang selamat dari gagal jantung lebih banyak daripada pasien yang meninggal akibat gagal jantung.



Gambar 3.2 Sebaran Pasien Yang Meninggal/Selamat Dari Gagal Jantung

Gambar 3.2 merupakan sebaran pasien yang meninggal akibat gagal jantung dan selamat dari gagal jantung. Berdasarkan gambar dapat diketahui bahwa pasien yang selamat dari gagal jantung lebih banyak daripada pasien yang meninggal akibat gagal jantung. Tabel dibawah ini merupakan rincian jumlah pasien yang selamat dari gagal jantung dan meninggal akibat gagal jantung.

Tabel 3.2 Peristiwa Kematian Pasien Gagal Jantung

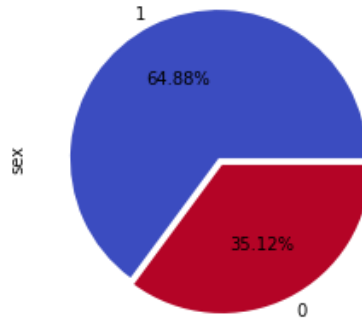
| Peristiwa Kematian | Jumlah |
|--------------------|-----------|
| Meninggal | 96 Orang |
| Selamat | 203 Orang |

Dalam tabel 3.2 dapat dilihat bahwa jumlah pasien yang selamat dari gagal jantung sebanyak 203 orang dan pasien yang meninggal akibat gagal jantung sebanyak 96 orang. Oleh karena itu dapat diketahui bahwa dataset tidak seimbang karena fitur ini merupakan variabel target dari penelitian ini, sehingga diperlukan penyeimbangan data agar data dapat mudah diolah oleh model sehingga output yang dihasilkan memiliki nilai yang baik.

B. Gender (*Sex*)

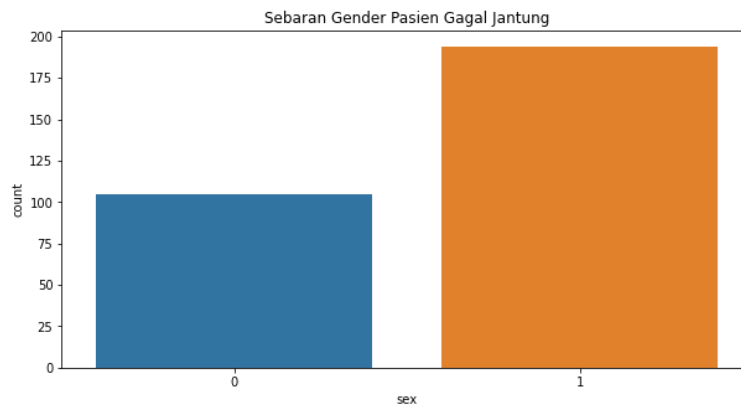
Berdasarkan informasi yang dirilis oleh Dinas Kesehatan Provinsi Kalimantan Barat diketahui bahwa laki-laki mempunyai resiko yang lebih besar daripada perempuan. Hal ini berhubungan dengan hormon estrogen yang dimiliki oleh para perempuan. Hormon estrogen memiliki karakteristik yang protektif kepada aterosklerosis. Setelah perempuan memasuki masa menopause, resiko

terhadap gagal jantung akan semakin tinggi dikarenakan tingkat hormon estrogen pada perempuan semakin turun [34].



Gambar 3.3 Persentase Gender Pasien Gagal Jantung

Berdasarkan gambar 3.3 diketahui sebanyak 64,88% pasien gagal jantung memiliki gender laki-laki dan sisanya 32,11% pasien gagal jantung memiliki gender perempuan.



Gambar 3.4 Sebaran Gender Pasien Gagal Jantung

Gambar 3.4 menunjukkan sebaran gender pasien gagal jantung, diketahui bahwa pasien gagal jantung yang berjenis kelamin laki-laki lebih banyak daripada pasien gagal jantung yang berjenis kelamin perempuan. Berikut adalah rincian jumlah pasien gagal jantung berdasarkan gender.

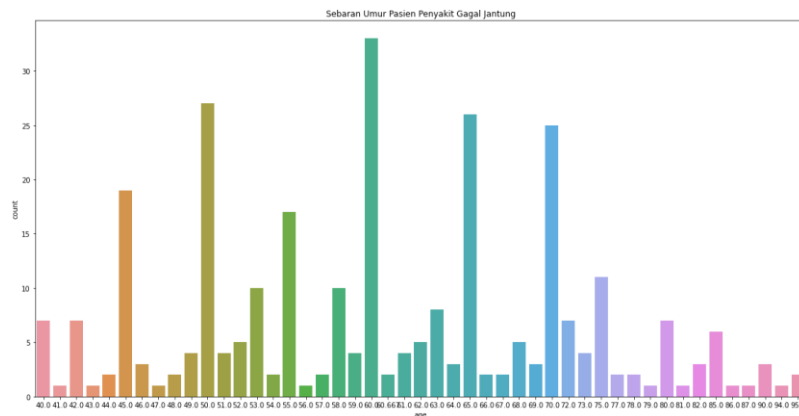
Tabel 3.3 Gender Pasien Gagal Jantung

| Gender | Jumlah |
|-----------|-----------|
| Laki-Laki | 194 Orang |
| Perempuan | 105 Orang |

Pada tabel 3.3 dapat dilihat bahwa pasien terdiri dari 105 wanita dan 194 pria. Disfungsi sistolik ventrikel telah dialami oleh 299 pasien gagal jantung oleh karena itu pasien-pasien tersebut dikelompokkan ke dalam kelas III atau IV tahapan gagal jantung oleh New York *Heart Association* (NYHA) [7].

C. Umur (*Age*)

Berdasarkan informasi yang dirilis oleh Dinas Kesehatan Provinsi Kalimantan Barat diketahui bahwa umur adalah faktor yang sangat mempengaruhi terjadinya gagal jantung, karena terjadinya proses aterosklerosis pada arteri koroner. Saluran arteri koroner digambarkan seperti saluran pipa air yang terbuat dari besi yang semakin lama akan timbul kerak pada dinding pipa yang menyebabkan tersumbatnya aliran air [34].

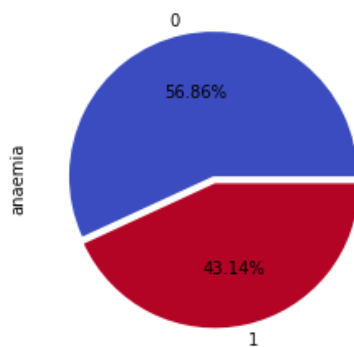


Gambar 3.5 Sebaran Umur Pasien Penyakit Gagal Jantung

Dari grafik yang tertera di gambar 3.5 dapat dilihat bahwa pasien dengan umur 60 tahun paling banyak mengalami gagal jantung. Sebanyak 33 orang pasien yang berumur 60 tahun pada grafik diatas.

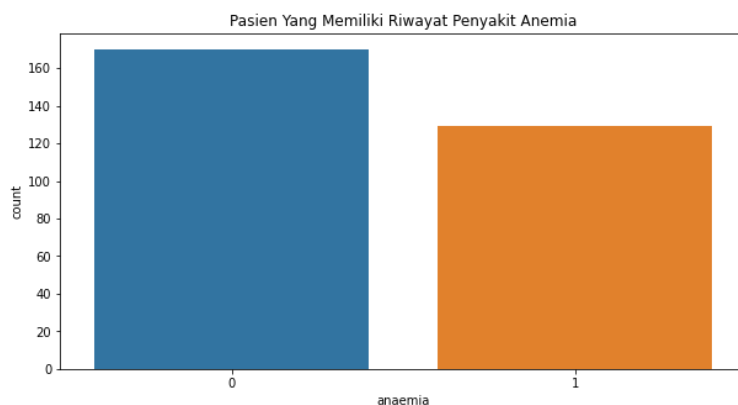
D. Anemia (*Anaemia*)

Berdasarkan informasi yang dirilis oleh Pusat Jantung Nasional diketahui bahwa saat seseorang mengalami anemia, oksigen yang diperoleh dari darah akan berkurang. Apabila situasi tersebut tidak segera di tindak lanjuti maka akan terjadi kerusakan pada organ yang terdapat di dalam tubuh, contohnya jantung [35].



Gambar 3.6 Persentase Pasien Gagal Jantung Yang Memiliki Anemia

Gambar 3.6 merupakan persentase pasien gagal jantung yang memiliki anemia dan yang tidak memiliki anemia. Diketahui bahwa pasien gagal jantung yang memiliki anemia sebanyak 43,14% sedangkan sisanya yaitu sebanyak 56,86% tidak memiliki anemia.



Gambar 3.7 Pasien Yang Memiliki Anemia

Gambar 3.7 merupakan sebaran pasien yang memiliki anemia. Berdasarkan gambar dapat dilihat bahwa pasien yang memiliki anemia jumlahnya sedikit daripada pasien yang tidak memiliki anemia. Tabel dibawah ini menjelaskan rincian jumlah pasien gagal jantung yang memiliki anemia dan pasien gagal jantung yang tidak memiliki anemia.

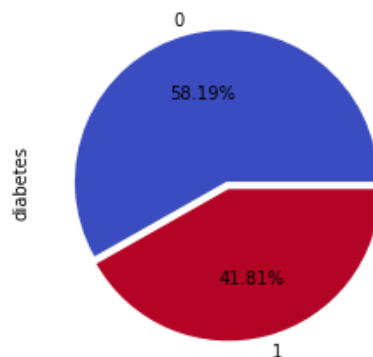
Tabel 3.4 Pasien Yang Memiliki Anemia

| Anemia | Jumlah |
|--------|-----------|
| Ya | 129 Orang |
| Tidak | 170 Orang |

Berdasarkan data pada tabel 3.4 diketahui sebanyak 129 orang pasien yang terkena gagal jantung memiliki anemia, sedangkan sisanya yaitu sebanyak 170 orang pasien yang terkena gagal jantung tidak memiliki anemia.

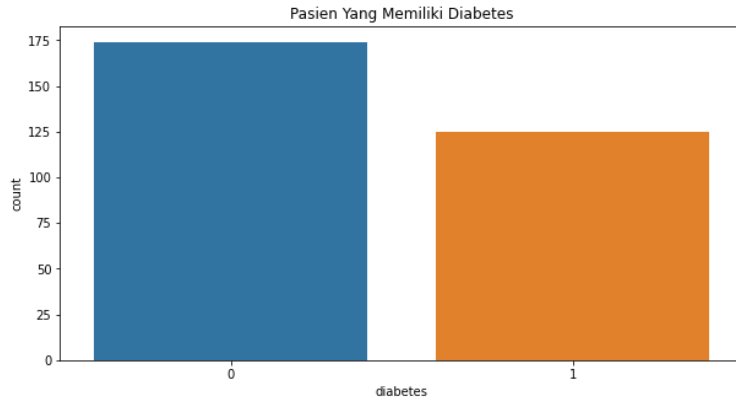
E. Diabetes

Berdasarkan informasi yang dirilis oleh Dinas Kesehatan Provinsi Kalimantan Barat diketahui bahwa diabetes menimbulkan resiko kepada penyakit jantung jika persentase glukosa dalam darah meningkat, lebih utamanya jika terjadi dalam periode yang lumayan lama dikarenakan glukosa tersebut bisa menjadi racun untuk tubuh, contohnya sistem kardiovaskuler. Pasien dengan diabetes lebih sering mengalami gangguan pada organ jantung saat umurnya masih muda. Diabetes yang tak terkontrol dengan persentase glukosa yang meningkat dalam darah condong meningkatkan persentase kolesterol. Proses metabolisme lemak dan degeneratif vaskular yang tak umum berperan dalam terjadinya perkembangan atheroma sehingga menyebabkan pembuluh darah arteri menjadi sempit atau disebut juga aterosklerosis [34].



Gambar 3.8 Persentase Pasien Gagal Jantung Yang Memiliki Diabetes

Gambar 3.8 merupakan persentase pasien gagal jantung yang memiliki diabetes dan pasien gagal jantung yang tidak memiliki diabetes. Dalam gambar tersebut pasien gagal jantung yang memiliki diabetes sebanyak 41,81% sedangkan sisanya yaitu pasien gagal jantung yang tidak memiliki diabetes sebanyak 58,19%. Tabel dibawah ini menjelaskan rincian jumlah pasien gagal jantung yang memiliki diabetes dan pasien gagal jantung yang tidak memiliki diabetes.



Gambar 3.9 Pasien Yang Memiliki Diabetes

Gambar 3.9 menjelaskan sebaran data pasien gagal jantung yang memiliki diabetes dan pasien gagal jantung yang tidak memiliki diabetes. Terlihat bahwa pasien gagal jantung yang memiliki diabetes berjumlah sedikit daripada pasien gagal jantung yang tidak memiliki diabetes. Berikut ini adalah rincian jumlah pasien yang memiliki diabetes dan yang tidak memiliki diabetes.

Tabel 3.5 Pasien Yang Memiliki Diabetes

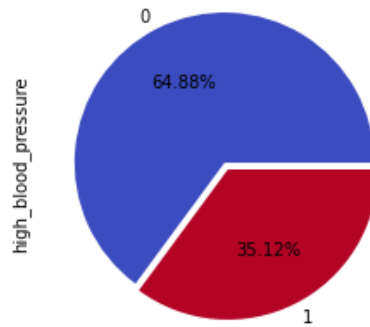
| Diabetes | Jumlah |
|----------|-----------|
| Ya | 125 Orang |
| Tidak | 174 Orang |

Berdasarkan data pada tabel 3.5 diketahui sebanyak 125 orang pasien yang terkena gagal jantung memiliki diabetes, sedangkan sisanya yaitu sebanyak 174 orang pasien yang terkena gagal jantung tidak memiliki diabetes.

F. *High blood pressure* (Tekanan Darah Tinggi)

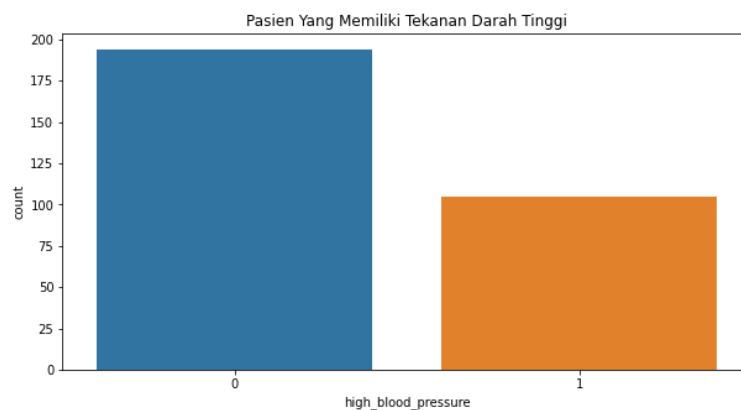
Berdasarkan informasi yang dirilis oleh Dinas Kesehatan Provinsi Kalimantan Barat diketahui bahwa resiko penyakit jantung dan pembuluh darah naik beriringan dengan meningkatnya tekanan darah dalam tubuh. Berdasarkan hasil penelitian membuktikan bahwa tekanan darah sistolik sebesar 130-139 mmHg dan tekanan darah diastolik sebesar 85-89 mmHg akan menyebabkan resiko penyakit jantung dan pembuluh darah 2 kali lipat daripada tekanan darah kurang dari 120 per 80 mmhg [34]. Kemudian Pusat Jantung Nasional juga menjelaskan bahwa tekanan darah diperlukan untuk memompa darah ke seluruh organ tubuh. Apabila tekanan darah tinggi dapat mengakibatkan jantung bekerja

ekstra untuk menyebarkan darah ke seluruh organ tubuh sehingga otot jantung akan semakin tebal. Apabila hal ini terjadi terus menerus akan membebani jantung sehingga jantung tidak sanggup untuk memompa darah. Otot jantung akan melemah dan dapat menjadi kaku. Apabila situasi tersebut tak segera di tindak lanjuti akan terjadi kerusakan pada jantung [35].



Gambar 3.10 Persentase Pasien Gagal Jantung Yang Memiliki Tekanan Darah Tinggi

Gambar 3.10 merupakan persentase pasien gagal jantung yang memiliki tekanan darah tinggi dan pasien gagal jantung yang tidak memiliki tekanan darah tinggi. Diketahui sebanyak 64,88% pasien gagal jantung yang tidak memiliki tekanan darah tinggi dan sisanya sebanyak 35,12% memiliki tekanan darah tinggi. Tabel dibawah ini merupakan rincian jumlah pasien gagal jantung yang memiliki tekanan darah tinggi dan pasien gagal jantung yang tidak memiliki tekanan darah tinggi.



Gambar 3.11 Pasien Yang Memiliki Tekanan Darah Tinggi

Gambar 3.11 menjelaskan mengenai pasien gagal jantung yang memiliki tekanan darah tinggi dan pasien gagal jantung yang tidak memiliki tekanan darah tinggi. Dalam gambar tersebut dapat dilihat bahwa pasien gagal jantung yang memiliki tekanan darah tinggi jumlahnya lebih sedikit daripada pasien gagal jantung yang tidak memiliki tekanan darah tinggi. Tabel dibawah merupakan rincian jumlah pasien gagal jantung yang memiliki tekanan darah tinggi dan pasien gagal jantung yang tidak memiliki tekanan darah tinggi.

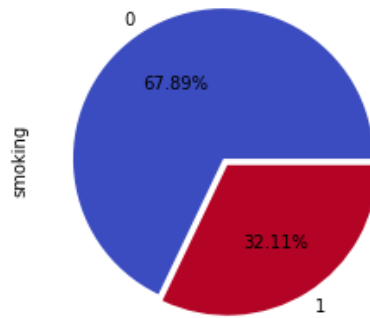
Tabel 3.6 Pasien Yang Memiliki Tekanan Darah Tinggi

| Tekanan Darah Tinggi | Jumlah |
|----------------------|-----------|
| Ya | 105 Orang |
| Tidak | 194 Orang |

Berdasarkan data pada tabel 3.6 diketahui sebanyak 105 orang pasien yang terkena gagal jantung memiliki tekanan darah tinggi, sedangkan sisanya yaitu sebanyak 194 orang pasien yang terkena gagal jantung tidak memiliki tekanan darah tinggi.

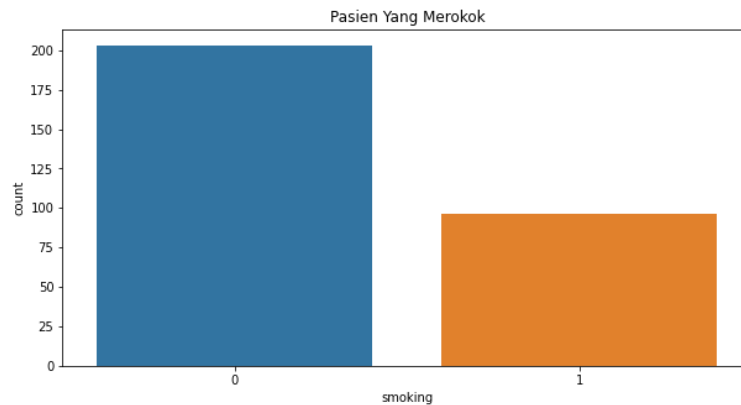
G. *Smoking* (Merokok)

Berdasarkan informasi yang dirilis oleh Dinas Kesehatan Provinsi Kalimantan Barat diketahui bahwa orang yang merokok memiliki resiko 2 hingga 3 kali lipat meninggal akibat penyakit jantung daripada orang yang tak merokok. Merokok dapat mempercepat kerusakan pada dinding pembuluh darah. Karbon monoksida mampu memicu hipoksia jaringan arteri, nikotin mengakibatkan mobilisasi katekolamin yang bisa mengakibatkan rusaknya dinding arteri. Glikoprotein tembakau mampu memicu reaksi hipersensitif pada dinding arteri [34].



Gambar 3.12 Persentase Pasien Gagal Jantung Yang Merokok

Gambar 3.12 merupakan persentase pasien gagal jantung yang merokok dan pasien gagal jantung yang tidak merokok. Dalam gambar tersebut dapat dilihat bahwa pasien gagal jantung yang tidak merokok memiliki persentase 67,89%, sedangkan pasien gagal jantung yang merokok memiliki persentase 32,11%.



Gambar 3.13 Pasien Yang Merokok

Gambar 3.13 merupakan sebaran pasien gagal jantung yang merokok dan pasien gagal jantung yang tidak merokok. Dalam gambar tersebut diketahui bahwa jumlah pasien gagal jantung yang merokok lebih sedikit daripada pasien gagal jantung yang tidak merokok.

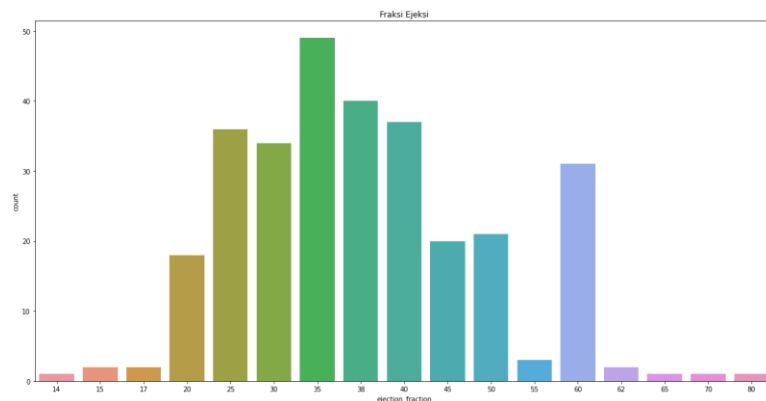
Tabel 3.7 Pasien Yang Merokok

| Merokok | Jumlah |
|---------|-----------|
| Ya | 96 Orang |
| Tidak | 203 Orang |

Berdasarkan data pada tabel 3.7 diketahui sebanyak 96 orang pasien yang terkena gagal jantung merokok, sedangkan sisanya yaitu sebanyak 203 orang pasien yang terkena gagal jantung tidak merokok.

H. *Ejection fraction* (Fraksi Ejeksi)

Berdasarkan informasi yang dirilis oleh Mayo Clinic diketahui bahwa fraksi ejeksi merupakan persentase darah yang keluar dari jantung setiap kali kontraksi. Pada saat jantung melakukan kontraksi maka akan memompa darah keluar dari ventrikel. Pada saat jantung rileks, darah akan masuk ke ventrikel. Fraksi ejeksi normal sekitar 50% hingga 75%, menurut American Heart Association. Batas terendah fraksi ejeksi dapat berkisar antara 41% dan 50% [36].



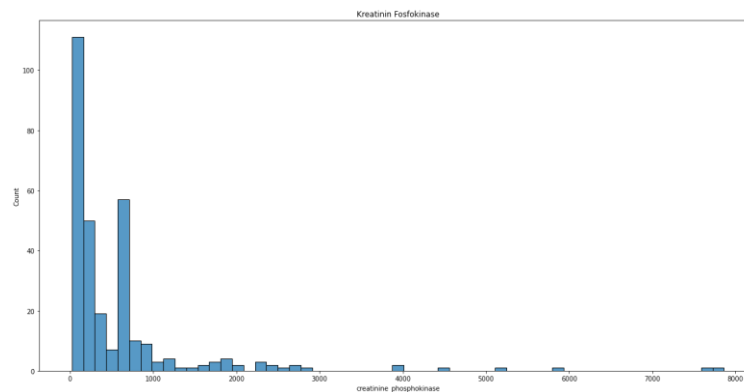
Gambar 3.14 Sebaran Fraksi Ejeksi Pasien Penyakit Gagal Jantung

Dari grafik yang tertera di gambar 3.14 dapat dilihat bahwa pasien dengan fraksi ejeksi sebesar 35% paling banyak mengalami gagal jantung. Sebanyak 49 orang pasien yang memiliki fraksi ejeksi sebesar 35%. Jumlah ini tentunya kurang dari batas terendah persentase fraksi ejeksi.

I. *Creatinine phosphokinase* (Kreatinin Fosfokinase)

Berdasarkan informasi yang dirilis oleh Dr. Fadhli Rizal Makarim pada website halodoc diketahui bahwa *creatinine phosphokinase* (CK) adalah protein yang diketahui sebagai enzim. *Creatinine phosphokinase* diketahui terletak di otot rangka, jantung, dan otak dengan jumlah yang lebih sedikit. Normal apabila CK pada darah memiliki jumlah yang sedikit. Apabila jumlahnya banyak kemungkinan rusaknya atau penyakit pada otot rangka, jantung, dan otak.

Persentase CK normal apabila jumlahnya 10 hingga 129 mikrogram per liter (mcg/L) [37].

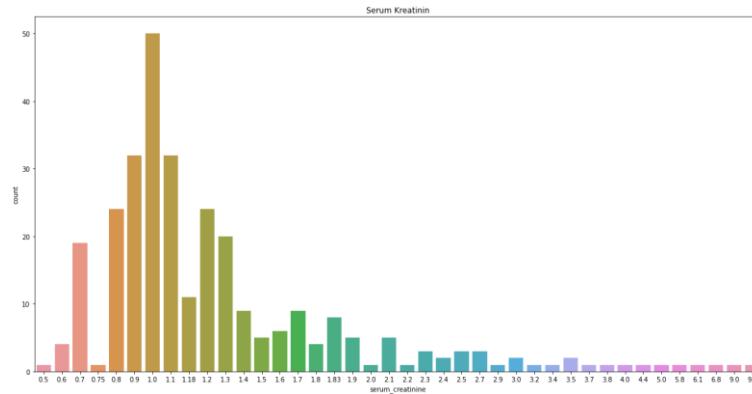


Gambar 3.15 Sebaran *Creatinine phosphokinase* Pasien Penyakit Gagal Jantung

Dari grafik yang tertera di gambar 3.15 dapat dilihat bahwa pasien dengan *creatinine phosphokinase* lebih dari 129 mikrogram per liter (mcg/L) sebanyak 210 orang. Jumlah ini tentunya melebihi batas normal kadar CK dalam darah. Sedangkan 89 orang lainnya memiliki kadar *creatinine phosphokinase* dibawah 129 mikrogram per liter (mcg/L).

J. Serum *creatinine* (Serum Kreatinin)

Berdasarkan informasi yang dirilis oleh Dr. Patricia Lukas Goentoro pada website [hellosehat](#) diketahui bahwa kreatinin merupakan limbah hasil metabolisme otot yang dipakai saat otot berkontraksi. Kreatinin dibentuk oleh kreatin. Kreatin adalah zat utama pada otot yang memiliki fungsi untuk menghasilkan energi. Kreatin harus difilter dulu oleh ginjal sebelum dikeluarkan dari tubuh dalam bentuk urine. Apabila ginjal melakukan kerjanya dengan baik maka kadar serum kreatinin akan sama. Apabila ginjal bermasalah maka persentase kreatinin akan naik sehingga terjadi penumpukan kreatinin dalam darah. Hal ini dapat mengakibatkan seseorang bisa mempunyai tekanan darah tinggi ataupun gagal jantung. Persentase kreatinin bisa dikatakan normal jika sebesar 0,6 - 1,2 mg/dL pada pasien pria dan 0,5 - 1,1 mg/dL pada pasien Wanita [38].

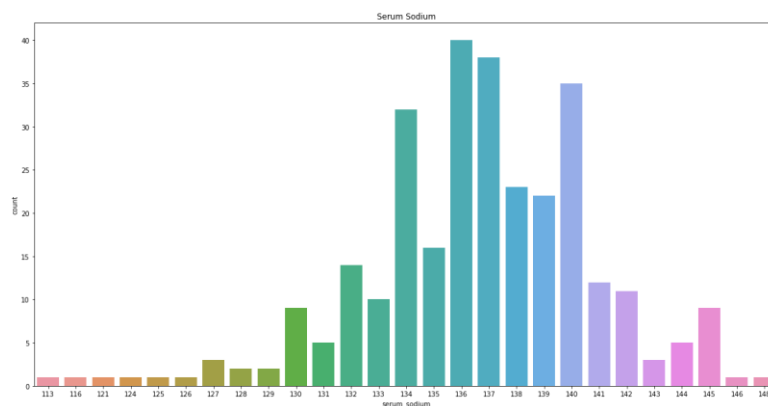


Gambar 3.16 Sebaran *Serum creatinine* Pasien Penyakit Gagal Jantung

Dari grafik yang tertera di gambar 3.16 dapat dilihat bahwa pasien dengan *serum creatinine* lebih dari 0,5 - 1,2 mg/dL sebanyak 101 orang. Sedangkan 198 orang lainnya memiliki *serum creatinine* dibawah 1,2 mg/dL.

K. *Serum sodium*

Berdasarkan informasi yang dirilis oleh Mayo *Clinic* diketahui bahwa sodium berfungsi untuk mengatur tekanan darah agar selalu normal, mensupport kinerja saraf dan otot, dan menyeimbangkan cairan dalam tubuh. Kadar sodium dalam darah secara normal sebesar 135 - 145 miliequivalents per liter (mEq/L). Hiponatremia terjadi pada saat kadar sodium dalam darah sangat sedikit atau di bawah 135 mEq/L [39].

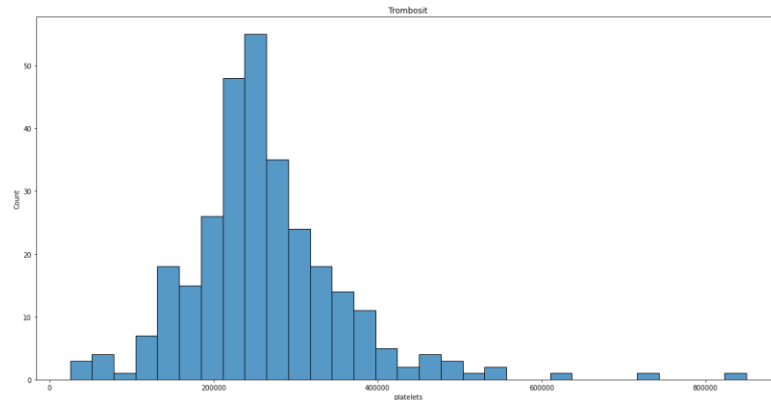


Gambar 3.17 Sebaran *Serum sodium* Pasien Penyakit Gagal Jantung

Dari grafik yang tertera di gambar 3.17 dapat dilihat bahwa pasien dengan *serum sodium* kurang dari 135 - 145 miliequivalents per liter (mEq/L) sebanyak 216 orang. Sedangkan 83 orang lainnya memiliki *serum sodium* dibawah 135 miliequivalents per liter (mEq/L).

L. *Platelets* (Trombosit)

Berdasarkan informasi yang dirilis oleh *Cleveland Clinic* diketahui bahwa trombosit merupakan sel darah yang berfungsi untuk menghentikan pendarahan. Normalnya jumlah trombosit pada orang dewasa sebesar 15000 hingga 45000 trombosit per mikroliter darah. Trombositopenia terjadi apabila sumsum tulang tak cukup membentuk trombosit. Apabila seseorang menderita trombositopenia maka darah yang keluar pada saat terjadi pendarahan akan banyak dan pendarahan akan susah untuk dihentikan. Seseorang dengan trombositopenia berat mempunyai resiko gagal jantung dikarenakan sedikitnya darah yang ke jantung [40].

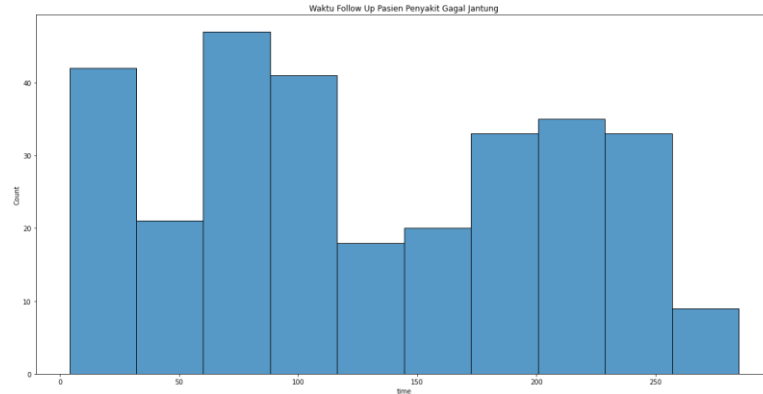


Gambar 3.18 Sebaran Trombosit Pasien Penyakit Gagal Jantung

Dari grafik yang tertera di gambar 3.18 dapat dilihat bahwa pasien dengan trombositopenia sebanyak 27 orang. Sedangkan 272 orang lainnya memiliki jumlah trombosit yang normal.

M. *Time* (Periode Masa Tindak Lanjut)

Berdasarkan informasi yang dirilis oleh Pusat Jantung Nasional diketahui bahwa pokok masalah penanganan komprehensif gagal jantung di masyarakat merupakan *follow up* pasien gagal jantung ketika menjalani perawatan dari dokter dan kepatuhan pasien saat menjalani perawatan gagal jantung yang dianjurkan dokter karena sudah tak mengalami gejala gagal jantung oleh karena itu seringkali pasien melakukan semaunya sendiri untuk menghentikan protokol terapi yang disarankan oleh dokter [41].



Gambar 3.19 Sebaran Periode *Follow Up* Pasien Penyakit Gagal Jantung

Dari grafik yang tertera di gambar 3.19 dapat dilihat bahwa 182 pasien melakukan *follow up* kurang dari 150 kali dan 117 pasien melakukan *follow up* lebih dari 150 kali.

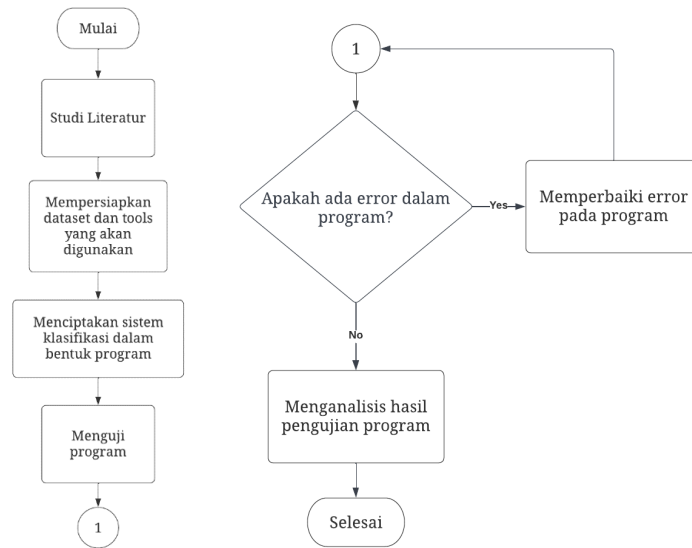
3.2 TOOLS

Tools yang digunakan dibedakan menjadi dua yaitu *software* dan *hardware*. Untuk *software* menggunakan *google colab* yang berfungsi untuk analisa dan visualisasi data dengan bahasa pemrograman *Python*.

Sedangkan *hardware* yang dipakai adalah Laptop Sony Vaio dengan spesifikasi layar 14 inch, intel core i3 3217U @ 1.80GHz, memory RAM 8 GB, dan SSD 128 GB.

3.3 ALUR PENELITIAN

Gambar dibawah ini merupakan alur penelitian dari penelitian ini yang meliputi tahap awal, menengah, dan akhir.



Gambar 3.20 Alur Penelitian

Berikut ini adalah penjelasan lebih lanjut mengenai alur penelitian yang terdapat pada gambar 3.20.

1. Tahap awal: Studi literatur dan mempersiapkan dataset dan *tools* yang akan digunakan.
2. Tahap menengah: Menciptakan sistem klasifikasi dalam bentuk program dan menguji program. Dalam penelitian ini, peneliti membuat skema penggunaan 12 fitur hingga 1 fitur. Berikut adalah rinciannya:
 1. Skema 12 fitur menggunakan fitur *ejection_fraction*, *serum_creatinine*, *age*, *serum_sodium*, *creatinine_phosphokinase*, *platelets*, *time*, *anaemia*, *diabetes*, *high_blood_pressure*, *sex*, *smoking*.
 2. Skema 11 fitur menggunakan fitur *ejection_fraction*, *serum_creatinine*, *age*, *serum_sodium*, *creatinine_phosphokinase*, *platelets*, *time*, *anaemia*, *diabetes*, *sex*, *smoking*.
 3. Skema 10 fitur menggunakan fitur *ejection_fraction*, *serum_creatinine*, *age*, *serum_sodium*, *creatinine_phosphokinase*, *platelets*, *time*, *anaemia*, *sex*, *smoking*.

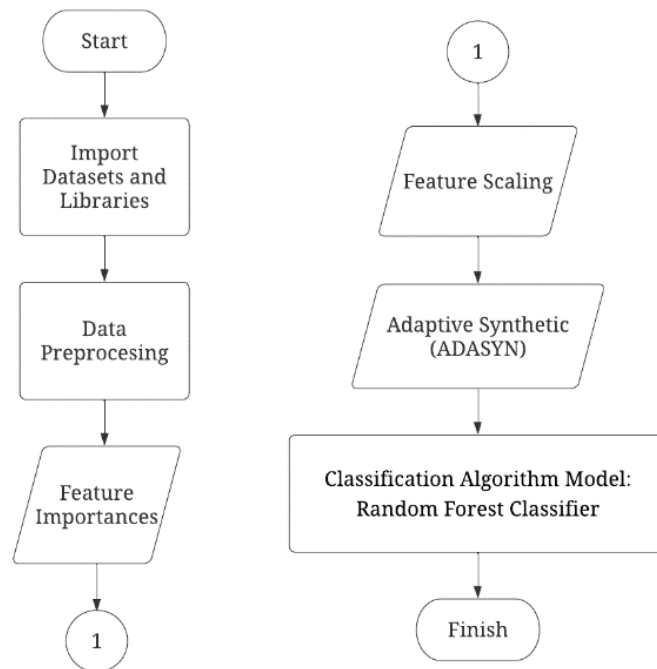
4. Skema 9 fitur menggunakan fitur *ejection_fraction*, *serum_creatinine*, *age*, *serum_sodium*, *creatinine_phosphokinase*, *platelets*, *time*, *anaemia*, *sex*.
5. Skema 8 fitur menggunakan fitur *ejection_fraction*, *serum_creatinine*, *age*, *serum_sodium*, *creatinine_phosphokinase*, *platelets*, *time*, *sex*.
6. Skema 7 fitur menggunakan fitur *ejection_fraction*, *serum_creatinine*, *age*, *serum_sodium*, *creatinine_phosphokinase*, *platelets*, *time*.
7. Skema 6 fitur menggunakan fitur *ejection_fraction*, *serum_creatinine*, *age*, *creatinine_phosphokinase*, *platelets*, *time*.
8. Skema 5 fitur menggunakan fitur *ejection_fraction*, *serum_creatinine*, *age*, *creatinine_phosphokinase*, *time*.
9. Skema 4 fitur menggunakan fitur *ejection_fraction*, *serum_creatinine*, *age*, *time*.
10. Skema 3 fitur menggunakan fitur *ejection_fraction*, *serum_creatinine*, *time*.
11. Skema 2 fitur menggunakan fitur *serum_creatinine*, *time*.
12. Skema 1 fitur menggunakan fitur *time*.

Skema - skema tersebut didapatkan dari *Feature Importance* yang akan dijelaskan di bab 4. Setelah sistem berhasil diciptakan dalam bentuk program maka akan diuji untuk mendapatkan hasil yang mendekati data aslinya. Apabila terdapat *error* maka program akan diperbaiki dan diuji hingga tidak terdapat *error* dan mendapat hasil yang lebih baik. Dan apabila program yang diuji berhasil maka akan di analisis ditahap berikutnya.

3. Tahap akhir: Menganalisis hasil pengujian program.

3.4 ALUR PERANCANGAN PROGRAM

Gambar dibawah ini merupakan alur perancangan program pada penelitian ini yang meliputi *import* dataset dan pustaka, data *preprocessing* yang terdiri dari 3 tahap, dan membangun model *Random Forest*.



Gambar 3.21 Alur Perancangan Program

Berikut ini adalah penjelasan lebih lanjut mengenai alur perancangan program yang terdapat pada gambar 3.21.

1. *Import* Dataset dan Pustaka.

Proses ini bertujuan untuk memasukkan dataset [Heart failure prediction / Kaggle](#) [7] berupa *file* csv dan pustaka yang dibutuhkan untuk proses analisis seperti *pandas*, *matplotlib*, *seaborn*, *sklearn*, dan *imblearn*. *Pandas* merupakan pustaka *open source* yang menyediakan struktur data dan analisis data yang mudah dipakai dan memiliki kinerja yang tinggi untuk bahasa pemrograman *python* [42]. *Matplotlib* adalah *library python* yang berfungsi untuk membuat grafik 2 dimensi dengan kualitas tinggi. Dengan *Matplotlib*, membuat *plots*, *histograms*, *spectra*, *bar charts*, *errorcharts*, *scatterplots*, akan lebih mudah dilakukan [42]. *Seaborn* dibangun di atas *matplotlib* sebagai *library* pondasi utama atau menjadi pelengkap *matplotlib*. *Seaborn* dilengkapi dengan fitur - fitur yang mempermudah visualisasi data [43]. *Scikit-learn* merupakan modul *python*

yang mengintegrasikan berbagai algoritma pembelajaran mesin canggih untuk masalah skala menengah *supervised* dan *unsupervised* [44]. *Imbalanced-learn* atau disebut juga sebagai *imblearn* adalah perpustakaan *open source* yang mengandalkan *scikit-learn* dan menyediakan alat ketika berhadapan dengan klasifikasi dengan kelas yang tidak seimbang [45]. Berikut adalah *scriptnya*:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

from sklearn.preprocessing import MinMaxScaler
from imblearn.over_sampling import ADASYN
from sklearn.model_selection import train_test_split
from sklearn.model_selection import KFold
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import plot_confusion_matrix
from sklearn.metrics import accuracy_score, precision_score, recall_score,
f1_score, classification_report
```

2. Preprocessing Data

Preprocessing adalah proses mengubah data *inputan* menjadi data dengan format yang pas dan siap untuk diproses ke tahap berikutnya. Hal yang dilakukan pada saat *preprocessing* data dalam penelitian ini diantaranya: pengecekan data, *Feature Importances*, *feature Scaling*, dan ADASYN. Tujuan dari *preprocessing* data ialah untuk meningkatkan performansi dan hasil yang hendak dicapai. Berikut ini adalah *script* untuk mengecek data:

```
print(df.shape)
df.head(11)
df.info()
df = df.rename(columns={"DEATH_EVENT": "target"})
df.isnull().sum()
df.duplicated().sum()
df.describe()
```

Berikut ini adalah *script Feature Importances*:

```
X = df.drop(['target'], axis=1)
y = df['target']
from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier()
model.fit(X,y)
print(model.feature_importances_)
feat_importances = pd.Series(model.feature_importances_, index=X.columns)
plt.figure(figsize=(10,5))
feat_importances.nlargest(12).plot(kind='barh')
plt.title("Fitur yang Paling Mempengaruhi Target")
plt.show()
```

Berikut ini adalah *script feature Scaling*:

```
final_df = df[['ejection_fraction', 'serum_creatinine', 'age', 'serum_sodium',
'creatinine_phosphokinase', 'platelets', 'time', 'anaemia', 'diabetes',
'high_blood_pressure', 'sex', 'smoking']]
scaler = MinMaxScaler()
X = scaler.fit(final_df)
X = scaler.transform(final_df)
y = df['target']
X1 = pd.DataFrame(data=X, columns=["smoking", "sex", "anaemia", "diabetes",
"high_blood_pressure", "ejection_fraction", "serum_creatinine", "age",
"serum_sodium", "creatinine_phosphokinase", "platelets", "time"])
X1.head(11)
fig, (ax1, ax2)= plt.subplots(nrows = 1, ncols = 2, figsize = (20,5))
sns.kdeplot(data=df, ax=ax1)
ax1.set_title('Before Scaling')
sns.kdeplot(data=X1, ax=ax2)
ax2.set_title('After Scaling')
fig, (ax1, ax2) = plt.subplots(ncols=2, figsize=(12, 5))
ax1.set_title('Before Scaling')
sns.kdeplot(data=df, x = 'ejection_fraction',ax=ax1)
```

```
ax2.set_title('After Scaling')
sns.kdeplot(data=X1, x = 'ejection_fraction',ax=ax2)
plt.show()
```

Berikut adalah *script* ADASYN:

```
df['target'].value_counts()
kward_params = {'linewidth': 1, 'edgecolor': 'black'}
colors = ['red' if v == 0 else 'green' if v == 1 else 'blue' for v in y]
plt.scatter(X[:, 0], X[:, 1], c=colors, **kward_params)
sns.despine()
plt.suptitle("Before ADASYN")
pass
resample = ADASYN(sampling_strategy='all', random_state=42)
X, y = resample.fit_resample(X,y)
y.value_counts()
kward_params = {'linewidth': 1, 'edgecolor': 'black'}
colors = ['red' if v == 0 else 'green' if v == 1 else 'blue' for v in y]
plt.scatter(X[:, 0], X[:, 1], c=colors, **kward_params)
sns.despine()
plt.suptitle("After ADASYN")
pass
```

3. Classification Algorithm Model

Algoritma *Random Forest* akan mengolah data yang didapatkan dari keluaran ADASYN. Hasil akhir *Random Forest* akan menampilkan nilai akurasi, *precision*, *recall*, dan *f1-score* yang dibantu oleh *K-Fold Cross Validation* dan *Confusion Matrix*. Berikut adalah *script Random Forest*:

```
classifier_rf = RandomForestClassifier()
accuracy = []
for train_index, test_index in kf.split(X, y):
    X_train, X_test = X[train_index], X[test_index]
    y_train, y_test = y[train_index], y[test_index]
    classifier_rf.fit(X_train,y_train)
    y_pred = classifier_rf.predict(X_test)
```

```
ac = accuracy_score(y_test,y_pred)
accuracy.append(ac)
print(accuracy)
```

3.5 PENGUJIAN PROGRAM

1. *K-Fold Cross Validation*

Pada metode *K-Fold Cross Validation* data akan dipisah menjadi dua bagian yaitu data *training*/data latih dan data *testing*/data uji. Total data setelah di *oversampling* akan dibagi sesuai dengan banyaknya lipatan/*fold* yang ditentukan, dalam penelitian ini *fold* berjumlah $10/k = 10$. Berikut adalah *script K-Fold Cross Validation*:

```
kf = KFold(n_splits=10, shuffle=True, random_state=42)
cnt = 1
for train_index, test_index in kf.split(X, y):
    print(f'Fold:{cnt}, Train set: {len(train_index)}, Test set:{len(test_index)}')
    cnt += 1
```

2. *Confusion Matrix*

Hasil dari proses *K-Fold Cross Validation* akan diproses lagi dengan menggunakan *Confusion Matrix*. *Confusion Matrix* berguna menghitung nilai *accuracy*, *precision*, *recall*, dan *f1-score*, serta mengetahui nilai *true positive*, *true negative*, *false positive*, dan *false negative* dari suatu model algoritma yang dievaluasi. Berikut adalah *script* dari *Confusion Matrix*:

```
plot_confusion_matrix(classifier_rf, X_test, y_test)
plot_confusion_matrix(classifier_rf, X_train, y_train)
plot_confusion_matrix(classifier_rf, X, y)
print(classification_report(y_test, y_pred))
```