**MENENTUKAN RANGE HARGA GAWAI BERDASARKAN SPESIFIKASI DENGAN METODE KLASIFIKASI**

**RANDOM FOREST DAN SVM**

**TUGAS AKHIR**

Digunakan Sebagai Syarat Pengerjaan

Tugas Akhir Mata Kuliah *Machine Learning*

**Oleh Kelompok 7 :**

**ARIEF NAUVAN RAMADHA NIM.2041720134**

**BINTANG ADIYATMA A.P. NIM.2041720130**

**MUCHAMAD RIZAL G.A. NIM.2041720015**

**TI-3G**



**POLITEKNIK NEGERI MALANG**

**JURUSAN TEKNOLOGI INFORMASI**

**PROGRAM STUDI D4 TEKNIK INFORMATIKA**

**2022**

# DAFTAR ISI

[DAFTAR ISI ii](#_heading=h.gjdgxs)

[DAFTAR GAMBAR iii](#_heading=h.30j0zll)

[DAFTAR TABEL iv](#_heading=h.1fob9te)

[BAB 1 PENDAHULUAN 1](#_heading=h.3znysh7)

[1.1 Latar Belakang 1](#_heading=h.2et92p0)

[1.2 Rumusan Masalah 2](#_heading=h.tyjcwt)

[1.3 Tujuan](#_heading=h.4d34og8) 2

[1.4 Manfaat](#_heading=h.1ksv4uv) 2

[BAB II METODE PELAKSANAAN](#_heading=h.r0tknzbt0eun) 3

[2.1 Landasan Teori](#_heading=h.m3nk8yrrskrl) 3

[2.1.1. Relasi Harga dan Spesifikasi Handphone](#_heading=h.z337ya) 3

[2.1.2 Metode SVM(Support Vector Machine)](#_heading=h.3j2qqm3) 3

[2.1.3 Metode Random Forest](#_heading=h.4i7ojhp) 5

[2.2 Diagram Alur](#_heading=h.2xcytpi) 7

[2.2.1. Tahapan Pelaksanaan](#_heading=h.1ci93xb) 7

[2.2.2. Diagram Alur SVM](#_heading=h.qsh70q) 7

[2.2.3 Diagram Alur Random Forest](#_heading=h.1pxezwc) 9

[BAB III HASIL DAN PEMBAHASAN](#_heading=h.khion2c9e5zj) 10

[3.1 Dataset](#_heading=h.v9ov9gvwtp5o) 10

[3.2 Perhitungan Manual](#_heading=h.ktc5m685ja4h) 10

[3.2.1 Perhitungan Manual SVM](#_heading=h.hsd5m7ixxx35) 10

[3.2.2 Perhitungan Manual Random Forest 1](#_heading=h.90ztlapyqcd1)7

[3.3 Implementasi Kode Program 2](#_heading=h.flffi71ldnbl)0

[3.3.1 Kode Program Klasifikasi 2](#_heading=h.2bn6wsx)0

[3.3.2 Kode Program Modeling 2](#_heading=h.sa9lsuywg8bw)5

[BAB IV KESIMPULAN DAN SARAN](#_heading=h.2iq8gzs) 28

[4.1 Kesimpulan](#_heading=h.frmvn1jw3pym) 28

[4.2 Saran](#_heading=h.xvir7l) 28

[DAFTAR PUSTAKA 3](#_heading=h.iqek9g4v47pn)0

# DAFTAR GAMBAR

[Gambar 2.1 Diagram Pelaksanaan](#_heading=h.3whwml4) 7

[Gambar 2.2 Diagram Alur SVM](#_heading=h.3as4poj) 8

[Gambar 2.3 Diagram Alur Random Forest](#_heading=h.147n2zr) 9

[Gambar 3.1 Persiapan Data 2](#_heading=h.46r0co2)1

[Gambar 3.2 Check Dataset Info 2](#_heading=h.2lwamvv)1

[Gambar 3.3 Check Dataset Describe 2](#_heading=h.111kx3o)2

[Gambar 3.4](#_heading=h.3l18frh) Visualisasi Data Heatmap Correlation Matrix [23](#_heading=h.3l18frh)

[Gambar 3.5 Drop Column 2](#_heading=h.206ipza)4

[Gambar 3.6 Pembuatan Variabel x dan y 24](#_heading=h.4k668n3)

[Gambar 3.7](#_heading=h.2zbgiuw) Pemisahan Data Test dan Data Train [24](#_heading=h.2zbgiuw)

[Gambar 3.8](#_heading=h.1egqt2p) Pembagian Panjang Data [2](#_heading=h.1egqt2p)5

[Gambar 3.9.](#_heading=h.3ygebqi) Standarisasi [2](#_heading=h.3ygebqi)5

[Gambar 3.10](#_heading=h.2dlolyb) Mencari hypermeter Random Forest Grid Search [2](#_heading=h.2dlolyb)6

[Gambar 3.11](#_heading=h.sqyw64) Mencari Akurasi RandomForest [2](#_heading=h.sqyw64)6

[Gambar 3.12](#_heading=h.3cqmetx) Mencari hypermeter svm gridsearch [2](#_heading=h.3cqmetx)7

[Gambar 3.13](#_heading=h.1rvwp1q) Mencari Akurasi SVM [2](#_heading=h.1rvwp1q)7

# DAFTAR TABEL

[Tabel 3.1](#_heading=h.ihv636) Dataset Mobile Price10

[Tabel 3.2](#_heading=h.32hioqz) Menentukan Nilai Bobot [1](#_heading=h.32hioqz)1

[Tabel 3.3](#_heading=h.41mghml) Menentukan Jumlah dari Range Harga [1](#_heading=h.41mghml)1

[Tabel 3.4](#_heading=h.2grqrue) Encode Price Range [1](#_heading=h.2grqrue)2

[Tabel 3.5](#_heading=h.vx1227) Menentukan Nilai Bobot(w) [1](#_heading=h.vx1227)3

[Tabel 3.6](#_heading=h.3fwokq0) Menentukan Nilai Bias [1](#_heading=h.3fwokq0)4

[Tabel 3.7](#_heading=h.1v1yuxt) Menentukan Hasil Testing [15](#_heading=h.1v1yuxt)

[Tabel 3.8](#_heading=h.4f1mdlm) Membandingkan hasil testing dengan encode price range [16](#_heading=h.4f1mdlm)

[Tabel 3.9](#_heading=h.snbg9of48p41) Menghitung Akurasi dengan Confusion Matrix [1](#_heading=h.snbg9of48p41)6

[Tabel 3.10](#_heading=h.cfao4oiq3lda) Data spesifikasi gawai menurut price range 0 [17](#_heading=h.cfao4oiq3lda)

[Tabel 3.11](#_heading=h.u8stg3sv5kao) Data spesifikasi gawai menurut price range 1 [17](#_heading=h.u8stg3sv5kao)

[Tabel 3.12](#_heading=h.sr4o168id4d0) Data spesifikasi gawai menurut price range 2 [18](#_heading=h.sr4o168id4d0)

[Tabel 3.13](#_heading=h.2u6wntf) Data spesifikasi gawai menurut price range 3 [18](#_heading=h.2u6wntf)

[Tabel 3.14](#_heading=h.19c6y18) Penentuan Root dan Node. [19](#_heading=h.19c6y18)

[Tabel 3.15](#_heading=h.3tbugp1) Price range berdasarkan ram > 1500. [19](#_heading=h.3tbugp1)

[Tabel 3.16](#_heading=h.28h4qwu) Price range berdasarkan ram <= 1500 [19](#_heading=h.28h4qwu)

[Tabel 3.17](#_heading=h.nmf14n) Gini index ram > 1500 [20](#_heading=h.nmf14n)

[Tabel 3.18](#_heading=h.37m2jsg) Gini index ram > 1500 [20](#_heading=h.37m2jsg)

# BAB 1 PENDAHULUAN

## Latar Belakang

Perkembangan teknologi hingga terciptanya gawai tentu saja membawa dampak positif bagi masyarakat. Dengan adanya gawai, siapapun sangat mudah mengakses informasi terkini tanpa perlu membeli koran atau pergi ke perpustakaan. Selain itu komunikasi dengan orang-orang di luar kota maupun luar negeri sekalipun jadi lebih mudah dan mudah. Melepas rindu pada orang tercinta tak butuh waktu lama lewat penggunaan gawai.

Harga adalah suatu nilai tukar, dimana digunakan dalam pemasaran dan bisnis yang paling efektif. Harga menjadikan patokan pertama dalam pembelian dan penjualan. Setiap pelanggan pada saat ingin membeli sebuah ponsel pasti memikirkan spesifikasi ponsel tersebut yang akan dibelinya sesuai dengan estimasi harga yang disiapkan.

*Artificial Intelligence* (kecerdasan buatan) yang membuat mesin mampu menjawab pertanyaan secara cerdas sekarang ini dalam bidang rekayasa yang sangat luas. Machine Learning memberi kita teknik terbaik untuk kecerdasan buatan seperti klasifikasi, regresi, pembelajaran terawasi dan pembelajaran tanpa pengawasan dan banyak lagi. Berbagai jenis algoritma pemilihan fitur tersedia untuk memilih fitur yang terbaik dan meminimalkan kumpulan data. Karena ini adalah masalah pengoptimalan maka banyak teknik yang digunakan untuk mengoptimalkan atau mengurangi dimensi dari dataset. Pada hal ini dibahas banyak fitur yang sangat penting dipertimbangkan untuk memperkirakan harga ponsel. Misalnya prosesor dari ponsel, besaran baterai, ukuran dan ketebalan ponsel juga merupakan faktor penting. Memori internal, kualitas kamera, dan kualitas video harus dipertimbangkan. Sehingga menjadi latar belakang kami yaitu dengan menggunakan dataset untuk memprediksi harga produk yang tersedia dan meluncurkan baru.

Dengan adanya permasalahan diatas, maka diperlukan sebuat sistem cerdas yang mampu mengklasifikasikan gawai menurut range harga dan spesifikasi dari gawai tersebut. Klasifikasi merupakan sebuah proses pengolahan data yang dilakukan untuk menemukan suatu model atau menjelaskan dan membedakan konsep dari kelas data, bertujuan untuk memperkirakan suatu kelas dari objek tertentu yang tidak diketahui kelasnya. Penelitian kali ini metode Support Vector Machine (SVM), yaitu merupakan sistem pembelajaran dengan menggunakan ruang hipotesis yang berupa fungsi-fungsi linear didalam sebuah fitur yang memiliki dimensi tinggi dan dilatih menggunakan algoritma pembelajaran yang berdasarkan teori optimasi.

Terdapat beberapa metode dalam melakukan pengklasifikasian salah satunya adalah metode Random Forest. Dimana Random Forest dapat meningkatkan akurasi karena adanya pemilihan secara acak dalam membangkitkan simpul anak untuk setiap node (simpul diatasnya) dan diakumulasikan hasil klasifikasi dari setiap pohon (tree), kemudian dipilih hasil klasifikasi yang paling banyak muncul.

## Rumusan Masalah

Rumusan masalah pada penelitian ini dapat disusun sebagai berikut: Dari latar belakang diatas, maka dapat dirumuskan dengan membuat beberapa pertanyaan, adapun pertanyaan itu adalah sebagai berikut :

1. Bagaimana menentukan metode Random Forest sesuai untuk permasalahan “Menentukan Range Harga Gawai Berdasarkan Spesifikasi” ?
2. Bagaimana menentukan metode SVM (Support Vector Machine) sesuai untuk permasalahan “Menentukan Range Harga Gawai Berdasarkan Spesifikasi” ?
3. Bagaimana pengguna/*user* mengetahui informasi khusus mengenai produk gawai dengan spesifikasi yang bagus?
4. Bagaimana cara mengklasifikasikan range harga gawai menurut spesifikasi dari gawai tersebut?

## Tujuan

Tujuan yang diharapkan dari penelitian adanya potensi banjir tersebut yaitu :

1. Memastikan bahwa Random Forest dapat meningkatkan akurasi karena adanya pemilihan secara acak dalam membangkitkan simpul anak untuk setiap node (simpul diatasnya) dan diakumulasikan hasil klasifikasi dari setiap pohon (tree), kemudian dipilih hasil klasifikasi yang paling banyak muncul.
2. Menentukan Metode SVM dapat menawarkan akurasi yang lebih tinggi dibandingkan dengan metode klasifikasi lainnya seperti regresi linier berganda ketika satu set data yang sangat besar tersedia, sehingga SVM membuat menangani data dimensi tinggi lebih baik dan menghindari masalah yang kurang pas.
3. Membuat range antara 0-3 dengan melakukan klasifikasi berdasarkan harga gawai dan spesifikasi yang ditawarkan sehingga pengguna/*user* dapat dengan mudah menerima informasi tersebut.
4. Dapat mengklasifikasikan range harga gawai menurut spesifikasi dari gawai.

## Manfaat

Manfaat yang diharapkan dari penelitian menentukan range harga gawai tersebut adalah :

1. Penelitian ini dapat digunakan sebagai cara menentukan jangkauan harga berdasarkan spesifikasi yang dimiliki.
2. Sebagai pemenuhan tugas akhir dari mata kuliah machine learning.

# BAB II METODE PELAKSANAAN

## Landasan Teori

### 2.1.1. Relasi Harga dan Spesifikasi Handphone

Handphone merupakan alat telekomunikasi elektronik bersifat dua arah yang mudah untuk dibawa kemana saja dan mempunyai kemampuan untuk bisa mengirim pesan baik berupa suara, gambar dan informasi. Dalam kehidupan di era sekarang, manusia tentu saja hampir tidak pernah lepas dari yang namanya handphone. Apalagi sekarang dengan semakin berkembangnya teknologi sehingga membuat handphone memiliki berbagai manfaat dan kegunaan sekaligus. Tidak hanya digunakan untuk alat komunikasi.

Dengan banyaknya manfaat yang dimiliki, membuat banyak orang ingin membeli handphone. Karena kebutuhan ini, membuat pasar handphone sangat luas. jika kita lihat pada zaman sekarang, dengan harga murah kita sudah bisa memiliki handphone. Namun jika kita membutuhkan handphone dengan performa atau untuk mencukupi kebutuhan yang kita inginkan, kita harus mengeluarkan uang lebih banyak. Dengan harga yang lebih mahal, kita juga dapat handphone dengan spesifikasi yang tinggi dan sebaliknya. Oleh karena itu perlu mengetahui jangkauan harga untuk menentukan spesifikasi yang dibutuhkan.

### 2.1.2 Metode SVM (Support Vector Machine)

1. **Pengertian**

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 bersama rekannya Bernhard Boser dan Isabelle Guyon. SVM (Support Vector Machine) merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi (seperti Support Vector Classification) dan regresi (Support Vector Regression). Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matang dan lebih jelas secara matematis dibandingkan dengan teknik-teknik klasifikasi lainnya. SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan linier maupun non linear.

1. **Tipe SVM**
2. **SVM Linier**

SVM linear digunakan untuk data yang dapat dipisahkan secara linear, yang berarti jika sebuah dataset dapat diklasifikasi menjadi dua kelas dengan menggunakan sebuah garis lurus tunggal, maka data tersebut disebut sebagai data yang dapat dipisahkan secara linear, dan classifier yang digunakan disebut sebagai Linear SVM classifier.

1. **SVM Non Linear**

SVM non-linear digunakan untuk data yang dapat dipisahkan secara non-linear, yang berarti jika sebuah dataset tidak dapat diklasifikasi menggunakan garis lurus, maka data tersebut disebut data non-linear dan classifier yang digunakan disebut sebagai Non-linear SVM classifier.

1. **Model Persamaan**

Model Persamaan SVM Sederhana adalah seperti berikut ini :

Data pada suatu dataset diberikan variabel 𝑥𝑖 , sedangkan untuk kelas pada dataset diberikan variabel 𝑦𝑖 . Metode SVM membagi dataset menjadi 2 kelas. Kelas pertama yang dipisah oleh hyperplane bernilai 1, sedangkan kelas lainnya bernilai -1.

(1)

(2)

Dimana :

* 𝑋𝑖 = data ke -i
* 𝑊 = niai bobot support vector yang tegak lurus dengan hyperplane
* 𝑏 = niai bias
* 𝑌𝑖 = 𝑘𝑒𝑙𝑎𝑠 𝑑𝑎𝑡𝑎 𝑘𝑒 – 𝑖

Bobot vector (w) adalah garis vektor yang tegak lurus antara titik pusat kordinat dengan garis hyperplane. Bias (b) merupakan kordinat garis relative terhadap titik kordinat. Persamaan (3) merupakan persamaan untuk menghitung nilai b, sedangkan persamaan (4) merupakan persamaan untuk mencari nilai w.

(3)

(4)

Keterangan :

𝑏 = nilai bias

𝑤. 𝑥 + = nilai bobot untuk kelas data positif

𝑤. 𝑥 − = nilai bobot untuk kelas data negatif

𝑤 = bobot vektor

𝛼𝑖 = nilai bobot data ke-i

𝑦𝑖 = kelas data ke-i

𝑥𝑖 = data ke-i

Nilai-nilai a dan b dapat dihitung dengan menggunakan Rumus dibawah ini :

a = (Σy) (Σx²) – (Σx) (Σxy)

. n(Σx²) – (Σx)²

b = n(Σxy) – (Σx) (Σy)

. n(Σx²) – (Σx)²

1. **Urutan langkah - langkah SVM**

Berikut ini adalah Langkah-langkah dalam melakukan Analisis Regresi Linear Sederhana :

1. Tentukan Tujuan dari melakukan Analisis metode SVM.
2. Lakukan Pengumpulan Data
3. Mencari persamaan hyperline
4. Identifikasi nilai bobot(*w*) dan bias(b)
5. Menggunakan eliminasi dan substitusi untuk mencari nilai bobot(*w*) dan bias
6. Melakukan pengujian data dengan memasukan nilai bobot dan bias ke dalam persamaan hyperline
7. mengecek akurasi dengan confusion matrix

### 2.1.3 Metode Random Forest

1. **Pengertian**

Random Forest adalah teknik pembelajaran mesin yang digunakan untuk memecahkan masalah regresi dan klasifikasi. Ini menggunakan pembelajaran ensemble, yang merupakan teknik yang menggabungkan banyak pengklasifikasi untuk memberikan solusi untuk masalah yang kompleks. Metode random forest (RF) merupakan metode yang dapat meningkatkan hasil akurasi, karena dalam membangkitkan simpul anak untuk setiap node dilakukan secara acak. Metode ini digunakan untuk membangun pohon keputusan yang terdiri dari root node, internal node, dan leaf node dengan mengambil atribut dan data secara acak sesuai ketentuan yang diberlakukan. Root node merupakan simpul yang terletak paling atas, atau biasa disebut sebagai akar dari pohon keputusan. Internal node adalah simpul percabangan, dimana node ini mempunyai output minimal dua dan hanya ada satu input. Sedangkan leaf node atau terminal node merupakan simpul terakhir yang hanya memiliki satu input dan tidak mempunyai output.

1. **Tipe Random Forest berdasarkan input**

Metode Random Forest digolongkan menjadi beberapa tipe berdasarkan jenis inputan. Berikut ini penjelasannya.

Random Forest Classifier

Random Forest bagus untuk klasifikasi. Dapat digunakan untuk membuat prediksi kategori dengan beberapa nilai yang mungkin dan dapat dikalibrasi untuk probabilitas output. Satu hal yang perlu diwaspadai adalah overfitting. Random Forest rawan terjadi overfitting, terutama ketika bekerja dengan dataset yang relatif kecil. Perlu di curigai jika model data dapat membuat prediksi yang "terlalu bagus" pada set uji menggunakan Random Forest. Salah satu cara overfitting adalah menggunakan fitur yang benar-benar relevan dalam model data yang digunakan.

Random Forest Regressor

Random Forest dapat digunakan sebagai regresi dengan memperluas 'tree' sepenuhnya sehingga setiap daun memiliki tepat satu nilai. Breiman menyarankan untuk membuat regresi random forest dengan cara memperluas pohon secara acak. Kemudian sebuah prediksi secara sederhana mengembalikan variabel respon individu dari distribusi dapat dibangun jika 'forest' cukup besar. Satu peringatan bahwa perkembangan 'tree' sepenuhnya dapat menutupi atau melebihi kapasitas: jika itu terjadi, intervalnya akan sia-sia, seperti prediksi. Hal yang diharapkan adalah sama seperti akurasi dan presisi.

1. **Model Persamaan**

Model Persamaan

Dimana Y adalah himpunan kasus dan p(c|Y) merupakan proporsi nilai Y terhadap kelas c.

Dimana Values (a) merupakan semua nilai yang mungkin dalam himpunan kasus a. Yv adalah subkelas dari Y dengan kelas v yang berhubungan dengan kelas a. Ya adalah semua nilai yang sesuai dengan a.

1. **Langkah - langkah**

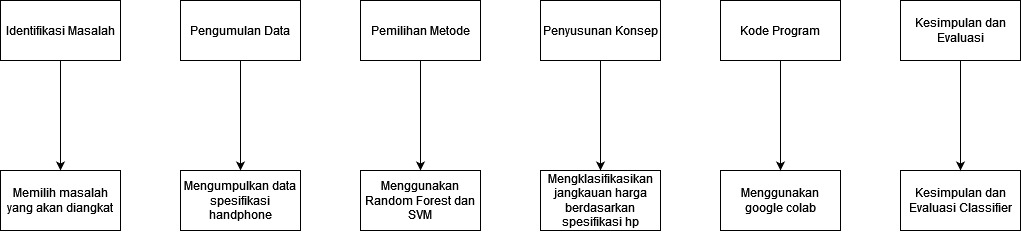
Random Forest bekerja dalam dua fase pertama adalah membuat hutan acak dengan menggabungkan N pohon keputusan, dan kedua membuat prediksi untuk setiap pohon yang dibuat pada fase pertama. Proses Kerja dapat dijelaskan dalam langkah-langkah dan diagram di bawah ini:

* Langkah-1: Pilih titik data K acak dari set pelatihan.
* Langkah-2: Bangun pohon keputusan yang terkait dengan titik data yang dipilih (Subset).
* Langkah-3: Pilih angka N untuk pohon keputusan yang ingin Anda bangun.
* Langkah-4: Ulangi Langkah 1 & 2.
* Langkah-5: Untuk poin data baru, temukan prediksi dari setiap pohon keputusan, dan ketetapan poin data baru ke kategori yang memenangkan suara mayoritas.

## 2.2 Diagram Alur

### 2.2.1. Tahapan Pelaksanaan

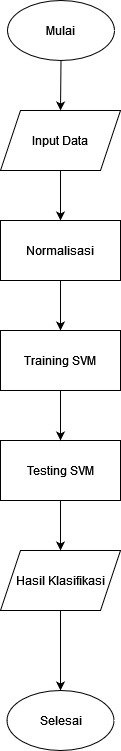
Tahapan pelaksanaan dari penelitian ini memuat enam tahapan dimulai dari identifikasi masalah hingga tahap pengujian dan evaluasi yang digambarkan pada diagram berikut ini :



Gambar 2.1. Diagram Pelaksanaan

### 2.2.2. Diagram Alur SVM

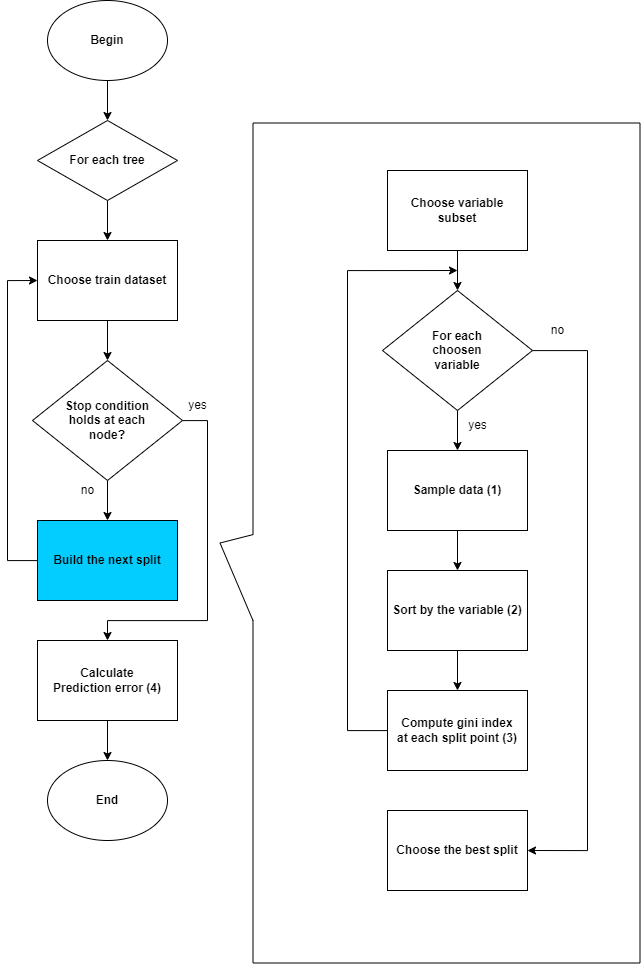
Berikut merupakan alur dari metode regresi linier sederhana.



Gambar 2.2. Diagram Alur SVM

Berdasarkan Gambar 2.2, tahap pertama yaitu memasukkan dataset dengan format .xls yang mana data dari penelitian ini berupa data mobile price classification yang berjumlah 2000 data. Data tersebut berasal dari kaggle. Tahap kedua yaitu proses training SVM yang terdiri dari matriks Hessian yang hasilnya akan digunakan untuk perhitungan berikutnya yaitu menghitung nilai δαi dan nilai αi. Selanjutnya menghitung sequential training yaitu melakukan perhitungan terhadap testing SVM, langkah pertama dalam proses testing adalah menghitung nilai bias, kemudian menghitung nilai K(xi , xtest) yaitu dengan melihat nilai terbesar pada kelas positif dan negatif dengan menggunakan fungsi kernel yang digunakan. Setelah nilai K(xi , xtest) didapatkan langkah selanjutnya adalah menghitung nilai f (x)test , kemudian jika sudah didapat nilai f (x)test maka sudah dapat diklasifikasikan berdasarkan hasil dari nilai f(x)test tersebut. Jika nilai 0 maka data tersebut masuk pada jangkauan harga murah, nilai 1 maka data tersebut masuk pada jangkauan harga normal, nilai 2 maka data tersebut masuk pada jangkauan harga mahal sedangkan jika bernilai 3 maka data tersebut masuk pada jangkauan harga sangat mahal.

### 2.2.3 Diagram Random Forest



Gambar 2.3. Diagram Alur Random Forest

Pertama - tama menentukan data training yang diambil dari dataset dan menentukan apakah data tersebut berbentuk numerik atau tidak. Jika tidak, maka pada data training dilakukan penghitungan probabilitas dan dibuat tabel peluangnya. Jika ya, maka pada data training menghitung rata-rata dalam satu parameter dan menghitung standar deviasi dalam parameter. Setelah itu hasilnya dimasukkan ke dalam tabel peluang.

# BAB III HASIL DAN PEMBAHASAN

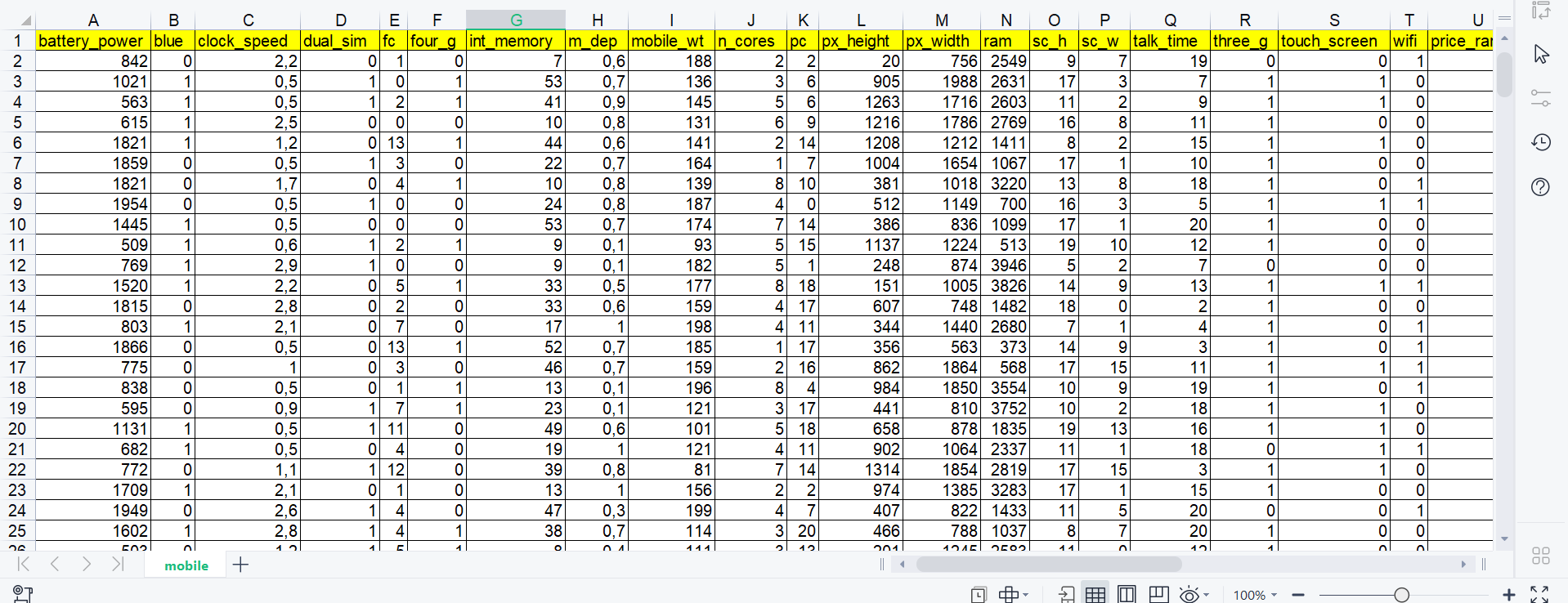
## 3.1 Dataset

Dataset yang digunakan pada penelitian ini terdiri dari 21 kolom, yaitu :

1. Battery Power (Total dalam mAh)
2. Bluetooth (dengan Ya = 1, Tidak = 0)
3. Clock Speed (kecepatan microprocessor)
4. Dual Sim (dengan Ya = 1, Tidak = 0)
5. Front Camera (megapixels)
6. 4G (dengan Ya = 1, Tidak = 0)
7. Internal Memory (Gigabytes)
8. Mobile Depth
9. Mobile Weight
10. Number of cores of processor
11. Primary Camera (megapixels)
12. Pixel Resolution Height
13. Pixel Resolution Width
14. RAM
15. Screen Height
16. Screen Width
17. Talk Time
18. 3G (dengan Ya = 1, Tidak = 0)
19. Touch Screen (dengan Ya = 1, Tidak = 0)
20. Wifi (dengan Ya = 1, Tidak = 0)
21. Price Range

Dalam dataset tersebut dapat dilihat pada tabel berikut :

Tabel 3.1 Dataset Mobile Price



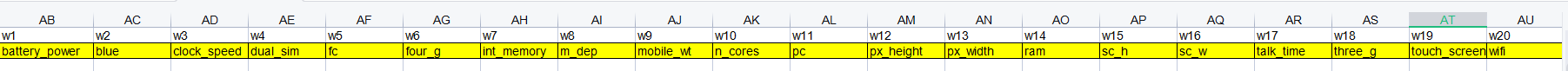
## 3.2 Perhitungan Manual

### 3.2.1 Perhitungan Manual SVM

1. **Menganalisis Nilai Bobot (*W*)**

Dalam penggunaan method SVM, harus menggunakan bobot(*w*) yang mana diambil dari fitur dalam dataset yang telah dibuat seperti berikut ini :

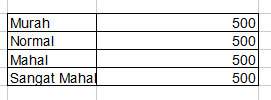
Tabel 3.2 Menentukan Nilai Bobot



Karena dataset yang digunakan memiliki dua puluh kolom maka fitur yang dimiliki pun terdiri dari 20 nilai bobot. Nilai bobot ini tidak tergantung pada jumlah kolomnya saja melainkan relasi kolom satu dengan lainnya. Nilai bobot disini masih belum memiliki nilai, nantinya nilai bobot ini akan terisi dengan menggunakan subtitusi dan eliminasi.

Berdasarkan hasil analisis di atas, terdapat dua puluh kolom yaitu berisi tentang spesifikasi handphone seperti battery power, front camera dan lain sebagainya. Berdasarkan hasil analisis dapat diketahui bahwa jumlah dari masing-masing range harga sama yaitu berjumlah 500.

Tabel 3.3 Menentukan Jumlah dari Range Harga

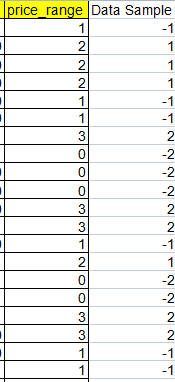


1. **Menganalisa Bias (*b*)**

Untuk melakukan perhitungan SVM, selain memerlukan nilai bobot(*w*) perlu juga menentukan nilai dari bias. Karena pada saat nantinya dibuat, nilai bias dan nilai bobot akan dibuat sistem persamaan linear dimana y akan menjadi nilai dari klasifikasi dataset. jika melihat dari dataset, nilai y memiliki empat nilai yaitu 0,1,2 dan 3. Dimana nilai-nilai tersebut mengartikan klasifikasi jangkauan harga handphone. Keempat nilai ini akan dijadikan data sampel sebagai berikut :

* 0 akan menjadi -2
* 1 akan menjadi -1
* 2 akan menjadi 1
* 3 akan menjadi 2

Tabel 3.4 Encode Price Range



Penggunaan data sampel, dimaksudkan agar memudahkan dalam melakukan perhitungan. Penentuan nilai bias disini masih hanya sekedar menentukan negatif atau positif nilai bias tersebut, contoh jika data sampel bernilai negatif satu maka pada sistem persamaan linear, nilai b akan bernilai negatif dan sebaliknya. Nantinya dalam menentukan nilai bias menggunakan substitusi dan eliminasi.

1. **Menghitung Persamaan dan menentukan nilai bobot dan bias**

Dengan menggunakan persamaan dalam melakukan perhitungan SVM ini didapatkan data sampel dua puluh persamaan yang mana didapat dari unsur sistem persamaan linear yang harus memiliki jumlah matriks yang sama dalam melakukan perhitungannya. Karena pada dataset memiliki dua puluh kolom maka harus berjumlah dua puluh juga pada barisnya. Adapun beberapa contoh persamaan yang didapat dari dataset yaitu :

* w1.x1+w2.x2+w3.x3+w4.x4+w5.x5+w6.x6+w7.x7+w8.x8+w9.x9+w10.x10+w11.x11+w12.x12+w13.x13+w14.x14+w15.x15+w16.x16+w17.x17+w18.x18+w19.x19+w20.x20 -b > -1
* w1.x1+w2.x2+w3.x3+w4.x4+w5.x5+w6.x6+w7.x7+w8.x8+w9.x9+w10.x10+w11.x11+w12.x12+w13.x13+w14.x14+w15.x15+w16.x16+w17.x17+w18.x18+w19.x19+w20.x20 -b > -1

Nilai x disini merupakan nilai dari dataset. Jadi jika nilai dari battery

power adalah 2314 maka dapat dituliskan menjadi (2314)*w*1. Dengan

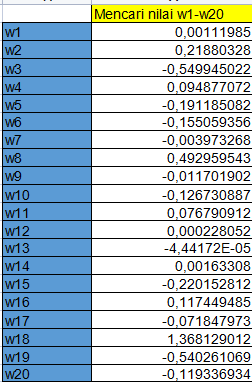
melakukan substitusi dan eliminasi didapatkan nilai *w*1-*w*20 adalah sebagai

berikut :

Rumus untuk menentukan nilai w1-w20

=MMULT(MINVERSE(A2:T21);V2:V21), melakukan eliminasi matriks 20x20 agar mendapat nilai bobot(*w*)

Tabel 3.5 Menentukan Nilai Bobot(w)

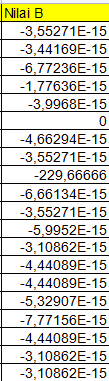


Setelah mendapatkan nilai bobot (*w*), selanjutnya mencari nilai bias dengan cara substitusi nilai bobot ke dua puluh persamaan sehingga mendapat bias sebagai berikut :

Rumus untuk menentukan nilai bias

=((A2\*Y2)+(B2\*Y3)+(C2\*Y4)+(D2\*Y5)+(E2\*Y6)+(F2\*Y7)+(G2\*Y8)+(H2\*Y9)+(I2\*Y10)+(J2\*Y11)+(K2\*Y12)+(L2\*Y13)+(M2\*Y14)+(N2\*Y15)+(O2\*Y16)+(P2\*Y17)+(Q2\*Y18)+(R2\*Y19)+(S2\*Y20)+(T2\*Y21))+1 , melakukan substitusi nilai bobot(*w*) ke dalam persamaan hyperline, nilai +1 didapat dari persamaan hyperline yang pertama dimana memiliki nilai = -1. Karena dipindah ruas maka nilai -1 menjadi +1. Konsep ini diterapkan disemua persamaan hyperline untuk mencari nilai bias.

Tabel 3.6 Menentukan Nilai Bias

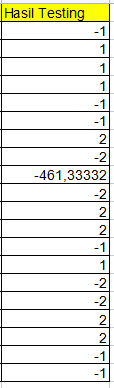


Setelah mendapatkan nilai bobot serta bias, masukan semua nilai tersebut ke dalam persamaan sehingga menghasilkan hasil tes sebagai berikut :

Rumus untuk menentukan nilai w1-w20

=((A2\*Y2)+(B2\*Y3)+(C2\*Y4)+(D2\*Y5)+(E2\*Y6)+(F2\*Y7)+(G2\*Y8)+(H2\*Y9)+(I2\*Y10)+(J2\*Y11)+(K2\*Y12)+(L2\*Y13)+(M2\*Y14)+(N2\*Y15)+(O2\*Y16)+(P2\*Y17)+(Q2\*Y18)+(R2\*Y19)+(S2\*Y20)+(T2\*Y21))+Z2, Menjumlahkan semua nilai yang sudah disubtitusikan ke dalam persamaan hyperline.

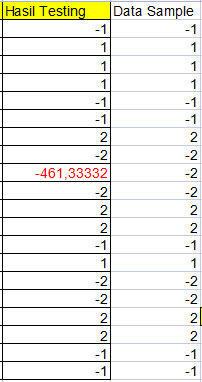
Tabel 3.7 Menentukan Hasil Testing



1. **Menganalisa dan Menghitung Akurasi**

Dari perhitungan manual yang sudah dilakukan, ternyata terdapat perbedaan hasil dari data model dengan hasil testing sebagai berikut :

Tabel 3.8 Membandingkan hasil testing dengan encode price range

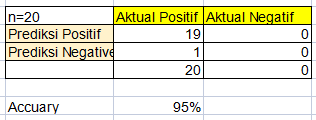


Untuk menghitung akurasi yang didapatkan, pada perhitungan manual ini menggunakan confusion matrix akurasi sebagai berikut :

Rumus untuk menentukan akurasi confusion matrix

=(AB25+AC26)/(AB25+AC25+AB26+AC26)

Tabel 3.9 Menghitung Akurasi dengan Confusion Matrix



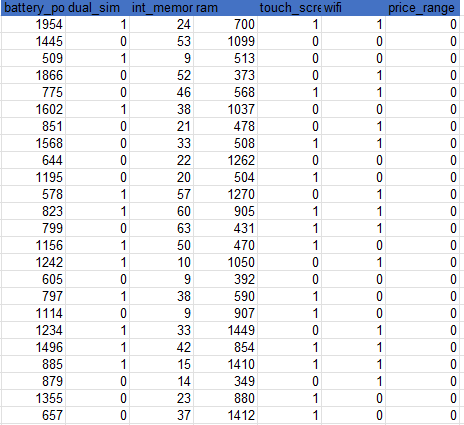
Dengan menggunakan confusion matrix dapat diambil akurasi sebesar 95% berdasarkan 20 sampel yang digunakan.

### 3.2.2 Perhitungan Manual Random Forest

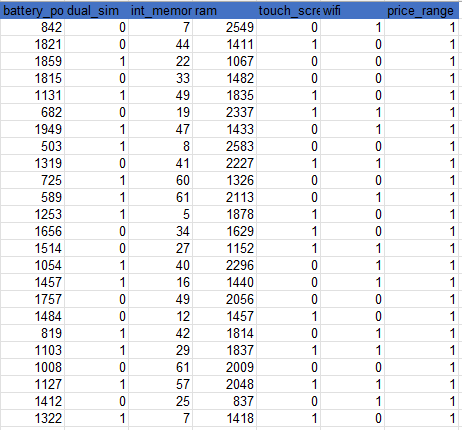
1. **Menentukan filtering dataset**

Data dilakukan filtering berdasarkan range harga antara 0-3, yang memiliki tujuan untuk untuk mempermudah dalam membaca dataset dan melakukan perhitungan manual dari data yang berjumlah 2000 baris.

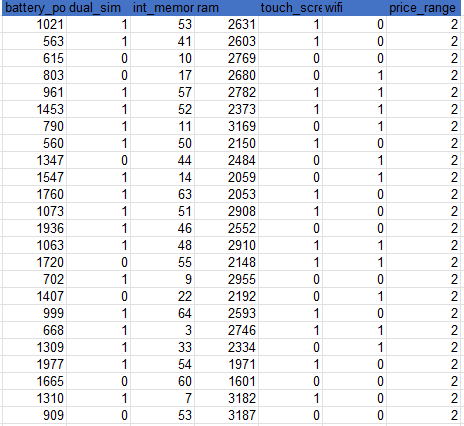
Tabel 3.10 Data spesifikasi gawai menurut price range 0



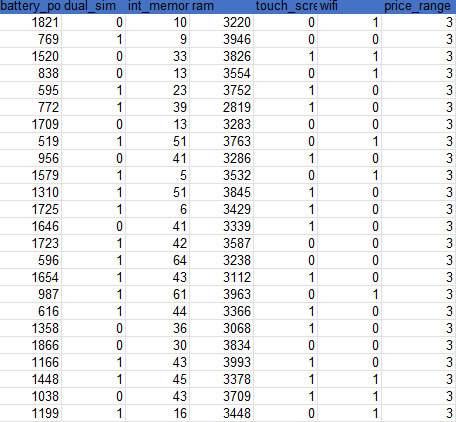
Tabel 3.11 Data spesifikasi gawai menurut price range 1



Tabel 3.12 Data spesifikasi gawai menurut price range 2



Tabel 3.13 Data spesifikasi gawai menurut price range 3



1. **Menentukan nilai pembanding**

Pada tahap ini adalah menentukan pembanding yang digunakan untuk menjadi acuan dari range harga gawai, yang dimana menggunakan kolom ram sebagai pembanding range harga. Kemudian dari kolom pembanding tersebut dilakukan sortir menjadi 2 yaitu ram > 1500 dan ram <= 1500.

Tabel 3.14 Penentuan Root dan Node.



1. **Menghitung jumlah data berdasarkan klasifikasi**

Setelah dilakukan perbandingan price range berdasarkan ram maka dilakukan klasifikasi berdasarkan dataset untuk ram dengan besaran kurang dari sama dengan 1500 dan ram dengan besaran lebih dari 1500.

Tabel 3.15 Price range berdasarkan ram > 1500.



* Rumus mencari nilai probabilitas range 0 kondisi “ram > 1500” : =COUNTIF(Sheet1!L2:L501;">1500"), dimana menggunakan dataset yang sudah dilakukan filter pada langkah sebelumnya .
* Rumus mencari nilai probabilitas range 1 kondisi “ram > 1500” : =COUNTIF(Sheet1!T2:T502;">1500"), dimana menggunakan dataset yang sudah dilakukan filter pada langkah sebelumnya .
* Rumus mencari nilai probabilitas range 2 kondisi “ram > 1500” : =COUNTIF(Sheet1!AB2:AB502;">1500"), dimana menggunakan dataset yang sudah dilakukan filter pada langkah sebelumnya .
* Rumus mencari nilai probabilitas range 3 kondisi “ram > 1500” : =COUNTIF(Sheet1!D2:D502;">1500"), dimana menggunakan dataset yang sudah dilakukan filter pada langkah sebelumnya .

Tabel 3.16 Price range berdasarkan ram <= 1500.

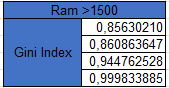


* Rumus mencari nilai probabilitas range 0 kondisi “ram <= 1500” : =COUNTIF(Sheet1!L2:L501;">1500"), dimana menggunakan dataset yang sudah dilakukan filter pada langkah sebelumnya .
* Rumus mencari nilai probabilitas range 1 kondisi “ram <= 1500” : =COUNTIF(Sheet1!L2:L501;">1500"), dimana menggunakan dataset yang sudah dilakukan filter pada langkah sebelumnya .
* Rumus mencari nilai probabilitas range 2 kondisi “ram <= 1500” : =COUNTIF(Sheet1!AB2:AB502;"<=1500"), dimana menggunakan dataset yang sudah dilakukan filter pada langkah sebelumnya .
* Rumus mencari nilai probabilitas range 3 kondisi “ram <= 1500” : =COUNTIF(Sheet1!D2:D502;"<=1500"), dimana menggunakan dataset yang sudah dilakukan filter pada langkah sebelumnya .

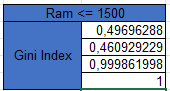
1. **Menghitung Gini Index**

Pada tahap selanjutnya yaitu dengan cara perhitungan *gini index* yang dihitung dengan mengurangkan jumlah probabilitas kuadrat dari setiap kelas dari satu. Ini lebih menyukai partisi yang lebih besar dan sangat mudah diterapkan. Sederhananya, ini menghitung probabilitas fitur tertentu yang dipilih secara acak yang diklasifikasikan secara tidak benar. Maka dari perhitungan yang telah dilakukan diperoleh *gini index* seperti berikut:

Tabel 3.17 Gini index ram > 1500.



Tabel 3.18 Gini index ram > 1500.



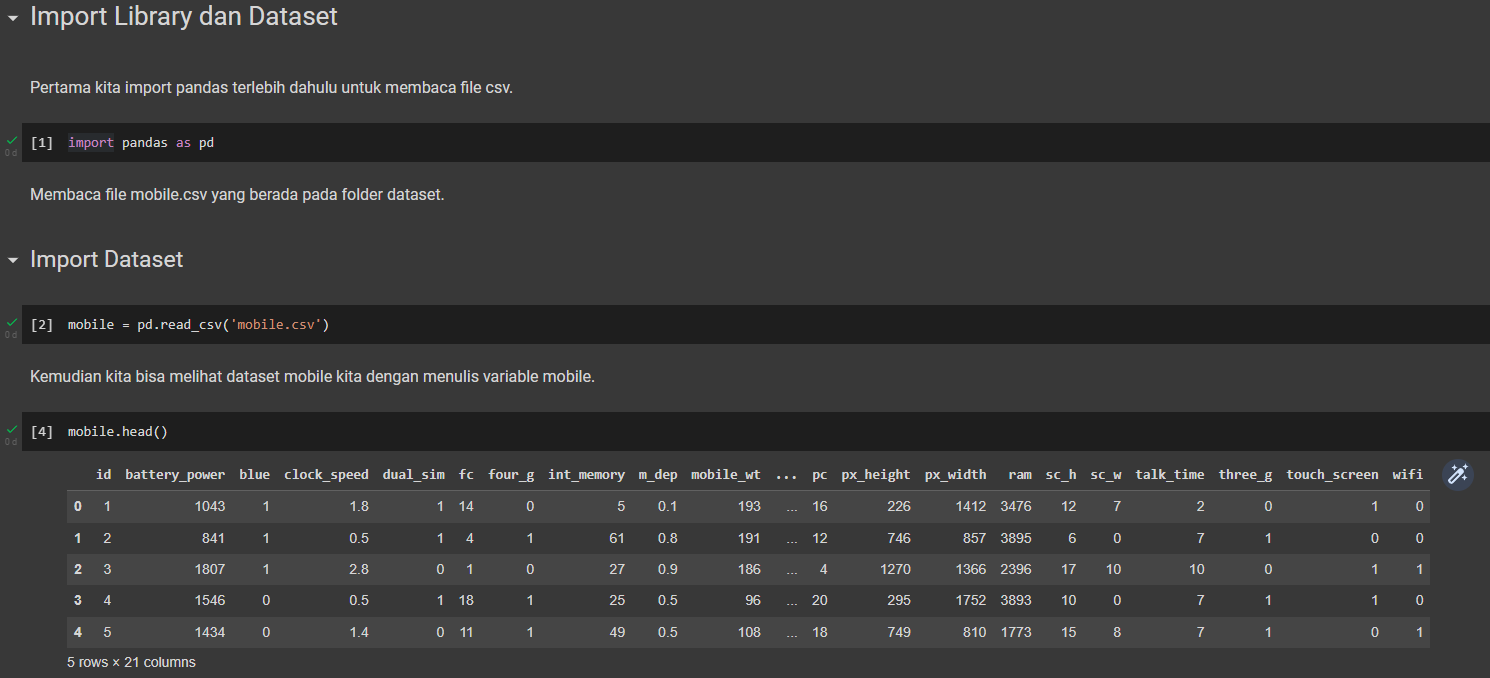
1. **Hasil**

Dari perhitungan manual yang sudah dilakukan dengan menggunakan metode random forest maka didapatkan bahwa hasil dari perhitungan Weighted Gini Index dari node kiri (ram > 1500) adalah 0,881 dan node kanan (ram <= 1500) adalah 0,493.

## 3.3 Implementasi Kode Program

### 3.3.1 Kode Program Klasifikasi

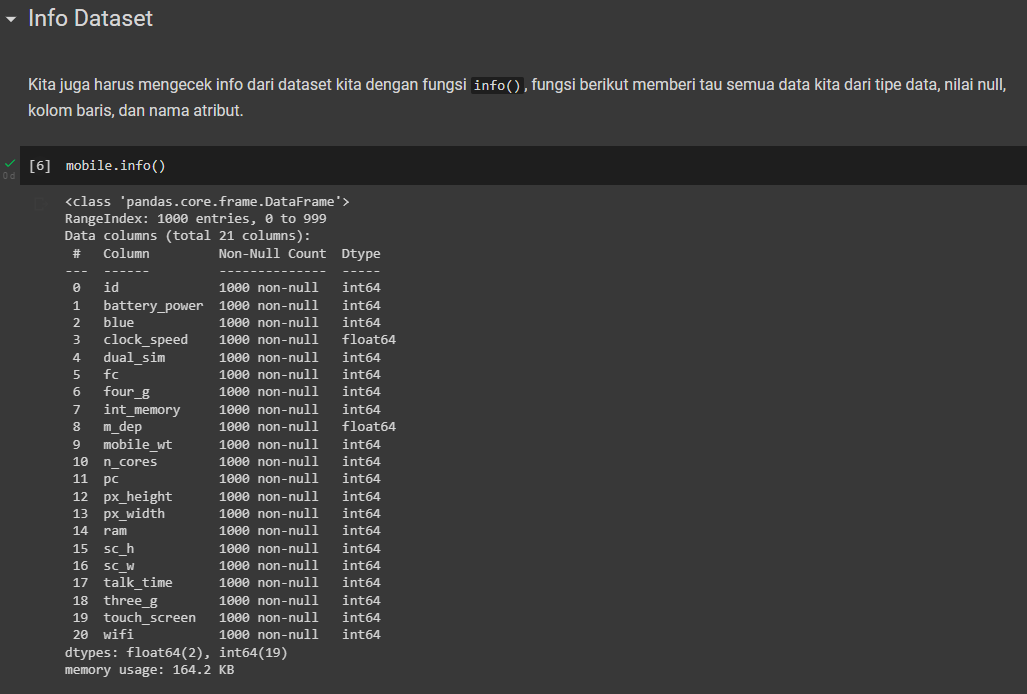
1. Mempersiapkan Data



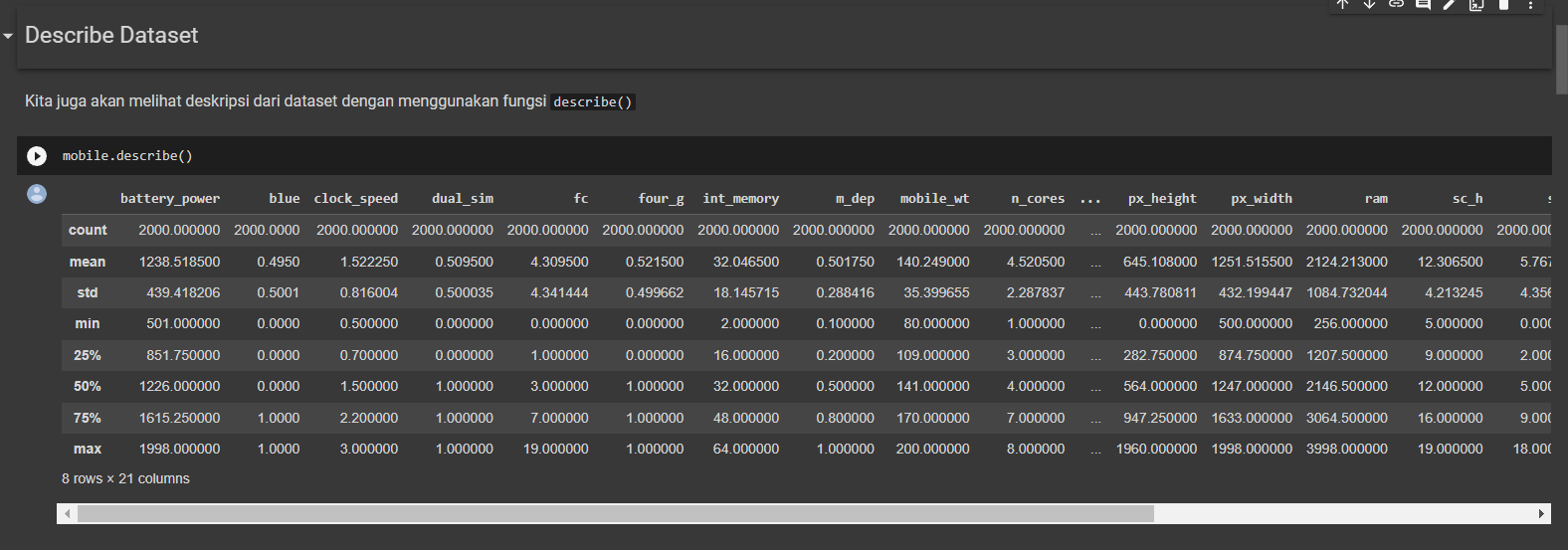
Gambar 3.1 Persiapan Data

Melakukan import library dan dataset “mobile.csv”. Lalu cek data dengan menampilkan 5 urutan teratas dataset dengan menggunakan *.head()* . import library pandas digunakan untuk membaca sebuah data seperti .csv.

1. Memahami Data



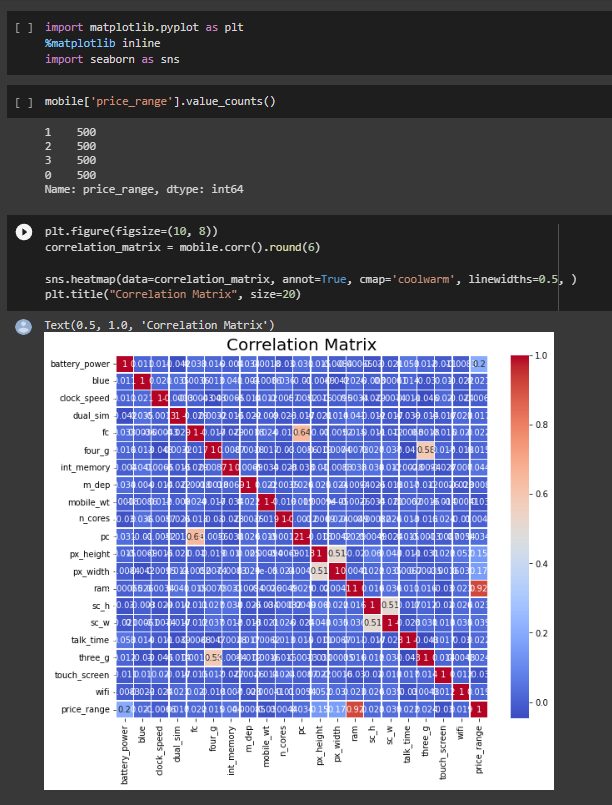
Gambar 3.2 Check Dataset Info



Gambar 3.3 Check Dataset Describe

Memahami dataset yang tersedia dengan mengecek info data dan deskripsi data.

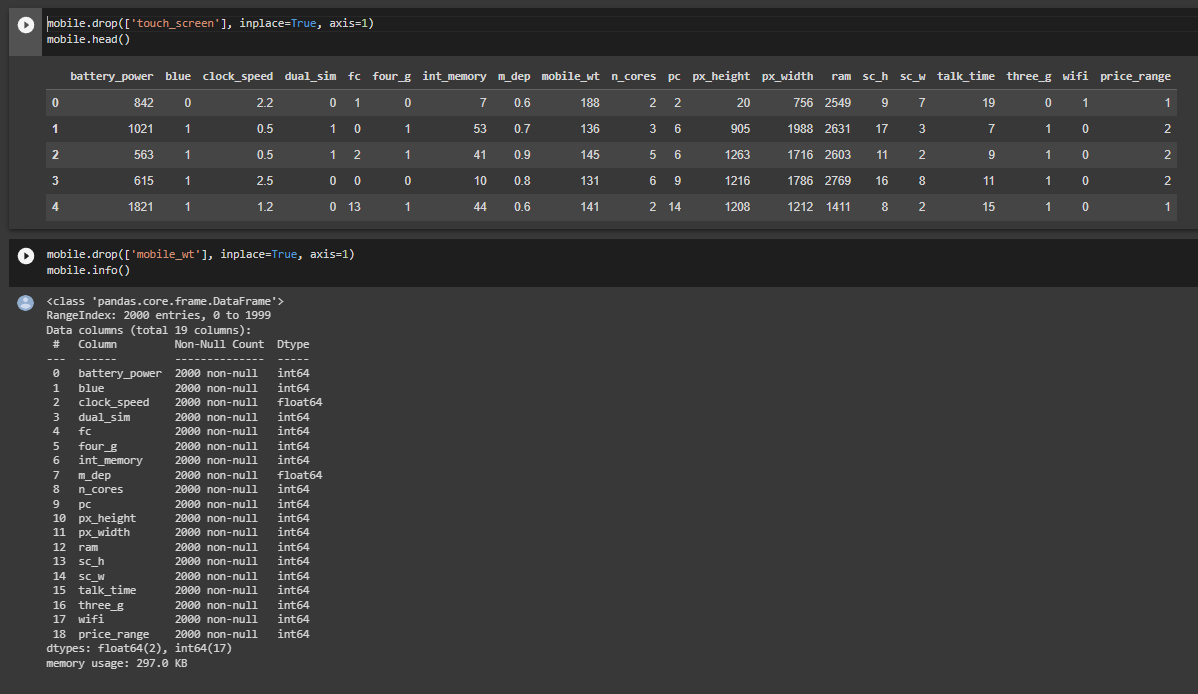
1. Visualisasi Korelasi Data Antar Kolom Dengan Correlation Matrix



Gambar 3.4 Visualisasi Data Heatmap Correlation Matrix

Pada visualisasi heatmap tersebut, spesifikasi handphone dengan korelasi kuat yaitu RAM dengan nilai 0,92. sedangkan import library matplotlib digunakan untuk melakukan visualisasi data seperti membuat plot grafik untuk satu sumbu atau lebih. Setiap sumbu memiliki sumbu horizontal (x) dan sumbu vertikal (y).

1. Melakukan drop column



Gambar 3.5 Drop Column

Melakukan Drop column “touch\_screen” dan column “mobile\_wt”. Karena memiliki korelasi paling kecil di antara yang lain. Karena ingin memaksimalkan proses train dengan data yang powerfull, data yang memiliki sangat sedikit korelasi bisa mempengaruhi proses train.

1. Tahapan split data train dan test
2. Membuat variabel x dan y



Gambar 3.6 Pembuatan Variabel x dan y

Membuat variabel x dan y berdasarkan analisis korelasi dari kegiatan sebelumnya yaitu :

x : berisi semua data kecuali data pada kolom price\_range

y : data pada kolom price\_range

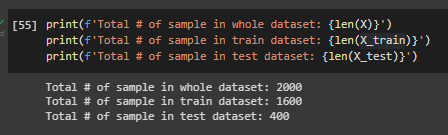
1. Pemisahan Data Testing dengan Data Training



Gambar 3.7 Pemisahan Data Test dan Data Train

Fungsi dari import library train\_test\_split diatas digunakan untuk memisahkan data train dan data test dengan proporsi data test : 0,2. dimana digunakan untuk mengevaluasi performa model data test dan data train.

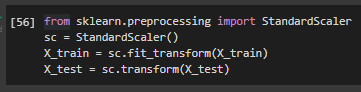
1. Pembagian data x\_train dan x\_test



Gambar 3.8 Pembagian Panjang Data

Berikut merupakan pembagian data dengan x\_train dan x\_test. Dengan hasil di atas kita bisa menyimpulkan, bahwa :

1. data train = 1600
2. data test = 400
3. Standarisasi

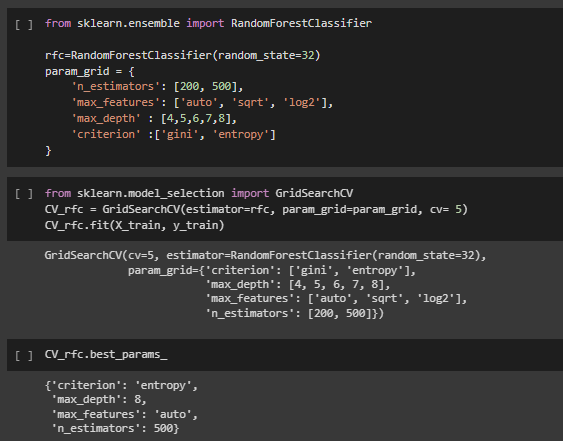


Gambar 3.9 Standarisasi

Kegunaan dari import library StandardScaller untuk melakukan Standarisasi agar data memiliki rentang yang sama antara 0 dan 1. Ini juga akan membantu membantu pada saat proses train nanti.

### 3.3.2 Kode Program Modeling

1. **Random Forest**
2. Mencari hypermeter random forest terbaik menggunakan grid search.

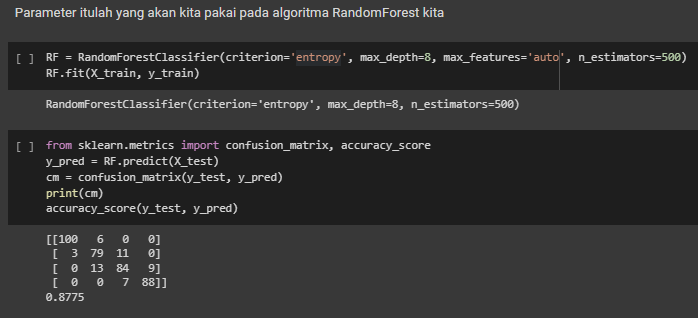


Gambar 3.10 Mencari hypermeter Random Forest Grid Search

Melakukan import library dan dataset “mobile.csv”. Setelah melakukan GridSearch mendapatkan hasil :

* criterion : ‘entropy’
* max\_depth : 8
* max\_feature : auto
* n\_estimator : 500

1. Mencari akurasi Random Forest

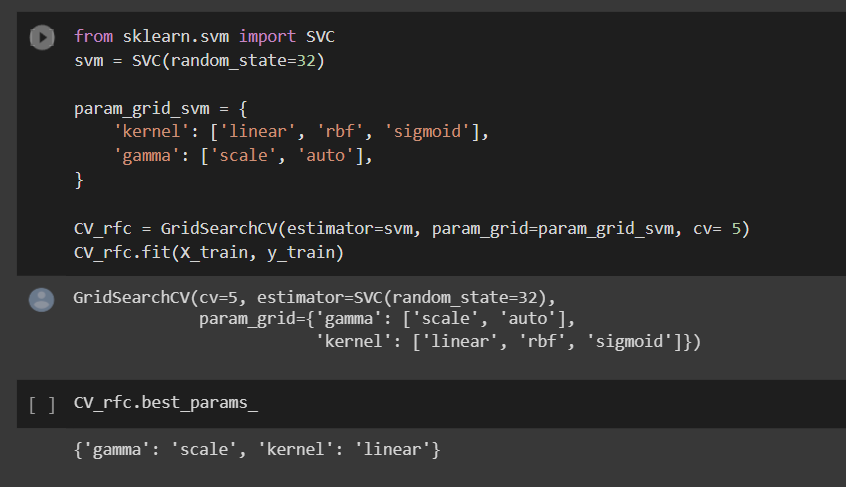


Gambar 3.11 Mencari Akurasi RandomForest

Dengan RandomForest kita mendapatkan akurasi sekitar 87% yang mana terbilang cukup baik.

**B. SVM**

1. Mencari hypermeter svm terbaik menggunakan gridsearch

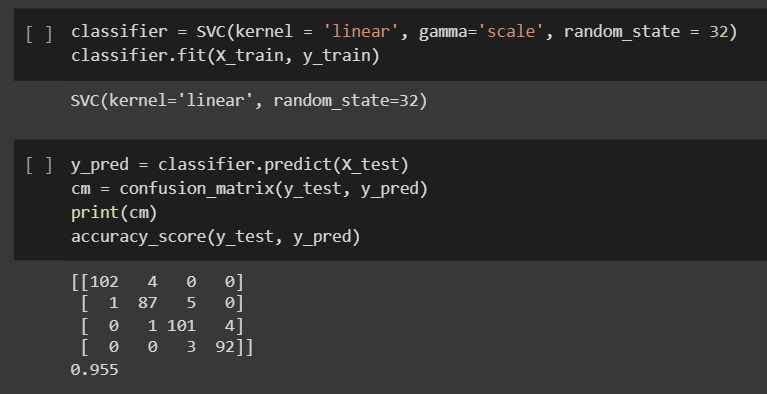


Gambar 3.12 Mencari hypermeter svm gridsearch

Melakukan import library dan dataset “mobile.csv”. Setelah melakukan GridSearch mendapatkan hasil :

* Kernel : Linear
* gamma : auto

1. Mencari Akurasi SVM



Gambar 3.13 Mencari Akurasi SVM

Dengan SVM mendapatkan akurasi sekitar 95% cukup baik, dimana validasi untuk mencari akurasi yang lebih akurat.

# BAB IV KESIMPULAN DAN SARAN

## 4.1 Kesimpulan

Berdasarkan penelitian tugas akhir ini telah dilakukan implementasi dengan dua metode machine learning pada dataset *mobile price classification*. Dataset yang telah dipilih berdasarkan topik yang akan dibahas, dan dataset tersebut diperoleh dari Kaggle. Dalam penelitian ini diambil topik Menentukan Jangkauan Harga Handphone berdasarkan Spesifikasi, berikut beberapa kesimpulan yang berhubungan dengan tujuan penelitian :

1. Penelitian ini dapat mengklasifikasikan jangkauan harga handphone dari spesifikasi handphone dengan tahapan pelaksanaan seperti berikut :
2. Identifikasi masalah
3. Pengumpulan data
4. Pemilihan metode
5. Penyusunan konsep
6. Pembuatan kode program
7. Evaluasi dan kesimpulan
8. Penelitian ini menggunakan dua metode yaitu SVM dan Random Forest Classifier.
9. Penelitian ini berhasil mengimplementasikan dua metode tersebut untuk menentukan jangkauan harga handphone berdasarkan spesifikasi yang dimiliki.
10. Pada metode pertama yaitu SVM, menggunakan persamaan hyperline yang nantinya akan mencari nilai bobot(*w*) serta nilai bias dengan menggunakan perhitungan sistem persamaan linear yaitu eliminasi dan substitusi. Setelah mendapatkan nilai yang diinginkan selanjutnya substitusi nilai tersebut ke dalam persamaan hyperline. Hasil dari persamaan hyperline akan dijadikan data dalam menghitung akurasi menggunakan confusion matrix.
11. Pada metode kedua yaitu Random Forest, yang mana mempunyai training data yang panjang, kemudian dilakukan pengambilan sedikit data training tersebut menggunakan subsample yang akan menghasilkan sebuah forest. Random Forest menggunakan dataset yang kemudian dilakukan pembagian menjadi data training dan data testing. Dari data training dilakukan pembagian lagi menjadi algoritma tree, kemudian dari tree tersebut didapatkan mayoritas keputusan yang akan menentukan hasil dari random forest.

## 4.2 Saran

Saran yang diperlukan untuk pengembangan program lebih lanjut antara lain :

1. Penambahan tampilan proses pelatihan sehingga pengguna dapat melihat proses yang terjadi.
2. Penggunaan metode klasifikasi data yang lain untuk perbandingan.
3. Penggunaan data yang lebih bervariasi.
4. Penambahan visualisasi data yang lebih bervariasi.

# DAFTAR PUSTAKA

Arifiyani, N. (2020). *Price Mobile Classification*, 1-16.

Khan, Z., & Asim, M. (2018). International Journal of Computer Applications. *Mobile Price Class prediction using Machine Learning Techniques*, 6-11.

Parapat, I. M., Furqon, M. T., & Sutrisno. (2018). Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. *Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi*, 6.

Pipalia, K., & Bhadja, R. (2020). International Journal for Research in Applied Science & Engineering Technology. *Performance Evaluation of Different Supervised Learning Algorithms for Mobile Price Classification*, 8.

Pudaruth, S. (2014). Computer Science and Engineering Department. *Predicting the Price of Used Cars using Machine Learning Techniques*, 12.

Siburian, V. W., & Mulyana, I. E. (2018). Prosiding Annual Research Seminar. *Prediksi Harga Ponsel Menggunakan Metode*, 4.

Srimuang, W., & Intrasothonchun, S. (2015). 12th International Joint Conference on Computer Science and Software Engineering (JCSSE). *Classification Model of Network Intrusion using Weighted Extreme Learning Machine*, 190-194.